Revisiting Adam Smith and the Division of Labor: New Evidence from U.S. Occupational Data, 1860–1940 *

Nicholas A. Carollo[1], Elior D. Cohen[2], and Jingyi Huang[3]

[1]Federal Reserve Board of Governors
[2]Federal Reserve Bank of Kansas City
[3]Brandeis University

October 2025

Abstract

Using novel occupational data from the U.S. between 1860 and 1940, we evaluate three of Adam Smith's core propositions about the relationship between the division of labor, market size, innovation, and productivity. We first document significant growth in occupational diversity during this period using new measures of occupational specialization that we construct from workers' self-reported job titles in the decennial Census. Consistent with Smith's hypotheses, we find strong empirical evidence that occupational specialization increases with the extent of the market, is facilitated by technological innovation, and is ultimately associated with higher labor productivity. Our findings also extend Smith's narrative by highlighting the role of organizational changes and innovation spillovers during the Second Industrial Revolution. These results speak to the enduring relevance of Smith's insights in the context of an industrializing economy characterized by large firms, complex organizational structures, and rapid technological change.

Key Words: Division of labor, occupations, productivity growth, technological change
JEL Classifications: N11, O14, J24, D24

*The views expressed in this paper are those of the authors and do not represent the views or policies of the Federal Reserve System or its staff. We thank Nicolas Ziebarth and an anonymous referee for helpful comments and suggestions that improved this paper. We also thank James Lee for sharing city-industry level Census of Manufactures data and Enrico Berkes for sharing data on the establishment of Carnegie libraries. Contact: nick.carollo@frb.gov, elior.cohen@kc.frb.org, jingyi.huang.econ@gmail.com.

"The greatest improvements in the productive powers of labour, and the greater part of the skill, dexterity, and judgment, with which it is anywhere directed, or applied, seem to have been the effects of the division of labour."

—Adam Smith, *The Wealth of Nations*, Book I, Chapter I

1 Introduction

The relationship between the division of labor, the extent of the market, innovation, and labor productivity that Adam Smith described in *The Wealth of Nations* has long been considered a cornerstone of economic thought. Smith's insights underpin modern theories of industrial organization, agglomeration, and innovation, and his thesis that improvements in the productive powers of labor arise from increasing specialization is widely accepted. Despite the influence of Smith's work, however, there is surprisingly little evidence on how the division of labor itself evolves or its broader economic implications.[1] A principal limitation has been the absence of systematic data on job tasks, which has impeded efforts to quantify changes in the division of labor across markets and industries over time.

In this paper, we revisit Adam Smith's insights on the division of labor using novel occupational data from the full-count U.S. Census between 1860 and 1940—a period of rapid industrialization and technological transformation. We construct highly granular measures of the division of labor based on the original occupational titles recorded in Census manuscripts, which specify each worker's "profession, occupation, or trade." These data reveal thousands of distinct job titles, which we use to trace the evolution of the division of labor across industries and local labor markets. We then link our occupational data to a broad range of city, county, and industry characteristics to evaluate three of Smith's core propositions: (1) that the division of labor increases with the extent of the market, (2) that it is enabled by technological innovation, and (3) that it enhances labor productivity. In doing so, we provide new systematic evidence on how the division of labor evolved in an industrializing economy and how Smith's insights manifest in the historical record.

Throughout this paper, we use the term occupational specialization to describe how job tasks are allocated across workers, that is, the extent of the division of labor (Becker and Murphy, 1992; Duranton and Jayet, 2011). Specialized jobs, in our context, involve fewer tasks than less specialized ones, but not necessarily greater skill or expertise. While the Census data do not directly record task content, we argue that the diversity of job titles we identify is a good proxy for specialization, an idea that is central to both Smith's view of the division of labor, as well as modern empirical work in labor economics (Atalay et al., 2024; Autor et al., 2024).[2] Supporting this assumption, we

[1] Most evidence on the relationship between specialization and labor productivity is indirect, showing for example, that larger cities and markets are more productive (Ades and Glaeser, 1999; Rosenthal and Strange, 2004), that larger firms exhibit more job differentiation (Chandler, 1977; Lamoreaux, 2019), and that innovative industries and regions adopt more specialized tasks (Jaffe et al., 1993; Atalay et al., 2024).

[2] Although the distinction between an occupation and a job title is somewhat arbitrary, for clarity we use the term "job title" whenever we mean a worker's specific, self-identified "profession, occupation, or trade" (e.g., "comptometer

link our data to the first edition of the Dictionary of Occupational Titles and show that descriptions of task content for defined titles tend to be shorter, on average, in places where we observe a greater number of titles in a given industry.

The cleaned occupation strings from the full-count Census reveal a dramatic increase in occupational diversity over the 80 years we study, with the number of distinct job titles rising from approximately 800 in 1860 to more than 6,000 by 1940. This growth, however, was uneven across sectors. Manufacturing and service industries exhibited significant increases in job title variety, whereas agriculture remained relatively stable despite its large employment share. We also document a pronounced decline in the prevalence of generic titles such as "clerk," "helper," and "laborer" beginning around 1880, indicating that job titles became increasingly differentiated and descriptive over time. While Smith's discussion of the division of labor centers around the reallocation of production tasks, our data reveal that administrative jobs also contributed to the growth of occupational specialization in the industrializing United States, reflecting more fundamental changes to firm scale and organization than those Smith observed in 18th-century Britain.

We next provide empirical support for Smith's theory that the division of labor increases with the extent of the market. Counties with larger populations, better domestic trade access, and higher urbanization rates recorded more job titles, indicating a more specialized workforce. These relationships hold even within narrowly defined industries—accounting for differences in sectoral composition over time and across markets—and when controlling for growth in product variety within manufacturing industries. Among the county-level characteristics we examine, population is the strongest predictor of occupational specialization, followed by domestic trade access. A one standard deviation increase in population is associated with an 80% increase in job titles within a county, while a one standard deviation increase in trade access is associated with a 14% increase. We additionally document a strong positive relationship between local industry size and the division of labor, suggesting that large-scale industrial production is crucial to specialization.

To further explore the link between market structure and specialization, we use data from the Census of Manufactures (CM) to study the role of firm size in shaping the division of labor, building on the idea that larger firms may employ a more specialized workforce and sort into larger markets (Tian, 2021). Even within a selected sample of relatively large urban markets, we find that industries with larger average establishment sizes exhibit greater occupational specialization. Specifically, a 1% increase in the average number of workers per firm is associated with a 0.04% increase in the number of reported job titles. Moreover, holding total industry employment constant, industries with a larger number of establishments display less specialization. These estimates suggest that, while the aggregate scale of an industry matters, large firms are the main drivers of specialization, consistent with the observation that the division of labor stems not only from the allocation of production tasks but also from the emergence of increasingly complex production processes and organizational structures (Becker and Murphy, 1992).

operator") and "occupation code" whenever we mean an aggregate, Census-defined category that combines related job titles (e.g., "office machine operators").

Smith also emphasized the reinforcing relationship between technological innovation and the division of labor. We examine this connection through two complementary channels: technological progress at the industry level and innovation spillovers at the city level. For the technological progress channel, we use data on the universe of utility patents issued in the United States between 1836 and 1940 and instrument for industry-level patenting with the arrival of breakthrough innovations identified by Kelly et al. (2021). The results show that industries with greater technological advancement experienced larger increases in occupational specialization: a 1% increase in new patents in the preceding decade (as predicted by breakthroughs) is associated with a 0.35% increase in the number of distinct job titles.

For the innovation spillover channel, we exploit spatial variation in the diffusion of knowledge by instrumenting city-level patenting with the historical construction of Carnegie libraries (Berkes and Nencka, 2024). Cities with higher predicted innovation intensity developed a more specialized workforce, with a 1% increase in new patents in the past decade associated with a 0.29% increase in the number of job titles. Together, these findings show that technological innovation—both within industries and through local spillovers—was a key driver of the expanding division of labor in the industrializing United States, linking Smith's original insights to modern theories of agglomeration and knowledge diffusion (Bloom et al., 2013; Jaffe et al., 1993; Moretti, 2004).

Last, we show that occupational specialization is positively associated with labor productivity. Using city-industry level data from the Census of Manufacturers, we find that value added per worker is strongly correlated with our measures of occupational variety. The results are both economically and statistically significant: a 1% increase in the number of titles is associated with a 0.2% increase in value added per worker, even after controlling for other determinants of specialization that may affect productivity through additional channels. This finding provides new empirical support for Smith's famous assertion that specialization is a key determinant of the productive powers of labor. That this holds true more than century after he wrote the *The Wealth of Nations* is a testament to the enduring relevance of Smith's insights.

Related literature. An important contribution of our work is the construction of a novel, highly granular occupation dataset that captures virtually the entire U.S. workforce over nearly a century. By leveraging the original occupation strings recorded in the Census, we provide one of the most detailed and comprehensive measures of the division of labor to date. This universal coverage, combined with the long temporal dimension—from 1860 to 1940—uncovers hidden detail in the U.S. labor market and offers new evidence on the dynamics and evolution of occupational specialization across locations and industries during the Second Industrial Revolution.

Our findings contribute to several branches of literature at the intersection of economic history, labor economics, and urban economics. First, they speak directly to the literature on the industrialization of the U.S. economy. Past research focuses on factors such as new technologies (Rosenberg and Trajtenberg, 2004; Fiszbein et al., 2024), market integration (Haines and Margo, 2008), and the rise of large-scale production (Chandler, 1977; Lamoreaux, 2019). While most researchers acknowledge the importance of occupational specialization, data constraints have historically limited empirical analysis of the division of labor.[3] Our findings provide direct evidence that specialization was an important driver of productivity growth during this period.

Similar to our approach, several other studies also leverage data on occupations and job titles to measure the division of labor in both modern (Atalay et al., 2024; Duranton and Jayet, 2011; Tian, 2021) and historical contexts (Ades and Glaeser, 1999; Chilosi et al., 2025). To the best of our knowledge, however, we are the first to do so for the late 19th- and early 20th-century United States, a period characterized by rapid economic growth and structural change. Our paper is most closely related to that of Chilosi et al. (2025), who analyze self-reported occupations from registered wills to document Smithian specialization in Britain between 1500 and 1800. We complement their findings by providing strong support for Smith's hypotheses on the determinants and productivity effects of the division of labor despite the very different contexts and time periods.

Finally, our study speaks to work on the geography of job tasks (Atalay et al., 2024; Michaels et al., 2019; Moretti, 2012) and urban agglomeration (Duranton and Puga, 2005; Rosenthal and Strange, 2004) by tracing the relationship between size (county, market, and industry) and the division of labor over time. Classical theories of agglomeration argue that larger cities facilitate more specialized labor markets through improved matching, input sharing, and knowledge spillovers. We provide new historical evidence for these theories, showing that larger cities indeed adopted more specialized occupations and achieved higher productivity.

2 Data

At the core of our project are new measures of occupational specialization for the late 19th- and early 20th-century United States, derived from workers' self-reported "profession, occupation, or trade" in the decennial Census. These data allow us to identify thousands of distinct job titles, which we argue are a useful proxy for the allocation of tasks across locations and industries and thus the division of labor. We also incorporate data on the extent of the domestic market, technological innovation, and manufacturing productivity to empirically assess Smith's classic theories on the determinants and economic implications of the division of labor.

2.1 Measuring Occupational Specialization

Throughout this paper, we use the term "occupational specialization" to refer to the distribution of job tasks across workers, that is, the extent of the division of labor. Importantly, we do not see occupational specialization as necessarily synonymous with skill or expertise, though some specialized occupations certainly require a great deal of job-specific human capital. Rather, we focus on the relationship between job titles and the range of tasks workers perform, an idea central to modern labor economics that was anticipated by Smith's seminal work.

[3] One notable exception is Sokoloff (1984), who uses the share of women and children employed by the firm as a proxy for the division of labor. More recently, Atack et al. (2019) use the "Hand and Machine Labor" study, comparing matched pairs of firms producing the same output but with different production processes.

The Wealth of Nations opens with an example of the division of labor in the pin-making industry that demonstrates this idea. Smith famously observes that production can be divided into 18 distinct tasks, from drawing the wire to grinding the heads to, finally, sorting and packing the pins. Similarly, he describes the production of a woolen coat as involving numerous specialized roles, including "the sorter of the wool, the wool-comber or carder, the dyer, the scribbler, the spinner, the weaver, the fuller, the dresser, with many others." In addition to the reallocation of production tasks, Smith also offers a somewhat broader view of the division of labor resulting from workers specializing in a narrower range of products or services:

> "A country carpenter deals in every sort of work that is made of wood; a country smith in every sort of work that is made of iron. The former is not only a carpenter, but a joiner, a cabinet-maker, and even a carver in wood, as well as a wheel-wright, a plough-right, a cart and waggon-maker. The employments of the latter are still more various."

Smith's observations reveal a crucial insight for our empirical analysis. Workers engaged in more specialized tasks often have distinct job titles, and hence a larger variety of titles in a location or industry may be informative about the division of labor, even if task allocation is not directly observed. Following this insight, we measure occupational specialization between 1860 and 1940 using data from the complete-count U.S. Census restricted-use version, which is available through IPUMS USA (Ruggles et al., 2024). Specifically, we clean and standardize the variable "OCC-STR," a text field that records each respondent's job title exactly as it was given on their original enumeration form.[4]

Census occupation records. The Census began recording the "profession, occupation, or trade" of both men and women in 1860. This question was open-ended, allowing respondents to report their job titles freely rather than selecting from a prespecified list. Only during the subsequent processing stage did clerical staff assign these responses to standardized classification codes for tabulation and publication. When conducting interviews, enumerators were instructed to record responses with as much detail as possible. For example, instructions prepared for the 1870 Census stated that:

> "The inquiry 'Profession, Occupation, or Trade' is one of the most important questions on this schedule. Make a study of it. Take especial pains to

[4]Our data use agreement with IPUMS limits us to a 20% extract of this database, which we select by focusing on employed civilian adults between the ages of 15 and 64 who reported an occupation. We also exclude common titles such as "clerk," "farmer," and "teacher" to meet the sample size restriction while retaining the most informative job titles. Although we do not observe the title strings for workers excluded from our 20% extract, these individuals must, by construction, have held one of the titles on the list that we prespecified. We therefore assign titles to these workers (i.e., clerks, farmers, etc.) based on their OCC1950 code, allowing us to compute our specialization measures using the universe of U.S. employment with negligible loss of granularity.

> avoid unmeaning terms, or such as are too general to convey a definite idea of occupation" – Census Office (1870, p. 13).

The document then provides nearly two full pages of instructions related to the occupation field alone, emphasizing that generic titles (e.g., agent, clerk, or mechanic) should be accompanied by additional detail whenever possible. Although specific instructions were revised prior to each Census, all of them consistently emphasized the importance of capturing precise and descriptive job titles. As a result, these records offer a unique source to study the historical evolution of the U.S. labor market in granular detail.

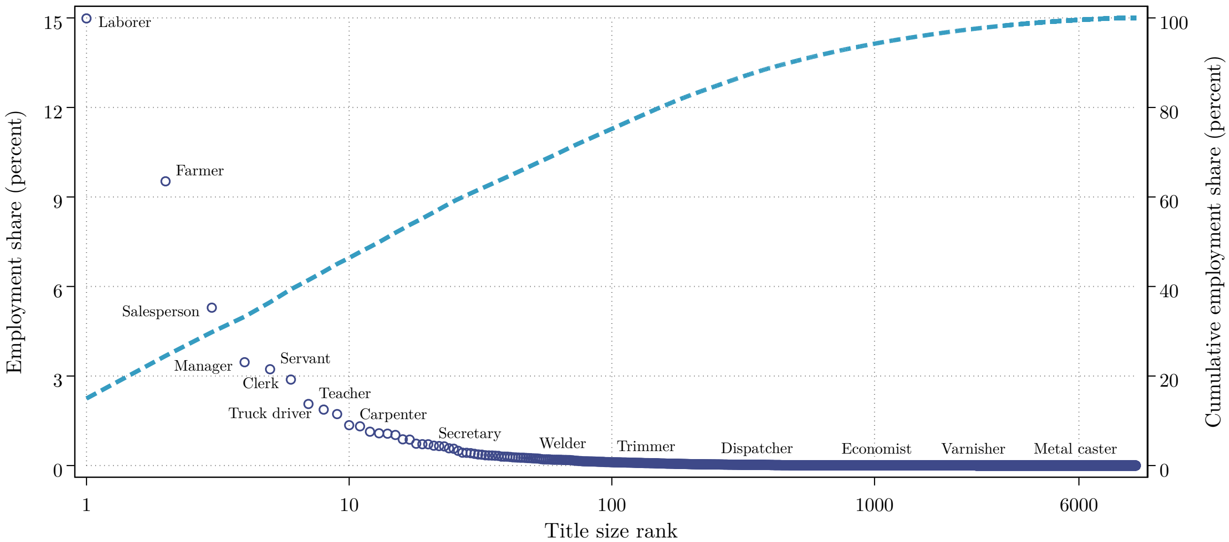
Job titles were digitized along with other text fields from the original Census forms and are available through the "OCCSTR" variable provided by IPUMS. In the Appendix, we detail our process for cleaning and standardizing this unharmonized text field, reducing millions of unique strings to approximately 10,000 standardized titles. Although we cannot classify every string appearing in the data, we can assign a standardized title to more than 95% of workers in each Census year we analyze. When constructing our specialization measures, we exclude individuals who did not report an occupation, reported only an industry, or had an unidentifiable title (mostly due to incomprehensible text).

Leveraging standardized job titles rather than statistical occupation codes (e.g., "OCC1950") offers two key advantages for measuring occupational specialization. First, it allows for much more granular identification of job titles. While some OCC1950 codes like "chiropractors" or "loom fixers" closely align with a single job title, many others group together hundreds of distinct jobs, obscuring meaningful variation in task content.[5] Second, statistical classifications, including OCC1950, are not fully harmonized across all census years, introducing inconsistencies that complicate the measurement of changes in specialization over time. For example, the IPUMS data for 1940 have no workers classified as "accountants and auditors," who are instead combined with "bookkeepers" and "cashiers" in this year. By working directly with the raw text data, our approach mitigates these classification issues and provides a more consistent measure of occupational specialization over time.

Industry classification. A key question we address in this paper is how the division of labor varies across and within industries. Here, we rely on the IPUMS "IND1950" classification. The main reason for this choice is that the Census only began asking specifically about industry in 1910 (though some workers in earlier years still provided one), meaning that industry strings are less comparable than occupation strings before and after 1910. In contrast, IPUMS assigns an IND1950 code to most workers in all Census waves, likely using a combination of job title and industry strings, Census classification codes, and assumptions about which titles occur only in certain industries for years before 1910. Although this imputation almost certainly introduces some measurement error

[5]Whereas OCC1950 is limited to 283 categories, our use of OCCSTR yields approximately 10,000 standardized titles, offering a significantly more precise measure of occupational specialization. For example, the category "operatives, not elsewhere classified" accounts for 40% of total manufacturing employment in 1940 and includes around 2,000 distinct titles reported by 100 or more workers.

Figure 1: Distribution of Employment Across Job Titles, 1940 Census



Notes: This figure plots the distribution of job title employment shares relative to their log rank using data from the 1940 Census. The sample excludes workers who we cannot assign to a standardized title.

in workers' industry affiliation, we show that our main findings are little changed when restricting the sample to 1910 through 1940, where industry classifications are the most reliable.[6]

Location classification. We use either counties or cities to define the geographic boundaries of local labor markets. Counties offer the most granular geographic unit for which data are consistently collected and comparable over time. Moreover, many of the key data sources used in our analysis—such as market access and patent data—are reported at the county level, making counties a natural and practical choice for geographic analysis. To address changes in geographic boundaries over the study period, we harmonize all data to 2010 county boundaries using crosswalks from Ferrara et al. (2024) for aggregate county-level statistics and from the Census Place Project (Berkes et al., 2023) for variables computed using IPUMS microdata. In analysis using data from the Census of Manufactures, we shift to using cities to align with the CM's geographic coverage.

Specialization measures. Our first occupational specialization measure is the number of unique job titles present in a location or industry.[7] In addition to the sample restrictions discussed above,

[6] In the 1940 data, OCC1950 and IND1950 were assigned based on 1940-vintage occupation and industry codes—which were also recorded on the enumeration forms—rather than the original free-form occupation and industry strings. As a result, some industry and occupation codes that are defined in the 1950 classification and appear in previous census waves drop out of the 1940 sample entirely. We therefore combine a small number of industry codes to improve the comparability of industry categories between 1910 and 1940, though we continue to refer to our aggregated classification as IND1950 for simplicity.

[7] We do not attempt to consolidate synonyms in our analysis (e.g., attorney and lawyer), which could overstate the division of labor. On the other hand, job titles will underestimate the division of labor if workers with generic titles like "clerk" or "laborer" are actually performing highly specific tasks. Our analysis therefore speaks to relative rather than absolute differences in the division of labor across time, locations, and industries. We discuss the prevalence

we limit our attention to titles that were reported by 100 or more workers nationally. We adopt this restriction primarily to reduce measurement error from misspecified titles.[8]

As shown in Figure 1, the distribution of employment is highly skewed toward the most common titles. Even as of 1940, the year in which job title variety is the highest, the top 100 titles in our data still employ around 75% of all workers nationally. This skewness poses a potential issue for using unweighted counts of unique titles as a specialization measure, as it may overstate variety in settings where employment is concentrated in just a few roles. Therefore, we construct two other measures of specialization that account for how employment is allocated across job titles. First, we follow Ades and Glaeser (1999) and define a Dixit–Stiglitz variety index:

$$Variety_{jt} = \left(\sum_{o \in \mathcal{O}} e_{ojt}^{\frac{1}{2}} \right)^2, \tag{1}$$

where e_{ojt} is the employment share of title o in year t and industry (or location) j. The index increases when there are more distinct job titles, but it also accounts for scale such that titles with a greater employment share contribute more to the index, reflecting their economic weight and also making it more robust to measurement issues related to rare or misspecified titles.

Alternatively, we use employment shares to compute Shannon entropy, defined as

$$Entropy_{jt} = - \sum_{o \in \mathcal{O}} e_{ojt} \ln(e_{ojt}). \tag{2}$$

Shannon entropy is a useful measure of the division of labor because it increases with both the number of distinct titles and the evenness of the employment distribution across those titles. It captures not only how many different jobs exist but also how balanced the labor force is across them. The negative log term in the entropy index places relatively greater weight on less common titles, as these carry more "information" about the complexity and granularity of the workforce.[9]

2.2 Relationship to Task Content

The implicit relationship between job titles and task content is a central assumption for our analysis. Although we do not observe the tasks that individual workers perform, as a validation exercise we show that our specialization measures are still informative about how tasks are structured, in aggregate, across industries and geography.

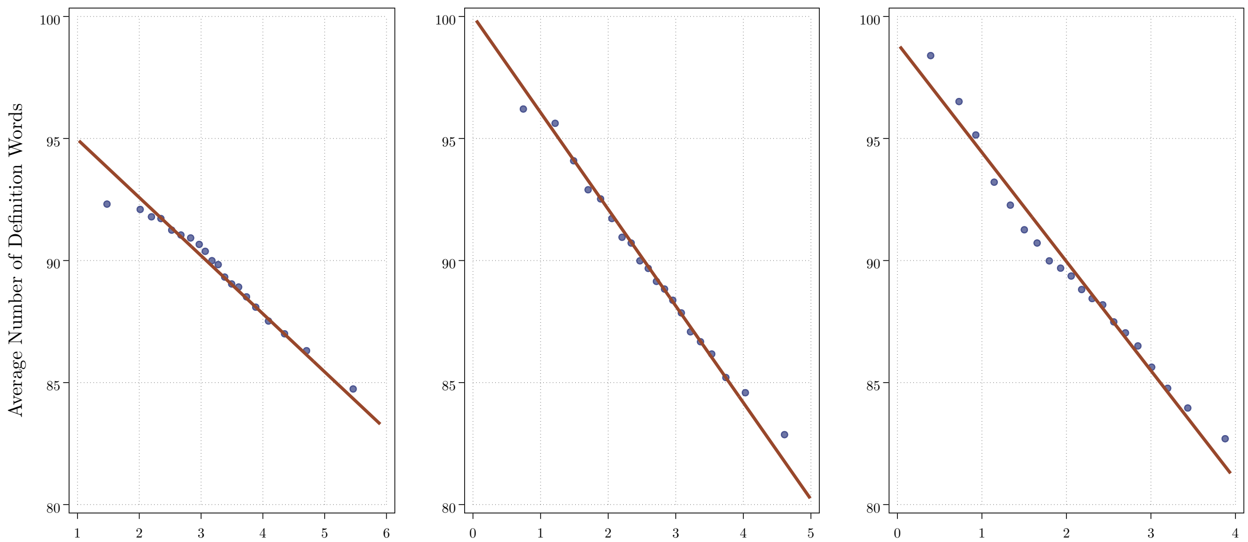
We first link the standardized titles that we observe in the Census to definitions of these jobs

and interpretation of generic titles further in Section 3.3.

[8] Either because of transcription errors or edits made during data cleaning, we find a very small number of titles in years where they should not exist (e.g., "aviator" in 1880).

[9] Shannon entropy is a central concept in information theory (Shannon, 1948) and has been applied in the economics literature to measure product variety (Straathof, 2007), linguistic fractionalization (Desmet et al., 2009), and surname diversity (Posch et al., 2025). As a benchmark, we calculate our three specialization measures using first and last names from the IPUMS 1% Census samples in place of job titles. We limit the sample to names that occurred at least five times in the data. Figure A2 plots trends in name "specialization" over time. For all three specialization metrics, aggregate occupational specialization is similar in scale to that of first names in the population.

Figure 2: Relationship Between Specialization Measures and Definition Length



(a) Log Job Titles **(b)** Log Variety Index **(c)** Entropy

Notes: This figure displays binned scatter plots of the average length of title definitions by county-industry-year cell on our three title-based specialization measures, using data from 1860 to 1940. We use data on the total number of words appearing in each occupation's definition from the first edition of the Dictionary of Occupational Titles (DOT). For each cell, we construct the average number of definition words per worker for the subset of workers reporting a title that we can link to a DOT definition. The fitted lines shown in these figures are equivalent to the regression $Words_{jct} = \alpha_{jc} + \beta Specialization_{jct} + \delta_{jdt} + \epsilon_{jct}$, which includes the same set of fixed effects as used in our main analysis. Here, α_{jc} is an industry-county fixed effect and δ_{jdt} is an industry-year fixed effect that varies across nine census divisions d. The sample is limited to county-industry-year cells with at least 50 workers who report a valid job title.

from the Dictionary of Occupational Titles (U.S. Department of Labor, Employment and Training Administration, 1939). The DOT provides textual descriptions of around 17,000 titles, though only a subset of these appear in the Census microdata, and conversely, not all of the standardized Census titles have an obvious DOT equivalent.[10] We then count the number of words appearing in the DOT's definition of each title. Because the DOT was specifically designed to provide detailed descriptions of occupations, longer definitions plausibly reflect a broader range of tasks than shorter ones. Finally, for each county, industry, and year cell, we compute the average length of title definitions among the set of workers whose title can be linked to the DOT.

Figure 2 displays binned scatter plots of the average number of definition words per job in county-industry-year cells on each of our occupational specialization measures. To produce these figures, we control for the same set of fixed effects included in our preferred county-industry regression specification so that the identifying variation is equivalent to that of our main analysis (see

[10]We focus on near-exact matches between the Census and the DOT, allowing for minor differences in word order and spacing. We also exclude "classification" and "term" titles from our analysis. These DOT categories include generic titles such as clerk, farmer, and laborer, which are not accompanied by detailed task descriptions but instead list only specific titles that may appear within these broad categories. Our DOT-based measures are therefore constructed using a subset of about 2,000 matched titles that are highly informative about task content but do not cover the universe of Census employment. This is the main reason we do not use the DOT as one of our primary specialization measures.

equation (5)). In all three cases, we find that the title-based specialization measures are negatively correlated with the average number of DOT words per job. That is, counties and industries with a greater variety of job titles also tend to have jobs that perform a narrower range of tasks, on average.[11] While we cannot link all Census titles to a DOT definition, as a validation exercise, this finding is consistent with the connection Smith draws between task content and job titles and lends support to our use of the latter as a proxy for the division of labor.

2.3 Additional Data

Extent of the market. We measure the extent of the market for counties and cities using three variables: total population, the urbanization rate, and domestic market access based on a gravity trade model from Donaldson and Hornbeck (2016).[12] Specifically, market access for county c at time t is defined as

$$Access_{ct} = \sum_{d} \tau_{cdt}^{-\theta} N_{dt}, \tag{3}$$

where τ_{cdt} is the minimum trade cost between origin county c and destination county d, θ is trade elasticity, and N_d is the total population of county d. We follow Donaldson and Hornbeck (2016) and use $\theta = 8.22$, which was derived from their structural model estimation. Market access can change in response to reductions in trade costs (primarily driven by the expansion of the U.S. railroad network) or to shifts in the geographic distribution of the population. We hold trade costs fixed at their 1920 levels when computing market access in 1930 and 1940, assuming that transportation networks remained relatively unchanged after 1920.

Although the three measures of market size we adopt are positively correlated, each captures somewhat different information about the extent of the market. County population reflects both local demand for goods and the thickness of the labor market. The trade-based market access measure captures the degree of domestic market integration. Finally, the urbanization rate is also important to control for, as it reflects the concentration of the population in urban areas, which is closely linked to agglomeration economies and opportunities for specialization. Urban areas, however, need not be large. Throughout our sample period, the Census defines most urban areas as cities or incorporated places with 2,500 or more inhabitants.[13]

Patents. We use the full universe of utility patents issued between 1836 and 1940 to measure technological innovation at the county and industry level. We begin by matching the 2.2 million

[11] In Table A4 we repeat this exercise using only verbs—which may better reflect actions, processes, and procedures—and when limiting our analysis to post-1910—when industry information is more reliable. In all cases, the results are qualitatively similar to those shown in Figure 2.

[12] We focus on the domestic market, as international trade accounted for a small fraction of total production during this period. For example, between 1930 and 1940, exports were less than 3.4% of total GDP.

[13] Although these variables are closely related, each provides independent identifying variation. The variance inflation factors for these three variables in our county panel ranges from 1.3 to 1.6, well below the levels at which multicollinearity becomes a concern. After residualizing these variables on county and year fixed effects—which is our main source of identifying variation in these variables—the VIFs range from 1.0 to 1.2.

patent records from Autor et al. (2024) with inventor location data from HistPat (Petralia et al., 2016). We allocate patents equally to inventors' counties when a patent lists more than one inventor. Although the patent records include standardized "Cooperative Patent Classification" (CPC) codes, they do not directly align with our historical industry classification (IND1950). To bridge this gap, we use the probabilistic crosswalk developed by Goldschlag et al. (2020), which maps CPC codes to the 2007 North American Industrial Classification System (NAICS). We then manually match NAICS codes to IND1950 codes by comparing industry titles and adjusting the probability weights accordingly. This procedure allows us to compute a probability-weighted count of patents for each industry over time.

Census of Manufactures. We use aggregated city-by-industry level data from the U.S. Census of Manufactures (CM) to measure firm size and manufacturing productivity (Lee, 2015). The data report the number of establishments, value added, and employment for each city-industry-year cell. To align the CM records with specialization measures derived from the population Census, we impose several sample restrictions. First, we use only decennial CM reports between 1910 and 1940, the period for which we have the most reliable industry data in the population Census. Second, while the original CM records include data for over 100 cities, both the set of cities covered by the survey and the minimum size of firms included in the city-industry aggregates vary from year to year. To ensure comparability over time, we restrict the sample to a balanced panel of city-industry pairs observed in all four census years.[14]

3 Trends in Occupational Specialization

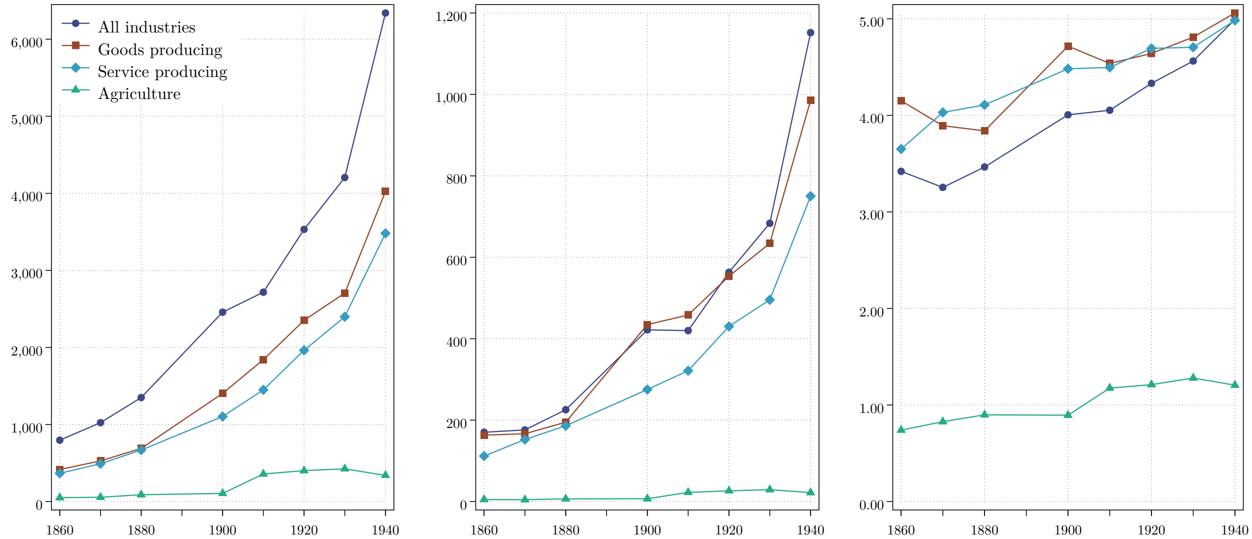
We begin by documenting nationwide trends in the number and variety of job titles workers reported to the Census, which reveals a dramatic and broad-based increase in occupational diversity during our sample period. We also explore how the growth and decline of specific titles reflects distinct sources of occupational specialization, including task reallocation, technological progress, and the emergence of large-scale industrial firms.

3.1 Aggregate Trends

Figure 3 shows that the variety of job titles appearing in Census records rose substantially over the 80 years covered by our data. Across all industries, the count of standardized titles rose from about 800 in 1860 to more than 6,000 by 1940. Moreover, the growth of specialization accelerated over time, with the largest increase between 1930 and 1940. Our variety and entropy measures—which account for relative employment shares across titles—show similar growth, affirming that this trend captures economically significant changes to the U.S. occupation structure and is not especially sensitive to the particular metric we choose.

[14] Our analysis sample includes 55 cities and 40 industries. The median city reports 7 industries.

Figure 3: Trends in Occupational Specialization, 1860–1940



(a) Job Titles **(b)** Variety Index **(c)** Entropy

Notes: To reduce measurement error, the sample used to construct each series is limited to job titles with at least 100 observations nationally in each year and industry category. Goods producing industries include construction, manufacturing, and mining. All other industries except agriculture are service producing.

We find a similar increase in occupational specialization across goods- and service-producing industries, indicating that the division of labor was not a development exclusive to the manufacturing sector. Indeed, the increasing number of job titles underscores how our measures of specialization are closely related to the emergence and diffusion of "new work" (Autor et al., 2024). As we discuss further below, these aggregate trends reflect a variety of factors including the reallocation of existing tasks across workers (what Smith documented in his pin factory example), new opportunities for specialization afforded by novel technologies and production processes, the need for specialized clerical and administrative workers to manage increasingly complex firms, as well as the introduction of new industries and products requiring different task inputs.

In contrast to the patterns we observe in other sectors, we find comparatively little growth in the number and variety of agricultural titles, consistent with Smith's claim that "[t]he nature of agriculture [...] does not admit of so many subdivisions of labour [...] as manufactures." Strikingly, Chilosi et al. (2025) find similar divergent trends between agriculture and other sectors in early modern England using data on job titles appearing in registered wills. In other words, Smith's hypothesis that the division of labor is more likely to emerge in manufacturing than agriculture holds true across both the United States and Britain from 1550 to 1940.

3.2 Specialization Within Industries

The increase in aggregate specialization is also apparent across major industry groups. As shown in Table 1, most sectors reported only a few dozen job titles in 1860, but even those with the least

Table 1: Number of Job Titles by Year and Industry Supersector

	1860	1870	1880	1900	1910	1920	1930	1940
Agriculture, forestry, fishing	51	56	89	106	358	402	425	340
Mining	9	27	39	75	227	272	316	346
Construction	55	85	89	160	388	419	556	790
Manufacturing	358	423	584	1,218	1,599	2,089	2,335	3,434
Transportation, communication, utilities	131	180	249	414	635	844	951	1,115
Wholesale and retail trade	66	100	137	224	542	701	816	1,106
Finance, insurance, real estate	17	31	46	68	166	225	318	385
Business and repair services	34	48	64	103	179	257	299	333
Personal services	60	67	88	164	391	451	495	498
Entertainment and recreation	7	11	22	46	106	142	218	297
Professional and related services	40	56	87	153	274	383	513	787
Public administration	46	68	97	135	254	358	425	699

Notes: To reduce measurement error, the sample used to construct this table is limited to job titles with at least 100 observations in each year and industry category.

growth in specialization had adopted several hundred titles by 1940.[15] This trend reflects a broad-based increase in occupational differentiation across sectors, not limited to industries undergoing major shifts in their output or product mix.

To illustrate these changes more concretely, Figure 4 depicts the evolution of job titles in three detailed industries: “apparel and accessories”, “iron and steel”, and “pulp, paper, and paperboard mills.” Each rectangle represents a distinct job title reported in that industry, and its size is proportional to the employment share of the title in a given census year. Consistent with the broader trends in manufacturing shown in Table 1, the number of unique titles, as represented by the number of rectangles in each figure, grew substantially from 1860 to 1940.

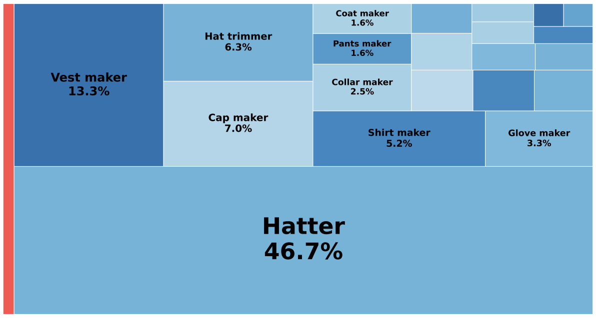
More importantly, the figure is informative about three distinct sources of occupational specialization. First, we find evidence that production was increasingly segmented into discrete tasks. For example, in 1860, titles like glove maker and shirt maker in the apparel industry suggest that workers were engaged in the entire production process. By 1940, workers were more likely to report titles such as cutter, finisher, presser, and sewer that recall a specific stage of the production process instead. Similar patterns also appear in the iron and paper industries, where titles shift from boilermaker to welder and from paper maker to sorter and finisher. Second, titles such as power-machine operator, sewing machine operator, and punch-press operator (though not labeled in the figure) imply the adoption of new technologies that demand specialized workers. Third, the growth of administrative jobs—both the number of distinct titles and their employment share—highlights how increasing specialization in the production process coincided with broader changes in the organizational structure of modern manufacturing.[16]

[15] While some industries—including agriculture, construction, and personal services—saw an unusually large jump in the number of titles between 1900 and 1910 that may be related to the Census introducing a separate industry question, most series show no clear discontinuity. We also find significant growth in specialization between 1910 and 1940, the period for which we have the most consistent industry data. Taken together, these patterns suggest that the trends we document are not primarily driven by changes in how occupation and industry information were recorded by the Census.

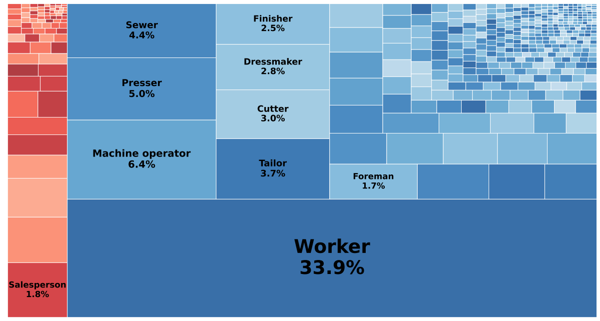
[16] Some of these patterns may partially reflect measurement issues related to industry reporting. As noted earlier,

Figure 4: Examples of Increasing Job Title Variety

(1) Apparel and Accessories

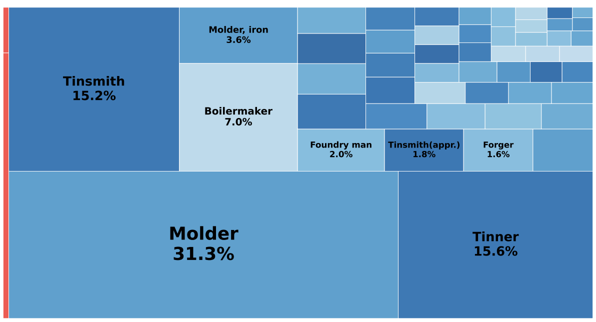


(a) 1880 (22 Titles)

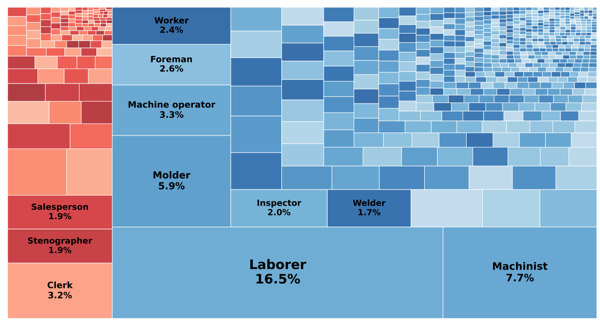


(b) 1940 (479 Titles)

(2) Iron and Steel

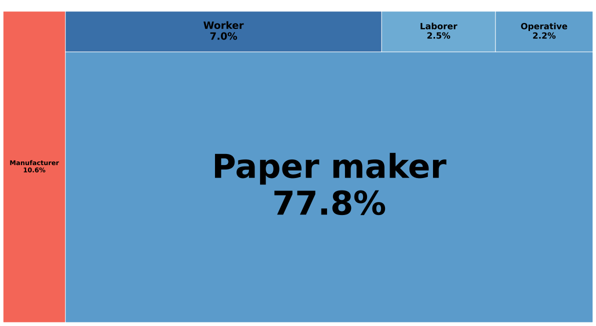


(c) 1880 (52 Titles)

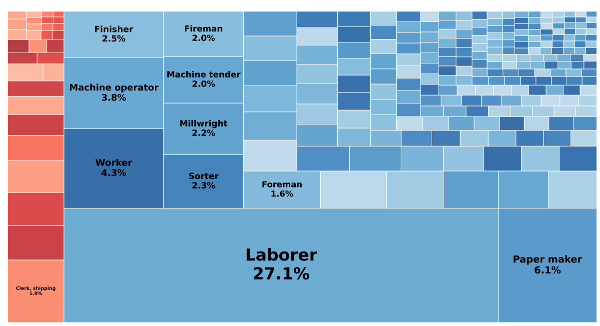


(d) 1940 (717 Titles)

(3) Pulp, Paper, and Paperboard Mills



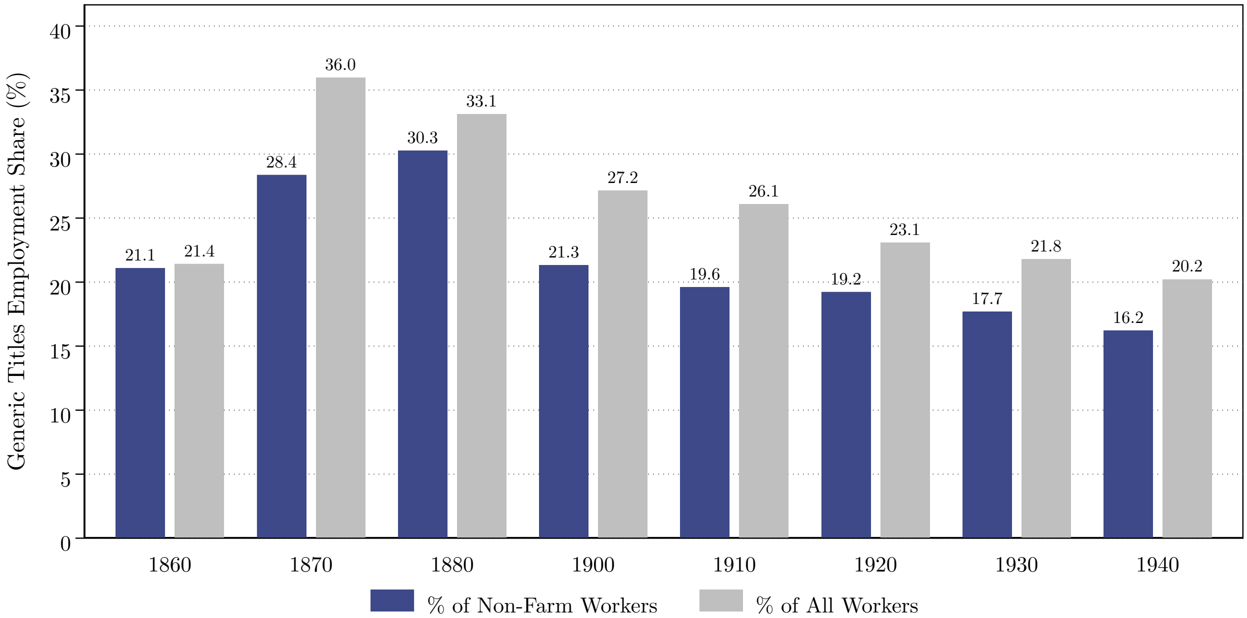
(e) 1880 (5 Titles)



(f) 1940 (256 Titles)

Notes: Sample includes industry-occupation cells with at least 50 observations in a given year. The size of each rectangle corresponds to the title's relative employment share. Administrative titles are colored in red hues, while non-administrative titles are colored in blue hues.

Figure 5: Employment Share of Generic Titles, 1860–1940



Notes: The figure displays the overall and non-farm employment share of the top 12 most frequent job titles other than farmer from 1860 to 1940 (laborer, worker, clerk, agent, keeper, hand, operator, buyer, apprentice, helper, owner, maker).

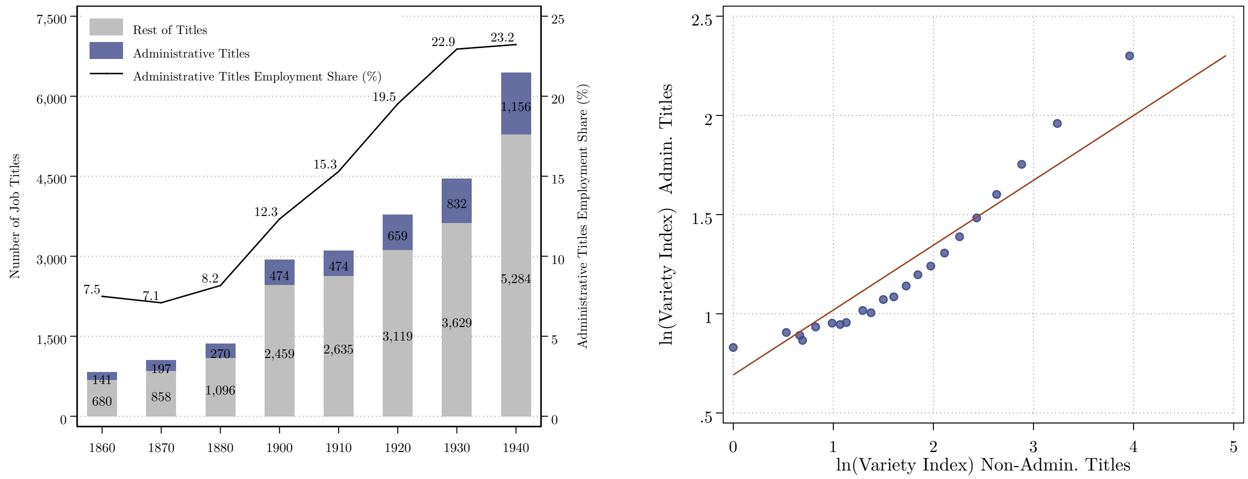
3.3 Generic Titles

The examples in Figure 4 show that, from 1860 to 1940, generic titles including worker, laborer, and clerk became more prevalent in certain industries. To understand whether titles that are relatively uninformative about the scope of workers' tasks became more prevalent economy-wide, we calculate the employment share of the 12 most common generic titles we find in the data across all industries (laborer, worker, clerk, agent, keeper, hand, operator, buyer, apprentice, helper, owner, and maker). We exclude the title "farmer" from this category to avoid conflating job title differentiation with the transition from an agricultural to an industrial economy, though we include farmers in the total employment denominator when computing shares.

Figure 5 shows that, across all industries, the share of workers reporting generic titles increased from 21% in 1860 to 36% in 1870. The initial increase coincided with the earlier phase of mechanization, when the U.S. economy transitioned from artisanal production to modern manufacturing. Worker and laborer replaced hatter and blacksmith. This temporary increase was followed by a steady decline in the employment share of generic titles between 1870 and 1940. This is also evident when restricting attention to non-farm employment only, where the generic title employment share fell from a peak of 30% in 1880 to just 16% by 1940. This finding suggests that, as the

the Census only began collecting industry information systematically in 1910, and prior to that year, industry affiliations were imputed based on job titles and other available information. Workers with titles closely tied to specific industries—such as glove maker or hatter—could be readily be classified. However, generic titles like worker or operator are less obviously linked to any single industry.

Figure 6: Administrative and Non-Administrative Titles



(a) Trends in Administrative Titles

(b) Admin vs. Non-Admin Titles

Notes: Panel (a) displays the number of administrative versus non-administrative job titles that occur at least 100 times between 1860 and 1940. The solid line represents the employment share of administrative job titles. Panel (b) displays binned scatter plots of the log variety index of administrative versus non-administrative titles at the county-industry level. The fitted line represents the regression coefficient β in the regression $ln(Variety)^{admin}_{jct} = \beta ln(Variety)^{non-admin}_{jct} + Employment_{jct} + PopDens_{ct} + \alpha_{jc} + \delta_{jdt} + \epsilon_{jct}$. This regression controls for industry size and population density and includes county-by-industry (α_{jc}) and industry-by-division-by-year (δ_{jdt}) fixed effects.

modern production processes developed, workers were increasingly assigned—and came to identify with—increasingly differentiated and descriptive job titles, consistent with increasing occupational specialization and greater division of labor. Importantly, the decline in generic titles predates the introduction of industry codes in 1910, so this pattern is not simply an artifact of how the census collected occupation and industry information.

3.4 Administrative Titles

While Smith clearly recognized the importance of task reallocation and technological change, he primarily focused on the production process itself. However, as noted by Chandler (1984), in the late 19th- and early 20th-century U.S., industrial firms with large production scale and expansive distribution networks employed teams of administrators for "close and constant coordination and supervision of material flows." We observe this shift in the industry case studies shown in Figure 4, where we see the growth of titles such as manager, foreman, supervisor, and inspector that specialize in administrative tasks. This was accompanied by an increase in clerical work performed by bookkeepers, salespeople, shipping clerks, and timekeepers who managed the complex transactions conducted by large industrial firms.

In aggregate, administrative titles—including managerial, clerical, and sales jobs—continued to grow across industries over time. Panel (a) of Figure 6 plots the number of administrative job titles and their corresponding employment share from 1860 to 1940. While administrative titles accounted for only 7.5% of employment in 1860, by 1940, more than one in five workers were employed in an administrative job. In addition, the growth of administrative titles coincided with the growth of non-administrative titles. Panel (b) presents a binned scatter plot of job title variety for administrative versus non-administrative occupations at the county-industry level. The positive correlation indicates that county-industry cells with a higher level of occupational specialization in the production process also had a more diverse set of administrative titles.

Thus, compared to Smith's observations in 18th-century England, the growth of a more specialized workforce in the turn of the century U.S. reflected a more fundamental change in the economy. As continuous-mass production replaced artisanal workshops, firms developed a dedicated administrative staff to manage coordination tasks separately from production tasks. The expansion of administrative titles can therefore be considered another form of division of labor, though one that was less relevant in Smith's context than it was for later industrialization.

4 Determinants of Occupational Specialization

We next examine the determinants of occupational specialization across locations and industries, focusing on two channels proposed by Smith: market size and technology. Consistent with his theory, we find robust empirical evidence that size—predominantly, but not exclusively, local population and total employment in the industry—leads to greater occupational specialization across counties and cities. We also find that workers in more innovative industries and cities—as measured by patenting activity—are more specialized relative to other workers who were less directly exposed to the technological frontier.

4.1 The Extent of the Market

In the third chapter of *The Wealth of Nations*, Smith asserts that the division of labor is limited by the extent of the market: "When the market is very small, no person can have any encouragement to dedicate himself entirely to one employment, for want of the power to exchange all that surplus part of the produce of his own labour [...]." He then uses concrete examples to illustrate how the size of the market shapes the extent of task specialization, a central theme in his theory of the division of labor. In the case of services, such as porters, specialization is limited by the size of the local population; non-tradable activities require a dense enough market to support constant employment. As he notes, "even an ordinary market-town is scarce large enough to afford him [the porter] constant occupation."

In contrast, in the goods-producing sector—such as nail-making—where output can be traded beyond the immediate locality, Smith recognizes the importance of transportation technology in expanding market access: "as by means of water-carriage, a more extensive market is opened to every sort of industry than what land-carriage alone can afford it." In both cases, Smith underscores that the extent of the market—whether limited by population size or enabled by infrastructure and transportation technology—is an important determinant of the scope of specialization in the labor market.

Following Smith, we analyze how relevant the extent of the market was for the growth and geography of occupational specialization in the United States. In many ways, our setting provides an ideal context for evaluating Smith's theory. First, between 1860 and 1940, the U.S. population grew from 30 million to 130 million, largely due to immigration. At the same time, the share of the population living in urban areas rose from 20% to nearly 60%, substantially increasing local demand. Finally, the expansion of the railroad network and improvements in transportation technology significantly enhanced domestic market access to trade, particularly for firms operating in the Northeast and Midwest. In this section, we show that all three of these factors are strongly associated with the growth of the division of labor across cities and counties. We also find a robust link with local industry size and the average number of employees per establishment—an important margin of firm-level scale that enables a finer internal division of labor and supports the emergence of more specialized occupational roles.

County-level evidence. We begin with cross-sectional comparisons. Figure 7 displays binned scatter plots of the log number of titles present in a county against three measures of market size. To construct these figures, we control for year-by-census-division fixed effects to allow for heterogeneous trends across geographic regions (e.g., due to agricultural suitability or westward expansion).[17] The fitted regression lines can therefore be interpreted as the average cross-sectional relationship between the division of labor and the extent of the market across counties within the same census division. In all three panels, we find that counties with larger markets exhibit substantially higher job title variety, which we interpret as consistent with Smith's hypothesis that broader markets foster a greater division of labor. For example, Cook County, Illinois, which includes Chicago, the largest city in the Midwest, recorded nearly three times as many job titles in 1940 as Dane County, Wisconsin, which includes the smaller Midwestern city of Madison.

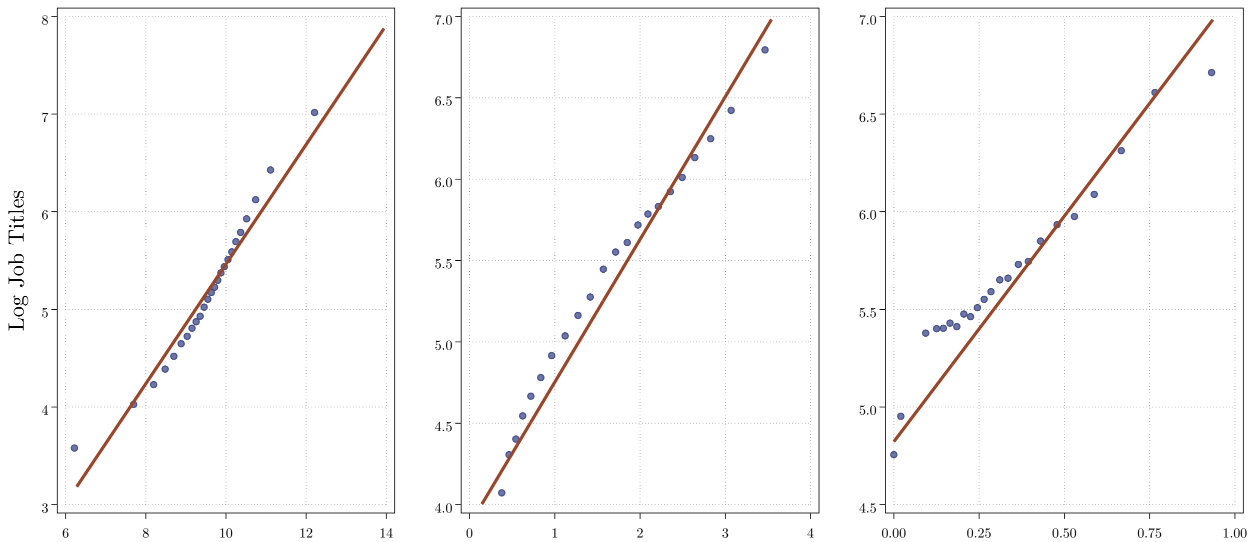
Moving beyond cross-sectional comparisons, the panel structure of our data allows us to examine how occupational specialization evolves within counties over time. We estimate regressions of the form

$$Y_{ct} = \alpha_c + \mathbf{X}_{ct}\Gamma + \delta_{dt} + \epsilon_{ct}, \tag{4}$$

where Y_{ct} is a specialization measure for county c at time t and $\mathbf{X}_{ct}$ includes our three market size measures: log county population, market access (which we normalize by its sample standard deviation), and the urban population share. We control for county fixed effects α_c and census-division-by-year fixed effects δ_{dt} in our preferred specification. Including county fixed effects in this regression nets out any time-invariant county characteristics (such as proximity to a major port or other natural endowments) that may be correlated with market size and the division of labor. This further implies that our estimates are identified from changes in market size within counties over time.

[17] Controlling for regional heterogeneity only materially affects specifications that include our market access measure, as discussed further below.

Figure 7: Cross-Sectional Relationship Between Specialization and Market Size



(a) Log Population **(b)** Market Access **(c)** Urban Share

Notes: This figure uses county-level data and displays binned scatter plots of the log number of distinct job titles present in each county on three measures of the extent of the market (log population, market access, and urban share, respectively). Market access is normalized by its sample standard deviation. The fitted regression lines are equivalent to the regression $\ln(Titles_{ct}) = \delta_{dt} + \beta X_{ct} + \epsilon_{ct}$, where X_{ct} is the market size measure and δ_{dt} is a census-division-by-year fixed effect.

To disentangle the independent effects of different aspects of market size, we jointly include population, market access, and urbanization in a single regression specification. This approach accounts for the positive correlations among these variables, which may confound their individual effects when analyzed separately. Table 2 shows that all three variables are independently and positively associated with specialization, though local population growth has the largest effect. Column (1), which does not control for regional heterogeneity, shows that a 1% increase in population is associated with a 0.5% increase in the number of titles reported in the county, holding all other variables constant. Put differently, a one standard deviation increase in log population is associated with an 80% increase in job titles. A one standard deviation increase in domestic market access has a smaller but still economically meaningful effect on specialization, increasing the number of titles reported by 16%. Finally, a 10 percentage point increase in the urban population share (conditional on total population) increases the number of titles by only 2%.

Columns (3) and (5) show similar effects of market size on the job title variety and entropy measures. While population remains the strongest predictor of specialization, market access and urbanization are somewhat more influential when using measures that account for the distribution of employment across titles than they are for the count of job titles. One possible explanation is that these measures capture not only the presence of certain job titles but also how evenly employment is spread across them—dimensions of specialization that may be more sensitive to the scale and density of product markets and urban environments.[18] Finally, in even-numbered columns of the

[18]For example, the entropy measure places relatively more weight on titles with a small employment share. Since

Table 2: Determinants of Occupational Specialization, County Level 1860–1940

	Log Job Titles		Log Variety Index		Entropy	
	(1)	(2)	(3)	(4)	(5)	(6)
Log total population	0.515***	0.508***	0.347***	0.332***	0.160***	0.149***
	(0.011)	(0.011)	(0.010)	(0.010)	(0.009)	(0.010)
Market access	0.160***	0.317***	0.204***	0.415***	0.163***	0.392***
	(0.017)	(0.029)	(0.018)	(0.029)	(0.019)	(0.030)
Urban population share	0.212***	0.205***	0.534***	0.480***	1.017***	0.895***
	(0.043)	(0.045)	(0.043)	(0.044)	(0.047)	(0.047)
County-year observations	21,886	21,886	21,886	21,886	21,886	21,886
R-squared (within)	0.520	0.492	0.331	0.315	0.179	0.169
County and year FE	✓	✓	✓	✓	✓	✓
Division × year FE		✓		✓		✓

Notes: Market access is normalized by its sample standard deviation. Standard errors are clustered at the county level, and significance levels are indicated by *** 1%, ** 5%, and * 10%.

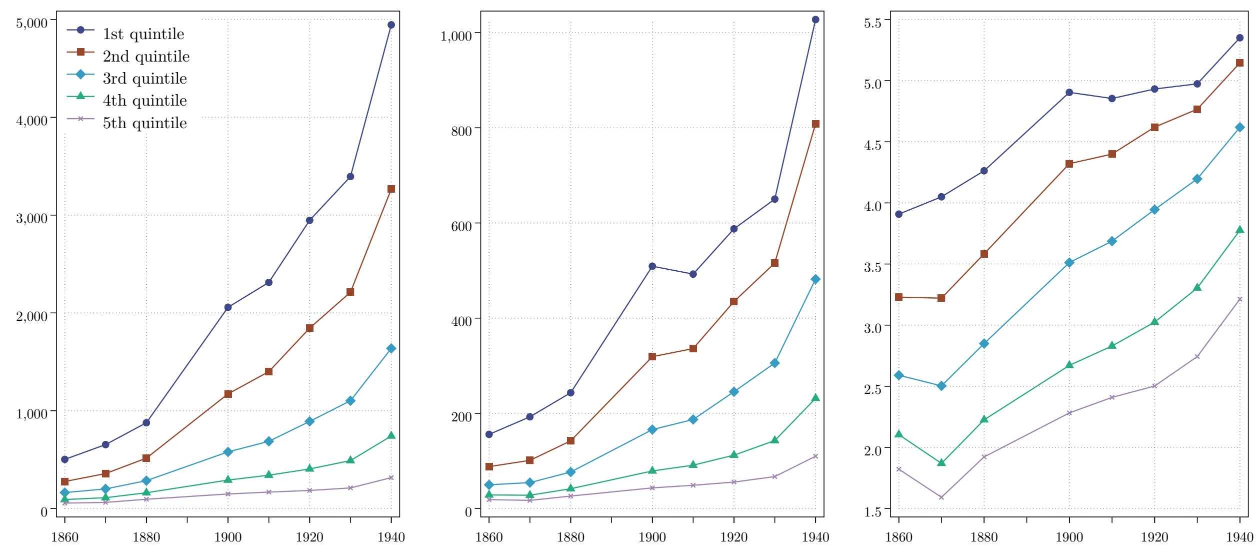
table, we interact the year fixed effects with census division indicators, which limits comparisons to counties within the same region (our preferred specification). Accounting for regional heterogeneity has little effect on our point estimates for population and urbanization, but the coefficients on market access roughly double in this specification. This is likely because growth in market access is highly spatially correlated, with larger changes occurring in the Midwest and South than in the West, as shown in Figure A1.

The finding that county population emerges as a key determinant of occupational specialization has important implications for geographic heterogeneity in the trends documented in Section 3. To explore this, we partition counties into quintiles by their 1940 population, with each quintile containing approximately 20% of the U.S. population in that year. We then plot the average of our specialization measures across counties in each population quintile in Figure 8. The figures show that the growth in aggregate specialization observed between 1860 and 1940 is driven mostly by more populous counties. The number and variety of titles we find in the bottom quintile of counties barely grew at all, and the smallest counties had fewer titles on average in 1940 than the largest counties did in 1860. This divergence highlights how large areas increasingly pulled ahead in terms of occupational complexity, contributing to a widening gap in the extent of specialization across local labor markets.

Within-industry evidence. One potential concern is that the relationship between our county-level specialization measures and market size may reflect compositional effects. Specifically, more populated counties or those with greater market access may attract a larger variety of industries that require different task inputs, and thus different job titles. Similarly, at least some of the growth in the number and variety of job titles we observe over time is related to the emergence

these occupations are likely to be niche services in urban markets, the urban population share appears to have a greater effect when we measure specialization using entropy.

Figure 8: Trends in Occupational Specialization by Population Quintile



(a) Job Titles **(b)** Variety Index **(c)** Entropy

Notes: Counties are partitioned so that each quintile contains approximately 20% of the total U.S. population in 1940. Quintiles are sorted from largest (1st) to smallest (5th), with specialization measures averaged across counties within each quintile.

of entirely new industries and products (e.g., aircraft, automobiles, and electrical appliances). To control for differences in industry composition, we compare changes in occupational specialization within county by three-digit industry cells (the finest level of industry detail consistently available in the data). Below, we additionally show that our findings are robust to controlling for an even more granular measure of product variety constructed from Census industry strings.

We use a county-by-industry panel to estimate a modified version of equation (4):

$$Y_{jct} = \alpha_{jc} + \mathbf{X}_{ct}\Gamma + \delta_{jdt} + \epsilon_{jct}. \tag{5}$$

Here, Y_{jct} is a specialization metric, now defined for county c, industry j, at time t. As before, $\mathbf{X}_{ct}$ includes the three market size measures. The fixed effect α_{jc} implies that our estimates are identified from changes within county-industry cells, and δ_{jdt} controls for industry-specific trends across census divisions. This specification accounts for any differences in three-digit industry composition across counties, as all comparisons are restricted to the same industry. For the same reason, it also absorbs changes in job title variety explained by emergence of new industries like motor vehicle manufacturing.

Table 3 shows the results. Column (1) reports a specification analogous to column (2) of Table 2. Comparing the number of titles within county-industry over time, we continue to find a strong relationship between the extent of the market and the number of titles. Notably, the point estimates on county population are only slightly attenuated relative to those at the county level, suggesting that composition effects are not the primary driver of our findings. We find similar results using variety and entropy measures, where the coefficients on local population are

Table 3: **Occupational Specialization, County-Industry Level 1860–1940**

	Log Job Titles		Log Variety Index		Entropy	
	(1)	(2)	(3)	(4)	(5)	(6)
Log total population	0.440***	0.165***	0.310***	0.188***	0.176***	0.165***
	(0.017)	(0.009)	(0.013)	(0.010)	(0.008)	(0.009)
Market access	0.165***	0.142***	0.140***	0.130***	0.057***	0.057***
	(0.024)	(0.017)	(0.021)	(0.019)	(0.017)	(0.016)
Urban population share	0.231***	0.176***	0.238***	0.214***	0.161***	0.159***
	(0.044)	(0.029)	(0.038)	(0.032)	(0.025)	(0.025)
Log industry employment		0.410***		0.181***		0.016***
		(0.004)		(0.004)		(0.004)
County-industry-year obs.	290,040	290,040	290,040	290,040	290,040	290,040
R-squared (within)	0.193	0.427	0.123	0.178	0.042	0.043
County × industry FE	✓	✓	✓	✓	✓	✓
Industry × division × year FE	✓	✓	✓	✓	✓	✓

Notes: The sample includes only county-industry cells with at least 50 workers who report valid job titles. Market access is normalized by its sample standard deviation. Standard errors are clustered at the county level, and significance levels are indicated by *** 1%, ** 5%, and * 10%.

also comparable to the county aggregate estimates. Adjusting for industry composition attenuates the market access and urbanization coefficients somewhat more, though these also remain strongly correlated with occupational specialization.

We proceed by examining whether the total number of employees in an industry is related to its division of labor. In the even-numbered columns, we include the log of total industry employment in the county as an additional explanatory variable. In these specifications, the effect of county population on the number of titles and the variety index become noticeably smaller. Specifically, a 1% increase in county population is associated with a 0.17% increase in the number of titles, instead of 0.44% without controlling for industry size. Controlling for industry size does not materially change the effect of market access or urbanization. Meanwhile, holding market size constant, a 1% increase in industry employment is associated with a 0.41% increase in the number of titles. These results suggest one potential mechanism for the influence of population: more populous counties are more likely to have larger industries, which in turn exhibit greater occupational specialization, possibly reflecting economies of scale in production.

Evidence on firm size. Our industry case studies in Figure 4 suggest that many specialized job titles are related to either mechanization (machine operator) or the hierarchical organization of the workforce (foreman, sales manager), both of which are more likely to occur in larger firms. Given the results above, we examine whether larger industries have higher degrees of specialization because they have larger firms. Although we lack firm-level microdata, our city-industry data from the CM allows us to compute average firm size across cities by dividing the total number of employees in an industry by the number of establishments. As discussed in Section 2.3, the CM data are a selected sample of large cities. Therefore, the results in this section are informative

Table 4: Occupational Specialization, City-Industry Level 1910–1940

	Log Job Titles		Log Variety Index		Entropy	
	(1)	(2)	(3)	(4)	(5)	(6)
Log total population	0.566***	0.563***	0.449***	0.443***	0.379***	0.372***
	(0.052)	(0.056)	(0.042)	(0.042)	(0.089)	(0.086)
Market access	-0.066	-0.044	-0.018	0.021	0.035	0.091
	(0.048)	(0.049)	(0.074)	(0.060)	(0.115)	(0.093)
Log average firm employment	0.035***		0.022*		0.006	
	(0.010)		(0.011)		(0.018)	
Log total employment		0.037***		0.024**		0.009
		(0.009)		(0.010)		(0.016)
Log number of establishments		-0.022*		0.001		0.027
		(0.011)		(0.014)		(0.022)
City-industry-year obs.	1,916	1,916	1,916	1,916	1,916	1,916
R-squared (within)	0.281	0.284	0.195	0.206	0.077	0.090
City × industry FE	✓	✓	✓	✓	✓	✓
Industry × year FE	✓	✓	✓	✓	✓	✓

Notes: These regressions are weighted by the total number of individuals observed in the population census for each industry-city-year cell. Market access is normalized by its sample standard deviation. Log industry employment and establishments measure the number of workers and establishments at the city-industry level, respectively. Standard errors are clustered at the city level, and significance levels are indicated by *** 1%, ** 5%, and * 10%.

about large urban areas but are not necessarily generalizable to the average city or county.

It is worth noting that our analyses rely on two distinct sources of employment data. For the county-industry analysis presented in Table 3, we use industry classifications reported in the population census. In contrast, for the city-industry analysis below, we use employment counts reported by firms in the CM. Beyond providing data on the number of firms, the CM offers an additional advantage: industry employment counts from firms are likely more accurate than self-reported affiliations from individuals. For instance, a truck driver employed by a meatpacking plant might self-identify as working in "trucking service" in the population census, whereas in the CM data, this same worker would be correctly classified as part of the "meat products" industry.

We estimate regressions equivalent to equation (5), using city-industry as the unit of analysis.[19] Table 4 shows the results, with columns (1), (3), and (5) presenting the main estimates from this specification, replacing total industry employment with mean employment per establishment. After controlling for mean firm size, we find that local population size remains the strongest and most important determinant of occupational specialization. In contrast, market access becomes statistically insignificant, likely due to the limited variation in access among the large cities in the sample post-1910 when railroad networks were already established. A 1% increase in average firm size is associated with a 0.04% increase in the number of titles, or, equivalently, a one standard

[19] In specifications combining our specialization measures with data from the CM, we weight observations by the total number of individuals observed in the population census for each industry-city-year cell. Given the selected sample and discrepancies in industry employment counts between the two surveys, we place greater weight on larger cities and industries, where estimates are likely to be measured with less error. The unweighted results are qualitatively similar; however, the coefficients on market access become statistically significant when weights are excluded.

deviation increase in log mean firm size corresponds to a 5% increase in the number of titles within the industry. This finding is consistent with our prior that firms with a larger workforce tend to develop more complex organizational structures and thus higher degrees of specialization, a finding also documented for modern Brazil by Tian (2021).

Columns (2), (4), and (6) include total industry employment and the number of establishments as separate explanatory variables. As before, we find that aggregate industry size matters, as total employment is positively associated with occupational specialization. More importantly, holding total industry employment constant, an increase in the number of firms is associated with less specialization. In other words, industries become more specialized when their growth is driven by larger firms expanding their workforce rather than by a proliferation of smaller establishments.

4.2 Robustness Checks

Controlling for product variety. As noted above, our county-industry specification implicitly accounts for changes in occupational specialization driven by the emergence of new 3-digit industries. Even within industries, however, increasing product variety could explain at least some of the residual variation in job titles. Similarly, the positive correlation between specialization and market size might be driven by larger markets and industries producing a wider range of goods.

To further separate the division of labor from product innovation, we construct a measure of product variety growth in manufacturing industries using the INDSTR variable from the restricted-use IPUMS data between 1910 and 1930.[20] We first clean the raw industry text using the same procedure that we apply to occupation strings. Next, we strip words such as "factory," "mill," and "works" from the cleaned industry strings to isolate—as best we can—product names embedded within workers' self-reported industry. Finally, we compute a Dixit-Stiglitz variety index over the cleaned product categories. Although our data is not rich enough to provide a comprehensive catalog of all goods, this measure allows us to proxy for relative differences in product variety within and across industries at a level much more granular than 3-digit industry codes.[21]

Table 5 shows that our main market access results are robust to controlling for product variety. Columns (1), (3), and (5) reproduce our preferred specification for the manufacturing industry from 1910 to 1930. The results are quantitatively similar to those from our full sample, although industry employment matters somewhat more relative local population and, as in Table 4, we find no effect of market access on occupational specialization in the early 20th-century. Consistent with different products requiring different task inputs, columns (2), (4), and (6) show that greater product variety within an industry is associated with a larger number and variety of job titles. Importantly, however, this additional control has little effect on the estimated relationship between

[20] Digitized industry strings are not yet available for the full workforce in 1940. We restrict attention to manufacturing industries for this analysis, as industry strings are less informative about changes in the variety of services.

[21] Within the apparel and accessories industry, for example, our approach allows us to differentiate between products such as hats, pants, shirts, and neckties. However, we note that, as with occupations, workers also report general industry terms such as "clothing." Although industry strings are usually not specific enough differentiate between even finer product categories like suit pants and overalls, occupational task content is also likely to differ less when comparing increasingly similar goods.

Table 5: Occupational Specialization, Controlling for Product Variety
County-Industry Level, Manufacturing Only, 1910–1930

	Log Job Titles		Log Variety Index		Entropy	
	(1)	(2)	(3)	(4)	(5)	(6)
Log total population	0.071***	0.054***	0.120***	0.097***	0.184***	0.159***
	(0.014)	(0.014)	(0.020)	(0.020)	(0.027)	(0.026)
Market access	-0.044	-0.041	-0.046	-0.048	-0.017	-0.025
	(0.037)	(0.035)	(0.052)	(0.051)	(0.067)	(0.067)
Urban population share	0.226***	0.215***	0.324***	0.304***	0.355***	0.324***
	(0.051)	(0.049)	(0.072)	(0.070)	(0.092)	(0.089)
Log industry employment	0.507***	0.509***	0.286***	0.289***	0.044***	0.049***
	(0.004)	(0.004)	(0.006)	(0.006)	(0.008)	(0.008)
Log product variety		0.080***		0.104***		0.113***
		(0.006)		(0.008)		(0.010)
County-industry-year obs.	34,037	33,324	34,037	33,324	34,037	33,324
R-squared (within)	0.566	0.574	0.187	0.200	0.018	0.027
County × industry FE	✓	✓	✓	✓	✓	✓
Industry × division × year FE	✓	✓	✓	✓	✓	✓

Notes: The sample includes only county-industry cells with at least 50 workers who report valid job titles. Market access is normalized by its sample standard deviation. Standard errors are clustered at the county level, and significance levels are indicated by *** 1%, ** 5%, and * 10%.

occupational specialization, the extent of the market, and industry employment. This finding is more consistent with our interpretation of the latter as reflecting the reorganization of production tasks than it is with a shift toward greater product variety within industries.

Instrumenting for population. The strong relationship observed between population size and the division of labor also raises concerns regarding reverse causality: greater specialization could drive up demand for labor by enabling more efficient production, potentially attracting a larger population through higher wage incentives or improved job opportunities. To address potential reverse causality between population size and occupational specialization, we instrument for county population using an immigration-based shift-share design widely used in the labor and urban economics literature (e.g., Card, 2001; Saiz, 2007). This instrument predicts the log of the foreign-born population in each county for each year from 1860 to 1940, based on the 1850 distribution of immigrants by country (region) of origin and the subsequent national inflows of immigrants from each country between 1860 and 1940.[22] Formally, we construct the predicted foreign-born population for each county c in each year t as

$$\widehat{FBpop}_{ct} = \sum_i \omega_{ic1850} Immig_{i-ct}, \tag{6}$$

[22] We use 18 mutually exclusive countries/regions to construct the instrument: Asia, Central Europe, Eastern Europe, Greece, Italy, Portugal, Russia, Spain, Germany, Ireland, Scandinavia, the United Kingdom, Western Europe, Canada, Caribbean, Latin America, Mexico, and the rest of the world.

Table 6: Occupational Specialization with Immigration Instrument, 1860–1940

	Log Job Titles		Log Variety Index		Entropy	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Log total population	0.167***	0.211***	0.191***	0.242***	0.166***	0.224***
	(0.010)	(0.014)	(0.011)	(0.016)	(0.009)	(0.015)
Market access	0.146***	0.141***	0.131***	0.125***	0.055***	0.049***
	(0.017)	(0.017)	(0.020)	(0.020)	(0.017)	(0.018)
Urban population share	0.169***	0.117***	0.206***	0.145***	0.154***	0.085***
	(0.030)	(0.034)	(0.033)	(0.039)	(0.025)	(0.032)
Log industry employment	0.407***	0.394***	0.178***	0.163***	0.014***	-0.002
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
County-industry-year obs.	274,166	274,166	274,166	274,166	274,166	274,166
R-squared (within)	0.424	0.423	0.176	0.175	0.042	0.040
Cragg-Donald F		38529.6		38529.6		38529.6
Kleibergen-Paap F		290.7		290.7		290.7
County × industry FE	✓	✓	✓	✓	✓	✓
Industry × division × year FE	✓	✓	✓	✓	✓	✓

Notes: Columns (1), (3), and (5) present OLS regression estimates, where the outcome variables are the three measures of labor specialization. Columns (2), (4), and (6) report IV estimates using the log of predicted foreign-born population as an instrument for log total county population. The instrument is constructed by interacting the 1850 distribution of foreign-born residents across counties with decadal national inflows of immigrants from each country of origin, excluding those who settled in the county in question. The sample is restricted to county-industry cells with at least 50 observations reporting valid occupation titles. Market access is normalized by its sample standard deviation. Standard errors are clustered at the county level, and significance levels are indicated by *** 1%, ** 5%, and * 10%.

where ω_{ic1850} represents the share of all immigrants from region i who were residing in county c in 1850, and $Immig_{i-ct}$ denotes the total number of immigrants arriving in the United States from region i between years $t-10$ and t, excluding those who ultimately settled in county c.

The exclusion restriction for this instrument follows the standard logic of the shift-share design. The regional distribution of immigrants in 1850 was likely shaped by transportation access, networks, and agricultural opportunities and not by expectations about industrial structure or occupational complexity that emerged later. Moreover, national inflows of immigrants from different origins are also plausibly exogenous to subsequent county-level specialization. Finally, by excluding immigrants who settled in county c, we reduce mechanical correlations with local outcomes, further limiting concerns that the instrument reflects endogenous population growth.[23]

Table 6 compares the ordinary least squares (OLS) results with the corresponding IV estimates. First, the instrument appears strong, as evidenced by high Cragg–Donald and Kleibergen–Paap F-statistics. Second, the IV estimates for log total population are statistically significant. Notably, the IV coefficients are a bit larger than the OLS estimates—an unexpected result given that one may expect unobserved positive demand shocks to drive both higher population growth and greater specialization. One possible explanation is treatment effect heterogeneity, as the IV approach

[23] Our identification argument is the same as in the existing shift-share literature (e.g., Card, 2001; Saiz, 2007). It provides plausibly exogenous variation in county population growth. We do not claim a novel exclusion argument relative to the earlier literature but rather adapt this well-established approach to our historical setting.

Table 7: Occupational Specialization Robustness and Heterogeneity, 1860–1940

	Pre-1910	Post-1910	Goods	Services
Outcome: Log Variety Index	(1)	(2)	(3)	(4)
Log total population	0.127***	0.138***	0.148***	0.227***
	(0.015)	(0.011)	(0.010)	(0.015)
Market access	0.122***	-0.046*	0.159***	0.103***
	(0.022)	(0.026)	(0.021)	(0.021)
Urban population share	0.120***	0.244***	0.392***	0.073**
	(0.032)	(0.028)	(0.037)	(0.033)
Log industry size	0.035***	0.255***	0.186***	0.173***
	(0.006)	(0.004)	(0.004)	(0.007)
County-year observations	66,668	207,376	113,264	176,776
R-squared (within)	0.047	0.157	0.171	0.189
County × industry FE	✓	✓	✓	✓
Industry × division × year FE	✓	✓	✓	✓

Notes: The sample includes only county-industry cells with at least 50 observations with valid occupation titles. Market access is normalized by its sample standard deviation. Standard errors are clustered at the county level, and significance levels are indicated by *** 1%, ** 5%, and * 10%.

identifies a local average treatment effect for counties affected by the instrument. However, the IV and OLS estimates are quite similar in magnitude, and we cannot reject the null hypothesis that they are equal.

Additional robustness checks. Another potential concern is that because the census did not collect industry information before 1910, the imputation of industry for earlier years may introduce non-random measurement error. For example, certain job titles such as glass blower and tinsmith can be easily matched to unique industries, while others such as worker and salesperson do not have obvious industry affiliations. This may lead us to under count specialization in certain industries before 1910. Columns (1) and (2) of Table 7 report the estimation results for equation (5) using data before and after 1910 separately. Despite the imputation, the estimated coefficients for the two periods are largely comparable to the pooled sample. One notable difference, however, is that the effect of domestic market access is stronger for the pre-1910 period and turns weakly negative afterward. As noted above, this is likely due to the fact that railroad networks expanded dramatically before the turn of the century but construction slowed thereafter.

Another important consideration is that while Smith's hypothesis regarding market access is most directly applicable to tradable goods, Figure 3 shows that service industries also exhibited considerable increases in specialization. Columns (3) and (4) of Table 7 report the estimation results separately for goods- and service-producing industries. As expected, local market size, measured by total population, has a stronger association with specialization in the service sector, whereas domestic market trade access is more influential for goods-producing industries. Nevertheless, higher market access is still associated with higher levels of specialization even for the service sector. This may reflect complementarities across industries: while service providers like advertisers or lawyers cannot ship their products on the railroad, they might support manufacturing firms that serve broader markets. These patterns align with the literature on urban agglomeration, which emphasizes inter-industry linkages, knowledge spillovers, and input-output relationships as key drivers of specialization and productivity in dense or highly connected locations (e.g., Ellison et al., 2010; Duranton and Puga, 2004; Moretti, 2010).

4.3 Specialization and Technological Innovation

Smith also discusses the relationship between technological innovation and the division of labor. While he did not explicitly claim that innovation causes occupational specialization, he does note that firms in which labor is most subdivided also appeared to use a greater variety of tools and machinery created by "the ingenuity of the makers." Building on this observation, we examine whether workers in more innovative markets tend to be more specialized and whether this relationship is plausibly causal.

We conceptualize two distinct channels through which innovation can shape the division of labor. The first channel operates through technological progress within industries. Industry-specific innovations—such as new machinery or production techniques—can fragment production processes into more specialized tasks, requiring distinct occupational roles. We capture this channel using the volume of patenting activity at the industry level, which reflects the rate of technological advancement directly affecting production.[24] The second channel arises from local spillovers and agglomeration effects. Innovations clustered in certain counties or cities can diffuse across sectors, fostering complementary specializations even outside the original innovating industries. Knowledge spillovers, input sharing, and the local circulation of skilled labor can all generate more specialized occupations. In this sense, local innovation contributes to specialization not only through direct technological change but also indirectly through interdependent firms and workers.

One natural concern is that the division of labor and innovation are endogenous choices that may be simultaneously determined. For instance, high labor costs may spur the creation of labor-saving technology while also prompting firms to reorganize tasks. In addition, Smith notes that "the invention of all those machines [...] seems to have been originally owing to the division of labour." In other words, a specialized workforce may inspire the creation of new tools and machinery. We address these concerns using two complementary identification strategies corresponding to the two channels above.

Technological Progress. First, for the industry-level analysis, we estimate the following regression:

$$Y_{jt} = \beta_1 Patent_{jt} + \beta_2 EmpShare_{j,t-10} + \delta_j + \tau_t + \epsilon_{jt}, \tag{7}$$

[24] One caveat is that we cannot distinguish between patents that changed the production process (i.e., through new machinery or tools) from those that introduced new products to a given industry. This prevents us from including the product variety as a control in the analysis. One should view the estimated effects of "innovation" on occupational specialization as aggregate effects through both channels.

Table 8: Occupational Specialization and Exposure to Innovation

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Titles)	ln(Variety)	Entropy	ln(Titles)	ln(Variety)	Entropy
Panel A: Industry Level, Breakthrough Instrument						
asinh(New Patents)	0.409***	0.438***	0.454***	0.350*	0.426**	0.471***
	(0.142)	(0.128)	(0.127)	(0.182)	(0.163)	(0.149)
Obs.	733	733	733	733	733	733
R-squared (within)	0.066	0.079	0.082	0.065	0.079	0.082
Kleibergen–Paap F-statistic				5.146	9.406	9.191
Mean of dependent var.	3.380	2.731	2.128	3.380	2.731	2.128
SD of dependent var.	1.455	1.244	1.048	1.455	1.244	1.048
Industry FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Panel B: City-Level, Carnegie Library Instrument						
asinh(New Patents)	0.190***	0.162***	0.131***	0.294***	0.186**	0.082
	(0.007)	(0.007)	(0.007)	(0.097)	(0.094)	(0.102)
Obs.	7,666	7,666	7,666	7,666	7,666	7,666
R-squared (within)	0.177	0.138	0.081	0.123	0.135	0.069
Kleibergen–Paap F-statistic				9.242	3.898	0.659
Mean of dependent var.	5.090	4.357	3.523	5.090	4.357	3.523
SD of dependent var.	0.936	0.914	0.897	0.936	0.914	0.897
City FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

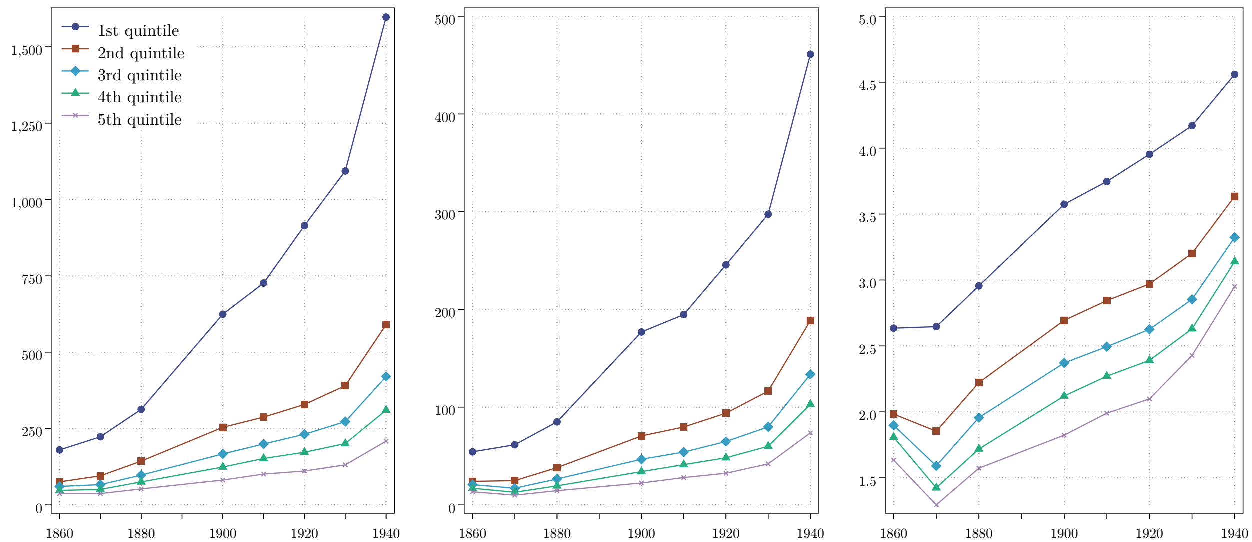
Notes: Panel A reports the regression coefficients using industry-level observations, controlling for industry employment share in the past census, industry and year fixed effects. In columns (4)–(6), we use the number of new patents for industry j in the past decades ($Patent_{jt}$) using the lagged number of breakthrough patents in industry j. Standard errors are clustered at the industry level. Panel B reports regression coefficients using city-level observations, controlling for city and year fixed effects. In columns (4)–(6), we instrument $Patent_{ct}$, the number of relevant patents since last census in city c at time t, with the Carnegie library grant instrument. Standard errors are clustered at the county level, and significance levels are indicated by *** 1%, ** 5%, and * 10%..

where Y_{jt} is a specialization measure for industry j in Census year t, $Patent_{jt}$ is the inverse hyperbolic sine of the number of new patents matched to industry j that were issued in the past decade, and $EmpShare_{j,t-10}$ is the lagged employment share of industry j, which controls for the possibility that larger industries may produce more patents. We use the occurrence of breakthrough patents defined by Kelly et al. (2021) as an instrument for the total number of patents matched to each industry. Breakthrough patents are both novel and impactful. They are dissimilar to past patents but inspire future innovation. We can therefore use the count of breakthrough patents as an instrument for the subsequent change in total patents, as the occurrence of breakthroughs is random and unanticipated, but they spur follow-up innovation.[25]

Panel A in Table 8 reports estimates of β_1. The results show that industries with more new patents experience faster growth in specialization. A 1% increase in the number of patents increases job title variety by 0.43%, or 0.35 standard deviations (column (5)). The results are comparable in scale when using other specialization measures and when comparing the OLS and IV estimates.

[25] Figure A3 confirms this pattern. Panel (a) plots the conditional correlation between the number of new patents in the past decade and the number of breakthrough patents in the subsequent decade, showing no clear relationship. Panel (b) plots the reverse—the number of breakthrough patents in the past decade and subsequent new patents in the industry—showing a positive and significant relationship between the two.

Figure 9: Trends in Occupational Specialization by Patents Quintile



(a) Job Titles **(b)** Variety Index **(c)** Entropy

Notes: Counties are partitioned so that each quintile contains approximately 20% of the stock of total U.S. patents in 1940. Quintiles are sorted from largest (1st) to smallest (5th), with the occupational specialization measures averaged across counties within each quintile.

Local Innovation. Next, we focus on the effects of local innovation on occupational specialization. In Figure 9, we partition counties into quintiles, each of which contains approximately 20% of the stock of U.S. patents in 1940. We then plot the average of our specialization measures across counties in each patenting quintile. Similar to our findings for population quintiles, we find that the growth in aggregate occupational specialization observed between 1860 and 1940 was driven mostly by high-innovation counties. Consistent with modern work on the local agglomeration effects of innovation (Jaffe et al., 1993), the descriptive evidence suggests that knowledge creation and diffusion within local economies played an important role in shaping the geography of occupational specialization during the Second Industrial Revolution.

To address endogeneity concerns, we follow Berkes and Nencka (2024) and use the construction of Carnegie libraries as an instrument for patents. Between 1883 and 1919, more than 1,500 cities applied for and received grants from the philanthropist Andrew Carnegie to build public libraries. Comparing the cities that accepted the grants with those that applied for but did not ultimately build libraries, Berkes and Nencka (2024) find that patenting increased by 10% following library construction.[26] We adapt their event study design as an instrumental variable to quantify the causal effect of local innovation on occupational specialization. The main identification assumption is that public libraries affect patenting activity by reducing the costs of accessing knowledge, but the construction of Carnegie libraries was determined by idiosyncratic local political preferences

[26] Berkes and Nencka (2024) note that not all recipient cities eventually built a library. Many cities rejected the Carnegie library grants due to his negative reputation, especially after Carnegie hired a private militia to violently break a steelworker strike. In other words, cities that rejected the grant for ideological reasons were otherwise comparable to those that accepted the grant.

rather than economic considerations. Further, libraries themselves are unlikely to affect the division of labor at the city level, except through their influence on local innovation.

We estimate the following city-level regression:

$$Y_{ct} = \beta_1 Patent_{ct} + \gamma_c + \tau_t + \epsilon_{ct}, \tag{8}$$

where Y_{ct} is the specialization measure in city c at time t, and $Patent_{ct}$ is the inverse hyperbolic sine of the number of new patents in city c issued in the past decade. The regression controls for city fixed effects γ_c and year fixed effects τ_t. We instrument the number of patents with z_{ct}, a dummy variable that equals 1 in the years after grant approval for cities that eventually built a library, and 0 otherwise. Following Berkes and Nencka (2024), we limit the sample to cities with less than 25,000 people and counties with less than 750,000 people in 1900. We also exclude cities that did not build a Carnegie library because a different philanthropist donated one instead.

Panel B in Table 8 presents the estimation results for equation (8). Among the 1,139 cities in the sample, those with more patents—as predicted by the establishment of a library—saw greater occupational specialization. Specifically, a 1% increase in patents led to a 0.19% increase in job title variety (column (5)). Note that the cities in the library sample are relatively small due to the sample restriction, with the median city population in 1900 being only 3,517. Nevertheless, even among these less populous cities, we still find that the more innovative places developed more specialized workforces.

Together, our findings in this section show that innovation influences the division of labor through both direct technological advancement within industries and indirect knowledge spillovers within local economies. The IV results additionally suggest that these relationships are not simply correlations driven by reverse causality or omitted local characteristics but instead reflect a causal link between innovation and the increasing complexity of the U.S. occupational structure. While Smith clearly recognized that innovation and the division of labor were closely related, our findings extend his theory by highlighting the role of general knowledge and innovation spillovers in facilitating greater specialization.

5 Productivity Implications

Smith famously claims that "the greatest improvements in the productive powers of labour [...] seem to have been the effects of the division of labour." Does Smith's observation from 18th-century England also apply to the American economy during the Second Industrial Revolution? We focus on a measure of labor productivity from the Census of Manufactures and estimate the following regression with a balanced city-industry panel:

$$Y_{jct} = \beta_s Specialization_{jct} + \mathbf{X}_{jct}\Gamma + \mu_{jc} + \gamma_{jt} + \epsilon_{jct}. \tag{9}$$

Here, Y_{jct} is the log of average value added per worker across all establishments for industry j in

Table 9: Labor Productivity and Occupational Specialization City-Industry Level 1910–1940

	Dependent Variable: Log value-added per worker						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log occupations	0.295***				0.219**		
	(0.074)				(0.090)		
Log variety index		0.259***				0.159*	
		(0.066)				(0.081)	
Entropy			0.147*				0.082
			(0.076)				(0.078)
Log city population				0.402***	0.278**	0.331**	0.371**
				(0.129)	(0.117)	(0.137)	(0.141)
Market access				0.287**	0.302**	0.290**	0.284**
				(0.123)	(0.127)	(0.130)	(0.130)
Log average firm employment				-0.040	-0.048	-0.044	-0.041
				(0.057)	(0.060)	(0.059)	(0.058)
City-industry-year obs.	1,916	1,916	1,916	1,916	1,916	1,916	1,916
R-squared (within)	0.025	0.016	0.009	0.041	0.050	0.046	0.043
City-Industry FE	✓	✓	✓	✓	✓	✓	✓
Industry-Year FE	✓	✓	✓	✓	✓	✓	✓

Notes: The dependent variable is log mean value added per worker for industry j in city c at year t. The sample is a balanced panel with only city-industry pairs with observations in all four Census years, from 1910 to 1940. Observations are weighted by the total number of individuals observed in the population census for each industry-city-year cell. Standard errors are clustered at the city-industry level, and significance levels are indicated by *** 1%, ** 5%, and * 10%.

city c at year t.[27] $Specialization_{jct}$ is one of our three specialization measures, and X_{jct} includes log city population, market access, and log mean firm employment as controls. The regression also includes industry-by-city fixed effects μ_{jc} and industry-by-year fixed effects γ_{jt}, implying that the estimates are identified from changes within city-industry cells over time, controlling for nationwide industry-specific trends.

Table 9 presents the results. The first three columns show a strong positive correlation between specialization and manufacturing productivity without controlling for any additional covariates. Specifically, column (1) shows that a 1% increase in the number of titles is associated with a 0.3% increase in value added per worker. To put this estimate into perspective, it implies that doubling the number of titles in an average industry raises productivity by 0.6 standard deviations. Alternative specialization metrics generate similar results.

While these estimates support the hypothesis that the division of labor increases labor productivity, they may also reflect agglomeration effects in larger cities or economies of scale in larger firms. For instance, firms in larger cities may also benefit from agglomeration effects through improved matching, knowledge spillovers, or better infrastructure. To account for these potential confounds, we include log city population, market access, and log firm employment in the regression. Note that

[27] Labor productivity is usually defined as output per hour worked. However, we have no data on hours in the Census of Manufactures. For that reason, we focus on average value added per worker.

the coefficients on specialization are identified from residual variation in these variables, holding market and firm size fixed.[28] As a benchmark, column (4) includes only market access and firm size controls. The estimates in columns (5)–(7) confirm that occupational specialization remains positively correlated with labor productivity even with additional controls, though the estimated coefficients are attenuated somewhat. A 1% increase in the number of titles coincides with a 0.2% increase in value added per worker, three-fourths the size of the estimate without controls in column (1). In other words, doubling the number of titles in an average industry coincides with a 0.45 standard deviation increase in labor productivity.

Overall, these results provide empirical support for Smith's observation that improvements in the productive powers of labor can be attributed to greater specialization of the workforce. While establishing a clean causal relationship between specialization and productivity is challenging due to the interdependent relationships we find between the division of labor, technological innovation, and market structure, our analysis nonetheless implies that occupational specialization played an important role in the economic history of the industrializing United States.

6 Conclusion

We use novel U.S. occupational data from 1860 to 1940 to empirically evaluate Adam Smith's core hypotheses on the division of labor. Strikingly, while Smith developed his theory in the context of 18th-century Britain, our findings demonstrate its enduring relevance more than a century later. As Smith proposed, we find that occupational specialization increases with both market size and innovation, and that greater division of labor coincides with higher productivity within narrowly defined industries across cities.

Beyond supporting Smith's original hypotheses, our analysis extends his theory of the division of labor along two dimensions. First, while Smith primarily focuses on the specialization of physical production processes, we show that organizational and procedural innovations also give rise to new, specialized occupations. This is particularly evident in the U.S. during the Second Industrial Revolution, when firms developed complex organizational structures to manage the growing scale of production and transaction volumes. Second, while Smith recognizes the effect of innovation on the division of labor, our findings also highlight the role of spillover effects, as local patenting is strongly associated with increased occupational specialization. This suggests that the division of labor responds not only to industry-specific technologies but also to the general knowledge and innovation environment in the local economy.

Our findings also underscore the importance of granular occupational data in revealing hidden structures and dynamics in historical labor markets. While our results affirm the continued relevance of Smith's core insights on the division of labor, they also demonstrate that the mechanisms driving occupational specialization are more complex and nuanced. In addition, we find persistent

[28] Because city population, market access, and firm sizes are themselves strongly correlated with occupational specialization, it again raises concerns about multicollinearity in this regression. However, after projecting out industry-city and industry-year fixed effects, no explanatory variable has a variance inflation factor larger than 1.5

and growing gaps in economic dynamism across regions. Large urban markets continued to outperform the rest of the country in innovation, specialization, and productivity. This raises important questions about the underlying causes and consequences of such persistent spatial inequality. While beyond the scope of this paper, it is worth further investigating the determinants of such gaps.

References

Ades, Alberto F., and Edward L. Glaeser. 1999. "Evidence on Growth, Increasing Returns, and the Extent of the Market." *The Quarterly Journal of Economics* 114 (3): 1025–1045.

Atack, Jeremy, Robert A. Margo, and Paul W. Rhode. 2019. ""Automation" of Manufacturing in the Late Nineteenth Century: The Hand and Machine Labor Study." *Journal of Economic Perspectives* 33 (2): 51–70.

Atalay, Enghin, Sebastian Sotelo, and Daniel Tannenbaum. 2024. "The Geography of Job Tasks." *Journal of Labor Economics* 42 (4): 979–1008.

Autor, David, Caroline Chin, Anna Salomons, and Bryan Seegmiller. 2024. "New Frontiers: The Origins and Content of New Work, 1940–2018." *The Quarterly Journal of Economics* 139 (3): 1399—1465.

Becker, Gary S., and Kevin M. Murphy. 1992. "The Division of Labor, Coordination Costs, and Knowledge." *The Quarterly Journal of Economics* 107 (4): 1137–1160.

Berkes, Enrico, Ezra Karger, and Peter Nencka. 2023. "The Census Place Project: A Method for Geolocating Unstructured Place Names." *Explorations in Economic History* 87 101477.

Berkes, Enrico, and Peter Nencka. 2024. "Knowledge Access: The Effects of Carnegie Libraries on Innovation." *The Review of Economics and Statistics* 1–45.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica* 81 (4): 1347–1393.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19 (1): 22–64.

Census Office. 1870. *Ninth Census of the United States: Instructions to Assistant Marshals.* Census Office, Department of the Interior.

Chandler, Alfred D. 1977. *The Visible Hand: The Managerial Revolution in American Business.* Harvard University Press.

Chandler, Alfred D. 1984. "The Emergence of Managerial Capitalism." *Business History Review* 58 (4): 473–503.

Chilosi, David, Giampaolo Lecce, and Patrick Wallis. 2025. "Smithian Growth in Britain before the Industrial Revolution, 1500–1800." CEPR Discussion Paper No. 20418.

Desmet, Klaus, Shlomo Weber, and Ignacio Ortuño-Ortín. 2009. "Linguistic Diversity and Redistribution." *Journal of the European Economic Association* 7 (6): 1291–1318.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A "Market Access" Approach." *The Quarterly Journal of Economics* 131 (2): 799–858.

Duranton, Gilles, and Hubert Jayet. 2011. "Is the Division of Labor Limited by the Extent of the Market? Evidence from French Cities." *Journal of Urban Economics* 69 56–71.

Duranton, Gilles, and Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies." In *Handbook of Regional and Urban Economics*, Elsevier.

Duranton, Gilles, and Diego Puga. 2005. "From Sectoral to Functional Urban Specialization." *Journal of Urban Economics* 57 (2): 343–370.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review* 100 (3): 1195–1213.

Ferrara, Andreas, Patrick A. Testa, and Liyang Zhou. 2024. "New Area- and Population-based Geographic Crosswalks for U.S. Counties and Congressional Districts, 1790–2020." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 57 (2): 67–79.

Fiszbein, Martin, Jeanne Lafortune, Ethan G. Lewis, and José Tessada. 2024. "Powering Up Productivity: The Effects of Electrification on U.S. Manufacturing." Working Paper 28076, National Bureau of Economic Research.

Goldschlag, Nathan, Travis J. Lybbert, and Nikolas J. Zolas. 2020. "Tracking the Technological Composition of Industries with Algorithmic Patent Concordances." *Economics of Innovation and New Technology* 29 (6): 582–602.

Haines, Michael R., and Robert A. Margo. 2008. "Railroads and Local Economic Development." In *Quantitative Economic History: The Good of Counting*, Routledge.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108 (3): 577–598.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. "Measuring Technological Innovation over the Long Run." *American Economic Review: Insights* 3 (3): 303–320.

Lamoreaux, Naomi R. 2019. "The Problem of Bigness: From Standard Oil to Google." *Journal of Economic Perspectives* 33 (3): 94–117.

Lee, James. 2015. *Proximate Causes: The Influence of Agglomeration, Access to Finance, and Infrastructure on US Economic Growth, 1860-1990*. Ph.D. dissertation, Harvard University.

Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding. 2019. "Task Specialization in U.S. Cities from 1880 to 2000." *Journal of the European Economic Association* 17 (3): 754–798.

Moretti, Enrico. 2004. "Human Capital Externalities in Cities." In *Handbook of Regional and Urban Economics*, Volume 4. 2243–2291, Elsevier.

Moretti, Enrico. 2010. "Local Multipliers." *American Economic Review* 100 (2): 373–377.

Moretti, Enrico. 2012. *The New Geography of Jobs*. Houghton Mifflin Harcourt.

Petralia, Sergio, Pierre-Alexandre Balland, and David L. Rigby. 2016. "Unveiling the Geography of Historical Patents in the United States from 1836 to 1975." *Scientific Data* 3 (1): 1–14.

Posch, Max, Jonathan Schulz, and Joseph Henrich. 2025. "How Cultural Diversity Drives Innovation: Surnames and Patents in U.S. History." Working Paper.

Rosenberg, Nathan, and Manuel Trajtenberg. 2004. "A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States." *The Journal of Economic History* 64 61–99.

Rosenthal, Stuart, and William Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Economics*, Volume 4. 2119–2171, Elsevier.

Ruggles, Steven, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren. 2024. "IPUMS Ancestry Full Count Restricted Data: Version 4.0R." Dataset, Minneapolis, MN: IPUMS, https://doi.org/10.18128/D014.V4.0R.

Saiz, Albert. 2007. "Immigration and Housing Rents in American Cities." *Journal of Urban Economics* 61 (2): 345–371.

Shannon, Claude E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27 (3): 379–423.

Sokoloff, Kenneth L. 1984. "Was the Transition from the Artisanal Shop to the Nonmechanized Factory Associated with Gains in Efficiency?: Evidence from the U. S. Manufacturing Censuses of 1820 and 1850." *Explorations in Economic History* 21 (4): 351–382.

Straathof, Sebastiaan M. 2007. "Shannon's Entropy as an Index of Product Variety." *Economics Letters* 94 (2): 297–303.

Tian, Lin. 2021. "Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil." Working Paper.

U.S. Department of Labor, Employment and Training Administration. 1939. *Dictionary of Occupational Titles: First Edition.* U.S. Government Printing Office.

Appendix

Job titles. Due to spelling errors in the original enumeration forms and transcription mistakes, we observe millions of unique occupation strings throughout the eight Census waves included in our study. Nevertheless, only a small fraction of these appear more than a handful of times. To clean the data, we first convert all strings to lower case, strip non-alphabetic characters, and remove articles and prepositions. Next, we replace common spelling errors and abbreviations (ensuring that any corrected titles exist independently in the data). Finally, if a complete string occurs fewer than 100 times in the entire dataset but contains a word or words that occur more frequently, we remove the additional text.[1]

The processing steps described above reduce the number of unique occupation strings in our data from around 10 million to 30,000. We proceed by manually reviewing and further standardizing this smaller set of titles. This involves harmonizing gender-specific titles (salesman and saleswoman), replacing verbs with nouns (accounting to accountant), and correcting additional spelling errors and abbreviations that were not replaced in the automated step of our data cleaning. Additionally, we separate occupation and industry titles when both are reported and use the first title listed for multiple job holders. After this step, we are left with about 10,000 standardized occupation titles.

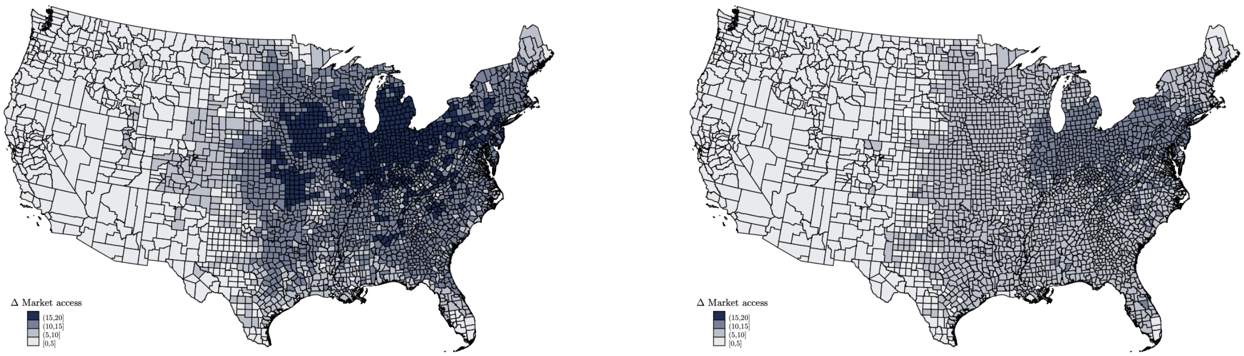
Census of Manufactures. The threshold for surveyed firms to be included in city-industry tabulations vary dramatically from year to year. For example, the 1910 data report city-industry pairs for cities with population above 50,000, while the 1930 data include only cities with population above 100,000. In addition, due to errors during digitization, some cells include extreme outliers. Therefore, we limit the sample to include only city-industry observations that occur in all four census years, while also excluding observations with per-worker value added above $10,000 or below zero.

Table A3 tabulates the mean and standard deviations of the variables we include in the city-industry level analyses. The sample includes 55 cities and 40 industries. This balanced panel represents a selected set of large cities, with the average city population increasing from 660,000 in 1910 to over 1 million in 1940. Nevertheless, while the sample restricts to industries with consistent reporting, it still covers a wide range of manufacturing sectors, ranging from bakery products, fabricated steel products, printing, electrical machinery and equipments, to knitting mills, grain-mill products, and glass products.

Evolution of Market Access Figure A1 shows changes in county market access from 1860 to 1900 and from 1900 to 1940. Since the market access measure was primarily based on railroad access, the changes are highly spatially correlated. The Midwest and South experienced the most growth, while Western states such as California did not experience much increase in domestic market access, despite a fast-growing local population.

[1]For example, if fewer than 100 total individuals report the title "labor economist" between 1860 and 1940, we reduce this string to the base title "economist." Although this results in some loss of detail, we emphasize that we trim only extremely rare titles. In our main analysis, we focus on titles reported 100 or more times in a given Census year.

Figure A1: Changes in Market Access, 1860–1940



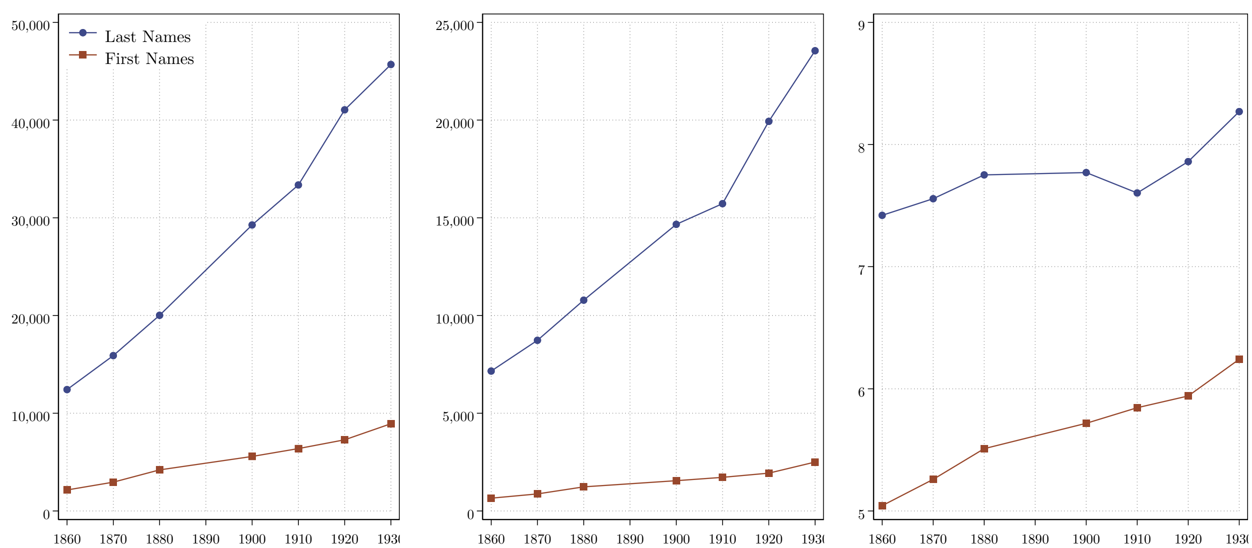
(a) Change in access, 1860–1900 **(b)** Change in access, 1910–1940

Table A1: Sample Used to Construct Standardized Job Titles

	1860	1870	1880	1900	1910	1920	1930	1940
Total employment	7,109,214	11,260,992	16,314,239	26,221,708	36,076,260	39,106,468	46,211,871	50,158,843
Occupations in main sample	819	1,053	1,364	2,931	3,107	3,776	4,459	6,438
Reports industry only (%)	0.74	1.13	1.80	2.19	1.14	1.28	0.91	1.22
Not classified (%)	0.40	0.31	0.50	1.18	3.12	2.12	2.76	2.93
Main sample (%)	98.11	98.00	97.18	96.01	95.25	96.07	95.88	95.52

Notes: Main sample includes occupation titles reported 100 or more times each year, excluding observations not assigned to a standardized title or reporting only an industry.

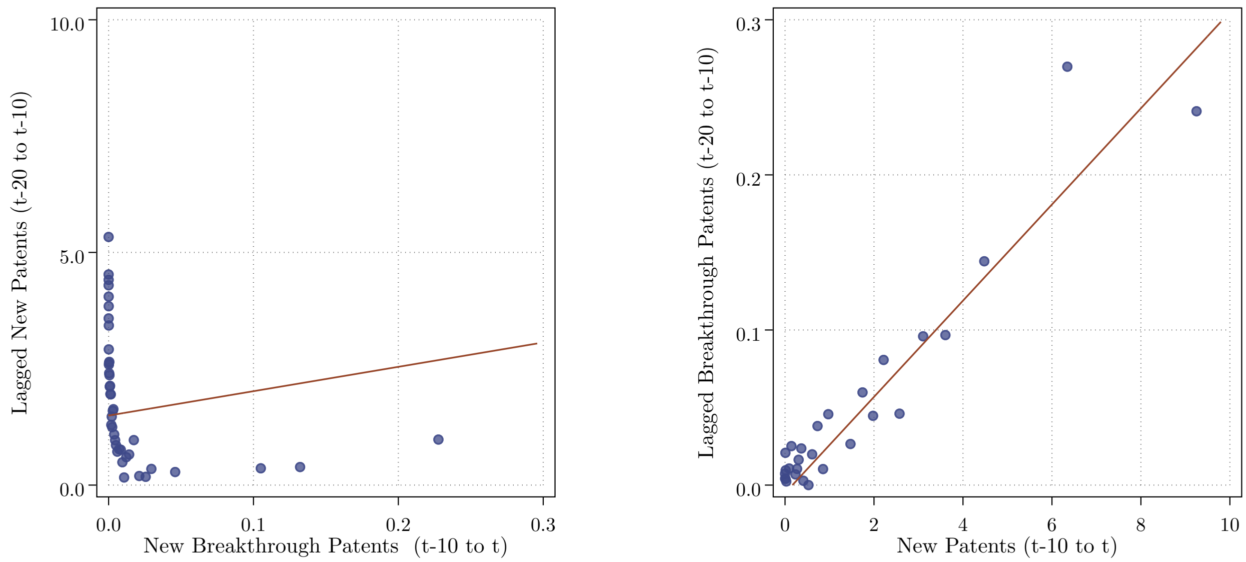
Figure A2: "Specialization" Measure of Names 1860–1930



(a) Names **(b)** Variety Index **(c)** Entropy

Notes: We construct the three "specialization" measures using first and last names from the 1% Census sample from IPUMS. We limit to names with at least five observations nationally in each year.

Figure A3: Stage Zero: Past Breakthrough Predicts Future Patents



(a) **(b)**

Notes: Panel (a) presents conditional correlations between past patent flow and breakthrough patents by regressing the number of new patents (by industry) in the past decade on breakthrough patents, controlling for year and industry fixed effects. Panel (b) presents the conditional correlation between breakthrough patents and subsequent patent flows by regressing the number of breakthrough patents in the past decade on the number of new patents, controlling for year and industry fixed effects.

Table A2: Summary Statistics for County Level Variables

Panel A: County-Aggregate	1860	1900	1910	1940	Total
Total population (000s)	10.12	24.36	29.64	42.42	26.63
	(24.27)	(65.50)	(85.40)	(142.93)	(91.00)
Market access (millions)	0.25	1.27	1.66	2.36	1.38
	(0.22)	(0.59)	(0.66)	(0.91)	(1.00)
Urban population share	0.04	0.13	0.17	0.23	0.14
	(0.13)	(0.21)	(0.23)	(0.25)	(0.22)
Cumulative county patents	9.44	190.95	277.15	617.39	273.73
	(96.56)	(1,540.97)	(2,176.72)	(4,687.78)	(2,705.78)
Panel B: County-Industry					
Number of industries	5.55	12.21	15.45	23.59	15.04
	(8.16)	(14.33)	(17.83)	(24.69)	(19.04)
Industry size (employment 000s)	1.31	2.75	3.27	2.49	2.55
	(1.45)	(2.15)	(2.64)	(2.18)	(2.29)

Table A3: Summary Statistics for Census of Manufactures City-Industry Panel

	1910	1920	1930	1940	Total
Total population (000s)	420.61	378.22	698.59	624.72	503.02
	(817.91)	(785.59)	(1,186.38)	(1,133.95)	(971.62)
Number of occupations	96.06	105.86	146.02	184.51	128.47
	(110.07)	(133.13)	(167.38)	(188.36)	(153.20)
Number of establishments	121.14	98.61	101.97	98.52	104.17
	(529.80)	(1,030.18)	(385.92)	(248.22)	(705.41)
Average firm employment	42.88	43.13	60.78	75.89	53.56
	(91.86)	(168.67)	(124.90)	(650.33)	(326.36)
Value added per worker	3,288.34	2,944.54	4,516.50	5,589.78	3,895.28
	(1,612.21)	(1,360.53)	(1,881.77)	(2,255.83)	(2,036.69)

Notes: The sample is a balanced panel with city-industry pairs that occur in all four census years. Value added per worker is adjusted to 1920 dollars.

Table A4: Relationship Between Specialization Measures and Definition Length

	Full Sample			Post-1910 Sample		
Total words (mean = 90)	(1)	(2)	(3)	(4)	(5)	(6)
Log Occupations	-2.381***			-2.515***		
	(0.139)			(0.174)		
Log Variety Index		-3.962***			-4.069***	
		(0.151)			(0.182)	
Entropy			-4.469***			-4.603***
			(0.155)			(0.173)
Unique verbs (mean = 13)						
Log Occupations	-0.272***			-0.288***		
	(0.019)			(0.023)		
Log Variety Index		-0.446***			-0.467***	
		(0.020)			(0.023)	
Entropy			-0.506***			-0.539***
			(0.021)			(0.022)
County-industry-year observations	267,874	267,874	267,874	204,845	204,845	204,845
County × industry FE	✓	✓	✓	✓	✓	✓
Industry × division × year FE	✓	✓	✓	✓	✓	✓

Notes: The sample includes only county-industry cells with at least 50 workers who report valid job titles. Standard errors are clustered at the county level, and significance levels are indicated by *** 1%, ** 5%, and * 10%.