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

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

This version: November 24, 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 their adoption? We address 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. We follow individual workers over time by linking the 100% count Census of Population for 1920, 1930, and 1940 and study 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 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 U.S. Our results uncover the following insights: (1) on average, electricity has no effect on employment and earnings within the population at large; (2) causes substantial re-allocation of workers from the farm to the factory; (3) causes a significant upward movement in the earnings distribution for transitioning farm workers; (4) leaves average manufacturing wages unaffected despite a net increase in manufacturing employment, suggesting a significant increase in both the supply and demand for low-skill manufacturing labor.*

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 and wage structure has recently attracted considerable attention both among researchers and in policy circles. For example, **?** estimate that 47% of all current employees in the U.S. are susceptible to replacement by a computer. Similarly, **?** document a "hollowing out" of the skill distribution concurrent with widespread computer adoption since the 1970s and argue that a key contributor to the increased wage inequality witnessed in recent decades was technological change. In U.S. history, **?** 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.

However, despite the relevance of these concerns, there is little direct empirical evidence on the trajectories of individual workers directly affected by these new technologies. Rather, the literature has predominantly focused on analyses at an aggregate level, using repeated cross sections of data. This data limitation makes it inherently difficult to identify where "hollowed out" individuals end up or where newly attracted workers in booming industries came from, making a determination of the ultimate welfare implications of technological change difficult.[2]

This paper focuses on the most recent "technological revloution" 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. **?**, **?**, and **?**, 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

[2] A few recent papers do analyze longitudinal data (see, for example, **?**), but these studies are limited to using small, possibly biased samples—such as the small Panel Survey of Income Dynamics (PSID).

uniform timing across areas. While a number of existing contributions have studied the impacts of electrification on the labor market, these studies were either restricted to the manufacturing sector or use broad industry and state level evidence (e.g. **????**).

We provide important additional insights by expanding on these contributions among several dimensions: first, we use the 100% count of the decennial Census of Population for the years 1920-1940—a period of rapid electrification in the U.S.—and link individuals across Census years to create an individual-level panel that covers all industries, occupations, and narrow geographies in the U.S. Following the same workers over time is the key to studying the trajectory of workers within and between industries, occupations, and geographies in response to the rapid adoption of electrical machinery. Moreover, to study the changing demands for the tasks performed in the occupations affected by electrification, we combine this longitudinal dataset with high-quality information on detailed occupational characteristics from the historical *Dictionary of Ocupational Titles* (DOT). The construction of this dataset represents a significant innovation in its own right.

Second, we use newly digitized, detailed maps of the high-voltage electricity grid for the period 1908-1946 in order to measure the process of electricity adoption at the county level. Previous studies have measured electricity adoption with the amount of purchased electrical horsepower by manufacturing firms (e.g. **?**) which restricted analysis to the manufacturing sector only. In contrast, our detailed maps of electrical power lines, combined with our comprehensive individual-level panel, allows us to study the effects of electrification within and across all sectors and geographies—including farming, with its tremendous share of employment at that time.

Finally, using the geographic expansion of the electricity grid to measure electrification, rather than the quantity of purchased electricity by manufacturing firms, has an additional advantage: it allows us to exploit geographic variation in the potential to develop hydro-electric power—a mix of land gradient and stream flow—as plausibly exogenous variation to identify the causal impact of electrification on individual-level trajectories. This is an important aspect of our study, as geographical variation in the adoption of electricity is prone to a number of endogeneity concerns:

for example, 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 a spurious correlation between electricity adoption and wage upgrading that is due to the omitted variable of productivity or credit access. Our identification strategy is therefore an important improvement upon previous correlational studies.[3]

Based on the above setup and data we we provide a number of new and important insights into the effects of electrification in the U.S. We start with showing that, on average for the entire U.S. and all industries, individuals who live in counties that become electrified are neither more nor less likely to lose their job or to find a new job. This is in stark contrast to earlier studies and our own alternative estimates indicating a significant increase in employment within the manufacturing sector only.

To decompose this discrepancy between our aggregate effects and the effects within manufacturing, we estimate the transition probabilities between five major sectors, and find that electrification causes a net positive flow from agriculture into manufacturing and construction. These result identify electrification as one important cause for to move from the farm to the factory in the pre-WWII United States, but raise a number of complementary questions: are former farm workers displaced by electrical farm machinery or are they attracted by higher factory wages? Do factory workers face downward pressure on their wages, or does new electrical machinery allow for sufficient expansions in scale to offset the influx of farm workers into factory jobs?

We provide answers to these questions by first showing that farm workers transition into occu-

[3]Similar identification strategies have recently used by **?** and **?** in slightly different but related contexts.

pations with a higher "occupational score"—a ranking of occupations based on median earnings in 1950 and provided by IPUMS—, suggesting a potentially welfare improving reallocation of former farm workers to factory jobs. We then use county level data from the U.S. Census of Manufacturers to estimate the impact of electricity on both employment and wages within the manufacturing sector. Interestingly, we find a significant increase in employment but no effect on wages.

Taken together with our estimated re-allocation of farm workers to the factory, this suggests two major impacts on the manufacturing sector. First, electricity causes an influx of low-skill labor into the manufacturing sector, shifting the supply of manufacturing labor outward. Simultaneously, the efficiency gains due to new electrical machinery must also have increased the demand for low-skill labor in the manufacturing sector, resulting in a net increase in manufacturing employment but leaving wages constant.

Finally, we investigate changes in the task-mix performed by individuals based on high quality information on the tasks involved within each detailed occupation. Interestingly, we find that, despite substantial reallocation of workers across sectors, the task-mix performed by each workers remains unaffected. This suggests that farm workers moving to factory jobs did not need to upgrade their skills in order to receive the higher manufacturing wages made possible by the newly adopted electrical machinery, such as conveyor belts or continuous batch processing equipment (**?**).

In sum, we argue that our results are fully consistent with earlier evidence restricted to the manufacturing sector but shed additional light on three important complementary developments: first, electricity induced massive cross-sectoral re-allocation, far and foremost a move from the farm to the factory. Second, former farm workers did not upgrade their skill set but were nevertheless able to enjoy higher wages due to improved electrical machinery. And finally, wages in the manufacturing sector remained stagnant, due to a simultaneous outward shift in the supply *and* demand for low-skill manufacturing workers. Thus, the overall adoption of electrical machinery may mask more subtle disaggregate patterns where its labor-saving aspect releases labor from agriculture and its labor-augmenting aspect increases labor demand in manufacturing.

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 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 (**?**). In previous work (**?**) 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 **?** 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. **?** 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, **?** 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, **?** and **?** 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.[4]

To the best of our knowledge, the only previous work on job mobility and wage changes using longitudinal, historical data is **?**. 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

[4]Previous literature on worker displacement over this historical period has been limited to case studies, e.g. **?** for the printing industry.

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 economy as a whole.

Another related strand of the literature concerns the aggregate effects of electricity adoption. **?** 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.

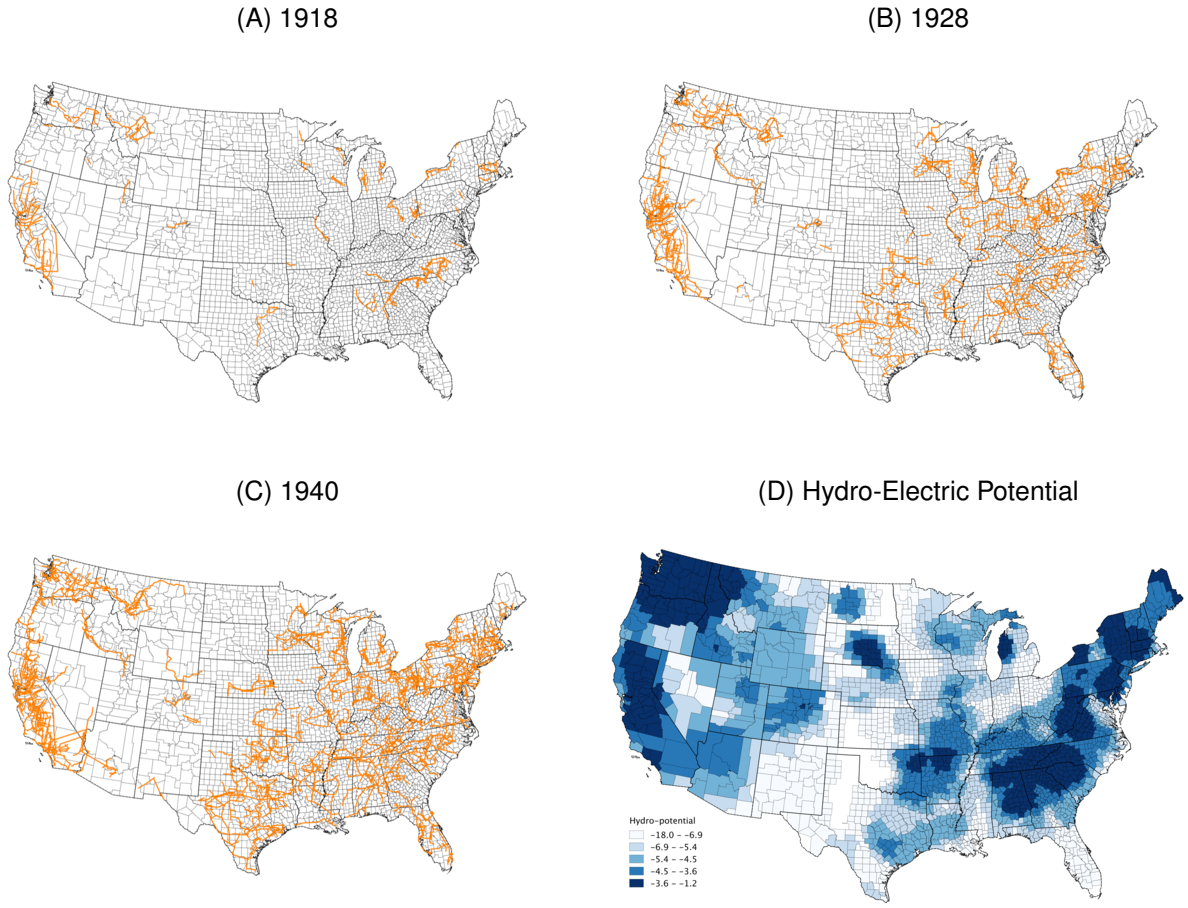
3. Measuring Electricity Adoption During 1920-1940

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. To overcome this hurdle, we assemble a novel and 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 high-voltage electricity grid (above 60kV) 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 high-voltage 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

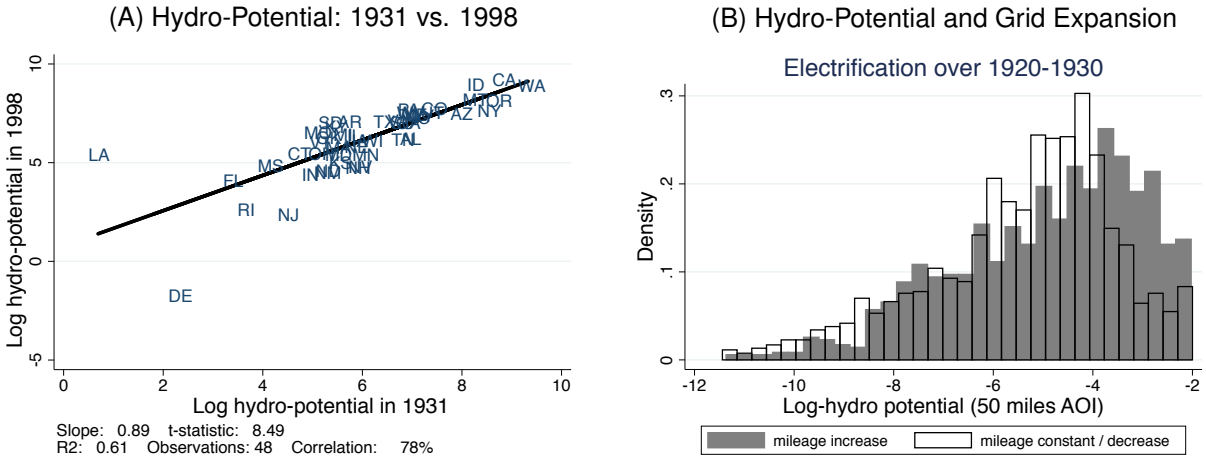
Perhaps the most challenging hurdle in existing studies was the lack of credible exogenous variation in the adoption of electricity across firms or industries. The geographic variation in grid expansion displayed in panels A-C of Figure 1 obviously also suffers from a variety of endogeneity concerns. 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.

In 1998, the **?** published 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 (**?**). 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, the geographic variation in this measure is essentially static over time. Panel A of Figure 2 confirms

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.

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.

Table 1 investigates this relationship more formally and reports our first-stage regressions for both the period 1920-1930 (panel A) and 1930-1940 (panel B). Specification 1 regresses an indicator variable for grid expansion on our log hydro-potential measure. Specifications 2-3 add region, division, and state fixed effects. For all four specifications we see a positive, highly significant correlation between hydro-potential with an F-Statistic of over 30 (1920-1930) and 24 (1930-1940) in the specification with 49 state fixed effects.

Table 1: Hydro Potential & the Electric Grid

	(1)	(2)	(3)	(4)	(5)
A. Increase in Grid Milage 1920-1930 (1=yes/0=no)					
Log Hydro Potential	0.0369***	0.0293***	0.0193***	0.0288***	0.0314***
	(0.00423)	(0.00453)	(0.00463)	(0.00526)	(0.00689)
Urban Density					0.127***
					(0.0124)
Obs.	2975	2975	2975	2975	2849
R^2	0.025	0.036	0.116	0.182	0.624
F-Stat.	77.79	42.26	17.58	30.22	20.76
Fixed Effects		4 regions	9 divisions	49 states	49 states
B. Increase in Grid Milage 1920-1930 (1=yes/0=no)					
Log Hydro Potential	0.0346***	0.0218***	0.0183***	0.0258***	0.0272***
	(0.00432)	(0.00458)	(0.00475)	(0.00543)	(0.00721)
Urban Density					0.127***
					0.0126
Obs.	2975	2975	2975	2975	2886
R^2	0.022	0.051	0.108	0.165	0.601
F-Stat.	66.87	23.61	15.59	24.09	14.25
Fixed Effects		4 regions	9 divisions	49 states	49 states

Notes: The table reports regressions of an indicator variable for the increase in electricity grid milage at the county level on our log hydro potential variable, log population density, as well as region, division, and state fixed effects. Panel A reports results for 1920-1930 and panel B results for 1930-1940. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

As our main goal is to study the effect of a access to electricity (supply of electricity) on various labor market outcomes, it is important to know to what degree higher density areas were naturally demanding more electricity. To account for this possibility, specification 5 controls for log population density and we see that the coefficient on hydro-potential is virtually unaffected by the inclusion of this variable. However, the predictive ability of this main specification (R^2) increases from around 17-18% to over 60%. This suggests that these two measures together with state fixed effects explain the majority of the variation in electric grid expansion.

Thus, we conclude that, after controlling for population density (capturing the higher demand for electricity), 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).

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 **?** to the prime-age (15-65 years old) white male population between 1920 and 1930. We obtained a match rate around 15%, 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 **?** and has been followed closely by subsequent authors. **?**, 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.

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 **?**. 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.

Wages and Employment in the Manufacturing Sector. Finally, we use the Census of Manufacturers to estimate average firm-level wages and employment within the manufacturing sector at the county

level. This additional datasource allows us to draw important comparisons between earlier work that exclusively focused on the manufacturing sector (e.g. **?**). In particular, it will allow us to contrast our aggregate effects for the united states, based on individual level trajectories, to the much more aggregate estimates that focus exclusively on the manufacturing sector.

6. Empirical Analysis

Our main empirical analysis investigates the differential change in various individual level occupational outcome variables between individuals residing in counties that do or do not experience a significant expansion in the electric grid. That is, we seek to estimate individual level regressions models of the form

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

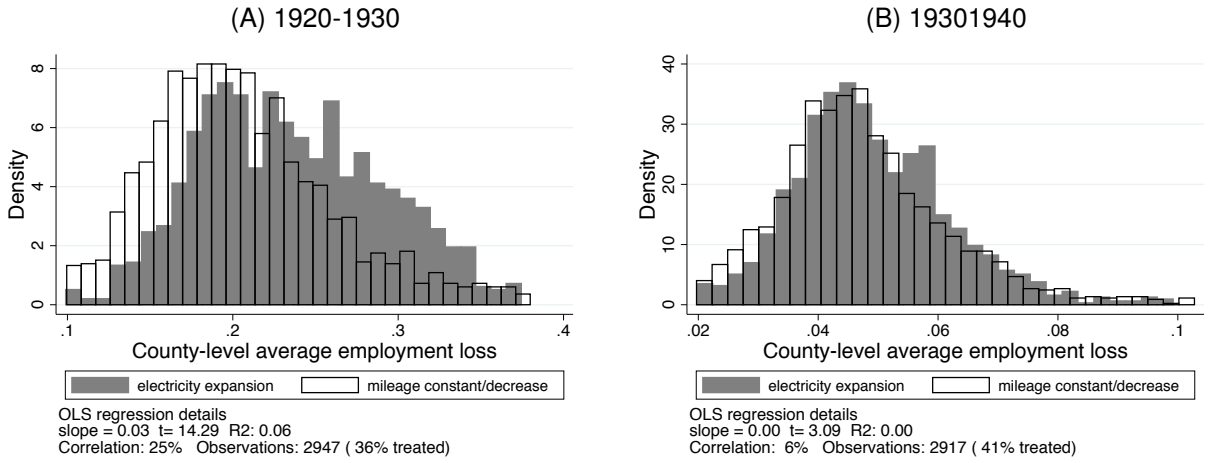
where i indexes workers, and j indexes counties, and t is the initial Census year. We match individuals across two consecutive Censuses to maximize the match rate. That is, for $t = 1920$, we work with our linked 1920-1930 Census panel; for $t = 1930$ with the linked 1930-1940 panel. Accordingly, $\Delta y_{i,j,t}$ 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. We argue in Section 4 that this instrument is both valid and strong.

6.1. Employment Loss/Gain

Our first measure of occupational change is the probability of job loss or gain. That is, we ask whether individuals who initially reside in counties with large electric grid expansions are more

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.

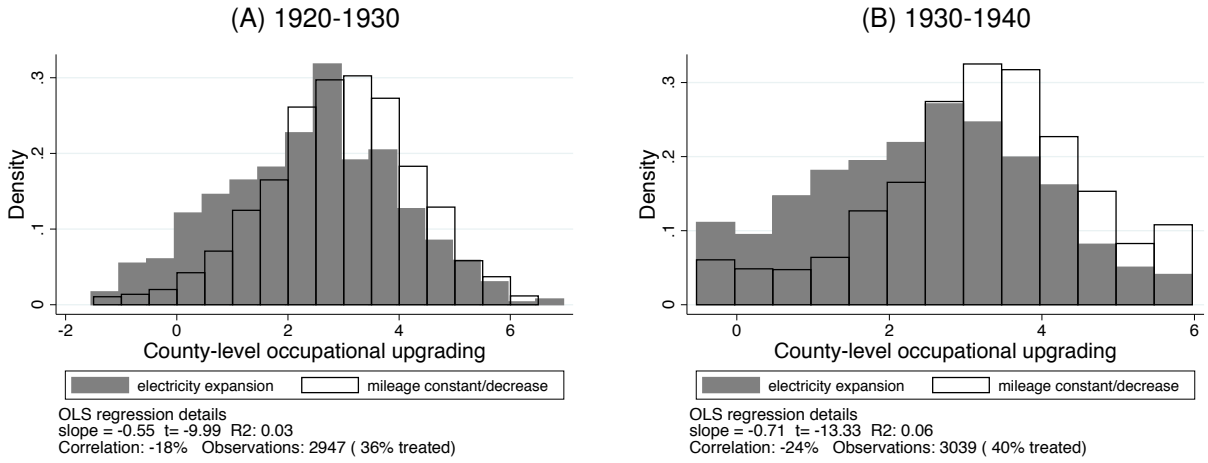
likely to lose their job, compared to those who live in counties without gird expansion. Figure 3 shows that, at least for the period 1920-1930, counties that experienced grid expansions had systematically larger job loss rates, as the gray density shifts toward the right. The corresponding OLS estimate of the differential impact on the mean job-loss probability is 0.03 and is statistically significant. That is, on average, individuals who, in 1920, lived in counties that experienced electric

Table 2: The Causal Effect of Electrification: Employmnet Loss/Gain

	(1)	(2)	(3)	(4)	(5)
A. Employment Loss					
1. 1920-1930					
Electric Expansion (0/1)	.0042**	-0.0001	-0.0021	-0.0025	-0.0011
	(.0018)	(.0012)	(.0015)	(.0016)	(.001)
Obs.	1355397	1355397	1355397	1355397	1341420
Counties	2872	2872	2872	2872	2847
First Stage F-Stat	5.2596	7.2677	7.6912	8.5459	12.364
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
2. 1930-1940					
Electric Expansion (0/1)	.0123**	.0159	.0115	.0208	.0206
	(.0061)	(.0108)	(.0105)	(.0163)	(.0134)
Obs.	3235395	3235395	3235395	3235395	3211050
Counties	2900	2900	2900	2900	2882
First Stage F-Stat	34.0056	12.1577	9.1805	5.862	7.7036
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
B. Employment Gain					
1. 1920-1930					
Electric Expansion (0/1)	-.0284*	-0.0013	0.0018	-0.0025	-0.006
	(.0157)	(.0057)	(.0056)	(.0076)	(.0074)
Obs.	7449	7449	7449	7449	7364
Counties	1407	1407	1407	1407	1391
First Stage F-Stat	5.4672	17.702	17.8658	12.0411	12.8029
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
2. 1930-1940					
Electric Expansion (0/1)	.0434	.2539	.2837	.3253	.4525
	(.0925)	(.2766)	(.2991)	(.5524)	(.5601)
Obs.	11214	11214	11214	11214	11127
Counties	1307	1307	1307	1307	1298
First Stage F-Stat	10.3367	1.7356	1.5965	.5554	1.0928
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes

Notes: The table reports IV estimates for regression model (1) using geographic variation in hydro-electric potential as an instrument. The table reports results for three outcome variables: the probability of employment loss (panel A), and the probability of employment gain (panel B). Standard errors are clustered on county and reported in parentheses below each coefficient. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Figure 4: Grid Expansion and occupational upgrading



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.

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, 2 reports IV estimates based on regression model (1) using geographic variation in hydro-electric potential as an instrument. In fact, the IV estimates confirm our concern of a spurious correlation and suggest that electrification, on average, neither had an effect on the probability that an individual loses a job nor on the probability that an initially jobless individual finds employment.

6.2. Occupational Change

While the above results suggest that electrification does not induce systematic movements out of employment, or into employment, they do not preclude that individuals may move into differ-

Table 3: The Causal Effect of Electrification: Occupational Upgrading/Downgrading

	(1)	(2)	(3)	(4)	(5)
A. 1920-1930: Occupational Upgrading (Change in IPUMS Occscore)					
Electric Expansion (0/1)	-0.8538	0.9101	2.1501	2.3732	0.3223
	(1.5682)	(1.78)	(2.1247)	(2.2178)	(.8732)
Obs.	1352364	1352364	1352364	1352364	1338418
Counties	2872	2872	2872	2872	2847
First Stage F-Stat	5.2755	7.2728	7.6946	8.5472	12.3873
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
B. 1930-1940: Occupational Upgrading (Change in IPUMS Occscore)					
Electric Expansion (0/1)	-.2743	1.7983	2.1832**	3.1179**	-.2434
	(.9283)	(1.2545)	(1.044)	(1.5062)	(.6323)
Obs.	3064060	3064060	3064060	3064060	3040910
Counties	2900	2900	2900	2900	2882
First Stage F-Stat	33.8499	12.0349	9.1072	5.9037	7.7539
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes

Notes: The table reports IV estimates for regression model (1) using geographic variation in hydro-electric potential as an instrument. The table reports results for the change in IPUMS occscore for the periods 1920-1930 (panel A) and 1930-1940 (panel B). Standard errors are clustered on county and reported in parentheses below each coefficient. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

ent occupations. In fact, previous work suggests that electrical machinery replaced a number of artisanal tasks in favor of low-skill assembly line work. 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 tasks in their job.

To measure occupational upgrading/downgrading 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

Table 4: 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
Agriculture	1	-.08**	0.02	.04**	.07***	0.01
Mining & construction	2	0.12	-0.11	0.1	0.03	-0.01
Mfg, durables	3	0.21	-.07*	.21*	-0.01	-.06*
Mfg, non-durables	4	.05*	-0.02	-0.01	.1*	-0.01
Transp. & utilities	5	-0.01	0.01	0.02	0.01	0.07

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

from / to		1	2	3	4	5
Agriculture	1	-.17**	.13***	0.02	.07***	-0.01
Mining & construction	2	.08**	0.01	-.06**	0.01	0
Mfg, durables	3	.13**	.11*	-.31**	0.02	-0.01
Mfg, non-durables	4	.11*	0.02	-0.09	0.19	0
Transp. & utilities	5	0.07	0.04	-0.02	0.03	-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.

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 3 suggests that, not only are the effects statistically insignificant, the point estimates even change sign. Again, this finding is robust to the inclusion of Census division and state fixed effects.

6.3. Sectoral Transitions

While we have so far shown that electrification, on average, neither caused job loss or gains, nor occupational up- or downgrading we investigate further why previous studies have found substantial changes along all of these dimensions when investigating exclusively the manufacturing sector. One aspect that these studies were unable to investigate are transitions in and out of the manufacturing sector, which, in principle, may even out any effects on employment and wages in the aggregated.

To investigate this possibility we next analyze the likelihood with which electrification induces an individual to move from one sector to another. Table 4 reports the casual impact of electrification on the probability with which a worker moves 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 8 percentage points less likely to remain in an agricultural job if they lived in a county that expanded its electricity grid over the period 1920-1930.

Table 4 reports a clearcut overall picture: on average, electrification induced workers to leave agriculture and the majority of these individuals are induced to move into the manufacturing sector. It is worth emphasizing that these results estimate the causal impact of electrification on the likelihood of moving sectors, suggesting that electrification was a strong amplifier for the movement from the farm to the factory, despite our estimates of a net zero effect in the aggergate.

6.4. Farm Workers

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, panel A of Table 5 confirms the aggregate net zero effect on employment also within individuals initially employed in farm occupations.

However, panel B 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 1930-1940. 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.

Table 5: The Causal Effect of Electrification: Farm Workers

	(1)	(2)	(3)	(4)	(5)
A. Farm Workers: Employment Loss					
1. 1920-1930					
Electric Expansion (0/1)	0.0004	0.0019	0.0001	-0.0012	-0.0035
	(.001)	(.0015)	(.0032)	(.0025)	(.0042)
Obs.	616492	616492	616492	616492	612298
Counties	2868	2868	2868	2868	2843
First Stage F-Stat	44.5296	17.6093	5.2849	10.1322	4.6488
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
2. 1930-1940					
Electric Expansion (0/1)	.0126	.0091	.007	.046	.0495
	(.0126)	(.0216)	(.0367)	(.0371)	(.045)
Obs.	1041785	1041785	1041785	1041785	1034992
Counties	2898	2898	2898	2898	2880
First Stage F-Stat	81.8328	26.0716	10.9643	11.1149	7.7606
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
B. Farm Workers: Occupational Upgrading (Change in IPUMS Occscore)					
1. 1920-1930					
Electric Expansion (0/1)	2.6452***	1.9992***	3.2423**	2.2206**	2.9004*
	(.4152)	(.5823)	(1.5515)	(.8834)	(1.6337)
Obs.	615450	615450	615450	615450	611261
Counties	2868	2868	2868	2868	2843
First Stage F-Stat	44.489	17.6057	5.2874	10.1403	4.6338
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
2. 1930-1940					
Electric Expansion (0/1)	2.449***	1.968***	2.2751**	2.1834**	2.1371*
	(.3537)	(.5565)	(.9483)	(.9406)	(1.1121)
Obs.	985799	985799	985799	985799	979341
Counties	2897	2897	2897	2897	2879
First Stage F-Stat	81.8661	25.9925	10.8077	11.1523	7.8346
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes

Notes: The table reports IV estimates for regression model (1) using geographic variation in hydro-electric potential as an instrument. The table reports results for three outcome variables: the probability of employment loss (panel A), the probability of employment gain (panel B), and the change in IPUMS occscore (panel B). Standard errors are clustered on county and reported in parentheses below each coefficient. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

6.5. 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 start to work in an industrial meat processing factory—performing in large part similar 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. Table 6 reports the IV results of the causal impact of electrification on this "task distance" measure, both for all workers (panels A.1 and B.1) and farm workers only (panels A.2 and B.2). Interestingly, we find that even for farm workers, the task mix performed on the job remains essentially unaffected after electrification. This is consistent with the idea that farm workers transition from performing low-skill farm tasks to low skill tasks at conveyor belts in newly electrified factories.

6.6. Employment & Wages in the Manufacturing Sector

Our analysis so far have has provided two main results: first, the aggregate structure of both wages and employment appear, on average, unaffected by electrification. At the same time, we see substantial re-allocation of former farm workers into the manufacturing sector that coincides with net gains in earnings for the individuals moving from the farm to the factory. How can these two facts be reconciled with one another and the existing evidence of a substantial correlations between electricity adoption and movements in the wage and employment structure within manufacturing (e.g. **??**)?

To address this question we employ county level data from the historical Census of Manufac-

Table 6: The Causal Effect of Electrification & Task Similarity

	(1)	(2)	(3)	(4)	(5)
A. 1920-1930: Task Similarity (DOT, Angular Proximity)					
1. All Workers					
Electric Expansion (0/1)	-.097***	0.0207	0.0476	0.0471	0.0004
	(.0322)	(.043)	(.0478)	(.0458)	(.0139)
Obs.	1295177	1295177	1295177	1295177	1282325
Counties	2872	2872	2872	2872	2847
First Stage F-Stat	5.6207	7.3665	7.6323	8.4225	12.3455
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
2. Farm Workers					
Electric Expansion (0/1)	-.046***	-.0258***	-.0309*	-.0173**	-0.0166
	(.0064)	(.0073)	(.0161)	(.0087)	(.0139)
Obs.	607146	607146	607146	607146	603029
Counties	2867	2867	2867	2867	2842
First Stage F-Stat	44.5474	17.65	5.3365	10.1614	4.6571
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
B. 1930-1940: Task Similarity (DOT, Angular Proximity)					
1. All Workers					
Electric Expansion (0/1)	-.0257	.0197	.0226	.0366*	-.011
	(.0191)	(.0215)	(.0159)	(.0215)	(.0098)
Obs.	2154855	2154855	2154855	2154855	2137601
Counties	2899	2899	2899	2899	2881
First Stage F-Stat	33.7303	13.6528	10.0343	7.0861	8.8235
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes
2. Farm Workers					
Electric Expansion (0/1)	-.0454***	-.021***	-.0096	-.0156	-.0104
	(.0058)	(.0067)	(.0093)	(.0103)	(.0111)
Obs.	769390	769390	769390	769390	763905
Counties	2896	2896	2896	2896	2878
First Stage F-Stat	74.6595	26.6902	12.0269	11.1648	8.0421
Fixed Effects		4 regions	9 divisions	49 states	49 states
Population Density					yes

Notes: The table reports IV estimates for regression model (1) using geographic variation in hydro-electric potential as an instrument. The table reports results for the change in angular proximity of tasks both for farm workers (panels A.1 and B.1) and farm workers only (panels A.2 and B.2). Standard errors are clustered on county and reported in parentheses below each coefficient. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

turers and analyze the causal effect of electrification on average employment and wage growth within the manufacturing sector. Table 7 reports our results and reveals two illuminating insights: first, electrification causes a net increase in employment growth, consistent with our finding of a substantial influx of farm workers into manufacturing. However, while we find that the former farm workers who move into manufacturing experience an increase earnings, we find no effect on average wage growth within manufacturing in neither period. This suggests that the newly adopted electrical machinery, such as the conveyor belt or batch processing equipment (**?**), must have caused a sufficiently large productivity gains to cause a substantial outward shift in the demand low-skill work, keeping wage growth unaffected.

7. Concluding Remarks

We have used a longitudinal, comprehensive dataset to trace the trajectory of American workers between 1920 and 1940. 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 hydro-electric potential.

Our main results suggest that electrification hat little to no effect on the population at large but caused substantial re-allocation of low-skill labor from farming into the manufacturing sector. This re-allocation caused a substantial job upgrading for the transitioning farm workers but left average wages in manufacturing unaffected. This suggest that the productivity gains due to new electrical machinery were substantial enough to cause an increase in the demand for low-skill conveyor belt work that exactly offset the additional supply of farm workers transitioning into manufacturing.

We therefore conclude that the "labor saving" nature of new equipment in the electrifying pre-WWII factory was likely exaggerated, as we find no effect on wages within manufacturing, despite a tremendous influx of farm labor. Given the sheer size of labor force in farming during the 1920s, it is conceivable that the sectoral re-allocation we estimate here may have been net welfare improving as we find substantial job upgrading within the transitioning farm workers without significant

Table 7: The Causal Effect of Electrification on Manufacturing Employment & Wages

	(1)	(2)	(3)	(4)	(5)
A. 1920-1930: Manufacturing Sector					
1. Wages					
Electric Expansion (0/1)	-0.316***	-0.184*	-0.224	0.0561	0.0500
	(0.0820)	(0.103)	(0.168)	(0.121)	(0.129)
Initial Population Density					0.0104
					(0.0148)
Obs.	1,847	1,847	1,847	1,847	1,847
Fixed effects		4 regions	9 divisions	49 states	49 states
2. Employment					
Electric Expansion (0/1)	0.361**	0.965***	1.401**	0.536	0.615*
	(0.145)	(0.313)	(0.617)	(0.344)	(0.373)
Initial Population Density					-0.134***
					(0.0429)
Obs.	1,847	1,847	1,847	1,847	1,847
Fixed effects		4 regions	9 divisions	49 states	49 states
A. 1930-1940: Manufacturing Sector					
1. Wages					
Electric Expansion (0/1)	0.317***	0.773***	0.615*	0.391	0.390
	(0.0689)	(0.290)	(0.334)	(0.246)	(0.256)
Initial Population Density					0.00100
					(0.0219)
Obs.	1,660	1,660	1,660	1,660	1,660
Fixed effects		4 regions	9 divisions	49 states	49 states
2. Employment					
Electric Expansion (0/1)	0.802***	1.855***	1.997**	1.391**	1.394**
	-0.184	-0.643	-0.859	-0.623	-0.615
Initial Population Density					-0.152***
					-0.0562
Obs.	1835	1835	1835	1835	1835
Fixed effects		4 regions	9 divisions	49 states	49 states

Notes: The table reports IV estimates for county level regression simlar to model (1) based on the Census of Manufacturers. We again exploit geographic variation in hydro-electric potential as our instrument. The table reports results for two outcome variables: percent growth in average firm level wages (panels A.1 and B.1) and percent growth in average firm level employment (panels A.1 and B.1). Standard errors are reported in parentheses below each coefficient. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

downward pressure on average wages within manufacturing or net gob gains/losses within the population at large.

References

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 collapsed by X.

Table A.8: Linking Workers Across Censuses: Match Rates

	max	occstr		occstr and indstr	
	rate	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%	-	-

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.8 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.