Tuesday, April 15, 2025
Access the code, data, and analysis at https://github.com/j-jayes/who-is-it

Praise the people or praise the place?*

Upper tail human capital in electrifying Sweden

Jonathan Jayes
Lund University Economic History Department
jonathan.jayes@ekh.lu.se

ABSTRACT This paper investigates the origins, characteristics, and career trajectories of high-skilled workers—engineers and business leaders—central to Sweden's electrification during the mid-20th century. Leveraging a novel dataset constructed from approximately 66,000 unique entries in the Vem är Vem? and Vem är Det? biographical dictionaries, I examine the backgrounds of individuals active in the earliest electrifying regions. While previous research found localized benefits and relative immobility for medium-skilled workers in parishes gaining early access to the "Western Line" power grid (Jayes, Enflo and Molinder, 2025), this study reveals a different dynamic for the high-skilled elite. Probit analyses comparing pioneers working in Western Line parishes before 1930 to their peers show they possessed a distinct profile: they were significantly more likely to have technical education, particularly from KTH (Royal Institute of Technology), specific career experience in the USA, and fathers from agricultural or production/transport/laborer backgrounds, suggesting different selection patterns or pathways to mobility. Furthermore, comparing migrants who moved into these parishes versus locals ("stayers") reveals that migrants were more likely to bring KTH degrees and general overseas experience, indicating that these locations with early access to electricity actively recruited specific external human capital. These findings nuance the "People vs. Place" debate, suggesting early electrification hubs ("the Place") attracted and relied upon high-skilled individuals ("the People") with a specific, internationally-informed technical skillset, blending external talent with the local workforce to drive technological advance.

*Thank you to seminar participants at the Copenhagen Business School Department of Strategy and Innovation PhD Seminar and the HEDG Group at the University of Southern Denmark for valuable feedback on this work.

I. Introduction

Economic history as a discipline is set to benefit in at least three ways from the nascent AI revolution. Novel sources of data become available as new tools make them readable to machines, analysis in new ways is possible with new kinds of algorithms, and the time-cost of asking certain kinds of questions decreases, opening up new avenues for research. In this paper, I leverage these benefits to explore a fundamental question concerning Sweden's industrial transformation: 'who were the high-skilled workers in electricity-related occupations, and where did they come from?'. I detail the process of structuring and analyzing novel data sources – extensive biographical dictionaries – to answer this question.

Consider the journey of Ivan Beckius, born in 1895, the son of a farmer in rural Västmanland. His path would lead him far from the farm. After technical school in Malmö in 1915, he earns an engineering degree from Stockholm's prestigious Royal Institute of Technology (KTH) in 1919. By 1920, he's a laboratory engineer at ASEA in Västerås – a city situated squarely on the newly established 'Western Line,' the backbone of Sweden's growing national power grid. Later, Beckius ventured across the Atlantic, spending 1923-1925 designing transformers for the industrial giant General Electric in Pittsfield, Massachusetts, and working briefly on power station construction in New York. He returned to ASEA, but his career continued its international trajectory, later managing transformer calculations and technical operations for ASEA in the UK. Beckius's trajectory—rooted in rural Sweden, forged at elite institutions, shaped by international experience, and central to the electrification effort—epitomizes the questions this paper explores: Who were the highly skilled individuals driving this technological revolution? What were their origins, education, and experiences? And in the interplay between emerging industrial centers and mobile talent, should we 'Praise the People' for their skills or 'Praise the Place' for its ability to attract them?

In the quest to understand the dynamics of economic development and technological advancement, previous research by this author and his supervisors examined the transformative impact of early electricity access in Sweden. "Power for progress: The impact of electricity on individual labor market outcomes" (**jayesPowerProgressImpact2025**) shed light on the transformative impact of early electricity access in Sweden. "Power for progress: The impact of electricity on individual labor market outcomes" revealed how the advent of electricity along the 'Western Line' led to significant positive economic outcomes for the general population: a boost in income levels, reduced inequality, and sustained employment, particularly benefiting lower-income workers with only primary education. That study also noted a tendency for these medium-skilled workers to remain in their birthplaces, suggesting localized benefits and perhaps limited mobility.

Building on these insights, the present paper shifts focus to the upper tail of the human capital distribution – the highly skilled engineers and business leaders like Beckius, whose expertise was crucial for designing, implementing, and managing the new electrical systems. Did the parishes that electrified first attract or cultivate a distinct type of high-skilled talent? Specifically, I ask: What were the characteristics of engineers and business leaders working in the early-electrifying Western Line parishes before 1930? How did their origins, education, and experiences compare to their high-skilled peers active elsewhere or later? Furthermore, within these hubs, did the locals ("stayers") differ significantly from the migrants who moved in ("movers-in")?

To answer these questions, this paper leverages newly combined and structured data from two comprehensive sets of Swedish biographical dictionaries, Vem är Vem? and Vem är Det?, yielding ~66,000 unique individual biographies from the mid-20th century, from which we identify ~48,000 engineers and business leaders. Digitizing and structuring this rich source using modern computational techniques allows us to analyze the backgrounds and career paths of this elite group in unprecedented detail.

My analysis reveals a complex picture that challenges simple narratives. I confirm the high geographic mobility of this elite group, contrasting sharply with the medium-skilled workers studied previously. More importantly, quantitative analysis shows that the high-skilled individuals working in Western Line parishes before 1930 possessed a distinct profile compared to their peers: they were significantly more likely to come from agricultural or working-class backgrounds, possess technical education (specifically from KTH), and have career experience from the United States. Further analysis comparing migrants and locals within these early hubs indicates that migrants were more likely than locals to bring KTH degrees and general overseas experience, suggesting that these places with early access to electricity actively recruited specific external talent to complement the local workforce.

These findings hold value for understanding the drivers of dynamism during Sweden's electrification and offer insights for contemporary policy aimed at fostering growth through technological change. The methodologies employed also provide a template for leveraging rich archival biographical data. The paper is laid out as follows: Section 2 places the research question in the context of the relevant literature. Section 3 details the biographical data sources and the methods used for digitization, structuring, and occupation classification. Section 4 presents the empirical strategy. Section 5 presents the results comparing the characteristics of high-skilled pioneers in early electrifying parishes, first to their peers and then comparing migrants versus locals within that group. Section 6 discusses the implications of these findings for the "People vs. Place" debate and our understanding of high-skilled labor's role in technological transitions. Section 7 concludes.

II. Context

This paper ties into three strands of literature; two in content and one in methodology. The first strand focusses on the use of individual level biographic data in economic history, at scale. The second strand focusses on the importance of human capital in economic development. The third strand focusses on the use of new tools to structure and analyze historical data.

The use of individual level biographic data in economic history at scale

Several recent papers have made use of biographic data in innovative ways. Ford et al. (2023) use biographies of high school graduates compiled at the time of their school reunions to create a far richer measure of human capital than the conventional measure, number of years of schooling, alone. Titled, "Not the Best Fillers in of Forms? The Danish and Norwegian Graduate Biographies and 'Upper Tail Knowledge' ", the authors explain that these biographies are "mini-CVs", containing information about the school leaver's grades, their occupational trajectory, and their family background. These are used to create an innovative approach to measuring upper tail knowledge.

Nekoei and Sinn (2021) titled "Herstory: the rise of self-made women" analyzes the historical prominence of self-made women using a specially created database. This database, formed by applying machine learning to Wikidata and Wikipedia, catalogues notable individuals throughout history, highlighting details like occupation and family background. Their unique approach reveals a significant increase in the number of prominent women, especially those who achieved success independently of their family connections, across various fields, starting with literature, since the 17th century. This research provides a fresh perspective on women's historical achievements and roles.

Importantly, these papers go beyond the use of just administrative records, which contain register-like data that economic historians are familiar with. Leveraging new kinds of sources in this way allows the authors to approach their research with different kinds of questions.

In this paper, I structure biographic data about elite individuals that is similar in structure to Ford et al. (2023), and use the career trajectories of these individuals to better understand the contribution of educated workers to the adoption of electricity across Sweden in the 20th century. The differentiator in this case is the scale - I capture 66,000 individuals in Sweden in an automated manner, or about one percent of the population of the country at the time. Also, I don't have access to the grade lists that Ford et al. (2023) do, so I cannot use the same kind of measure of human capital.

The importance of human capital in economic development

The question of where the high skilled workers in electricity related occupations in Sweden came from is important in order to understand the economic dynamism of that era. As such, it ties into a wealth of research on technological change and the labour market, which I review briefly here.

The historical adaptability of labor markets to technological change is well-documented. In their study of the U.S. labor market's response to the automation of telephone operation, Feigenbaum and Gross (2020) demonstrate how technological displacement in one sector led to increased demand in others, suggesting an inherent resilience in labor markets. This finding is particularly pertinent to this exploration of Sweden's electrification, as it indicates a potential for both displacement and opportunity in the face of technological change.

Claudia Goldin's extensive analysis of labor markets in the 20th century provides a comprehensive backdrop to this study (1994). Her work highlights critical shifts in labor participation, wage structures, and job security, reflecting the complex interplay between societal changes and labor market dynamics. These insights are crucial for understanding how shifts in human capital, like those during Sweden's electrification period, contribute to broader economic outcomes.

The impact of the Digital Revolution on labor markets, as reviewed in the Oxford Review of Economic Policy, is also salient to this study (Adams, 2018; Goos, 2018). These articles underscore the emergence of job polarization and the crucial role of policy interventions in ensuring equitable benefits from technological advancements. This perspective is instrumental in understanding the differential impacts of electrification in Sweden, especially in terms of job creation and labor market segmentation.

Moretti's exploration of the geographical clustering of talent and innovation in "The New Geography of Jobs" provides a crucial perspective on the spatial dynamics of economic development (Moretti, 2012). His findings about the importance of local ecosystems in fostering innovation and economic vitality resonate with the investigation of how early electrification in Swedish parishes influenced the distribution and impact of skilled labor. His concern, that gains to productivity are eaten up by increased cost of living (primarily through housing costs) when constraints prevail, is not evidenced in the first half of the 20th century in Sweden. However, his example of Silicon Valley – where high productivity and attractive jobs draw in people with high levels of skill, raising property prices - is becoming more concerning in today's relatively housing scarce urban centers.

New technologies require new skills. Mokyr's research provides insights into the importance of both artisans and engineers in the progression of the Industrial Revolution. His studies underscore the synergistic relationship between theoretical knowledge and practical expertise, essential in driving technological innovation and economic progress (Mokyr, 2017). In his examination of the socioeconomic elites of early modern Europe, Mokyr explains how their education and exposure to new ideas and sciences were pivotal in fostering various intellectual and technological advancements. This educated elite, through their changing culture and institutions, played a crucial role in creating an environment conducive to innovation (Mokyr, 2018).

Not every innovator needs higher education. Mokyr's perspective is crucial in understanding the dynamics of technological development and economic growth, emphasizing the collaborative efforts between well-educated scientists and highly skilled artisans. This interplay highlights the importance of practical skills, theoretical knowledge, and their combined impact on technological progress. For example, figures like metallurgist Henry Cort, who collaborated with scientists despite lacking formal scientific training, exemplify the productive synergy between different forms of expertise in this era (Mokyr, 2018).

In Sweden, Andersson and Molinder (Andersson and Molinder (2024)) investigated how migration influenced skill acquisition during industrialization (1880s–1930s). Using data from the Historical Swedish Population Panel (HISP) and linked censuses, they found that rural-urban migrants significantly increased their occupational income — a proxy for skills — after moving to cities. The impact of cities on skills had both static and dynamic effects. The static effect came from immediate access to higher-skilled jobs, while the dynamic effect reflected skill accumulation over time due to urban learning environments. Migrants with below-median initial incomes benefited the most, highlighting urban areas as drivers of upward mobility. Surprisingly, skill gains were similar across city sizes, including Stockholm and smaller urban centers. This suggests that the benefits of urbanization during Sweden's industrialization stemmed from the general characteristics of urban labor markets rather than city scale.

Heikkuri and Prado (Heikkuri (2024)) examined the impact of electrification on Swedish industry between 1913 and 1926. Using establishment-level data from industrial surveys, they found a positive correlation between electricity adoption, higher employment, and increased wages. Establishments that adopted electricity grew more rapidly in both employment and wages compared to those that did not, highlighting electrification as a key driver of economic advancement. While direct evidence on skills is lacking, Heikkuri observed occupational shifts from farm laborers to factory workers in electrified parishes. Although both occupations were considered unskilled, factory workers earned significantly higher wages. This aligns with the findings in **jayesPowerProgressImpact2025<empty citation>** and suggests that electrification influenced labor reallocation and improved earnings for unskilled workers. However, Heikkuri and Prado also caution that factors such as collective bargaining agreements may have played a role, underscoring the need for further research to determine electrification's impact on skill demand.

In this paper, I want to find out where the individuals came from who enabled the technological development that was associated with Sweden becoming richer and more equal. Did they come from the areas around where the technology was developed / adopted, learning skills on the job? Or did they get formal education at one of Sweden's universities and then bring these skills to the hubs of technology? Should we praise the people, or the place?

The use of new tools to structure and analyze historical data

The third strand of literature that this paper ties into is the use of new tools to structure and analyze historical data. Within this strand, there are perhaps two main use cases; problems of prediction and classification, and 'big data' gathering and analysis.

Relating to the former, Mullainathan and Spiess (2017) wrote a pathbreaking article that documents the use of machine learning as part of an economists toolkit in the context of prediction problems, differentiating it from traditional parameter estimation. The authors explain the use of new data types like satellite images and text, as well as machine learning's role in policy, estimation, and testing economic theories.

Some interesting papers that incorporate prediction and classification problems include Bandiera et al. (2020), who classify CEO behavior by collecting high-frequency diary information and then use a machine learning algorithm to classify CEOs into 'leaders' and 'managers' by the content of their meetings. Koschnick (2024) uses a machine learning topic model to classify each paper by the universe of all students at English universities in the seventeenth and early eighteenth century to calculate a measure of how innovative the paper was; how it differed from the papers before it in the field and how similar the papers afterwards were. Dahl et al. (2024) in a paper titled "Breaking the HISCO Barrier: AI and Occupational Data Standardization" apply a neural network to the task of classifying an occupational description and benchmark their results against human labelling to show that the neural network achieved comparable accuracy with human labelling and involved an order of magnitude fewer human hours. I make use of this tool in the classification of occupations in this paper.

'Big data' papers in economic history now abound, as surveyed in Gutmann et al. (2018) in a review titled *'Big Data' in Economic History.* Many of these papers construct and use high quality register or census-like data on individual outcomes. Notable examples include the Longitudinal, Intergenerational Family Electronic Micro-Database Project by Martha Bailey which focuses on family histories to understand long-term economic trends. These histories are collected from various census-like sources and innovative ML tools are used to construct the longitudinal links. Similarly, "The Making of Modern America: Migratory Flows in the Age of Mass Migration" by Bandiera et al. (2013) involved the digitization of 24 million records of migrant flows through Ellis Island in New York, and found that measured out-migration rates in the US were double the reported figures in the earliest decades of the 20th century. Eichengreen (2021) uses natural language processing (NLP) to understand the content of a parliamentary committee debate on the gold standard in South Africa. Clark and Cummins' families of England database contains 1.7m marriage records in England from the 19th to the 21st centuries and allows to authors to pry out social dynamics in family formation, as well as geographic sorting between the North and South of England (2018).

There is also a growing literature in which authors lay out the step by step processes required for economic historians to make use of these new tools. A great paper in this vein that bridges the gap between cutting-edge computer science literature and the use cases of applied economists is by Correia and Luck (2023). The paper "Digitizing Historical Balance Sheet Data: A Practitioner's Guide" explores the application of machine learning, particularly Optical Character Recognition (OCR), to digitize large-scale historical economic data. The authors highlight the limitations of off-the-shelf OCR software, mainly due to high error rates, and propose a combination of pre- and post-processing methods to enhance accuracy. They apply these methods to two extensive datasets of balance sheets and introduce "quipucamayoc," a Python package that unifies these techniques.

Amujala et al. (2023), in "Digitization and data frames for card index records" explain the entire process through which they digitize and structure loan records from bank cards that contain both machine written and hand written text in varying formats over time. Tha authors lower the barrier to entry for other researchers by explaining their use off-the-shelf technology from Amazon Web Services. Each step is explained along with tips for successful extraction of handwritten information.

Perhaps the most impressive of these kinds of papers is the *Layout Parser* from Melissa Dell and her lab. The team have produced a python library that can be fine tuned to parse document type, extract information, and use machine learning to correct common errors at the point of data extraction. Dell demonstrates the use of this tool by extracting firm performance data from Japanese reports on yearly firm output which are atypical from the kinds of tables or column text that off-the-shelf optical character recognition tools have been trained on (2020).

I hope that this paper can be useful for other researchers using similar kinds of sources as a prompt on where to start collecting and structuring their data at scale.

III. Materials and Methods

Biographical dictionaries

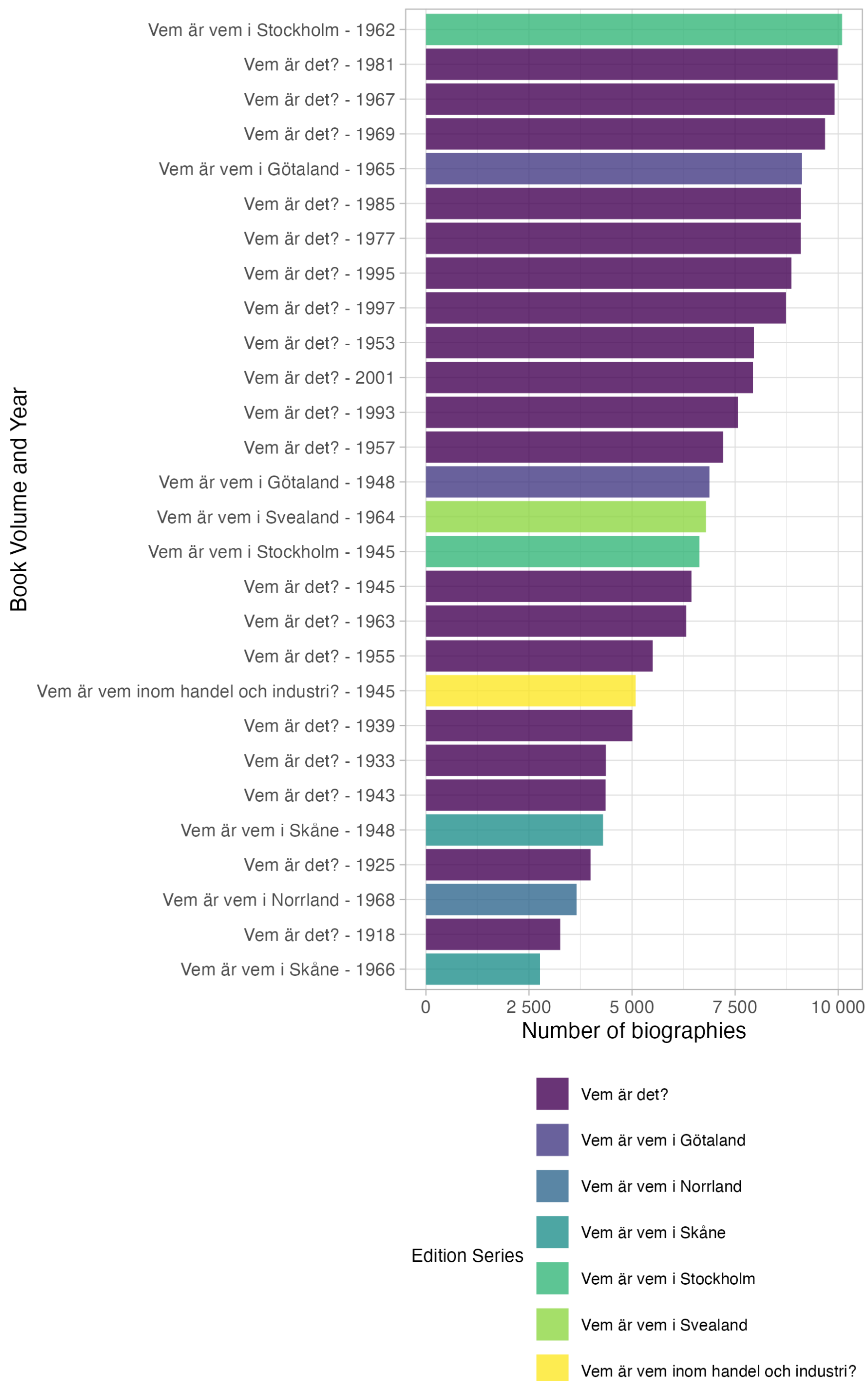
Vem är Vem? is a biographical dictionary, comprising a rich repository of information about notable individuals in Sweden. Published in two regional editions with a total of five volumes each, the first edition spanned from 1945 to 1950, and the second from 1962 to 1968, by the Bokförlaget Vem är Vem publishing house (Harnesky, 2013). An additional volume specifically focussed on individuals in industry and business was produced in 1945. This encyclopedia offers an invaluable snapshot of Swedish societal and professional landscapes during these pivotal periods.

The primary intention behind the creation of *Vem är Vem?* was to spotlight individuals who were at the peak of their careers, regardless of their age. This focus extends beyond traditional measures of influence, emphasizing the importance of those in influential positions or notable roles across relatively diverse sectors. As such, it serves as a crucial resource for understanding the professional and personal trajectories of around 66,000 individuals who shaped Swedish society in the mid-20th century.

It is worth noting that the criteria for inclusion was somewhat vague, and individuals could opt in to being included for a nominal fee. As a result, there are some individuals for whom not much information is included beyond biographic information, current location and profession. For others, there is a rich tapestry about their lives including records of career progression, business travel, technical writings and membership of civic organizations. The source does not capture a representative picture of Swedish society at the time, but rather those individuals with some level of social cachet or prestige, and a desire to be recorded in the biographical dictionaries as such.

Vem är Vem? is useful to economic historians thanks to its high quality digitization, with nine out of the 11 total volumes being made accessible online by librarians in Uppsala through *Projekt Runeberg*, as shown in Figure 1. This digitization has facilitated research, allowing for a broader exploration of the biographies and career paths of thousands of individuals. The encyclopedia's extensive coverage makes it a goldmine for researchers, historians, and anyone interested in the socio-economic history of Sweden during a period marked by significant change and development.

In the context of economic and historical research, *Vem är Vem?* serves as a unique tool. By providing detailed biographies and career information, it al-

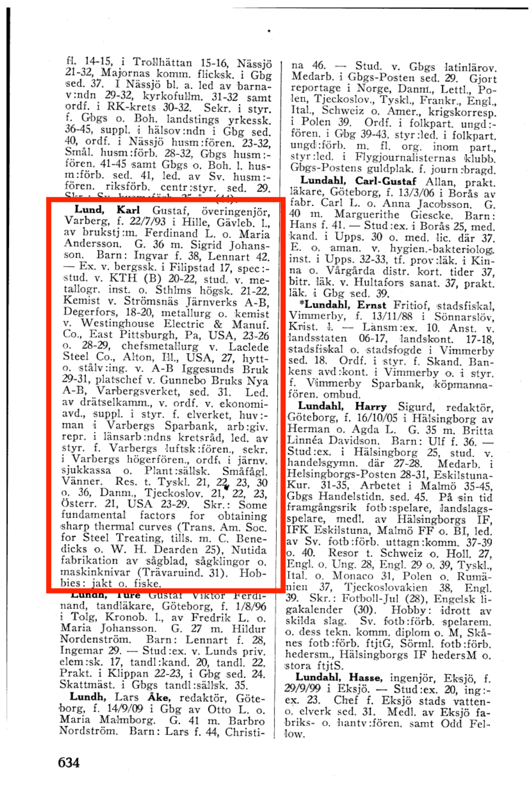


Source: *Vem är Vem?* and *Vem är Det?* (1945-1968) digitized by Harnesky (2013) and available at Projekt Runeberg.

Figure 1: Number of biographies in each volume of 'Vem är Vem?' and 'Vem är Det?'

lows for an in-depth analysis of the human capital that contributed to Sweden's economic and social evolution during the mid-20th century.

The biographic information about the individuals in the dictionaries are exemplified in Figure 2, which highlights the life of chemist and metallurgist Karl Gustaf Lund.



Source: Vem är Vem? Götalandsdelen utom Skåne 1948 page 634. Red box added by author.

Figure 2: A representative page of Vem är Vem?, highlighting the biography of Karl Gustaf Lund

The fields include:

1. **Education**: Lund's education at prestigious institutions such as the Royal Institute of Technology (KTH) indicates he had access to advanced technical knowledge. This level of education is critical for understanding the specialized skills that were necessary for innovation and advancement in electricity-related industries.

2. **Career Progression**: The text outlines Lund's career progression through various roles in metallurgy and chemical engineering. This trajectory can

illustrate how individuals applied their education in practice, contributing to industrial development. Tracking such careers can provide insight into the professional development paths that were common and valued in the sector at the time.

3. **International Experience**: His experiences in the United States reflect the cross-border exchange of knowledge and skills. This can show how international experiences contributed to the domestic industry by importing new ideas and practices, which is a key aspect of human capital development.

4. **Leadership and Management**: Lund's leadership positions, such as chairmanships and advisory roles, imply a combination of technical expertise and managerial acumen. The ability to lead and innovate within companies is a significant aspect of human capital that drives industry growth.

5. **Research and Innovation**: The reference to his translated research work indicates an engagement with cutting-edge technology and knowledge creation. Such contributions are the tangible outputs of human capital in action, pushing the industry forward through innovation.

6. **Professional Networks**: His involvement with societies and associations suggests a networked professional community, which is essential for the diffusion of innovative ideas and practices. These networks are often where knowledge is exchanged, partnerships are formed, and collaborations are initiated.

I use this biographical data to determine who the high skilled workers in electricity related jobs in Sweden were, and understand more about their career trajectories.

Source criticism

Obviously, the biographical dictionaries are not without their limitations. The criteria for inclusion in the dictionaries was somewhat vague, and individuals could opt in to being included for a nominal fee. It is not a representative sample of the Swedish population at the time. However, there should be some way to assess the representativeness of the biographical dictionaries. One way to do this is to compare the individuals in the biographical dictionaries to a different source of biographical information.

To assess the representativeness of the biographical dictionaries, I sampled a random selection of individuals from these dictionaries and manually checked whether they appear in the Swedish Biographical Lexicon (SBL).

The Swedish Biographical Lexicon (SBL), a personal history lexicon first published in 1917, now (as of spring 2024) includes entries from A to Södersten. It is a general inventory based on primary sources documenting important figures in Swedish society since about the 1500s, comprising approximately 27,200 pages with detailed data on individuals. The SBL aims to include individuals deemed significant by its editors, relying on broad and first-hand sources, without self-selection bias. This approach contrasts with the Vem är Vem? biographical dictionaries, where inclusion was voluntary, with criteria described as "somewhat vague," and individuals could pay to be featured. As a result, Vem är Vem? likely overrepresents those with a certain social standing and willingness to invest in being included, potentially limiting its representativeness of the wider society at the time.

I sampled 200 individuals from Vem är Vem? and found that only 38 (19%) were also present in the SBL, indicating that while Vem är Vem? contains a significant number of entries (about 66,000 individuals, representing roughly 1% of the Swedish population in the mid-20th century), it might overemphasize individuals of higher social status.

The level of detail provided in Vem är Vem? entries is inconsistent. Some individuals have extensive information on their careers, travels, publications, and organizational memberships, whereas others have only basic biographical information. This variability complicates the analysis of certain aspects of individuals' lives, particularly when attempting to classify career trajectories or sectors of work. However, while the amount of information might vary, we can clearly indicate when information is missing for individuals we analyze, and being transparent about these gaps is useful in understanding the limitations of my data.

Data collection strategy

Analyzing the biographical dictionaries required transforming the unstructured biographical text into a structured, machine-readable format suitable for quantitative analysis. This process is not complicated, but somewhat involved. It includes breaking each component of the source up (e.g. each biography), extracting the pertinent information from each record, storing each value with its associated key, and then saving this information in a way that is easy to analyze and aggregate.

The simplified process is laid out in Figure 3. The underlying code can be found on the GitHub repo linked at on the first page of this paper. I detail the third step, structuring the records, in the section below, and the remainder of the steps in the appendix.

Prior to the advent of Large Language Models (LLMs), this structuring of data from free text into key-value pairs was a task that required a large number of human hours to complete. It could be done either by putting the information into an excel sheet by hand, or writing rules to extract the information from the text. The first approach limits the number of observations a researcher can collect on her own, and the second approach quickly turns into the first.

The biographical dictionaries are written in a specific way, with many abbreviations and contracted names for common field titles and values. Due to the number of abbreviations, acronyms and contractions (for example, `Gävleb. l.` is the contraction of *Gävleborg län* or Gävleborg county in Figure 2), while it might be possible to take a simple rules based approach to replacing these contractions with their complete Swedish text, and then looking with regular expressions for specific terms relating to each piece of information, the number of rules soon balloons to an unreasonable figure, making the process unwieldy at best and impossible at worst. Writing a rule for every case necessitates as much

Data Collection Strategy

For biographical dictionaries

Step	Process	
1	Scrape book data from website	🌐➡️📚
2	Split records on each page of a book	📚➡️📄
3	Structure records with LLM	📄🔀📄
4	Augment data with coordinates	🔗🔍🌍🎓📚
5	Store data for analysis	🔗📚📂

Source: Author's own.

Figure 3: Illustration of data collection process

human involvement as would be required to manually structure the information - the first approach.

However, with the rapid advancements in LLM technology in the previous five years, and popular adoption of these tools through Chat-GPT and Microsoft's integration of GPTs into their products in the previous year, new tools mean this manual workload can be avoided to a large extent.

The structuring step involves sending a specific prompt to a LLM, along with the source material as text (rather than a scan of the book), and receiving back from the LLM a structured file with a key and value for each piece of information that I am interested in.

I make use of the computational backend of Chat-GPT, a model called GPT-4o-mini from OpenAI to structure the information from the dictionaries and catalogue into a JSON format that I can analyze, step 3, as shown in Figure 3. Many other LLMs, some open source, are available. I have chosen GPT-4o-mini as it is simple to interact with it in the programming language Python, and because doing so is relatively affordable for contained workloads such as this, when compared to hosting such a model on your own, beefy, costly computer.

To structure the 66,000 biographies in the dictionaries, it took about 15 hours of compute time, and cost 90 USD. When I timed myself extracting 10 biographies by hand, it took about 30 minutes. As such, the LLM method replaced 416 hours of human work with 15 hours of compute time, and 90 USD, plus the time invested in learning how to program the process.

By passing the text to the LLM, along with some context about what the model is being given, the model can behave like a skilled research assistant, reading the records, searching for the specific pieces of information requested, and outputting a structured file containing the information that we seek.

Prompting and context

To ensure accurate and consistent data extraction, the LLM was provided with specific instructions via a structured prompting strategy. This included defining the desired output format using a Pydantic schema, which specifies the required fields (e.g., name, birth date, education details, career history) and their data types. The prompt also provided context about the source material, including examples of common abbreviations, and instructed the model to return structured JSON output, which significantly improved the reliability of the extraction process compared to less constrained methods

A "system prompt" can be used to tell the model what kind of material it will be passed, and how it should respond. OpenAI suggest that users "ask the model to adopt a persona" in order to improve responses in a specific task in a guide on prompt engineering ("Prompt Engineering - OpenAI API", 2024).

I use a simple prompt to tell the model that it will be exposed to Swedish language biographical data:

```
system_prompt = "You are an expert on Swedish biographies in the 20th century."
```

Next I explain the kinds of information that I want it to extract, and the exact format that I want it in. I do this by specifying what is called a Pydantic data model (Pydantic, 2025). This model provides fields that the API must return and specifications for the kind of data in each field. An except of the data model is shown in Figure 4. The advantage of the Pydantic data model is that it allows me to specify the output in a way that is easy to read as a human, and a way in which OpenAI have built their API to accept. Since switching to this method of specifying the output, as opposed to a JSON schema, I have seen that the percentage of successfully extracted biographies has increased from 80% to 94%.[1]

```
class BioSchema(BaseModel):
    full_name: str = Field(..., description="Full name of the person")
    location: Optional[str] = Field(None, description="Location associated with the person")
    occupation: Occupation = Field(..., description="Occupation details of the person")
    birth_details: BirthDetails = Field(..., description="Details about the person's birth")
    education: Optional[List[EducationItem]] = Field(None, description="List of educational qualifications")
    career: List[CareerItem] = Field(..., description="Career history of the person")
    family: Optional[Family] = Field(None, description="Family details including spouse and children")
    publications: Optional[List[Publication]] = Field(None, description="List of publications")
    community_involvement: Optional[List[CommunityInvolvement]] = Field(None, description="Community roles and involvement")
    board_memberships: Optional[List[BoardMembership]] = Field(None, description="Board memberships held by the person")
    honorary_titles: Optional[List[HonoraryTitle]] = Field(None, description="List of honorary titles received")
    hobbies: Optional[List[str]] = Field(None, description="List of hobbies")
    travels: Optional[List[Travel]] = Field(None, description="Travel details")
    awards: Optional[List[str]] = Field(None, description="List of awards received")
    leadership_roles: Optional[List[str]] = Field(None, description="List of leadership roles held")
    languages_spoken: Optional[List[str]] = Field(None, description="Languages spoken by the person")
    military_service: Optional[str] = Field(None, description="Military service details")
    honors: Optional[str] = Field(None, description="Honors received by the person")
    death_date: Optional[str] = Field(None, description="Date of death")
```

Source: Author's own. For full schema, see the GitHub repo linked on the first page of this paper.

Figure 4: Excerpt of structuring schema

[1]This is an estimate based on a sample of 1000 biographies that I manually scored for accuracy.

Finally, I provide detailed context about the source and instructions for what I want the system to do. I include examples of the abbreviations and contractions that it will encounter, and inform the model as to what kind of output I am expecting in return.

```
prompt = f"""
You are an expert on Swedish biographies and will structure the
biographies of individuals from the 20th century biographical
dictionary 'Vem är Vem' that is provided below.

### Task:

1. Use the schema to organize the information in the biography.

2. Keep the biographic descriptions in Swedish and remove any
abbreviations based on your knowledge,
e.g. 'fil. kand.' is 'filosofie kandidat', and
'Skarab. l.' is 'Skaraborgs Län' etc.

3. For missing data in a required field,
include the field with a `None` value.

4. Ensure fields are correctly labeled and structured
as per the schema.

5. Put years in full based on context.
Put dates in DD-MM-YYYY format where possible.
"""
```

Example of structured biographic text

Following this process of structuring the records into a format with specified keys and values, I augment the data by geocoding locations in order to analyze geographic paths of individuals in the sample, and geographic clusters of firms.

Below I show the output of the data collection process, where the biographical dictionary entry on Swedish engineer and power station manager Axel Verner Nordell is shown in Figure 5 and some of the extracted information along with the geocoded coordinates are shown in Figure 6.

Nordell, Axel Verner, civilingenjör, fd. kraftverksdirektör, Motala, f. 15/8/81 i S. Möckleby, Kalmar l., av kyrkoh. Gustaf N. o. Almida Sellergren. G. 11 m. Agnes Hellgren. Barn: Inga f. 12, g. m. civ:ing. P. Rönström, Hans 14, civ:ing., Gösta 18, civ:ing., Ulla 20, g. m. civ:ing. H. Rönström. — Stud:ex. v. Lunds h. a. l. 99, avg:ex. fr. KTH (E) 04. Ritare v. ASEA i Malmö 04-05, ing. v. Elektr. A-B Holmia i Sthlm 05-07, v. Trollhätte kanal- o. vattenverk 07-09, distr:ing. v. stat. vattenf:verk 10-20, tf. chef f. Älvkarleby kraftv., Motalasektionen, 18, f. Motala kraftv. 19-20, kraftv:dir. v. stat. vattenf:verk, Motala kraftv., 20-47, pens. 47, därjämte verkst. dir. f. Motala Ströms Kraft A-B 30-47. Led. av kyrkofullm. sed. 32 o. av kyrkoråd sed. 31, kyrkvärd sed. 40, led. av o. ordf. i styr. f. Östergötl. Ensk. Banks avd:kont. i Motala sed. 22. Led. av Sv. tekn:fören. KVO2kl, RNO.

Source: Vem är vem inom handel och industri? 1944-45 page 380.

Figure 5: Raw information about Swedish engineer and power station manager Axel Verner Nordell

Key	Value
full_name	Nordell, Axel Verner
location	Motala, Östergötland
occupation	Civilingenjör, kraftverksdirektör
birth_date	15/08/1881
birth_place	S. Möckleby, Kalmar
birth_parents	Gustaf N. and Amanda Seillergren
parents_occupation	Kyrkoherde
birth_latitude	56.35646300000001
birth_longitude	16.420155
education_degree	Studentexamen
education_year	1899
education_institution	Lunds högre allmänna läroverk
education_latitude	55.7046601
education_longitude	13.1910073

Source: Author's own illustration of output

Figure 6: Extracted information about Axel Verner Nordell

Classifying occupations

In **jayesPowerProgressImpact2025<empty citation>**, we grouped occupations into three categories; direct electricity jobs (e.g. electricians), indirect electricity jobs that could benefit from electric motors (e.g. textile workers), and all other jobs. We made this classification based on the occupational title listed in the 1930 census alone. These titles are frequently used in economic history, along with a schema that classifies each title according to a list of possible titles and occupational descriptions. A widely used example is the Historical International Standard Classification of Occupations, or HISCO, defined originally by van Leeuwen et al. (2002). A wealth of mappings have been created that link an occupational string like *civilingenjör* to HISCO code 022, civil engineers, described as:

Workers in this unit group carry out research and advise on civil engineering problems, design projects and structures such as bridges, dams, docks, roads, airports, railways, waste disposal systems, flood control systems and industrial and other large buildings, and plan, organise and supervise their construction, maintenance and repair.

While this classification process used to involve a large amount of human hours, several new methods that use machine learning have emerged to make this process more efficient. Dahl et al. (2024) use a neural network to classify occupational strings using a supervised machine learning process based on the occupational titles in several waves of Danish census data. Merouani (2023) uses an unsupervised approach based on text embeddings to classify french occupational titles according to the HISCO schema, using a clever compartmentalization strategy to improve classification when individuals hold multiple titles.

In this paper I take a hybrid approach, making use of the additional data from the occupational trajectories of each individual; where they worked and what position they held at each firm or institution. This additional data helps me go beyond just a single occupational string and classify the sector that each engineer worked in. In many instances, the description of the individual in the biographical data relates to their education, rather than what they have done with it.

For example, while it is clear from Axel Nordell's occupational title, *Civilingenjör, kraftverksdirektör* (Master of Science in Engineering, power station director), that he worked in a directly electricity related occupation, this is not clear for Gustaf Fredrik Ambjörn, whose title is just *Civilingenjör*, based on the fact that he recieved the degree of *Civilingenjör* from the Royal Institute of Technology (KTH) in Stockholm.[2] When looking at his career trajectory, however, we can see that he began his career working at "Fore River Shipbuilding Co., Quincy, Mass., USA" and ended it working as a "Professor i praktiskt skeppsbyggeri" (professor in practical shipbuilding) at Chalmers Institute of Technology in Gothenburg. Evidently, examining his career trajectory we can discern the sector within which Ambjörn worked - using the information available to us in the biographical dictionaries can provide more information than the occupational title alone.

Classifying sectors for engineers

In order to programmatically assign the sector for each engineer (as engineers have the most variety of within occupation sectoral variation), I develop a two step process to classify first an individual's occupation based on their occupational title and then a sector based on their occupational trajectory.

In the first step, to get a HISCO code, I use the pipeline proposed by Dahl et al. (2024), named OccCANINE, that uses a supervised machine learning approach to classify an occupational string at the character level. Dahl et al. (2024) reports that their pipeline achieves a 95.5 percent agreement with the original source on observations on Swedish data. I use this pipeline and get a HISCO code for each occupational title in the biographical dictionaries, as well as for parents' occupations where available.

Having established the initial HISCO classification for occupational titles using the OccCANINE pipeline, the second step addresses the sectoral dimension, particularly crucial for engineers whose skills were applicable across a wide industrial and institutional landscape. This step leverages the detail contained within each individual's career trajectory – the sequence of employers and positions held over time – as recorded in the biographical dictionaries. The objective is to move beyond the nominal occupational title and assign a sector based on the substantive content of an individual's professional experience.

To achieve this programmatically, we employ techniques from natural language processing, specifically text embeddings, to represent both career paths and potential sectoral affiliations within a common high-dimensional vector space. The

[2] Swedish civil engineer degrees are usually translated instead as Master of Science in Engineering, rather than a direct translation to English as *Civil Engineer*.

core idea is to translate the textual descriptions of jobs, workplaces, and sectoral activities into numerical vectors such that semantic similarity corresponds to proximity in the vector space. For this purpose, we utilize the KB-BERT model ("KB/Bert-Base-Swedish-Cased · Hugging Face", 2023), a language model pre-trained on a substantial corpus of Swedish text, thereby capturing semantic relationships specific to the Swedish historical context. It is an adaptation of the breakthrough BERT model, introduced by Google Research in 2018 (Devlin et al., 2018). The advantage of this KB Lab model is that it has been trained on a selection of Swedish data, including books, news reports, and internet forums. Hence it is able to score the similarity of Swedish business descriptions and occupational titles.

Text embeddings are effective for clustering because they capture semantic meaning rather than relying on surface-level features like character composition. For example, while "steam engine" and "power station" are different in characters and literal meaning, they are semantically related in the context of industrial machinery and energy production. Text embeddings transform these phrases into numerical vectors that reflect this semantic similarity. When applied to clustering, this means that items with similar meanings, even if their literal expressions differ, are grouped together based on the contextual and conceptual similarities encoded in their embeddings. This capability makes text embeddings particularly powerful for organizing and categorizing text data in a way that aligns with human understanding and interpretation (Jurafsky & Martin, 2009).

In order to classify a sector for each engineer, I source the sectoral classifications from the 1947 version of Sveriges Handelskalender, a trade publication listing firms and their activities (Bonnier, 1947). This publication is a comprehensive directory of Swedish firms, providing insights into their sectors and activities. I extract the activities of each firm, and then cluster these into sectors using the text embedding model. This allows me to assign a sector to each firm, and then assign a sector to each engineer based on the similarity of their career experience to the sectors identified in the trade publication. I choose the 1947 version because the scanned quality is exceptionally high and digitization was easy, and because in the post 1930 period, the sectors that existed in the 1940s are likely to be similar to those that existed in the 1930s.

The process begins by vectorizing the career trajectories. For each engineer in our dataset, we first concatenate the textual information associated with their career path. This includes all listed job titles (e.g., konstruktör, driftsingenjör, verkställande direktör) and the names or descriptions of their workplaces (e.g., "ASEA, Västerås", "Kungliga Järnvägsstyrelsen", "L M Ericsson"). This concatenated string represents the textual fingerprint of an individual's career. We then feed this aggregated text into the KB-BERT model to generate a single dense vector embedding. This vector serves as a quantitative representation of the individual's entire professional trajectory within the semantic space defined by the language model. Concurrently, we derive vector representations for the 18 industrial and commercial sectors identified through the analysis of the 1947 Sveriges Handelskalender, as previously described. For each sector cluster (shown in Table 2), we create a representative textual description, potentially by aggre-

Table 2: List of 18 sectors identified in the 1947 version of Sveriges Handelskalender

Sectors
Book Printing and Publishing
Breweries, malt and soft drink factories
Building contractor, Concrete goods factory
Chemical-technical industry and jewellery
Chemical-technical industry, Chemical and pharmaceutical products, Paint and varnish factories
Clothing and textile industry
Electrical appliances and mechanical machinery
Food Industry
Furniture and decoration factories
Glass & Mirror Factories
Iron and Steel Foundries
Mechanical workshops for machine production (heavy industries)
Metal goods factory, Mechanical workshop (light industries)
Preserved and fresh food industries
Spinning, weaving and textile industry
Steel and metal industry
Textile and leather processing industry
Wood, sawmill and carpentry industry

Source: Author's own analysis based on the 1947 version of Sveriges Handelskalender.

gating the activity descriptions of the constituent firms or selecting keywords central to the cluster. This representative text for each of the 18 sectors is then processed through the same KB-BERT model to yield 18 distinct sector vectors. These vectors anchor the identified sectors within the same semantic space as the career trajectories.

With both individual career trajectories and sectors represented as vectors in the same high-dimensional space, we can quantify the similarity between an engineer's career and each potential sector. We employ cosine similarity, a standard metric in vector space analysis that measures the cosine of the angle between two non-zero vectors. A cosine similarity score close to 1 indicates a high degree of similarity in vector orientation (and thus, semantic content), while a score close to 0 indicates orthogonality or dissimilarity, and a score close to -1 indicates opposition. For each engineer's career trajectory vector, we calculate its cosine similarity with each of the 18 sector vectors. The engineer is then classified into the sector corresponding to the sector vector that yields the highest cosine similarity score with their career trajectory vector. This signifies that, within the semantic space captured by the KB-BERT model, the textual content describing the engineer's cumulative work experience is most closely aligned with the textual content representing that particular sector derived from the Handelskalender.

This vector-based approach allows for a classification that considers the entirety of an individual's recorded professional life, capturing shifts in roles and industries that a single occupational title might obscure. It provides a systematic and replicable method for leveraging the rich, albeit often unstructured, information contained in biographical career histories to map individuals onto a defined sectoral landscape, thereby facilitating a more detailed analysis of the deployment of engineering expertise within the Swedish economy during this period.

Data description

The core of the empirical analysis relies on variables extracted and derived from the combined *Vem är Vem?* and *Vem är Det?* biographical dictionaries. After digitizing and structuring the text using the methods described previously, I constructed a dataset containing key information for each unique individual identified as an engineer or business leader based on relevant HISCO code ranges (0200-0299 for engineers, 2100-2190 for managers/directors). These HISCO codes are assigned with OccCANINE from Dahl et al. (2024), not the sector specific classification I describe above. I limit the analysis to engineers and business directors, reducing the sample size to 47,000 biographies.

Basic demographic and geographic information forms the foundation. For each individual (`person_id`), I extracted the listed birth date (`birth_date`) and derived the birth decade (`birth_decade`), which serves as a crucial control variable (`C(birth_cohort)`) in the regression models, categorized into decades from 1880 to 1930. Birthplace location (`birth_location_name`) was geocoded to obtain latitude (`birth_location_lat`) and longitude (`birth_location_lon`). Using a spatial join with a historical parish map from 1920 (Junkka, 2024), I identified the corresponding birth parish (`birth_parish_code`, `birth_parish_name`) and determined if it was located along the Western Line (`birth_parish_is_western_line`), following the identification in **jayesPowerProgressImpact2025<empty citation>**. This variable is essential for identifying "stayers" in the analysis. Similar processing was applied to the listed current location (`current_location_name`, `lat`, `lon`) to determine if the individual resided in a Western Line parish at the time of the biography (`currently_lives_in_wl`), as well as whether any of their recorded workplaces were located in a Western Line parish.

To capture social origins, I extracted information on the individual's father, including name and occupation. The father's occupation string was then classified using the HISCO system, using OccCANINE (Dahl et al., 2024). For use in the regression analysis, I aggregated the father's HISCO code into broad major groups (`father_hisco_major_group_label`) based on the standard HISCO hierarchy (e.g., 'Professional/Technical', 'Administrative/Managerial', 'Agricultural/Fishing', 'Production/Transport/Laborer'), treated as a categorical variable in the regressions, where unknown is the base category.

The extensive education details listed in the biographies (`_education_raw`) were processed to create several variables. Based on keyword searches within the degree and institution fields (e.g., 'tekniska', 'KTH', 'handels', 'högskola'), I created binary indicators for whether an individual had any technical education (`edu_technical`), business education (`edu_business`), or other forms of higher education not classified as technical or business (`edu_other_higher`). Furthermore, using a standardized list of institution names derived from the raw entries, I created specific binary indicators for attendance at key institutions relevant to engineering and business: the Royal Institute of Technology (`studied_kth`), Chalmers University of Technology (`studied_chalmers`), the Stockholm School of Economics (`studied_hhs`), and any foreign university (`studied_foreign`).

Career trajectories, captured in the raw career entries (`_career_raw`), were used to derive indicators of international experience. By geocoding the workplaces of each individual's career and then examining the country codes associated with work locations, I created binary variables indicating whether an individual had any career experience outside Sweden (`career_has_overseas`) and specifically whether they had experience in the USA (`career_has_us`). The number of listed board memberships (`board_membership_count`) was also tallied directly from the biographical text. This information on board membership is useful to indicate the level of social capital and network connections an individual had, which is important in the context of the early electrification period.

Finally, several key variables were constructed specifically for the regression analyses. The primary grouping variable, `worked_wl_before_1930`, identifies the pioneers by checking if any career entry location (mapped to a parish using the spatial join) fell within a Western Line parish *and* had a start year before 1930. This forms the binary dependent variable (`dep_var_work_wl`) for the main probit model comparing pioneers to others. I also calculated the size of an individual's educational peer network (`edu_network_size`) based on the number of other individuals in the dataset graduating from the same primary educational institution within a defined time window (4 years), which serves as a control variable in the regressions. These structured variables derived from the rich biographical text allow for a quantitative investigation into the characteristics defining the high-skilled workforce during this pivotal period of technological change.

IV. Empirical Strategy

To investigate the role and characteristics of high-skilled labor during Sweden's early electrification, I employ a multi-step empirical strategy using the detailed biographical data compiled from the Vem är Vem? and Vem är Det? dictionaries. The goal is to first understand the general mobility patterns of this elite group and then to specifically identify the characteristics of those engineers and business leaders who were active in the "Western Line" parishes during the early phase before 1930, comparing them both to their peers and differentiating between locals and migrants within these hubs.

First, I establish the baseline geographic mobility for the high-skilled engineers and business leaders in my sample. Understanding the extent to which this group moved for education and career opportunities, particularly compared to the general population or the medium-skilled workers studied previously (**jayesPowerProgressImpact2025**), provides important context. I use descriptive statistics, calculating distances between birthplaces and subsequent education or work locations recorded in the biographies, to illustrate these patterns.

Second, I conduct the main analysis to determine whether the high-skilled individuals working in the Western Line parishes before 1930 – the pioneers active during the grid's establishment and early operation – possessed distinct observable characteristics compared to the broader sample of engineers and business leaders. This involves a two-part approach:

1. I begin with descriptive statistics, comparing the means and proportions of key background variables between the pioneer group (those with `worked_wl_before_1930 = TRUE`, N=1202) and all others in the engineer/businessman sample (N=47951). This provides an initial overview of potential differences in social origin (father's occupation), educational pathways (type of degree, institution), and international career experiences.

2. To isolate the partial correlation of each characteristic while controlling for confounding factors, particularly the significant difference in birth cohorts between the groups, I estimate a probit regression model. The model predicts the probability of an individual i belonging to the pioneer group ($WorkedWL_{pre1930,i} = 1$):

$$\Pr(WorkedWL_{pre1930,i} = 1|\mathbf{X_i}) = \Phi(\beta_0 + \mathbf{X_i}) \quad (1)$$

where Φ is the standard normal cumulative distribution function. The vector of individual characteristics, $\mathbf{X_i}$, includes categorical variables for father's HISCO major group (`C(father_hisco_major_group_label)`) and birth cohort (`C(birth_cohort)`), along with indicators for educational background (`edu_technical`, `edu_business`, `edu_other_higher`, studied_kth, studied_chalmers, studied_hhs, studied_foreign), career experience (career_has_overseas, career_has_us), board membership count (board_membership_count), and the size of their educational peer network (edu_network_size). The estimated coefficients reveal factors significantly associated with being an early actor in the Western Line parishes. Third (corresponding to Results Section 5.3), I delve deeper into the composition of the pioneer group itself to address the "People vs. Place" question more directly. I compare the characteristics of individuals who were *born* within a Western Line parish and worked there pre-1930 ("stayers," N=221) with those who were born elsewhere but *moved into* a Western Line parish to work before 1930 ("movers-in," N=876). For this, I estimate a second probit model, restricting the sample to only these 1,097 pioneers. The dependent variable is now $BornWL_i = 1$ if the pioneer i was born in a Western Line parish, and 0 otherwise. The model specification is analogous to the first, using the same vector of characteristics $\mathbf{X_i}$ to predict the likelihood of being a "stayer" versus a "mover-in" based on their backgrounds:

$$\Pr(BornWL_i = 1|\mathbf{X_i}, WorkedWL_{pre1930,i} = 1) = \Phi(\gamma_0 + \mathbf{X_i}) \qquad (2)$$

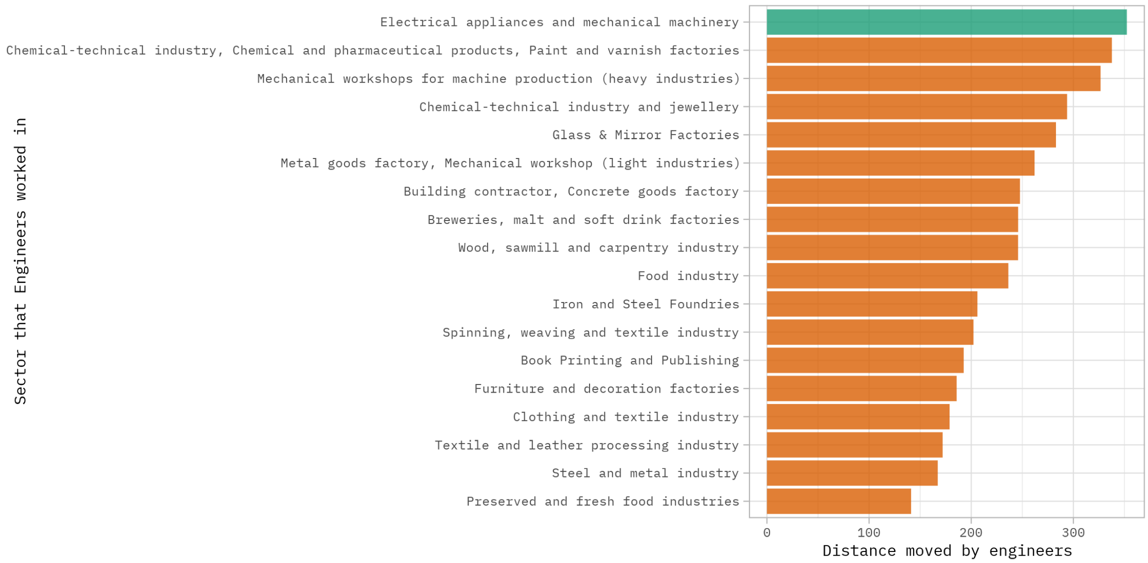
This allows me to identify which attributes, if any, significantly differentiated the local talent from the migrant talent within these crucial early electrification hubs. The results from these descriptive and regression analyses are presented in the following section.

V. Results

Geographic Mobility of High-Skilled Workers

A defining characteristic of the engineers and business leaders captured in the biographical dictionaries is their high degree of geographic mobility over the life course. In stark contrast to the findings for medium-skilled workers during the early electrification period, who tended to benefit locally and remain in their parishes of birth (**jayesPowerProgressImpact2025**), this high-skilled group frequently moved significant distances for education and career advancement. Calculating the distance between birthplace and subsequent career locations reveals the extent of this mobility.

Figure 7 plots the average distance moved from birthplace for engineers working in various industrial sectors identified in the biographical data. The distance is measured from birthplace to current location at time of biographical entry in the dictionaries. The distances are substantial across the board, far exceeding the average of 24km observed for working-age adults in the 1930 census. Notably, engineers working in sectors most closely associated with the Second Industrial Revolution's electromechanical transformation exhibit the greatest mobility. Those in 'Electrical appliances and mechanical machinery' moved an average of approximately 350km from their place of birth, significantly further than engineers in more traditional sectors like textiles or food industries, and over fourteen times the distance of the average 1930 worker. This pattern underscores that the individuals possessing the advanced technical and managerial skills crucial to electrification were part of a national, highly mobile labor market. Understanding who these individuals were, what characterized them, and how they were allocated across the landscape of technological change is therefore essential, forming the focus of the subsequent analysis.



Source: Author's own calculation of distances from birthplace to current location for engineers in the biographical dictionaries.

Figure 7: Average Distance Moved from Birthplace by Engineers, by Industrial Sector

Characteristics of High-Skilled Pioneers: Working in the Western Line Pre-1930

Having established the high mobility of the engineering and business elite, I now turn to examining the characteristics of those individuals who were present and working in the Western Line parishes before 1930. Were these early actors distinct from their high-skilled peers active elsewhere or later in the period covered by the biographical dictionaries? To investigate this, I first compare the background characteristics of the pioneer group (N=1,202 individuals identified as having worked in a Western Line parish prior to 1930) with the remaining engineers and business leaders in the sample (N=47,951).

Table 3 presents descriptive statistics comparing these two groups across several dimensions derived from the biographical data.

The descriptive comparison reveals potentially distinct patterns in social origins. Individuals working in the Western Line parishes pre-1930 were significantly more likely to have fathers whose occupations fell into the 'Production/Transport/Laborer' category (6.2% vs 4.1%, p=0.007) or the 'Agricultural/Fishing' category (3.7% vs 2.7%, p=0.029), compared to their

Table 3: Descriptive Statistics of Pioneers vs. Non-Pioneers

	Worked in Western Line Parish before 1930?			
	Worked WL Pre-1930	Did Not Work WL Pre-1930	Test Statistic	P-value
Birth Year				
Max	1931	1974		
Mean	1888.8	1906.6	t = -49.24	<0.001
Median	1889	1907		
Min	1839	1800		
N	1189	46489		
SD	12.1	18.8		
Father's HISCO Major Group				
Administrative/Managerial	60 (5.0%)	2907 (6.1%)	χ^2 = 2.19	0.139
Agricultural/Fishing	45 (3.7%)	1279 (2.7%)	χ^2 = 4.78	0.029
Clerical	2 (0.2%)	210 (0.4%)	χ^2 = 1.43	0.232
Production/Transport/Laborer	3 (0.2%)	41 (0.1%)	χ^2 = 1.93	0.164
Professional/Technical	1 (0.1%)	19 (0.0%)	χ^2 = 0.00	0.987
Sales	15 (1.2%)	683 (1.4%)	χ^2 = 0.15	0.699
Service	2 (0.2%)	126 (0.3%)	χ^2 = 0.13	0.718
Technical Education				
FALSE	451 (37.5%)	29334 (61.2%)		
TRUE	751 (62.5%)	18617 (38.8%)	χ^2 = 273.79	<0.001
Business Education				
FALSE	1071 (89.1%)	40380 (84.2%)		
TRUE	131 (10.9%)	7571 (15.8%)	χ^2 = 20.86	<0.001
Other Higher Education				
FALSE	1041 (86.6%)	31999 (66.7%)		
TRUE	161 (13.4%)	15952 (33.3%)	χ^2 = 209.26	<0.001
Studied at KTH				
FALSE	742 (61.7%)	36372 (75.9%)		
TRUE	460 (38.3%)	11579 (24.1%)	χ^2 = 125.69	<0.001
Studied at Chalmers				
FALSE	1092 (90.8%)	44980 (93.8%)		
TRUE	110 (9.2%)	2971 (6.2%)	χ^2 = 16.93	<0.001
Studied at HHS				
FALSE	1170 (97.3%)	45844 (95.6%)		
TRUE	32 (2.7%)	2107 (4.4%)	χ^2 = 8.04	0.005
Studied Abroad/Foreign Uni				
FALSE	1166 (97.0%)	46494 (97.0%)		
TRUE	36 (3.0%)	1457 (3.0%)	χ^2 = 0.00	0.999
Career Overseas				
FALSE	999 (83.1%)	43475 (90.7%)		
TRUE	203 (16.9%)	4476 (9.3%)	χ^2 = 76.81	<0.001
Career in USA				
FALSE	1100 (91.5%)	46313 (96.6%)		
TRUE	102 (8.5%)	1638 (3.4%)	χ^2 = 86.79	<0.001
Board Memberships (Count)				
Max	35	33		
Mean	1	1	t = 0.29	0.775
Median	0	0		
Min	0	0		
N	1202	47951		
SD	2.4	2.3		

Source: Processed biographical data (N = 49153 individuals included in comparisons). Tests compare 'Worked WL Pre-1930' vs 'Did Not Work WL Pre-1930'. Numeric: Welch's t-test. Categorical/Logical: Chi-squared test for proportions (χ^2). P-values < 0.001 shown as '<0.001'.

peers. This suggests that the pioneers might have been drawn from slightly less elite family backgrounds within this high-skilled sample, although further analysis is required.

The educational profile of the pioneer group also appears markedly different. They were substantially more likely to possess a technical education (62.5% vs 38.8%, $p<0.001$) and correspondingly less likely to have pursued business education (10.9% vs 15.8%, $p<0.001$) or other forms of higher education, such as law or humanities (13.4% vs 33.3%, $p<0.001$). This technical focus is further reflected in institutional attendance; pioneers were significantly more likely to have studied at KTH (38.3% vs 24.1%, $p<0.001$) or Chalmers (9.2% vs 6.2%, $p<0.001$), and less likely to have attended the Stockholm School of Economics (HHS).

Finally, patterns of international experience also differ descriptively. While there was no significant difference in the small proportion who studied at a foreign university, the pioneers were significantly more likely to have specific career experience in the USA (8.5% vs 3.4%, $p<0.001$). Interestingly, in this raw comparison, they were less likely to report general overseas career experience (16.9% vs 9.3%, $p<0.001$), hinting at the potentially distinct nature of the American connection.

These descriptive comparisons paint a picture of the high-skilled pioneers in early Western Line parishes as being an older cohort, potentially from less elevated social origins, with a strong concentration of technical training (particularly from KTH), and a notable prevalence of US experience.

However, these factors are interrelated. Most notably, the significant difference in birth cohorts might confound the interpretation of other characteristics that also changed over time, such as educational opportunities or the propensity for international experience. To disentangle these effects and assess which characteristics remain independently associated with being in the pioneer group, I estimate the probit regression model specified in the previous section (Equation 1), controlling simultaneously for background factors.

The results of this analysis, predicting the likelihood of working in a Western Line parish before 1930, are visualized in Figure 8 (detailed coefficients are available in Appendix). The model confirms several of the distinctions observed descriptively, even after accounting for birth cohort and other covariates.

Regarding social origins, the regression reveals a strong and statistically significant positive association between having a father in the 'Production/Transport/Laborer' category ($p<0.001$) and working in a WL parish pre-1930. A similar, though less strong, positive association exists for those with fathers in 'Agricultural/Fishing' ($p<0.10$). Conversely, having a father in 'Sales' is negatively associated ($p<0.05$). This reinforces the finding that individuals drawn to these high-skilled roles were disproportionately likely to originate from families outside the top administrative or professional strata, suggesting pathways for upward mobility or distinct selection mechanisms for these early technological ventures.

The educational profile remains highly significant. Possessing a technical education (`edu_technical`) strongly increases the likelihood of being in the pioneer group ($p<0.001$), while having 'Other Higher Education' (e.g., law, humanities)

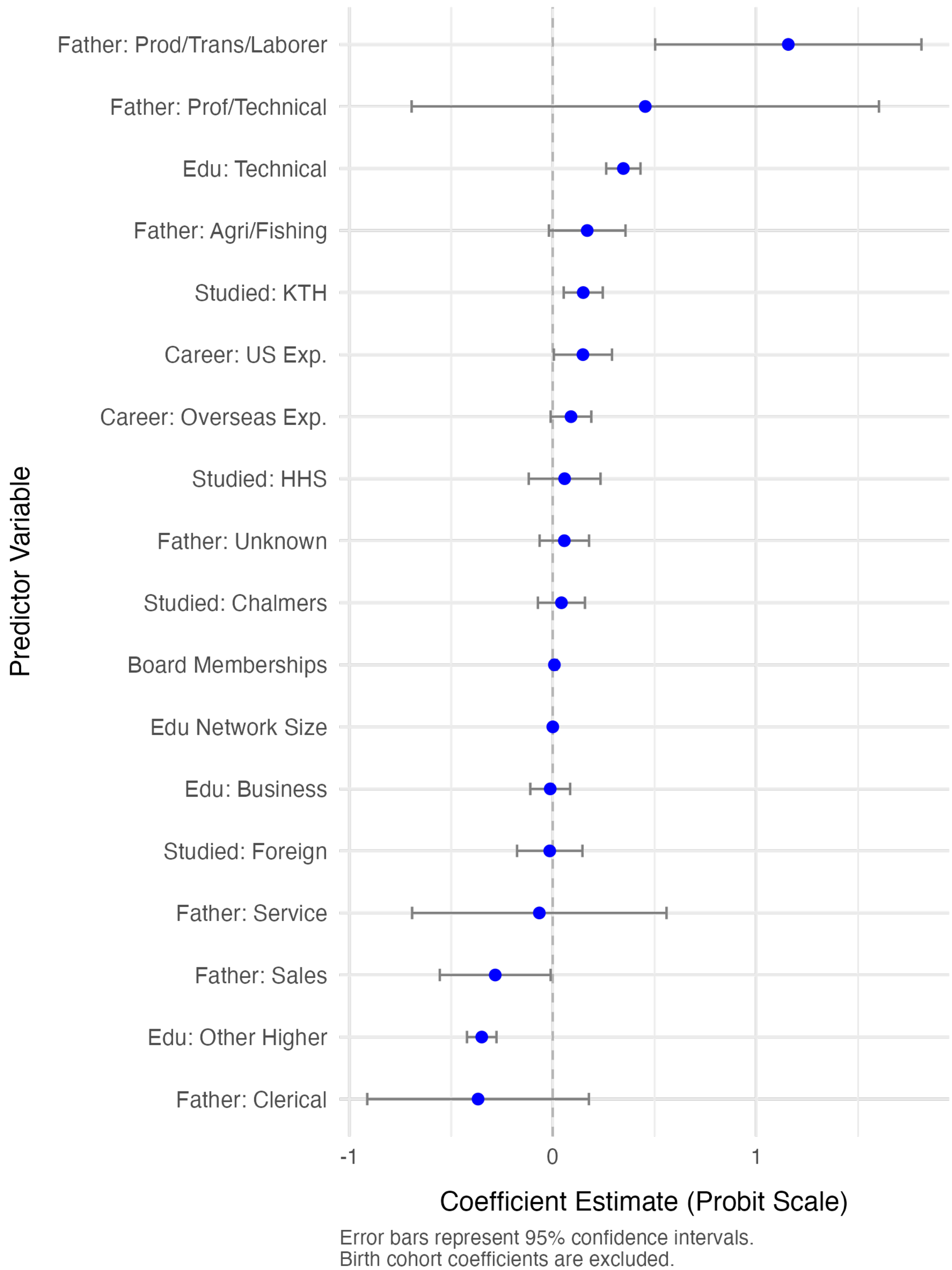


Figure 8: Probit Coefficients Predicting Work in Western Line Parish Pre-1930

significantly decreases it ($p<0.001$). Business education shows no significant independent effect. Drilling down into specific institutions, attendance at KTH (`studied_kth`) remains a significant positive predictor ($p<0.01$), confirming its central role in supplying talent for these roles, whereas studying at Chalmers, HHS, or a foreign university does not show a significant association in the regression context.

International experience also remains pertinent. Having career experience in the USA (`career_has_us`) is positively and significantly associated with being a pioneer ($p<0.05$). Interestingly, after controlling for other factors, general overseas experience (`career_has_overseas`) also shows a positive, albeit marginally significant, association ($p<0.10$). This contrasts with the raw descriptive statistics and suggests that while US experience held a particularly strong link, other forms of international exposure were not necessarily negatively correlated with involvement in these early projects once cohort and education are accounted for.

Finally, the analysis reveals a highly significant negative relationship between the size of an individual's educational peer network (`edu_network_size`) and the likelihood of working in a WL parish pre-1930 ($p<0.001$). This intriguing finding suggests that pioneers might have emerged from smaller cohorts or institutions, or perhaps followed less conventional paths than those embedded in larger peer groups at major universities, a point warranting further consideration. Overall, the probit analysis confirms that the high-skilled pioneers active in the Western Line parishes before 1930 possessed a distinct profile characterized by specific technical training often from KTH, a higher likelihood of US career experience, and origins more frequently rooted in agricultural or working-class backgrounds compared to their peers.

The significant positive association between prior career experience in the United States and working in these early electrifying parishes warrants closer examination. This was not simply time spent abroad; the biographical entries reveal that this experience was often acquired at the heart of American industrial and technological leadership. Figure 9 illustrates the geographic concentration of these US sojourns, primarily clustered in the industrial Northeast and Midwest. Individuals who later worked in the Western Line parishes pre-1930 spent time at major electrical manufacturers like General Electric and Western Electric, pioneers of mass production such as Ford Motor Co., steel giants like Carnegie Steel Co., and emerging technology firms like IBM. Others gained experience at the US operations of prominent Swedish multinationals, including SKF and Electrolux, or studied at leading American universities like Stanford and the University of Michigan, evidenced in Table 4.

Further insight into the nature of this American experience comes from classifying the types of US organizations where individuals in the early access to electrification group (Worked WL Pre-1930) and the control group gained their experience, as shown in Table 5. While interpretation requires some caution due to the number of individuals with specified US roles, a notable pattern emerges. Among those pioneers with US experience, a higher proportion gained it directly within the Electrical Industry (17.1%) compared to their peers in the control group (10.4%). This finding strongly reinforces the interpretation that the US

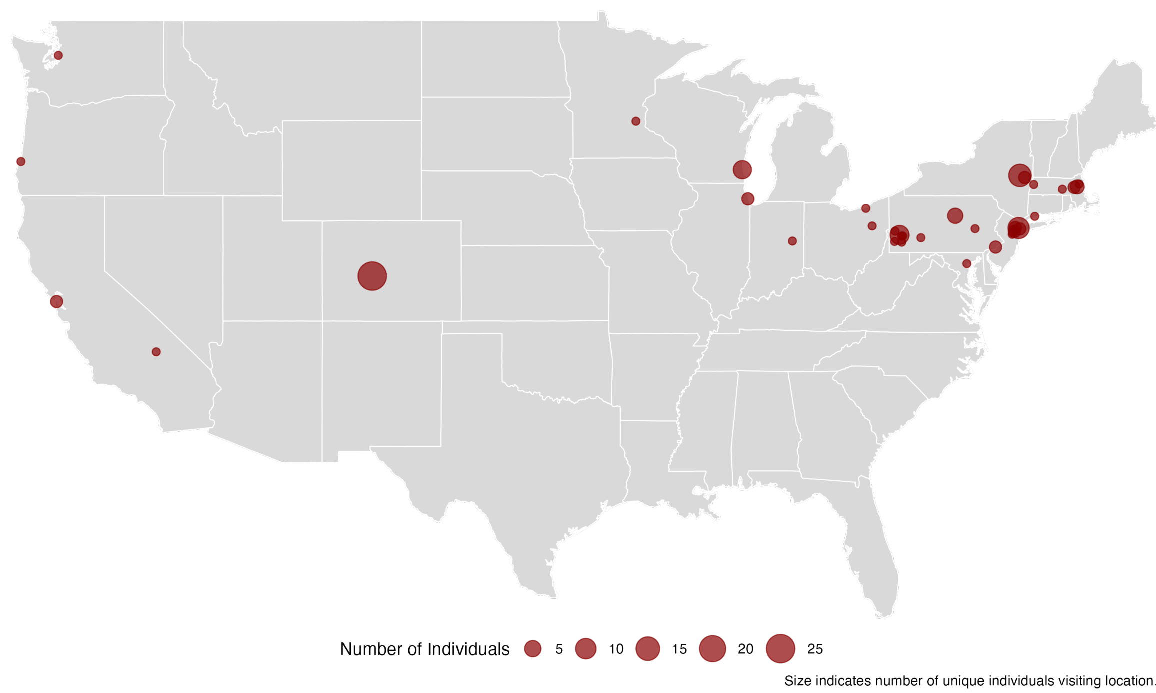


Figure 9: Map of US Locations for Engineers with Career Experience in the USA who Worked in Western Line Parishes Pre-1930

Table 4: 20 most common workplaces of US Experience for Engineers and Businessmen with Career Experience in the USA

USA Workplace
General Electric Co
Ford Motor Co
Svenska Handelskammaren
Carnegie Steel Co.
FN:s informationsdepartement
Utrikesdepartementet
Albert Bonnier Publishing House
Carnegie Steel Co
SKF Industries Inc
IBM USA
Lee Paper Co.
Western Electric Co
Joy Manufacturing Co
North Electric Co
Stanford University
Svenska ambassaden
University of Michigan
Electrolux Servel Corp
General Electric Company
Swedish American Prospecting Corp. N. Y., U. S. A.

Table 5: Distribution of US Organization Types for Individuals with US Career Experience

US Organization Type	Control Group	WL Pre-1930 Group
Electrical Industry	202 (10.4%)	20 (17.1%)
Automotive Industry	55 (2.8%)	0 (0.0%)
Manufacturing/Industry (Other)	457 (23.5%)	26 (22.2%)
University/Research	82 (4.2%)	2 (1.7%)
Consulting/Finance	38 (2.0%)	0 (0.0%)
Other/Unknown	920 (47.4%)	59 (50.4%)

experience valued in the context of early Swedish electrification was often directly related to the core technologies being implemented. Experience gained in general manufacturing or other industries was similarly prevalent in both groups, but the pioneers' US stints appear less frequently associated with university research, the automotive sector, or consulting and finance compared to the control group. This pattern suggests that the advantage conferred by American experience stemmed less from broad academic connections or diversified industrial knowledge, and more from targeted exposure to practices and technologies within the US electrical industry itself, knowledge directly transferable to the tasks awaiting these engineers and business leaders back in the pioneering Western Line parishes.

This pattern strongly suggests a process of targeted knowledge acquisition and transfer. The skills and practices related to electrical engineering, large-scale manufacturing, industrial organization, and potentially even business systems, honed at these specific American institutions and firms, were likely perceived as directly relevant and highly valuable for establishing and expanding Sweden's own electrical infrastructure and related industries. The fact that this specific US experience stands out as a significant predictor, more so than general overseas experience in the regression analysis, underscores the perceived importance of accessing the American technological frontier during this critical phase of Sweden's industrial development (Grönberg, 2003). These engineers and business leaders, therefore, acted as crucial conduits, bringing valuable, internationally-sourced human capital back to power Sweden's progress, particularly in the early electrifying regions along the Western Line.

Locals vs. Migrants within the Early Electrifying Group: Stayers and Movers-in

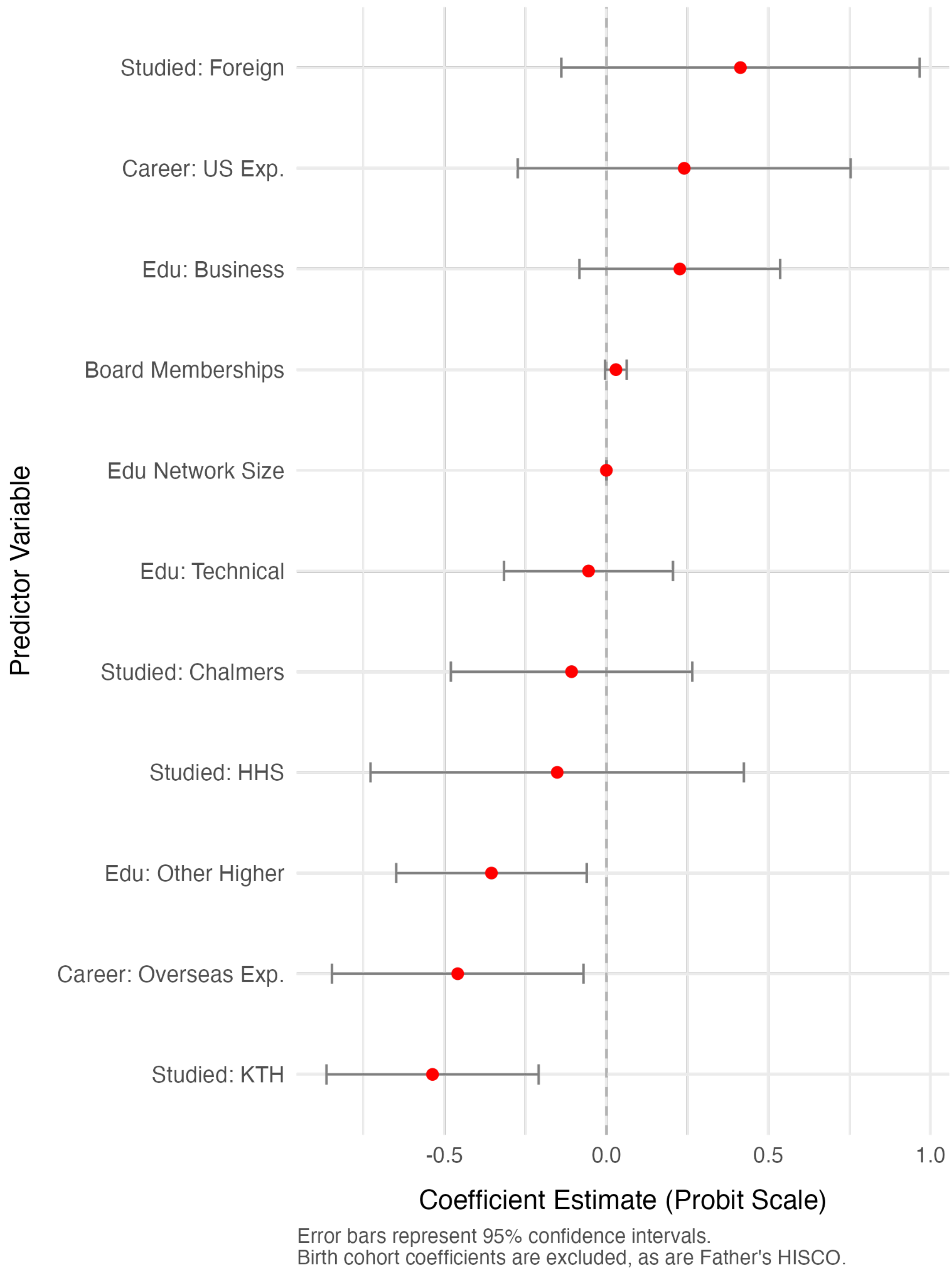
The previous analysis established that the high-skilled pioneers active in Western Line parishes before 1930 possessed a distinct profile compared to their peers elsewhere. To further refine the "People vs. Place" inquiry, I now examine whether observable differences existed within this pioneer group based on their origin relative to these key locations. Specifically, I compare those individuals who were both born in a Western Line parish and worked there before 1930 – the "Stayers" (N=221) – with those who were born elsewhere but moved into a Western Line parish for work before 1930 – the "Movers-in" (N=876). This comparison, conducted on the restricted sample of 1,097 engineers and businessmen, helps illuminate whether the necessary high-level human capital was primarily homegrown or imported during the initial phase of electrification in these hubs.

To assess this, I estimate a probit model where the dependent variable is a binary indicator (`born_in_wl`) taking the value 1 if the individual was a "Stayer" (born in a WL parish) and 0 if they were a "Mover-in" (born elsewhere). The model, shown in Equation 2, includes the same set of independent variables covering social origin, birth cohort, education, international experience, and network size as used in the previous analysis. The results identify characteristics significantly associated with being a local versus a migrant among these early actors.

The key findings from this regression are presented in Figure 10. Several significant differences emerge between the Stayers and Movers-in. Most notably, individuals born locally (Stayers) were significantly less likely to have studied at KTH (p=0.001) compared to those who moved in. Stayers were also significantly less likely to possess general overseas career experience (`career_has_overseas`, p=0.020) and less likely to hold 'Other Higher Education' degrees (p=0.018). Furthermore, the analysis suggests Stayers were less likely to have fathers occupied in 'Agricultural/Fishing' (p=0.017) than Movers-in. It is crucial, however, to note a limitation in interpreting social origins from this specific model: due potentially to the smaller number of Stayers, the coefficients for most other father occupation categories displayed extremely large standard errors, rendering them statistically unreliable. There is also marginal evidence ($p<0.10$) suggesting Stayers were slightly more likely to hold board memberships.

Perhaps just as revealing is what does *not* significantly distinguish the two groups within the pioneer cohort. Factors such as having a technical education overall (distinct from specific KTH attendance), business education, or having studied at Chalmers or HHS show no significant difference. Crucially, possessing career experience in the USA – a factor strongly associated with the pioneer group overall – does not significantly differentiate the Stayers from the Movers-in in this model (p=0.360).

These results lend weight to the interpretation that the pioneering "Place" – the Western Line parishes during early electrification – actively attracted specific forms of human capital that may have been less prevalent locally. The Movers-in appear to have disproportionately supplied the KTH training and general international exposure found among the pioneer workforce. This suggests a complementarity where external expertise, particularly from Sweden's leading technical university and through overseas experience, was essential and actively brought into these locations. The finding that US experience, while characteristic of the pioneer group as a whole, did not distinguish migrants from locals *within* that group further nuances the picture, perhaps indicating its broad importance for anyone involved in these frontier projects at the time, regardless of their geographic origin relative to the Western Line. The evidence points towards a dynamic where the specific demands of the "Place" were met by attracting



Source: Author's own analysis of the Vem är Vem? and Vem är Det? dictionaries. Note: Father's background and decade of birth omitted from figure for clarity.

Figure 10: Probit Coefficnet Plot Predicting Birth Origin (Stayer=1) among Engineers and Businessmen Working in Western Line Parishes Pre-1930

specialized “People” from beyond the parish boundaries to work alongside local talent.

VI. Discussion on People vs. Place

This paper set out to explore the characteristics of the high-skilled engineers and business leaders who were central figures during Sweden’s electrification, focusing particularly on the dynamics within the parishes that gained access to the national grid earliest via the Western Line. Building on previous findings that showed localized benefits and relative immobility for medium-skilled workers in these areas (**jayesPowerProgressImpact2025**), the core question became whether the dynamism observed in these early electrification hubs was primarily driven by attracting uniquely capable individuals (“Praising the People”) or whether the advantageous circumstances created by early electricity access (“Praising the Place”) allowed even standard high-skilled labor to thrive, perhaps by enabling local talent. The empirical results presented here suggest a more nuanced interplay between these two factors.

The analysis first reveals that the high-skilled pioneers working in Western Line parishes before 1930 were indeed observably distinct from their peers active elsewhere or later. They were significantly more likely to possess specific technical education, particularly from KTH, were more likely to have gained career experience in the USA, and disproportionately originated from agricultural or working-class family backgrounds relative to the broader elite sample, even after controlling for birth cohort effects. This finding challenges the simplest version of “Praising the Place”; these early hubs were not merely attracting generic high-skilled labor. Instead, the evidence suggests that the “Place” – characterized by nascent large-scale electrical projects – was associated with, and perhaps selected for, “People” possessing a particular combination of technical skills, specific international exposure, and potentially different social trajectories than the elite average.

Further dissecting the composition of this pioneer group by comparing locals (“Stayers”) with migrants (“Movers-in”) provides crucial insights. The analysis showed that Movers-in were significantly more likely than Stayers to hold degrees from KTH and to possess general overseas career experience. This strongly suggests that the early electrifying “Place” needed to actively import specific human capital attributes – particularly elite technical training from Sweden’s premier institution and broad international exposure – that were perhaps less prevalent or available among the locally born high-skilled individuals participating in these initial efforts. This finding tempers a narrative focused solely on nurturing local talent (“Praising the People” who were born there) and emphasizes the role of migration in fulfilling the specialized demands of these frontier locations.

The role of international experience, specifically from the United States, adds another layer to this picture. While US experience was a significant characteristic of the pioneer group overall when compared to their peers, it did not significantly differentiate the Movers-in from the Stayers within that group. This aligns compellingly with historical work, such as Grönberg (2003), detailing the targeted transfer of American technical and organizational knowledge back to Sweden by returning engineers during this era. The type of experience gained at firms like General Electric or Ford likely provided crucial know-how in electrical systems, mass production, and rational management pertinent to Sweden's industrial ambitions (Grönberg, 2003). The significance of US experience for the *entire* pioneer group, regardless of their origin relative to the Western Line, suggests its perceived value transcended the local/migrant divide for anyone involved at the cutting edge of electrification at that time, perhaps acting as a baseline requirement or a universally valued asset within these specific projects.

Ultimately, the dichotomy between "Praising the People" and "Praising the Place" appears insufficient to capture the dynamics observed for high-skilled labor during Sweden's early electrification. The evidence suggests a crucial *complementarity*. The "Place," endowed with early access to transformative infrastructure, created unique opportunities. However, realizing the potential of this Place required specific "People," characterized by advanced technical training (often from KTH) and relevant international experience (particularly from the USA). Critically, the analysis indicates that these necessary skills were not solely homegrown; the Place actively attracted essential human capital from outside, blending this imported talent with the local high-skilled workforce. The success of these pioneering electrification efforts likely depended significantly on the effective integration of this specific mix of specialized, mobile human capital with the opportunities afforded by the location.

VII. Conclusion

This paper has delved into the characteristics and origins of the high-skilled engineers and business leaders pivotal to Sweden's electrification, leveraging extensive biographical data from the Vem är Vem? and Vem är Det? dictionaries, structured and analyzed using contemporary computational techniques. Contrasting with previous findings for medium-skilled labor, this elite group exhibited high geographic mobility. Furthermore, the analysis revealed that the pioneers active in the earliest electrifying parishes along the Western Line before 1930 possessed a distinct profile compared to their peers, marked by specific technical training often from KTH, valuable experience gained in the United States, and a greater likelihood of originating from agricultural or working-class backgrounds. Comparing migrants versus locals within these hubs further suggested that external talent, particularly KTH graduates and those with overseas experience, played a crucial role, complementing the local high-skilled workforce.

Returning to the question of whether to "Praise the People or Praise the Place," the evidence points towards a vital interplay between the two. The unique opportunities presented by the pioneering "Place" attracted and required specific "People" possessing distinct skills and international experience, integrating imported expertise with local capabilities. This research demonstrates the power of applying new tools, such as large language models, to unlock rich, previously underutilized historical sources like biographical dictionaries, allowing for a granular analysis of the human capital driving technological and economic transformation.

By examining the individuals behind the aggregate trends, we gain a more nuanced understanding of the complex mechanisms shaping economic history.

Appendix

References

Adams, A. (2018). Technology and the labour market: The assessment. *Oxford Review of Economic Policy*, *34*(3), 349–361. https://doi.org/10.1093/oxrep/gry010

Amujala, S., Vossmeyer, A., & Das, S. R. (2023). Digitization and data frames for card index records. *Explorations in Economic History*, *87*, 101469. https://doi.org/10.1016/j.eeh.2022.101469

Andersson, J., & Molinder, J. (2024). Did Cities Increase Skills During Industrialization? Evidence from Rural-Urban Migration. *Uppsala Papers in Economic History*, (2024/12).

Bandiera, O., Prat, A., Hansen, S., & Sadun, R. (2020). CEO Behavior and Firm Performance. *Journal of Political Economy*, *128*(4), 1325–1369. https://doi.org/10.1086/705331

Bandiera, O., Rasul, I., & Viarengo, M. (2013). The Making of Modern America: Migratory Flows in the Age of Mass Migration. *Journal of Development Economics*, *102*, 23–47. https://doi.org/10.1016/j.jdeveco.2012.11.005

Bonnier. (1947). *Sveriges handelskalender*.

Clark, G. (2018, July). The big sort: The decline of northern England, 1780–2018.

Correia, S., & Luck, S. (2023). Digitizing historical balance sheet data: A practitioner's guide. *Explorations in Economic History*, *87*, 101475. https://doi.org/10.1016/j.eeh.2022.101475

Dahl, C. M., Johansen, T., & Vedel, C. (2024, April). Breaking the HISCO Barrier: Automatic Occupational Standardization with OccCANINE. https://doi.org/10.48550/arXiv.2402.13604

Dell, M., Zejiang, S., & Zhang, K. (2020). A Large Dataset of Historical Japanese Documents with Complex Layouts | Melissa Dell.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018, October). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Eichengreen, B. (2021). Gold and South Africa's Great Depression. *Economic History of Developing Regions*, *36*(2), 175–193. https://doi.org/10.1080/20780389.2021.1891879

Feigenbaum, J., & Gross, D. (2020, November). *Answering the Call of Automation: How the Labor Market Adjusted to Mechanizing Telephone Operation* (tech. rep. No. w28061). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w28061

Ford, N., Ranestad, K., & Sharp, P. (2023). Not the best fillers in of forms?: The Danish and Norwegian graduate biographies and 'upper tail knowledge'. *Not the best fillers in of forms?*

Goldin, C. (1994, June). *Labor Markets in the Twentieth Century* (tech. rep. No. h0058). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/h0058

Goos, M. (2018). The impact of technological progress on labour markets: Policy challenges. *Oxford Review of Economic Policy*, *34*(3), 362–375. https://doi.org/10.1093/oxrep/gry002

Grönberg, P.-O. (2003). Learning and Returning : Return Migration of Swedish Engineers from the United States, 1880-1940.

Gutmann, M. P., Merchant, E. K., & Roberts, E. (2018). "Big Data" in Economic History. *The Journal of Economic History*, *78*(1), 268–299. https://doi.org/10.1017/S0022050718000177

Harnesky. (2013). Vem är Vem?

Heikkuri, S. (2024). Technological change, skills, and occupational structure in Sweden, 1870-1950
Accepted: 2024-05-21T05:46:49Z.

Junkka, J. (2024, November). Junkka/histmaps.

Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.

KB/bert-base-swedish-cased · Hugging Face. (2023).

Koschnick, J. (2024). Breaking Tradition: Teacher-student effects at English Universities during the Scientific Revolution.

Merouani, Y. (2023, November). Innovation and Gender during Industrialization.

Mokyr, J. (2017). *A culture of growth: The origins of the modern economy*. Princeton University Press.

Mokyr, J. (2018). Bottom-up or top-down? The origins of the Industrial Revolution. *Journal of Institutional Economics*, *14*(6), 1003–1024. https://doi.org/10.1017/S174413741700042X

Table 6: Regression output for the probit model predicting working in a Western Line parish pre-1930

Variable	Estimate[1]	Std. Error	P-value	
Father: Agri/Fishing	0.169	0.096	0.078	.
Father: Clerical	−0.367	0.278	0.187	
Father: Prod/Trans/Laborer	1.157	0.334	0.001	**
Father: Prof/Technical	0.454	0.586	0.438	
Father: Sales	−0.283	0.139	0.041	*
Father: Service	−0.066	0.319	0.835	
Father: Unknown	0.057	0.062	0.363	
Edu: Technical	0.347	0.043	0.000	***
Edu: Business	−0.013	0.050	0.796	
Edu: Other Higher	−0.349	0.037	0.000	***
Career: Overseas Exp.	0.089	0.051	0.080	.
Career: US Exp.	0.148	0.073	0.041	*
Board Memberships	0.008	0.006	0.191	
Studied: KTH	0.149	0.049	0.002	**
Studied: Chalmers	0.042	0.059	0.471	
Studied: HHS	0.058	0.090	0.518	
Studied: Foreign	−0.015	0.082	0.853	
Edu Network Size	0.000	0.000	0.000	***

[1] Signif. codes: *** p<0.001, ** p<0.01, * p<0.05, . p<0.1
N = 47,715; Pseudo R-squared = 0.144;

Table 7: Regression output for the probit model differentiating between Western Line movers and stayers

Variable	Estimate[1]	Std. Error	P-value	
Father: Agri/Fishing	−0.730	0.306	0.017	*
Father: Clerical	−6.894	84,000.000	1.000	
Father: Prod/Trans/Laborer	−7.634	997,000.000	1.000	
Father: Prof/Technical	−6.048	5,426.209	0.999	
Father: Sales	−8.389	1,640,000.000	1.000	
Father: Service	6.643	84,000.000	1.000	
Edu: Technical	−0.055	0.133	0.680	
Edu: Business	0.227	0.158	0.151	
Edu: Other Higher	−0.355	0.150	0.018	*
Career: Overseas Exp.	−0.459	0.198	0.020	*
Career: US Exp.	0.240	0.262	0.360	
Board Memberships	0.029	0.017	0.086	.
Studied: KTH	−0.536	0.167	0.001	**
Studied: Chalmers	−0.107	0.190	0.573	
Studied: HHS	−0.152	0.294	0.606	
Studied: Foreign	0.414	0.282	0.142	
Edu Network Size	0.000	0.000	0.780	

[1] Signif. codes: *** p<0.001, ** p<0.01, * p<0.05, . p<0.1

N = 1,088; Pseudo R-squared = 0.098;

Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.
OCLC: ocn768167023.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *The Journal of Economic Perspectives*, *31*(2), 87–106.

Nekoei, A., & Sinn, F. (2021). HERSTORY: The rise of self-made women. *CEPR Discussion Papers*, (15736).

Prompt engineering - OpenAI API. (2024).

Pydantic. (2025). Models - Pydantic.

van Leeuwen, M. H. D., Maas, I., & Miles, A. (2002). *HISCO: Historical international standard classification of occupations*. Leuven University Press.
OCLC: ocm49628570.