Technological Revolutions and Occupational Change: Electrifying News from the Old Days

BY PAUL GAGGL, ROWENA GRAY, AND MIGUEL MORIN[1]

This version: October 31, 2015

Preliminary draft. Please do not circulate or cite without permission of the authors.

Abstract: *How do major technological breakthroughs affect workers in the wake of its adoption? We contribute to the understanding of this question by studying the impact of early electricity adoption, using newly digitized, detailed maps of the US electricity grid in 1918, 1928, and 1940. Moreover, we follow individual workers over time by linking individuals in the 100% count Census of Population for 1920, 1930, and 1940 to studty the impact of electricity adoption on employment, job and sectoral mobility, as well as earnings. To identify the causal impact on worker trajectories, we exploit the geography of hydro-electric potential, which is highly heterogeneous across U.S. regions and provides arguably exogenous variation in the incentive to adopt electricity. Unlike earlier studies, our analysis is neither limited to particular industries nor particular geographies and constitutes the first comprehensive analysis of individual worker trajectories in response to a major technological revolution in the US. Our preliminary results uncover a number of interesting insights: on average, electricity (1) replaces jobs; (2) causes substantial upward movement in the earnings distribution for farm workers; (3) causes individuals to perform substantially different tasks than before; (4) is a causal driver for the movement from the farm to the factory.*

Keywords: electricity, technical change, labor market, employment, occupations.

[1]Gaggl: University of North Carolina, Charlotte (pgaggl@uncc.edu); Gray: University of California, Merced (rgray6@ucmerced.edu); Morin: University of Cambridge, United Kingdom (mabm3@cam.ac.uk). Gray and Morin gratefully acknowledge the financial support by the Russell Sage Foundation (grant 85-15-12). We would like to thank Pol Antras, Steve Billings, Vasco Carvalho, Raj Chetty, Timothy Guinnane, Bill Janeway, Hamish Low, and Lisa Schulkind as well as participants at the the 2015 Big Data Conference and the macroeconomics lunch workshop at the University of Cambridge for extremely helpful suggestions and comments.

1. Introduction

The effect of technology adoption on the occupational structure has attracted considerable recent research by economists and economic historians. Frey and Osborne (2013) estimate that 47% of all current employees in the US are susceptible to replacement by a computer. Autor, Levy and Murnane (2003) documented the hollowing out of the skill distribution during computerization and thus argued that a key contributor to the increased wage inequality witnessed in recent decades was technological change. In U.S. history, Katz and Margo (2013) find that the middle of the skill distribution shrank during the development of the American economy in the 19th century. Both topics of skill obsolescence and wage inequality are a growing concern in the popular press and of major interest for policy makers. Despite the relevance of this question, we have little empirical evidence on the trajectories of individual workers displaced by a new technology. Rather, the literature is characterized by analyses at an aggregate level, using repeated cross sections of data, which do not easily allow the researchers to identify where the hollowed out individuals end up, making a determination of the ultimate welfare effects of technological change difficult. A few recent papers do analyze longitudinal data (see, for example, Cortes, 2015), but these studies are limited to using small, possibly biased samples—such as the small Panel Survey of Income Dynamics (PSID).

This paper focuses on the most recent period of general purpose technological change for which a large, representative longitudinal dataset of occupations, tasks and wages can be obtained: the pre-World War II period when U.S. factories and farms electrified. Gray (2013) and Devine (1983), among other work, have shown that electricity had wide-ranging implications for factory layout, scale and skill demand. This in part came from the fact that electricity was complementary to other important innovations such as production line technology. Electricity cost and adoption rates varied greatly across the U.S., as we show in more detail below, which makes this historical case an easier one in which to identify the causal impacts on the labor market, as compared to the computer case where adoption had a more uniform timing across areas. The historical case also offers the advantage of census records with names attached (only available up to 1940), which enables us to match across censuses and analyze our linked sample. In comparison to the existing historical literature, we also analyze the entire economy, rather than focusing simply on manufacturing.

Our project is the first to trace the evolution of American jobs at the individual level between 1920 and 1940 in response to electrification. We have access to the *universe* of census data for the U.S. population between 1900 and 1940 and currently utilize 1920-1940 which provides occupational information (the earlier years await completion by IPUMS). We link the 1920-1940 Censuses to produce our longitudinal dataset. Following the same workers over time is the key to studying the trajectory of workers in occupations vulnerable to replacement by electrical machinery. We match this longitudinal dataset to high-quality data on occupational characteristics from a historical *Dictionary of Occupational Titles* and to improved data on occupational wages at the state- or city-level from the Bureau of Labor Statistics. This dataset represents a significant innovation in its own right.

We then employ these data to identify the causal impact of electrification on the evolution of labor market outcomes for individual workers over the period 1920-1940. We exploit detailed geographic variation in the expansion of the electricity grid, using newly digitized maps of the US electricity grid, and ask whether individuals in more electrified counties were more likely to experience job loss, whether they moved to higher or lower paying occupations, and whether they were executing significantly different tasks. Moreover, we estimate the impact of electrification on transition probabilities between nine broad industrial sectors and particularly assess how much of the movement out of agriculture and into manufacturing can be explained by electrification.

At this point, we gauge the extent to which electrification impacted the occupational and wage status (as measured by the variable `occscore` provided by IPUMS). Future drafts will make a more complete use of the longitudinal nature of the data as well as the task and wage data that we have already collected, allowing us to identify better the mechanisms through which these occupational changes occurred. We will also explore heterogeneous effects depending on education level, age, and local economic conditions.

Geographical variation in the adoption of electricity is prone to a number of endogeneity concerns: a region's productivity and borrowing constraints affect the trade-off in learning new skills for workers or investing in electricity for firms. Moreover, more productive regions have a higher return to new skills and investments in new technologies, which leads firms to adopt electricity and workers to learn better skills. However, even if firms in these regions were barred from the electricity market and had not adopted electricity, workers likely would still upgrade their skills given initial conditions. This would likely result in no systematic difference between workers' outcomes with or without electricity adoption. In other words, a simple regression may reveal spurious a correlation between electricity adoption and wage upgrading that is due to the omitted variable of productivity or credit access.

To address these concerns, we use geographic variation in hydro-electric potential to instrument for electricity adoption. The intuition behind this instrument is that hydro-electric power was substantially cheaper than its alternative: coal powered generators. We show that this instrument is strong and use it to identify the causal impact of electricity adoption on the outcomes mentioned above.

In sum, our main results indicate that, on average, workers are more likely to become unemployed in response to electricity adoption. Moreover, we find no evidence for electricity induced occupational upgrading *on average*. However, our results indicate substantial occupational upgrading for individuals initially employed in farming related occupations. Moreover, we find that electrification causes a substantial shift in the task mix performed on the job. Finally, we estimate that electricity causes substantial movement from the agricultural sector to various manufacturing industries.

The remainder of this paper is organized as follows: we start with a brief literature review in Section 2, before discussing the expansion of the US electricity grid during 1920-1940 in Section 3. Section 4 focuses on the geography of hydro-potential and its potential to serve as an instrument for electricity adoption. We then describe our various sources for data on individuals, occupations, and wages in Section 5, before moving to our empirical analysis and preliminary results in Section 6. Section 7 concludes.

2. Related Literature

This project relates to two strands of the literature. First, recent research has focused on the "hollowing out" of the occupational structure over the past three decades (Autor et al., 2003). In previous work (Gray, 2013) has shown that electrification in United States manufacturing before 1940 led to a hollowing out of the skill distribution, whereby workers occupying jobs in the middle of the skill distribution (those specialized in dexterity tasks which usually required artisanal or apprenticed skill) lost out to those at the poles who were mainly clerical/managerial and manual workers. Electricity proved complementary to other technologies during this period, such as the assembly line, and so the implications of these results regarding the task distribution are that workers in craft occupations such as blacksmiths and carpenters, saw their demand within American factories decline while demand for raw manual and assembly line workers, performing simpler and smaller tasks, increased, along with that for timekeepers, supervisors and managers. Our improvement on Gray (2013) consists in expanding these insights to movements in relative wages and to the subsequent outcomes of middle-skill workers following electrification.

Other authors have explored the evolution of occupations following electrification. Bessen (2011) demonstrates historical task-biased technological change in the textile sector in the mid-nineteenth century. He identifies which tasks and therefore which workers benefited and which lost out as a result of mechanization. Using a variety of different data sources and a much coarser definition of skill, Katz and Margo (2013) also find a "hollowing out of the middle of the skill distribution in the 19th century, with monotonic skill upgrading dominating between 1920 and 1990. Their approach has the advantage of looking at the economy as a whole (rather than focusing solely on the manufacturing sector) but they identify only the broad correlations over the long run using fairly coarse data.

However, these studies are limited by the lack of a longitudinal dimension in three respects. First, they cannot observe the trajectory of displaced workers. The implications for policy are different depending on this trajectory: the living standards of middle-skill workers increase if they switch to the high end of the distribution and decrease otherwise. Second, these studies are silent on whether the hollowing out occurred at the extensive or intensive margin, with middle-wage workers losing their jobs or facing downward pressure on their wages which prompted them to change occupation. The policy implications are also different depending on this margin, e.g. offering retraining possibilities for middle-wage workers may be misguided if they are still employed. Third, repeated cross-sections induce compositional bias from aggregate-level shocks: the large migration away from the Dust Bowl in the 1930s resembled a shift away from agriculture in Oklahoma, for example, but migrants may have found a similar occupation in other states. For example, Salisbury (2014) and Stewart (2012) document the benefits of geographic mobility in terms of occupational holding or upgrading for the late 1800s. The longitudinal aspect of our dataset addresses all three issues and is the key to estimating precise outcomes of workers in the middle of the distribution before World War II. It is therefore a significant advancement of the literature on historical occupational change.[2]

To the best of our knowledge, the only previous work on job mobility and wage changes using longitudinal, historical data is Solon, Whatley and Stevens (1997). This paper uses detailed data from the personnel archives of two companies, A.M. Byers and the Ford Motor Company. They focus on the extent of wage adjustment over the business cycle. The longitudinal dimension is key to overcome typical measurement problems in the literature, mainly that worker assignments may adjust over the business cycle which may bias findings relating to wage adjustment, making wages appear flexible when in fact it is occupational status or worker quality that has adjusted. The focus is therefore quite different from our project, which is also much more representative of the US

[2] Previous literature on worker displacement over this historical period has been limited to case studies, e.g. Baker (2007) for the printing industry.

economy as a whole.

Another related strand of the literature concerns the aggregate effects of electricity adoption. Morin (2015) used the geography of the source of electric power as an instrument for electricity adoption and estimated its effect on the labor demand decisions of firms. He found that firms responded to cheaper electricity prices by increasing capital intensity, decreasing the labor share of income, and increasing labor productivity—the main predictions from the adoption of a labor-saving technology. Furthermore, firms also passed on cheaper electricity prices onto consumers but the demand for products was not sufficiently elastic: firms adjusted to higher productivity by firing workers instead of increasing production. A natural extension of this work is to follow the workers who were replaced by electrical machinery and observe their subsequent labor market outcomes.

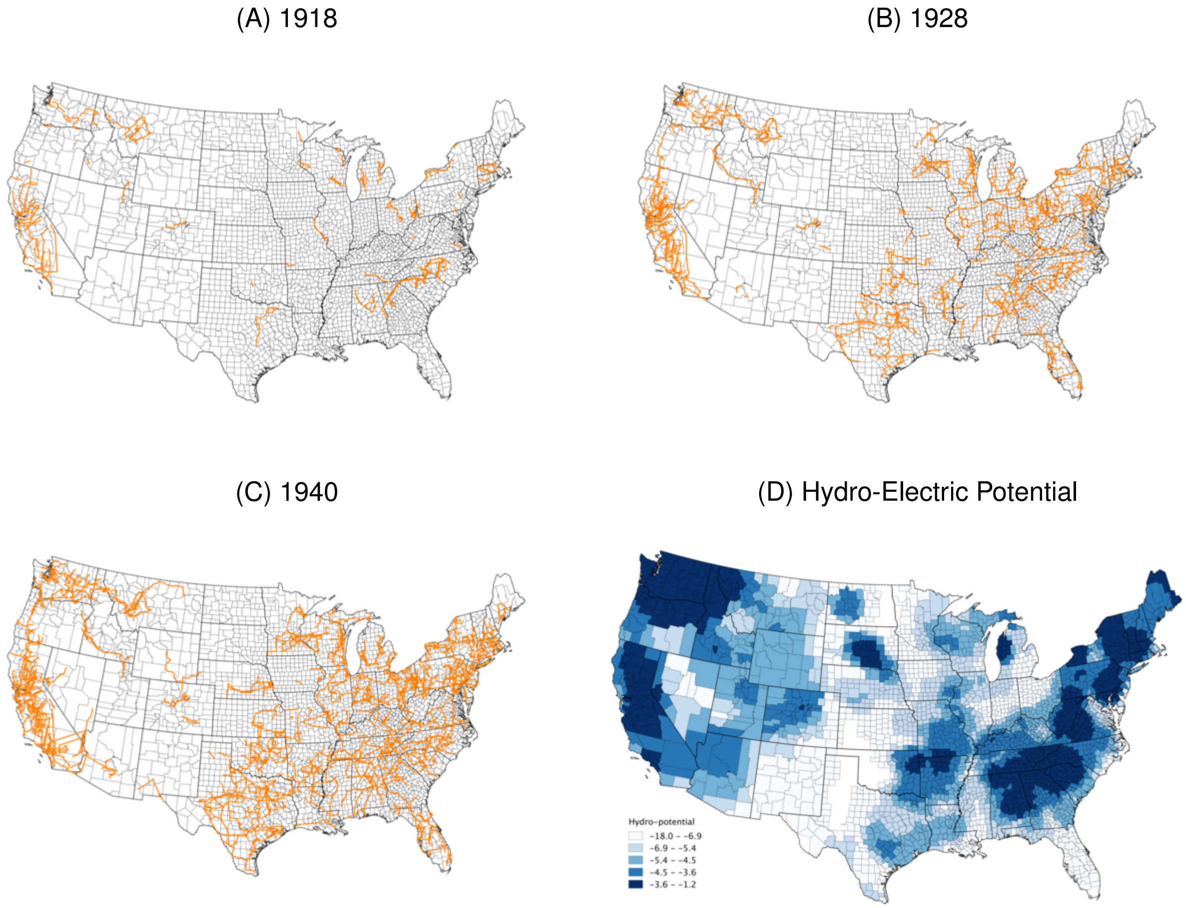
3. Measuring Electricity Adoption During 1920-1940

Due to data limitations, existing work studying the impact of electricity adoption on labor market outcomes was confined to either very aggregate analysis or was restricted to particular sectors—predominantly the manufacturing sector. Wo overcome this hurdle, we assemble a novel and fairly comprehensive measure of electricity adoption, based on detailed geographic variation in the physical location of electricity lines.

In particular, we digitize detailed maps of the electricity grid in 1918, 1928, and 1940, provided by the Edison Electric Institute (EEI), which are displayed in Figure 1. Panels A-C show a number of striking facts: First, in 1918 only a very few locations in the US were connected to the electricity grid. Second, both the period 1918-1928 and 1928-1940 see substantial expansions of the electricity grid. Finally, there is substantial geographic variation in the timing of electricity adoption.

These detailed maps of the electricity grid have several advantages over alternative ways to measure electricity adoption. The extant literature has so far focused on the usage of electricity

Figure 1: Expansion of the US Electricity Grid: 1918-1940



Notes: Panels A-C show digitized historical maps of the US electricity grid based on printed maps by the Edison Electricity Institute (EEI). Panel D illustrates hydro-electric potential as measured by the Idaho National Laboratory.

(measured in horsepower) based on surveys of manufacturing firms. This approach has the advantage of providing a true measure of electricity usage but has the severe disadvantage that it is based on small samples of firms within particular industries. While our maps do not reveal the actual usage of electricity, they have the benefit of providing a comprehensive account of the potential access to electricity within narrow geographies covering the entire mainland US. Specifically, this allows us to expand our analysis beyond the manufacturing sector and include rural areas as well as agriculture—the largest sector in terms of employment at the beginning of our sample.

4. Identifying the Causal Effect of Electricity Adoption

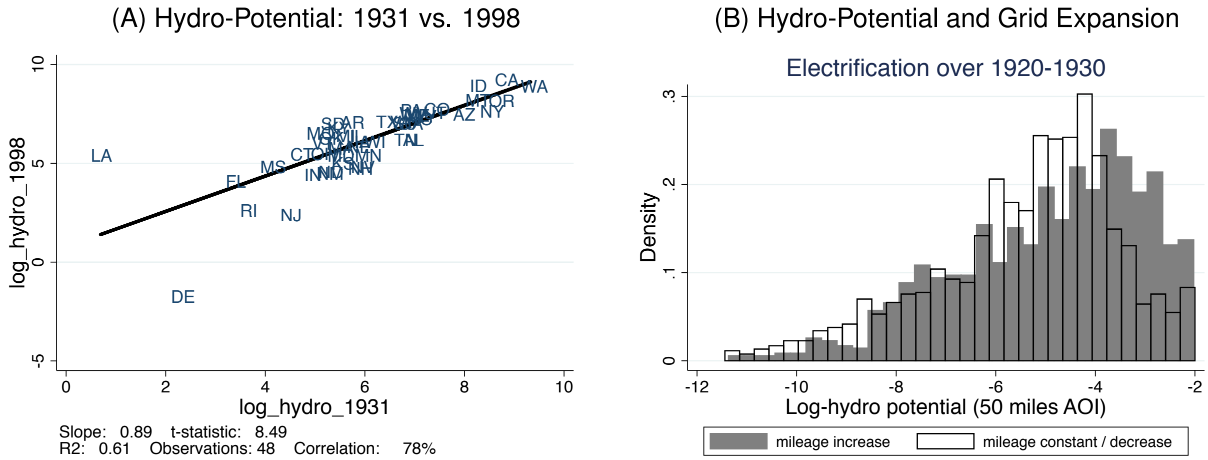
Perhaps the most challenging hurdle in existing studies was the lack of credible exogenous variation in the adoption of electricity across firms or industries. The geographic variation in grid expansion displayed in panels A-C of Figure 1 obviously also suffers from a variety of endogeneity concerns. More credit constrained regions may be slower to expand the electric grid; more highly educated areas may have greater incentives to exploit complementarities with electrical machinery and thereby attract electricity providers, etc.

However, our focus on geography—rather than firms or industries—allows us to exploit geographic variation in "hydro-electric potential" as a plausibly exogenous source of variation to instrument for the expansion of the electricity grid. The key argument behind this candidate instrument is the fact that the cost of producing hydro-electric power was substantially lower than that of operating coal powered generators. Thus, areas with greater hydro-electric potential—a combination of sufficient land grade and the presence of a river or stream—had an increased incentive to adopt electrical machinery relative to other areas. Moreover, since the physical location and grade of streams is arguably exogenous to the construction of power lines, this provides a plausible candidate instrument.

The Idaho National Laboratory published in 1998 an assessment of hydroelectric potential in all counties in the United States. This report was a 10-year effort to estimate the undeveloped hydropower capacity based not only on land gradient but also stream flow for 5,677 sites within the country (Severnini, 2012). To account for the ability to transmit power, we compute the total hydroelectric potential available 50 miles around a county divided by the area of influence. Panel D of Figure 1 illustrates the geographic variation in this measure. Furthermore, it is easy to see that this measure is highly correlated with the density and the growth of the electricity grid both during 1918-1928 and 1928-1940.

However, given that the detailed county level information on hydro potential was published in 1998 one may wonder how representative this variation is for the period 1920-1940. Interestingly,

Figure 2: The Geography of Hydro-Potential and Grid Expansion



Notes: Panel A illustrates a regression line of log hydro-electrc potential in 1998 on the same measure in 1931 at the state level. Panel B illustrates the density of log hydro-potential in counties with and without expansion in the milage of electrical lines over the period 1920-1930.

the geographic variation in this measure is essentially static over time. Panel A of Figure 2 confirms this fact by illustrating that the state-level correlation between hydro-potential in 1998 and 1931 is almost one.

While visual comparison of panels A-C and panel D of Figure 1 suggest a strong correlation between hydro-potential and the expansion of the electricity grid, we formally investigate this correlation here. To measure the expansion of the electricity grid we divide counties into ones that saw an increase of the total electricity line milage and ones that did not see any increase. Panel B of Figure 2 contrasts the density of hydro-potential within counties that saw an expansion of the electricity grid over 1920-1930 with that of counties who did not see an increase in the milage of power lines. In fact, a county level regression of an indicator variable for grid expansion suggests a strong positive correlation on average with an F statistic of over 50.

Thus, we conclude that hydro-electric potential provides a strong instrument for grid expansion that is likely to satisfy the exclusion restriction for an IV estimation approach.

5. Measuring the US Occupational Structure over 1920-1940

We construct and combine several data sources to produce a large and rich dataset that warrants a comprehensive analysis of the occupational structure from 1920 to 1940. The dataset consists of three parts: (1) longitudinal measures of occupational change and characteristics at the individual level from the full-count Census of Population; (2) task measures that describe what these historical occupations involved from the historical Dictionary of Occupational Titles (DOT); (3) wages at the occupation-, state- and city-level provided by the Bureau of Labor Statistics (BLS).

5.1. Longitudinal Census of Population

The Census of Population between 1920 and 1940 provides individual-level data on occupations and demographic information (names, ethnicity, age, gender, and birthplace etc). The 1940 census is the most detailed in terms of offering information on wage and salary income and years of education attained. The other census years contain information on the value or rent of the house, which we may use to proxy for income, and the information on occupation such as the IPUMS variable `occscore`, which we use for a ranking of occupational status. These measures will supplement our occupational wage dataset, which provides wages at the sub-national level and we detail below. Other variables of interest include the household composition of each individual, the household demographics, their detailed location and whether or not they are currently unemployed or, in 1940, whether they work for the Works Progress Administration or for a private employer. We already have access to the complete count, full censuses for 1900 through 1940. We implemented the matching procedure of Abramitzky, Boustan and Eriksson (2012) to the prime-age (18-60 years old) white male population between 1920 and 1930. We obtained a match rate around 20%, which is standard in this historical literature.

A growing body of research uses matching procedures to create longitudinal data samples. Matching without unique identifiers requires a trade-off between sample size and false matches: a more permissive match produces a larger sample size with less accurate matches. Traditionally, the literature has taken a conservative approach and limited sample construction to exact (or near exact) matches on the following criteria: last and first names; age; race and place of birth. This was the approach of Long and Ferrie (2013) and has been followed closely by subsequent authors. Collins and Wanamaker (2014), for example, report a match rate below 21%. The main concern is that these low match rates may yield unrepresentative samples and introduce bias: if richer or more educated people report their age more consistently over time for example, the sample would be biased towards these types.

5.2. Occupational task measures from the *Dictionary of Occupational Titles*

We have information from the only historical description of workplace tasks, the *Dictionary of Occupational Titles* (1956), a dataset that was first introduced in detail in Gray (2013). The dataset contains a description of over 4,000 detailed occupations divided into 45 categories. These categories include the level of verbal, numerical, strength, dexterity, managerial and clerical skills needed to do a particular job, as well as measures of the intelligence, training and education required, and of the physical demands of a job including the exposure to noise and whether a job must be completed primarily outdoors. These measures can be easily combined to an aggregate index of tasks, or can be used to conduct factor analysis to determine which tasks are most important in any particular study. The categories are very similar or identical to those present in the more recent Dictionary of Occupational Titles datasets (mainly the 1977 version which was updated in 1991 and used in Autor, Levy and Murnane, 2003, and elsewhere), making the historical task dataset very comparable to those used for modern studies of the evolution of the skill distribution of the U.S. labor force.

The longitudinal census data described above contains information on occupations which is currently in a rough format, corresponding to the occstring variable reported in some IPUMS samples, which gives the response listed on the original census cards when respondants were asked about their occupations. We built a concordance to link occupations in their current format to *occ1950*, which is a standardized definition of occupation used in IPUMS samples. This is described in an appendix below. Finally, the longitudinal samples could be linked up to the task data.

5.3. Wages at the occupation- and city-level

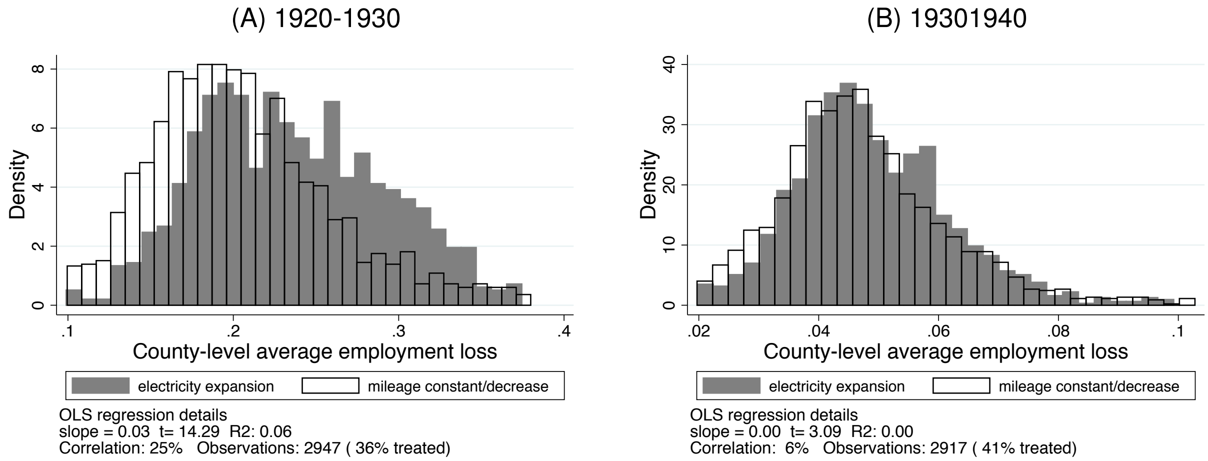
We match the task data from the *Dictionary of Occupational Titles* to the longitudinal information on occupations using a concordance built by Gray (2013) to produce a comprehensive understanding of task changes over the period 1920-1940. For the first time, we will also match it to occupational wage data at the state- or city-level to explore the overall impact on the wage distribution. Currently we have compiled data on occupational wages at the sub-national level using publications of the Bureau of Labor Statistics (covering the period 1900-1933) and the 1937 Ada Beney Cost of Living wage estimates. This dataset is the most detailed that can be constructed for the period and provides greater information on the evolution of relative wages for different classes of occupations over time. We argue that these are more information than the more aggregated measures used in Katz and Margo and the `occscore` measure used widely in the historical literature, which is informative but does not vary at the sub-national level. Our current wage sample allows us to look at the evolution of wages across X occupations during our sample period.

6. Empirical Analysis

Our baseline analysis investigates the differential change in various individual level occupational outcome variables between individuals residing in counties with and without electric grid expansion in a baseline year. That is, we seek to estimate individual level regressions models of the form

$$\Delta y_{i,j,t} = \beta_t EXP_{j,t} + X_{i,t} + Z_{j,t} + \epsilon_{i,j,t} \quad \text{for } t = 1920, 1930 \tag{1}$$

Figure 3: Grid Expansion and Employment Loss



Notes: The figures plot the density of county level employment-loss rates for counties with and without electric grid expansions. Panels A and B whot the reulst for the 1920-1930 and 1930-1940 samples, respectivel. For both periods we Windsorized the density to exlude outlier counties.

where i indexes workers, and j indexes counties, and t is the initial Census year. We match individuals across two consecutive Censuses to maximize the match rate. That is, for $t = 1930$, we work with our linked 1920-1930 Census panel; for $t = 1930$ with the linked 1930-1940 panel. Accordingly, $\Delta y_{i,j,t}$ is a stand-in for various measures of the individual level occupational change between two consecutive Censuses, $EXP_{j,t}$ and indicator variable for grid expansion in county j over the same period, while $X_{i,t}$ and $Z_{j,t}$ denote individual and location specific control variables.

As argued above, we believe that the geographical variation in $EXP_{i,t}$ is endogenous and we therefore instrument $EXP_{i,t}$ with geographical variation in the log of hydroelectric potential, as depicted in panel D of Figure 1.

Employment. Our first measure of occupational change is the probability of job loss. That is, we ask whether individuals who initially reside in counties with large electric grid expansions are more likely to lose their job, compared to those who live in counties without gird expansion. Figure 3 shows that, at least for the period 1920-1930, counties that experienced grid expansions had systematically larger job loss rates, as the gray density shifts toward the right. The corresponding OLS estimate of the differential impact on the mean job-loss probability is 0.03 and is statistically

Table 1: The Causal Effect of Electrification

	1920 - 1930			1930 - 1940		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Employment Loss						
Electric Expansion (0/1)	**0.0395*** **	**0.0304***	**0.0399*** **	**0.00914** **	**0.0110** **	**0.00977** **
	(0.0151)	(0.0157)	(0.0151)	(0.00444)	(0.00477)	(0.00455)
Constant	0.201***	0.218***	0.203***	0.0443***	0.0408***	0.0463***
	(0.0056)	(0.0073)	(0.00591)	(0.00188)	(0.00247)	(0.00166)
Observations	2770	2770	2770	2679	2679	2679
First Stage F-Stat	**56.87**	**51.55**	**56.79**	**41.23**	**37**	**39.7**
B. Occupational Upgrading (Change in IPUMS Occscore)						
Electric Expansion (0/1)	**0.0148**	**0.0656**	**0.025**	**0.189**	**0.423**	**0.238**
	(0.396)	(0.417)	(0.396)	(0.397)	(0.423)	(0.408)
Constant	2.725***	2.630***	2.776***	2.817***	2.302***	2.952***
	(0.147)	(0.194)	(0.156)	(0.167)	(0.221)	(0.151)
Observations	2768	2768	2768	2691	2691	2691
First Stage F-Stat	**58.58**	**53.21**	**58.53**	**51.34**	**46.91**	**49.24**
C. Angular Proximity of Initial and Final Occupation						
Electric Expansion (0/1)	**-0.0430*** **	**-0.0395*** **	**-0.0430*** **	**-0.0282*** **	**-0.0260*** **	**-0.0284*** **
	(0.00713)	(0.00739)	(0.00714)	(0.00633)	(0.00667)	(0.00642)
	-0.728	-0.668	-0.728	-0.569	-0.525	-0.573
Constnat	0.901***	0.895***	0.901***	0.884***	0.880***	0.883***
	(0.00269)	(0.00354)	(0.00283)	(0.00264)	(0.00348)	(0.00238)
Observations	2819	2819	2819	2790	2790	2790
First Stage F-Stat	**75.87**	**67.55**	**75.6**	**63.65**	**56.09**	**62.33**
Fixed effects		8 divisions	49 states		8 divisions	49 states

Notes: The table reports IV estimates for regression model (1) using geographic variation in hydro-electric potential as an instrument. Regressions (1) - (3) use the linked 1920-1930 panel and specifications (4) - (6) the 1930-1940 panel. The table reports results for three outcome variables: the probability of employment loss (panel A), the change in IPUMS occscore (panel B), and angular proximity of the initial and final occupation (panel C). Standard errors are reported in parentheses below each coefficient. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

significant. That is, on average, individuals who, in 1920, lived in counties that experienced electric grid expansions over the period 1920-1930 were 3 percentage points more likely to experience job loss. The less significant result for the period 1930-1940 is perhaps not surprising, given that this period includes the great depression in which electricity may not have had such a dramatic differential effect in light of the overall poor labor market performance.

While the density plots and OLS results in Figure 3 are indicative of a negative average employment impact on residents of newly electrified counties there are reasons to believe that these estimates are biased due to selection. To address this endogeneity problem, panel A of Table 1 reports IV estimates based on regression model (1) using geographic variation in hydro-electric potential as an instrument. The IV estimates resemble the general patterns observed in the reduced form evidence illustrated in Figure 3. Interestingly, while the point estimates during 1930-1940 are indeed small than during 1920-1930, our IV estimates suggest that the causal impact is significantly positive in both periods. That is, our estimates suggest that electrification caused, on average, a higher probability of job loss for individuals residing in the electricity adopting counties. Panel A of Table 1 further highlights that these estimates are robust to the inclusion of complete sets of Census division or state fixed effects.

Thus, these results suggest that electrification, on average, causes job loss within the initial residents of adopting counties. What happens to the individuals who actually remain employed? Do they remain in the same job? If so, do their wages go up or down? That is, are the tasks performed by the remaining workers valued more or less by the employer? To take a first stab at these questions we consider two additional measures of occupational change: first, we ask whether the individuals who are employed in both periods experience a "step up" or a "step down" in the earnings distribution. Second, we ask whether individuals who are employed in both periods are still performing the same job

Occupational upgrading. To measure occupational upgrading we now investigate the impact of electrification on an individual's the occupational score (IPUMS `occscore`)—a proxy for the occupational earnings distribution. In analogy to Figure 3, Figure 4 contrasts the density of the average individual level changes in the occupational score within counties that expanded their electrical grid with that in counties without grid expansion. These reduced form estimate suggest significantly less "job upgrading" among individuals who remain employed an originally resided in electrifying counties. However, in this case, it turns out that these effects disappear once we instrument with hydro-potential. In fact, Table panel B suggests that 1, not only are the effects statistically insignificant, the point estimates even change sign. Again, this finding is robust to the

Table 2: The Causal Effect of Electrification: Farm Workers

	1920 - 1930			1930 - 1940		
	(1)	(2)	(3)	(4)	(5)	(6)
Occupational Upgrading (Change in IPUMS Occscore) for Initial Farmers						
Electric Expansion (0/1)	**1.880*** **	**1.807*** **	**1.883*** **	**1.858*** **	**1.929*** **	**1.864*** **
	(0.36)	(0.378)	(0.361)	(0.426)	(0.46)	(0.432)
Constant	4.356***	4.479***	4.381***	5.964***	5.847***	6.002***
	(0.137)	(0.181)	(0.143)	(0.179)	(0.238)	(0.159)
Observations	2815	2815	2815	2776	2776	2776
First Stage F-Stat	**72.65**	**64.81**	**72.43**	**57.33**	**50.5**	**56.31**
Fixed effects		8 divistions	49 states		8 divistions	49 states

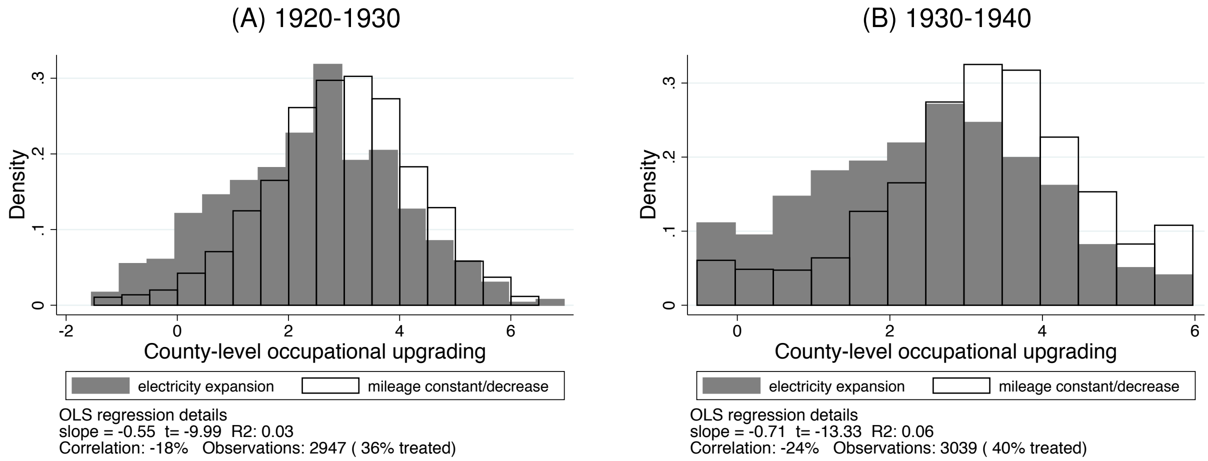
Notes: The table reports IV estimates for regression model (1) using geographic variation in hydro-electric potential as an instrument. Regressions (1) - (3) use the linked 1920-1930 panel and specifications (4) - (6) the 1930-1940 panel. The table reports results for in IPUMS occscore in analogy to panel B of Table 1 except we restrict the analysis to individuals initially in farm occupations only. Standard errors are reported in parentheses below each coefficient. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

inclusion of Census division and state fixed effects.

While these results suggest that electrification causes no systematic shift in the mean occupational score, it is still possible that there are "winners" and "losers" which cancel each other out. Our rich individual level sample allows us to focus even on narrow occupations and we therefore investigate the impact of electrification separately for occupations related to farming. Interestingly, Table 2 highlights that individuals who initially worked in farm occupations saw significant occupational upgrading. That is, farmers in electrifying counties who found a new job in the second period moved up approximately "two slots" (out of 60) on the occupational earnings distribution as measured by IPUMS's `occscore`. These results are not only robust to the inclusion of Census division and state fixed effects but are also effectively identical for the period 1920-1930 and 1920-1940. This suggests that electrification served as an amplifying factor for the widespread movement from the farm to the factory to during the 1920s and 1930s.

Job Similarity. Given that electrification appears to have caused at least some degree of movement between jobs, it is interesting to know how different the actual tasks performed by the worker are after the job transition. For example, does a hog farmer who previously used to slaughter his own hogs starts to work in an industrial meat processing factory—performing in large part similar

Figure 4: Grid Expansion and Employment Loss



Notes: The figures plot the density of county level employment-loss rates for counties with and without electric grid expansions. Panels A and B whot the reulst for the 1920-1930 and 1930-1940 samples, respectivel.

tasks that he used to perform as a hog farmer—or does he work in a cotton mill, performing completely different tasks. To analyze this question we match detailed task measures from the historical Dictionary of Occupational Titles (DOT) to the individual level Census data to compute the angular distance between the mix of tasks in each occupation. Intuitively, we represent an occupation as a vector of task characteristics from the DOT and compute the cartesian product of the initial an final occupation in order to ask "how different" the two occupations are. Panel C of Table 1 reports the IV results of the causal impact of electrification on this "task distance" measure. Notice that the dependent variable in this regression is zero if there is no change in the task mix, and is non-zero if there is a change in the task mix. Our estimates are all statistically significant and robust to the inclusion of Census division and state fixed effects.

Sectoral Transisions. Given the substantial amount of job transitions, we further analyze the likelihood with which electrification induces an individual to move from one sector to another. Table 3 reports the casual impact of electrification on the probability of workers moving from one sector to another. For example, looking at panel A, the coefficient in cell 1-1 suggests that workers formerly employed in the agricultural sector are 17 percentage points less likely to remain in an agricultural

Table 3: The Causal Effect of Electrification: Sectoral Transistions

A. Electricity Induced Sectoral Transition Probabilities during 1920-1930 (IV Estimates)

from / to		1	2	3	4	5	6	7	8	9
Agriculture	1	-0.17	0.06	0.08	0.07		-0.02			
Mining & construction	2	0.09		0.07	0.01					
Mfg, durables	3	0.22		0.22	0.07	-0.14			-0.26	
Mfg, non-durables	4						-0.11			
Transp. & utilities	5	-0.15		0.05						
Trade	6		0.05	0.03	0.05					
Fin., Ins., & RE	7							0.15		
Prof. & other services	8	-0.1		0.03	0.04					
Government	9				0.05		-0.12			

A. Electricity Induced Sectoral Transition Probabilities during 1930-1940 (IV Estimates)

from / to		1	2	3	4	5	6	7	8	9
Agriculture	1	-0.24	0.07	0.1	0.08		-0.03	0.01		0.02
Mining & construction	2		-0.09	0.05	0.06		-0.04			
Mfg, durables	3			0.28	0.14	0.04	-0.34	0.01		0.04
Mfg, non-durables	4				0.19		-0.24	0.02		
Transp. & utilities	5	-0.11		0.06	0.08					
Trade	6	-0.07		0.05	0.08				-0.04	
Fin., Ins., & RE	7				0.05					
Prof. & other services	8	-0.06		0.04	0.05		-0.06			0.03
Government	9				0.04					

Notes: The table reports IV estimates for regression model (1) using geographic variation in hydro-electric potential as an instrument. Each number reports the causal impact of electrification on the probability for a worker to move from industry A to industry B. All reported coefficients are significant at the 5% level.

job if they lived in a county that expanded its electricity grid over the period 1920-1930.

Table 3 reports a clearcut overall picture: electrification induced workers to leave agriculture (overall negative impact on column 1) and the majority of individuals are induced to move to the manufacturing sector (overall positive impact within columns 2, 3, and 4), regardless of their initial occupations. It is worth emphasizing that these results purely estimate the causal impact of electrification on the likelihood of moving sectors. This reinforces our earlier conclusion that electrification was a strong amplifier for the movement from the farm to the factory.

7. Concluding Remarks

We have used a longitudinal, comprehensive dataset following American workers between 1920 and 1930. To address concerns of self-selection of regions into the more productive electric technology (either through firms or workers), we instrument the adoption of electricity with the geography of ruggedness: more rugged regions are more likely to have hydroelectric power and to adopt electricity earlier; flatter regions are late adopters and catch up between 1920 and 1930. We find that hydroelectric power had little or no effect on the population at large but caused occupational change among laborers, giving substance to the usual claim that laborers were replaced by electrical machinery (Goldin and Katz, 2010).

We are still working on exploring other effects of electricity adoption, e.g. on occupational wages, the transferability of skills across jobs, and heterogeneous effects depending on a region's illiteracy and unionization or a worker's age and education.

We believe that looking at the past is a good guide to the future and that our final analysis using the panel from 1900-1940 will give a unique viewpoint for examining the historical evolution of the occupational structure. David (1990) argued that electrification and computerization followed similar patterns of technology adoption and productivity growth. Jovanovic and Rousseau (2005) consider "electricity and information technologies are probably the two most important general purpose technologies so far." Our work will strengthen this argument and shed light on the possible future of the employment structure.

References

Abramitzky R, Boustan LP, Eriksson K. 2012. Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review* **102**: 1832–1856.

Autor DH, Levy F, Murnane RJ. 2003. The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics* **118**: 1279–1333.
URL http://ideas.repec.org/a/tpr/qjecon/v118y2003i4p1279-1333.html

Baker E. 2007. *Displacement of Men by Machines - Effects of Technological Change in Commercial Printing*. Read Books. ISBN 9781406783674.
URL https://books.google.com/books?id=fUrnhC5pPxIC

Bessen J. 2011. Was Mechanization De-Skilling? The Origins of Task-Biased Technical Change. Working Papers 1101, Research on Innovation.
URL https://ideas.repec.org/p/roi/wpaper/1101.html

Collins WJ, Wanamaker MH. 2014. Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data. *American Economic Journal: Applied Economics* **6**: 220–52.
URL https://ideas.repec.org/a/aea/aejapp/v6y2014i1p220-52.html

Cortes GM. 2015. Where Have the Routine Workers Gone? A Study of Polarization Using Panel Data. Forthcoming in *Journal of Labor Economics*.
URL https://ideas.repec.org/p/man/sespap/1224.html

David P. 1990. The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox. *The American Economic Review* **80**: 355–361.

Devine WD. 1983. From shafts to wires: Historical perspective on electrification. *Journal of Economic History* **43**: 347–372.

Frey CB, Osborne M. 2013. The future of employment: how susceptible are jobs to computerisation? *University of Oxford Manuscript, access October* .

Goldin C, Katz L. 2010. *The Race between Education and Technology*. Harvard University Press.

Gray R. 2013. Taking technology to task: The skill content of technological change in early twentieth century United States. *Explorations in Economic History* **50**: 351–367.
URL https://ideas.repec.org/a/eee/exehis/v50y2013i3p351-367.html

Idaho National Engineering and Environmental Laboratory (INL). 1998. U.s. hydropower resource assessment. Technical report, Idaho Falls, ID.

Jovanovic B, Rousseau PL. 2005. General Purpose Technologies. In Aghion P, Durlauf SN (eds.) *Handbook of Economic Growth*. Elsevier, 1181–1224.

Katz LF, Margo RA. 2013. Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective. In *Human Capital in History: The American Record*, NBER Chapters. National Bureau of Economic Research, Inc, 15–57.
URL https://ideas.repec.org/h/nbr/nberch/12888.html

Long J, Ferrie J. 2013. Intergenerational Occupational Mobility in Great Britain and the United States since 1850. *American Economic Review* **103**: 1109–37.
URL https://ideas.repec.org/a/aea/aecrev/v103y2013i4p1109-37.html

Morin M. 2015. The labor market consequences of electricity adoption: concrete evidence from the Great Depression. *Manuscript, University of Cambridge* .

Salisbury L. 2014. Selective migration, wages, and occupational mobility in nineteenth century America. *Explorations in Economic History* **53**: 40–63.
URL https://ideas.repec.org/a/eee/exehis/v53y2014icp40-63.html

Severnini E. 2012. The Power of Hydroelectric Dams: Agglomeration Spillovers. *University of California at Berkeley, Manuscript accessed 30 August 2013* .

Solon G, Whatley W, Stevens AH. 1997. Wage Changes and Intrafirm Job Mobility over the Business Cycle: Two Case Studies. *ILR Review* **50**: 402–415.
URL https://ideas.repec.org/a/sae/ilrrev/v50y1997i3p402-415.html

Stewart JI. 2012. Migration to U.S. frontier cities and job opportunity, 1860–1880. *Explorations in Economic History* **49**: 528–542.
URL https://ideas.repec.org/a/eee/exehis/v49y2012i4p528-542.html

Appendix A. Details on matching occupations

This section briefly describes our method to link individuals across the 1920, 1930, and 1940 census. Each record i in time t has characteristics X_i (occupation string, industry string, quality flags) and a standardized occupation O_i (occ1950). We have this mapping for the 1940 complete count data and for the 1900-1940 1% or 5% samples. The aim is to extrapolate these mappings to the complete count for 1920 and 1930 (where we have occupation and industry strings, the original occupation and industry information entered on the census cards for each individual; 1900 and 1910 are pending). This will allow comparison of occupations over time and facilitate matching to the task data later. Each record has a weight h_i. In the original complete count data, $h_i = 1$, but without loss of generality, we consider the frequency table version of the complete count, collapsed by unique values of X_i and O_i, so h_i is the frequency of such occurrences.

To understand the approach, consider first only the string with the occupation, $X_i = \{occstr_i\}$. A given $occstr_i$ can match to a single O_i (it has no duplicates) or it can match to multiple O_i's (it has duplicates). The first case is the easiest: if it was all like this, we would have a perfect one-to-one cross-walk. To deal with the second, we consider only the most frequent match, i.e. the pair (occstr, occ1950) that is most representative of occstr. To do this, we sort the data in descending order of h_i and define coverage as:

$$coverage_X = \frac{\sum_{i \in first_X} h_i}{\sum_i h_i}$$

where $i \in first_X$ means that record i is the first record, or most frequent match, in the frequency

Table A.4: Linking Workers Across Censuses: Match Rates

	max rate	occstr		occstr and indstr	
		top match	top match +	top match	top match +
1900 5% sample	**49%**	49%	49%	-	-
1910 1% sample	**44%**	14%	14%	**44%**	44%
1920 1% sample	**40%**	13%	13%	**40%**	40%
1930 5% sample	**41%**	11%	11%	**39%**	39%
1940 1% sample	**35%**	10%	10%	**33%**	33%
1940 full count	**42%**	**20%**	17%	-	-

table collapsed by X.

The set of variables X can contain the four variables relating to occupation and industry: occstr, indstr, qocc and qind (the latter two are data quality flags for occstr and indstr).

Table A.4 below shows this coverage (the data quality flags qocc and qind add at most 1 percentage point and are omitted for clarity). The first column, "max" shows how much of population has non-missing information in occstr and in occ1950. Population is the denominator which facilitates comparison across years, so that measures are independent of how people who stayed at home were recorded, for example. This is the maximum match rate that we can achieve in a year. The second column shows how many individuals we keep by using only the most frequent match. The third column ("top match +") discards the fat left tail of the distribution, e.g. pairs of X-occ1950 that have less than N occurrences, where N depends on the sampling ($N = 1$ for the 1% sample, 5 for the 5% sample, and 100 for the complete count dataset).

The main thing to notice is that, aside from the 1940 complete count, the column "top match" is very close to "max": with (occstr, indstr) we can extract over 94% of the information available (94% = 33% / 35% in year 1940). Using a cross-walk compiled from the top match of the samples is the most efficient approach because it avoids having to manually match occstr_{1940} to $\text{occstr}_{t<1940}$ and still uses most of the information contained in occstr and indstr.

The second thing to notice is the low numbers for the 1940 full count: 20% as opposed to 33% from the 1% sample. In the worst case scenario, this could indicate that the IPUMS samples and the complete count dataset are not comparable and that the concordance built from the 1% or

5% samples does not apply to the full count. In the best case scenario, this discrepancy is due to indstr being missing from 1940 complete count. This is wrong: indstr is present in the 1% sample and should also be in the complete count. With indstr, the top match could be much higher. I am waiting to hear about this from IPUMS.

The 1900 sample is also missing indstr, but that doesn't affect the result, because that census only left space for occupation to be recorded by the enumerators, so that both occupation and industry information should be fully captured by occstr.

We inspected the top matches of each year manually and see no obvious bugs.