THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Essays on the Impacts of Weather Shocks and Electrification

Evidence from the early Twentieth Century U.S.

A thesis submitted to the Department of Geography and Environment of the London School of Economics for the degree of Doctor of Philosophy

Giorgia Cecchinato

June 30, 2023

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work. The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made.

This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 34,000 words, excluding appendixes, footnotes and bibliography.

Acknowledgements

I wish to express deep appreciation to my first supervisor Roger Fouquet who never stopped encouraging me during this journey. I appreciated his insights, but also his genuine generosity with time and patience. I also want to thank my second supervisor Giles Atkinson, who listened to every concern I had and always had a solution. Without their support, this thesis would not have been possible. I am also grateful to many people at the Geography and Environment Department, especially Ben Groom and Simon Dietz, who followed my PhD journey and ensured I was on track.

Given my economic history research focus, I have always felt a bit out of place, but everyone at the Grantham Institute always made me feel included. I really enjoyed my time sitting at the tenth floor office, I thank everyone I shared a meal or a coffee with. In particular, I want to thank my amazing friend and colleague Louise with whom I shared many fond memories.

Working and studying at LSE was only possible because of the support of the Economic and Social Research Council, so I thank them for allowing me to spend five - challenging - but extremely rewarding years in London.

My closest friends and family always made me smile, regardless of my regressions' results. Special mention goes to Augustina, Laila, Maria and Silvia, Jess, Liz and Max T. But also Robin, Carl, Volpe and Lupo. I am most grateful to my family for supporting my choices from the start.

Finally, and most importantly, I want to thank my best friend and partner Max, who consistently made every single struggle a little bit lighter.

Abstract

This thesis combines historical research with economic theory and machine learning methods to explore the impact of electricity access and weather variability on labour markets. Its main contribution is the creation of two large datasets of local electricity infrastructure in the U.S. during the early twentieth century. The thesis is organised into four chapters: an introduction, two chapters on electricity access and one on weather variability. The second chapter describes the creation of a new dataset about households' access to electric lights at the city and county levels. It looks at the differences in spatial patterns between the diffusion of arc and incandescent lights. The third chapter looks at the impact of rural electrification on local development in the Western U.S. between 1900 and 1930. Data are from a new dataset about transmission lines digitised from engineering journals. Identification is possible due to the peculiar geography of the Western U.S., where transmission lines crossed rural communities, resulting in exogenous variation in electricity access. This paper concludes that there was a small but significant impact on land values between 1900 and 1910. The fourth chapter looks at the impact of an increase in excessively high temperatures on the transition from agricultural to non-agricultural occupations. In line with the existing literature, I find that there is no effect on employment status but that areas with higher weather variability experience a more accentuated transition into the secondary sector. This thesis ends with a conclusion which summarizes the main findings of each chapter, suggests implications for current policy design, and highlights possible avenues for further research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis sits at the intersection of economic history, environmental economics, and spatial economics, with a particular focus on the United States in the 20th century. Utilizing economic analysis and informed by historical context, it seeks to understand the influence of electricity access expansion and climatic shocks on the local economy. Through three independent papers, this thesis unfolds as a multi-faceted narrative that contributes to the broader discourse on the role of technological and environmental factors in shaping economic development.

This research centers on the intersection of environment and technology and their role in spurring economic development. Two seminal works served as the foundation for my research interests. Thomas Hughes's "Networks of Power" (Hughes, 1993) provides a comprehensive understanding of the transformative role of electricity and its economic implications during the late 19th and early 20th centuries. This narrative of technological innovation and diffusion, with the history of the electrical transmission system at its core, inspired the investigation of the economic repercussions of electricity access expansion presented in this thesis.

In parallel, William Cronon's "Nature's Metropolis" (Cronon, 1991) uncovers the intricate relationship between urban and rural spaces, mediated by technology - the railroad - and environment - the arid western frontier. His focus on how the development of an urban center like Chicago depended on its geographical position, as well as on the development of connections between center and periphery, crystallized my interest in the interplay of geography and technology.

Both books make use of historical data to explore the importance of infrastructure, either in the form of the transmission or railroad system and to reflect on the impacts that absolute and relative location in space have on economic outcomes. My papers, on the other hand, make use of spatial variation to measure the impact that technology - access to electricity - or environment - weather shocks - have on local economic outcomes.

1.1 Summary of the three papers

This thesis contains three independent papers. The first two papers (Chapter 2 and chapter 3) look at the expansion of electricity access and its impacts, and the third paper (Chapter 4) examines the impact of weather shocks on the labour market. All three papers make use of data from the 20th century U.S.

The second chapter (2) introduces a new dataset about the diffusion of electric lights in the U.S.. Electric lights were among the first experiences households had with electricity. The early phase of the household electrification process occurred between

1900 and 1920 but data about the adoption of incandescent and arc lights before 1920 has been only available at a regional level. This paper provides a new county-level data set of the diffusion of incandescent and arc lights before 1920. The paper shows how counties in New England and the Pacific - especially California - led the electrification process in the U.S. This may well have given these two states a head-start in boosting welfare and economic development, with potentially long-term advantages.

The third chapter (3) looks at the process of rural electrification. Between 1900 and 1930, while the rest of the United States struggled to extend rural power lines, the Western States had already started the process of rural electrification. The aim of this paper is twofold: to provide a narrative for the expansion of electric services in the early 1900s and to investigate whether electricity access had an impact on the local economy. Using distance to the nearest hydroelectric power plant as a proxy for electrification rates, I find that there are no major effects, if not for an increase in land values between 1900 and 1910, possibly driven by the adoption of electric lights. I conclude that rural electrification impacted the rural economy only after the 1930s, possibly due to complementarity with new technologies and additional investment from the government.

The fourth chapter (4) explores the effects of weather shocks on employment outcomes. This paper aims to analyse the impacts of the U.S. Corn Belt droughts in the 1930s on the labour market, with the purpose of learning about mid- to long-term adaptation strategies of individuals hit by weather shocks. It first measures the impact of the droughts on corn yields between 1930 and 1940 using county-level data. Then, it uses full census-linked data for 2.2 million individuals from 1930 and

1940 to measure the aggregate impacts of the 1930s drought on the probability of being employed in 1940 and on wages. The results confirm that harmful degree days reduce employment in the agricultural sector and increase it in the non-agricultural sector. Also, it finds that a negative weather shock increased the number of people working for wages rather than as owners. For individuals who migrated to a different state, effects are also negative but less significant and smaller, with a positive impact on wages. Overall, the results suggest that adaptation has the potential to lead to better outcomes in the long run and highlight the importance of policies aiding individuals to adapt to the harmful impacts of climate change.

1.2 Economic history and microeconomics

This thesis aims to combine the perspective of economic history and the methods of modern microeconomics in three different ways. To conduct research in economic history effectively, it is crucial to assemble suitable primary sources and create new data when required. The process of discovering novel historical data provides researchers with essential information to understand long-term economic patterns. Examples of similar work are the studies by Donaldson and Hornbeck (2016) and Hornbeck (2012), in which the authors generated new quantitative data from primary sources using Geographical Information Systems (GIS). They digitized maps of railroads to measure market access or maps of counties affected or not by severe droughts to quantify the impacts of weather shocks. In a more recent study by Atack et al. (2022), the authors digitized 1899 data about hand and machine labor. In a similar vein, Chapter 2 of this thesis utilizes a new Optical Character Recognition (OCR) algorithm to parse primary sources into digital data and introduces a new dataset on the diffusion of electric lights in the United States. Chapter 3 utilizes GIS to digitize new maps about the transmission system of western US states.

Secondly, historical analysis is instrumental in understanding the nature of enduring processes like technological diffusion. By examining past phenomena and how they unfolded, we can discern patterns and mechanisms that persist over time. For example, the study of David (1990) who looked at the array of factors which needed to be in place for the dynamo to increase productivity. Or more recently, the study about capital-skill complementarity and its changes of Lafortune et al. (2019). This understanding is at the heart of chapter 3, which investigates the expansion of rural electrification in the U.S and analyzes this phenomenon as a long-term process which proceeded in a hetereogeneous and non-linear pattern in both space and time.

Finally, historical periods, with their unique circumstances, offer valuable testing grounds for economic theories. Historical events might be used as a source of exogenous variation, making it easier to isolate the variable of interest from other variables which might be influencing the outcome of interest. This approach reflects the quasi-experimental studies of Hornbeck (2012) who looked at the event of the Dust Bowl as an exogenous shock to measure the impact of climate change on land values. Similarly, Baker et al. (2018) who measured the impact of agricultural shocks - as proxied by the arrival of a pest - on educational attainment. In a similar vein, Chapter 4 leverages the U.S. Corn Belt droughts as an exogenous shock to test theories about labor market adaptation to environmental changes.

1.3 Technology adoption and adaptation in rural areas

Secondly, this thesis aims to contribute to the discussion on the drivers of rural development, either through technology adoption or through adaptation to weather shocks.

Technological change is essential to economic growth and development. Literature offers numerous instances where infrastructure investment and expansion have catalyzed local economies. Historical examples are the New Deal Rural Electrification Administration (Fishback et al., 2005) and the Tennessee Valley Authority project (Kline and Moretti, 2014), both of which demonstrate the positive impacts of infrastructure-driven technology adoption on local economies. More specifically, the linkage between energy access and economic development is well-established in the economic literature. A host of research, focusing particularly on the effects of electrification, underscores this relationship. In the realm of labor economics, access to electricity has been shown to affect labor demand and bargaining power. Specifically, rural regions in the U.S. witnessed increased labor demand following electrification (Kitchens and Fishback, 2015). Electrification also induced structural changes in employment patterns. Notably, studies from the U.S. (Gaggl et al., 2021) have documented a decrease in the employment share of the agricultural sector following electrification, indicating a transition towards non-agricultural employment. Furthermore, the long-term impacts of electrification have been substantiated by studies showing its persistent effects on economic growth and productivity. For instance, rural U.S. counties that gained early access to electricity experienced sustained economic growth compared to counties that gained access later (Lewis and

Severnini, 2017). Similarly, from a firm-level perspective, Fiszbein et al. (2020) reported a significant rise in labor productivity rates in the U.S. manufacturing sector with the advent of electricity. Chapter 2 and chapter 3 contribute to this line of research.

On the other hand, chapter 4 looks at another aspect of rural development: adaptation to extreme weather shocks. Studies which use historical data to explore this topic are numerous, the best examples being the research about the impacts of the Dust Bowl by Hornbeck (2012) and Long and Siu (2008) and the more recent study of Burke and Emerick (2016) looking at long-term changes of weather on U.S. agriculture. More specifically, scholarship from development economics looks at how households implement various strategies in response to extreme weather events (Rose, 2001). Temporary adjustments to consumption and saving patterns often ensue (Di Falco et al., 2011), alongside modifications in agricultural practices and input use (Aragón et al., 2021). Changes in trade and migration patterns also occur (Feng et al., 2015). A less explored aspect of adaptation, however, is sectoral reallocation (Jessoe et al., 2018; Branco and Féres, 2021), particularly the direct impact of weather on labour supply (Hill et al., 2021), which is what chapter 4 looks at.

Chapter 2

A County-Level Dataset of Electrification and Lighting Diffusion in the U.S., 1900-1920

2.1 Introduction

Between 1900 and 1920, the majority of people living in urban areas of the U.S. gained access to electricity in their own homes. Before then, electricity was considered an expensive alternative to gas and kerosene lamps, the leading lighting technologies at the time (Bright, 1949; Muller, 2016). Having access to electricity was a transformative experience for any household, of a magnitude similar to what was experienced in the factory (Fouquet and Pearson, 2006). Electric lights were cleaner and safer than gas lights and required less maintenance. Electric appliances significantly reduced the amount of drudgery experienced in housework, altering working relationships within the house (Cowan, 1983).

Until now, the only available data about electricity diffusion in the U.S. were derived from official state-level data [1] or short household surveys scattered around several government publications and engineering journals (e.g. Electrical World, 1912; 1920). It was only after 1930 that county-level data was collected. This was done solely for the purpose of informing the Rural Electrification Administration through local surveys conducted by the state. At this time, most urban households were already wired. At the national level, the first county-level record of dwelling with electrical appliances is from the 1940 Census. Researchers trying to estimate the impact of electricity diffusion in the early decades of the 20th century had to resort to proxies, such as the distance to power stations, while referring to the 1940 census data as validation [2].

Having discovered two detailed sources of information about electrification and lighting in the early twentieth century, in this paper, I describe the creation of a new data set about electricity diffusion at the county level for the years 1900, 1910 and 1920. It is anticipated that this new dataset will help scholars to better understand the process of electrification in the U.S. and enable them to analyse this technological diffusion with other key economic, social and political variables.

In the next section, I briefly outline the historical context by describing the process of electricity diffusion from the early 1880s to the 1920s. This is based on historical

---

[1] State-level data are available in the quinquennial U.S. special census of the electrical light and power industry between 1902 and 1937 (U.S. Bureau of the Census, 1907, 1912, 1917) digitized by John L. Neufeld and hosted online by the Economic History Association at this link.

[2] For example, the work by Severnini (2014); Lewis and Severnini (2017); Lewis (2018); Kitchens and Fishback (2015). There are two publications which have created their own dataset while I was drafting this thesis: Vidart (2020); Fiszbein et al. (2020).

examples and regional-level data and is discussed within the context of previous literature. Next, I go over the creation of the new dataset. First, I describe Powers and McGraw-Hill central station directories within the context of business directories as historical sources. Then, I explain the combination of algorithms and procedures used to extract information from primary sources. Finally, as a validation step, I compare my data with available sources and adjust for errors.

In section 2.4 I provide a brief analysis of the new data. In particular, county-level observations make the spatial variation in electricity diffusion much more precise: densely populated areas - such as the major cities of New England, California and the Midwest - had values for incandescent light diffusion of at least one order of magnitude bigger than the average county. California and other areas of the West were particularly exceptional, showing high numbers of per capita diffusion early in the century, possibly due to the lower cost of electricity generated by hydroelectric power plants. Arc lights per capita follow a different diffusion pattern, mostly correlated with the presence of industrial manufacturers. In section 2.5, I provide a summary of the potential uses of this new dataset, including studying technology diffusion, serving as a proxy for local development and examining the impact of electrification. Section 2.6 concludes.

2.2 Historical Background: The diffusion of electric lights in the U.S.

2.2.1 The beginnings: 1880 to 1900

Before electricity reached the house in the form of a convenient plug-in system, Americans had other early experiences with this new form of power. In the late 1870s, a very limited number of customers - approximately 600 lamps in manufacturers, mills and other commercial locations - used arc lights. Most of them were arc lamps manufactured by the Brush company and operated independently with an electric generator, the dynamo, also supplied by the same company (Passer, 1953). A decade later, following technical improvements which allowed the current produced by the dynamo to be more durable and reliable, arc lights started to substitute gas illumination in streets and public spaces. The Brush Electric Company opened its first central stations in 1879: in Cleveland, 12 arc lights were installed, and in San Francisco, 50 arc lights were used for various business establishments and street lights (Passer, 1953). In the following years, additional central stations opened in Detroit and New York, powering 16 and 20 arc lights respectively. By the end of 1880, there were approximately 6000 arc lights in operation in isolated plants and a few central stations, a number destined to increase more than ten-fold only two years later (see Table 2.1).

For the average American citizen, having access to electricity meant having a small central station located a short distance from the city centre and providing arc lights for streetlights and commercial businesses. Nye (1997) reports the case of Muncie, the classic example of a relatively large town in Indiana, with all the characteristics of the average American city[3]. In 1892, the city of Muncie decided to convert its system of gas lights to electricity by installing a system of 100 arc lights, which expanded to 132 in 1894 and 180 in 1901. While arc lights could compete with gas in street illumination, electricity was still too expensive compared to gas for the average household. In 1899, only 22 Muncie homes had electricity. Most houses were served by gas companies and in rural areas, people were still using kerosene lamps.

Also, arc lights were unsuitable for use in smaller indoor spaces for more than one reason. First, the light produced by the arc was excessively bright and of a cold white colour, making it unpleasant to look at, and very different from the warmer yellowish hue of the gaslight or the kerosene lamps, to which people were accustomed to (Bell, 1912). Second, the light was still generated by an open flame which consumed oxygen - occasionally hissing - produced heat, and flickered. Third, the mechanism to turn it on was relatively complex and allowed the lamps to be turned on and off only simultaneously, which was impractical for households.

Table 2.1: Total number of lights installed (000s), 1881-1902

	Arc Lights	Inc. Lights
1881	6	.
1883	75	.
1885	96	250
1886	140	.
1890	235	3000
1895	300	.
1902	386	18000

Source: Passer (1953)

This was why domestic consumers had to wait for the incandescent light bulb to be

---

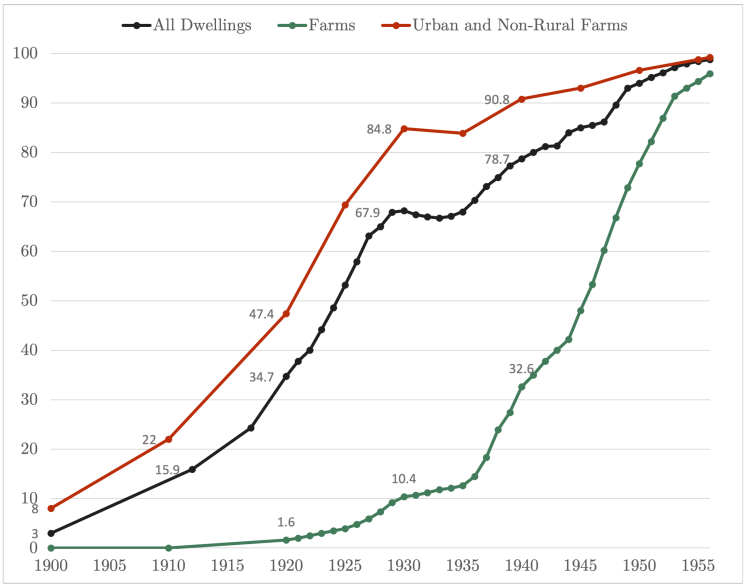
[3] Indeed, Muncie was renamed ”Middletown” in the famous 1929 sociological study by Robert Staughton Lynd and Helen Merrell Lynd ”Middletown: A Study in Modern American Culture”(Lynd and Lynd, 1929)

developed and commercialized before replacing gaslight in the home. After a decade of experimenting with making the light bulbs more efficient and reducing their costs, Edison opened the Pearl Street Central Station in New York in 1882, selling power to the offices of J.P. Morgan. Similar to the arc light business, in the following years, Edison's company started selling its electric system to private companies, operating it as an isolated plant, and to local franchises, operating it through a central station serving multiple customers. In 1885, only three years after Pearl Street, there were already 250 thousand incandescent lights in operation, increasing to 18 million in 1902. In those early years, the incandescent light bulb was "an item of conspicuous consumption". One light bulb cost \$1 - which at the time was equal to \$(2020)30 and half a day's common wages - and one-kilowatt hour of electricity cost 20 cents - equal to 620 cents (2020) (Nye, 1997). In those wealthy households, electric lamps were decorative novel objects displaying one's own status, but through these first niche markets, people started to understand the advantages of the incandescent bulb. In comparison to gaslight, this new form of illumination was safer and substantially cleaner, requiring less labour time to be managed, but also not generating soot soiling upholstery and carpets. As already mentioned, for middle-class and poorer households, especially in low-density rural areas, Edison's incandescent light bulb system was still too expensive, and they had to wait until the next century for that cost to fall(Nye, 1997).

2.2.2 The diffusion of incandescent lights: 1900 to 1920

At the national level, between 1900 and 1920, access to electricity in the house increased from 3% to approximately 35% (see Figure 2.1). The adoption of electric lighting was driven by a significant and consistent decrease in the cost to the

Figure 2.1: Electrification rates for Farm vs Urban Households (%)



Source: Lebergott (1976) and U.S. Bureau of the Census (1976), p. 827.

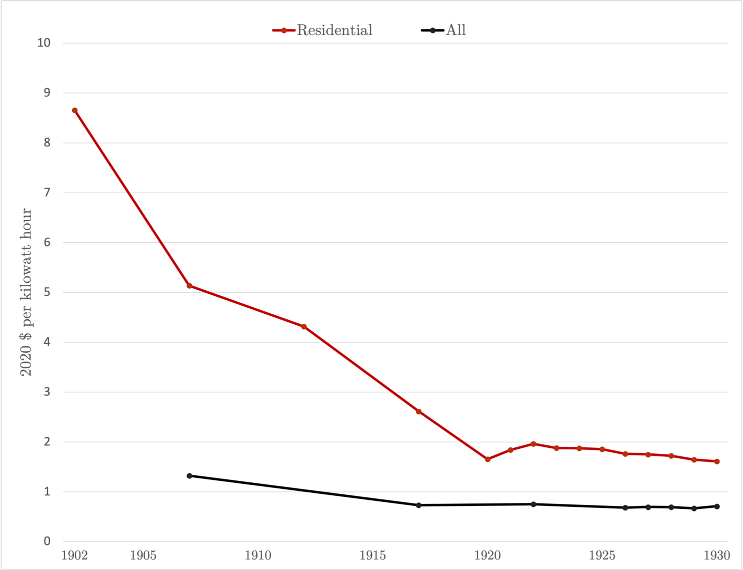
consumer. For those using incandescent lamps, this was influenced by the price of electricity, lamp efficiency, and the cost of the lamp itself. The residential price fell 80% between 1902 and 1920 to 180 cents (2020) per kilowatt hour (see Figure 2.2). (see Figure 2.2). This impressive reduction in price was mainly achieved thanks to improvements in generating and transmitting equipment, together with economies of scale. The improvement in the performance of the light bulbs is well summarized by Bright (1949) in observing the evolution in cost per thousand lumen-hours, which decreased by more than 90% between 1882 and 1923 (see Figure 2.3)[4].

Houses that adopted electric lighting were also likely to adopt other electric appliances soon after. This was because the installation of electric wiring for lighting

---

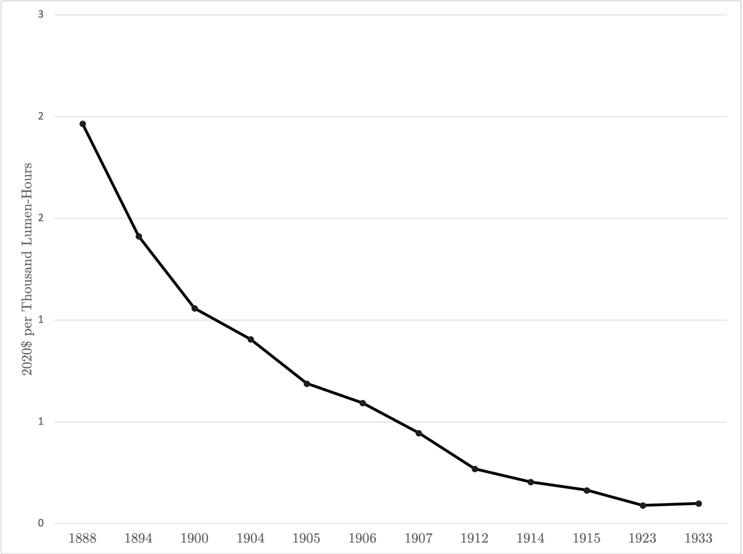
[4] A lumen is the amount of light given off through a unit solid angle (steradian) from a uniform point light source of one candle.

Figure 2.2: Residential vs Average prices of electricity



Source: U.S. Bureau of the Census (1976), p.827

Figure 2.3: Decrease in Cost of Lumen-Hours



Source: Bright (1949), p. 363

also made it possible to use electricity for other tasks. As a result, the adoption of electric lighting paved the way for the widespread use of electric appliances in homes, which became economically feasible after 1905. Once streets and retailers created the base load and the costs of building transmission systems decreased, utility companies needed households to fill their evening load with additional consumption. Therefore, they started concentrating their marketing effort on the "moderate means" consumer (Cowan, 1983). To that end, they increased their research into domestic electric appliances while also trying to keep their prices low. Among other factors, this was achieved by the standardization of the electric system to alternating current (AC)[5], which were reached nationwide around 1910.

At first, some companies marketed AC motors to be used to 'retrofit' sewing machines or washing machines. Later on, small appliances such as electric irons and fans started to be sold as upper-class paraphernalia. In 1910, after the AC standardization, smaller manufacturers entered the market selling new appliances to homes already wired by the major utilities. After a halt in expansion due to the First World War, between 1918 and 1920, consumers started to adopt new appliances across social classes. In a 1921 survey of electrified homes in Philadelphia, 64%/90% of the poorest/wealthiest homes had an iron, and 33%/84% had a vacuum cleaner. Coming right after the adoption of electric lights, appliances diffused first among higher-income households in urban areas, followed by lower-income and then eventually reached farms and rural households, most of which started to be wired in only after 1930.

---

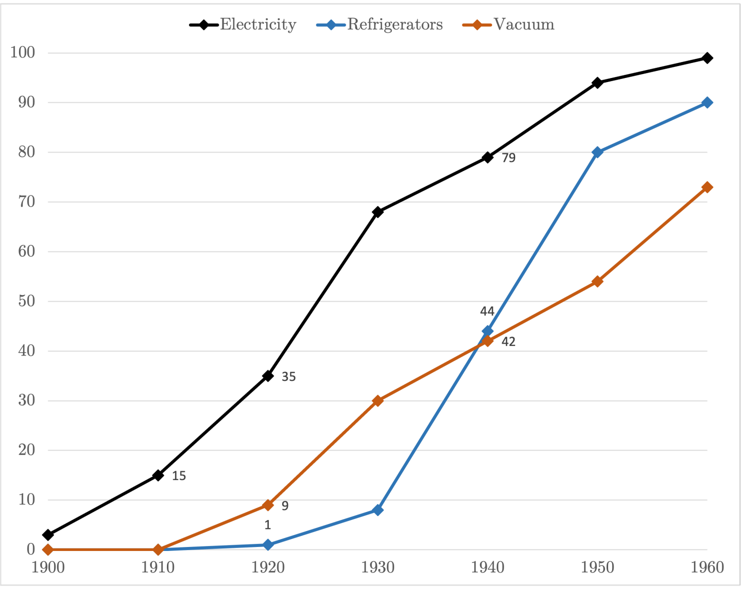
[5] Specifically, the nation was standardized to alternating current, generated at 60 cycles, transmitted at high voltages and then stepped down to 120 volts. The standardization brought an end to the so-called 'Current War', a period of time between the early 1900 and 1910 in which AC and DC current was competing for market dominance.

Table 2.2: Appliance ownership in 1300 electrified homes, Philadelphia, 1921

	Poor homes	Average homes	Modern homes	Better class
Iron	64	60	87	90
Vacuum cleaner	33	40	83	84
Washing machine	11	5	28	32
Fan	1	4	6	36

Source: C.J. Russell, Philadelphia Survey, from Nye (1997), p. 286.

Figure 2.4: Diffusion of Electrical Appliances (%)



Source: Lebergott (1976) and Historical Statistics of the United States (1976), p. 827.

2.2.3 Regional differences in the diffusion of electric lights

The information presented about the adoption of electric lighting is based on data. While this offers a general perspective, it may obscure regional variations in the spread of electric lights. Considering these differences reveals a more nuanced picture of how electric lighting was adopted in different areas and it is the main reason behind the development of this dataset. Information about regional variation in arc and incandescent lights is available from the quinquennial special census of the electrical light and power industry (see Table 2.3, also called 'The Electrical Census'). This series was collected by the U.S. Bureau of the Census between 1902 and 1937. It provides data about the equipment used, such as engines, turbines, and transmission lines, as well as the services provided, such as electric lighting, heating, and power for manufacturing and mining purposes[6]. This data is a valuable resource for comparing the use of electricity across different states and regions of the U.S. By summarizing the data in Table 4, a clear picture emerges of how electricity was adopted and utilized in various parts of the country.

As early as 1902, the East Coast of the United States was a leader in electrification. In terms of absolute numbers, the Middle Atlantic region (including New York City) had over 33% of the total count of incandescent lights (around 61,000), followed by the East North Central Region with 24% (around 43,000). However, when looking at relative numbers, New England emerges as the leader in the East Coast, with over 400 incandescent lights per thousand people. More interestingly, the per capita number of the Pacific region was even higher, with 637 incandescent lights per thousand people. The Mountain region also had relatively high numbers of lights per capita,

---

[6] The digitisation of this series has been done by John L. Neufeld and is available online at this link.

around 365, almost as high as the Middle Atlantic region.

On the other hand, southern states lagged behind by almost a magnitude: the South Atlantic region, for instance, had only 58 incandescent lights per thousand people. These regional differences in electrical light diffusion mirror structural differences in employment and per capita income of the different U.S. regions (Caselli and Coleman II, 2001).The North/South differences in incandescent lights per capita are similar to the North/South differences in the share of employment in agriculture. On the other hand, the North/West difference in electric lights diffusion mirrors North/West relative wages. In both cases, the dominance of the West is explained by the reliance on natural resources: high wages came from mining activity, while high diffusion of incandescent lights came from easy access to hydropower and cheap energy [7].

Although the available data show significant regional differences, these data are only at the state level. A more detailed dataset, such as the one developed in this paper, could provide a much more interesting picture. Indeed, in reality the distribution of lights was concentrated in counties with a highly urbanized population and until the end of 1920 the U.S. was divided into two realities: small towns with access to arc electric lights in main streets and some incandescent lights in retail shops and public buildings, and more densely populated urban centres with incandescent lights adopted in larger public buildings and spaces, a bigger number of shops and some wealthy households with a large number of arc lights lighting most streets. Thus, to explore a more granular picture of the diffusion of arc and incandescent lights, an alternative data set is required, which captures the difference in magnitude between different counties of the U.S.

---

[7] Note that the process of rural electrification is the subject of the second paper (chapter 3 in this thesis), and will not be discussed in great detail here.

Table 2.3: Regional differences in Inc. and Arc lights between 1902 and 1922

		Pop (mil)		Inc. Lights (000s)			Arc Lights (000s)		
Division	Year		%		%	p.c.		%	p.c.
East North Central									
	1902	15.99	21.18	4392.39	24.14	274.77	110507	28.65	6.91
	1912	18.25	20.09	18590.90	24.31	1018.64	138042	35.79	7.56
	1922	21.48	21.64	63818.41	24.39	2971.68	.	.	.
East South Central									
	1902	7.55	10.00	463.44	2.55	61.40	11328	2.94	1.50
	1912	8.41	9.26	2147.15	2.81	255.31	18365	4.76	2.18
	1922	8.89	8.96	6012.67	2.30	676.09	.	.	.
Middle Atlantic									
	1902	15.45	20.48	6135.97	33.73	397.03	122537	31.77	7.93
	1912	19.32	21.27	22096.04	28.89	1143.93	181582	47.08	9.40
	1922	22.26	22.43	87521.65	33.44	3931.59	.	.	.
Mountain									
	1902	1.67	2.22	612.26	3.37	365.60	9348	2.42	5.58
	1912	2.63	2.90	2239.28	2.93	850.30	9926	2.57	3.77
	1922	2.45	2.47	5199.65	1.99	2118.12	.	.	.
New England									
	1902	5.59	7.41	2425.24	13.33	433.70	47017	12.19	8.41
	1912	6.55	7.21	7955.77	10.40	1214.12	44682	11.58	6.82
	1922	7.40	7.46	29643.30	11.33	4005.36	.	.	.
Pacific									
	1902	1.90	2.52	1210.36	6.65	637.51	20764	5.38	10.94
	1912	3.05	3.36	7955.18	10.40	2607.99	30904	8.01	10.13
	1922	4.21	4.24	21641.62	8.27	5140.22	.	.	.
South Atlantic									
	1902	10.44	13.84	611.00	3.36	58.51	17183	4.46	1.65
	1912	12.19	13.43	3887.08	5.08	318.75	21389	5.55	1.75
	1922	9.75	9.82	13459.57	5.14	1380.83	.	.	.
West north Central									
	1902	10.35	13.71	1784.53	9.81	172.46	35022	9.08	3.38
	1912	11.64	12.81	8327.12	10.89	715.52	40860	10.59	3.51
	1922	12.54	12.64	25657.23	9.80	2045.34	.	.	.
West South Central									
	1902	6.53	8.65	545.27	3.00	83.47	11795	3.06	1.81
	1912	8.78	9.67	3285.68	4.30	374.03	19645	5.09	2.24
	1922	10.24	10.32	8746.78	3.34	853.99	.	.	.

Notes: Lights are measured in thousands lights per capita. Arc lights data for 1922 are unavailable because the technology had been superseded. Source: Electrical Census for the years 1907, 1912, 1917 (U.S. Bureau of the Census, 1907, 1912, 1917)

2.3 A new dataset

This section describes the creation of a new county-level dataset about the diffusion of incandescent and arc lights in the U.S. between 1900 and 1920. The data have been scraped from two historical business directories: the E.L. Powers Company Directory (Powers) and the McGraw-Hill Company Directory (Mcgraw). Both are publicly accessible online via the HathiTrust Digital Library website[8].

2.3.1 Business Directories

Industry-specific directories organize businesses by region and city and list them alphabetically. They provide useful information about each business, such as revenue, capital endowments and other financial data. This type of directory was used by businesses to find providers of specific products and to assess their reliability. The Powers and McGraw directories organized electric power central stations by state and city in alphabetical order. They also included useful information such as ownership, machinery, operations, and services for each of those stations. As a reference, figures 5.2 and 5.1 in Appendix A.2 show the first page and the table of contents of the 1900 Powers directory. In addition to the list of central stations for the United States, Canada and Mexico, each publication published a list of Electric Light Associations and a list of businesses selling electrical equipment.

Figure 2.5 shows a typical directory page. It has two columns. The names of the cities where the power plants are located are in capital bold letters. Following each city name is a number indicating the total population in that year. The columns also contain a series of coded details about the power plant, which can be decoded by

---

[8] For the interested reader, the original files are available online at Hathi Trust. In Appendix A, I listed each file and its respective HathiTrust link

referring to the 'List of Abbreviations' provided by the publishers, shown in figure 2.6. In the picture, I have highlighted with a red box the abbreviations I have used to decode the information I needed[9]. The McGraw directories have similar contents and are arranged in a similar fashion.

Figure 2.7 shows the similarities and differences between the directories, as well as how the data entries for a city changed over time. Every entry lists the population estimate - highlighted in grey -, the name of the companies owning the central station and its shareholders, and then various information about the equipment owned and the service provided. For some cities, like the one in our example, the city entered a contract with the company (city cont.), and details about that contract are listed at the end of the data entry. As an example, interpreting information about Muncie, we can read that in 1900 the city had 27 thousand people. There were two companies serving the city, one superintended (spt) by W.F. Warner and one owned by J.Boyce and three others. One has 167 arc lights in operation (167 arcs). One has 75 arc lights(75 arcs) in operation and 8 direct current arc lights (8 DC incandescent arc lights) in operation, as well as 2500 direct current incandescent lights.

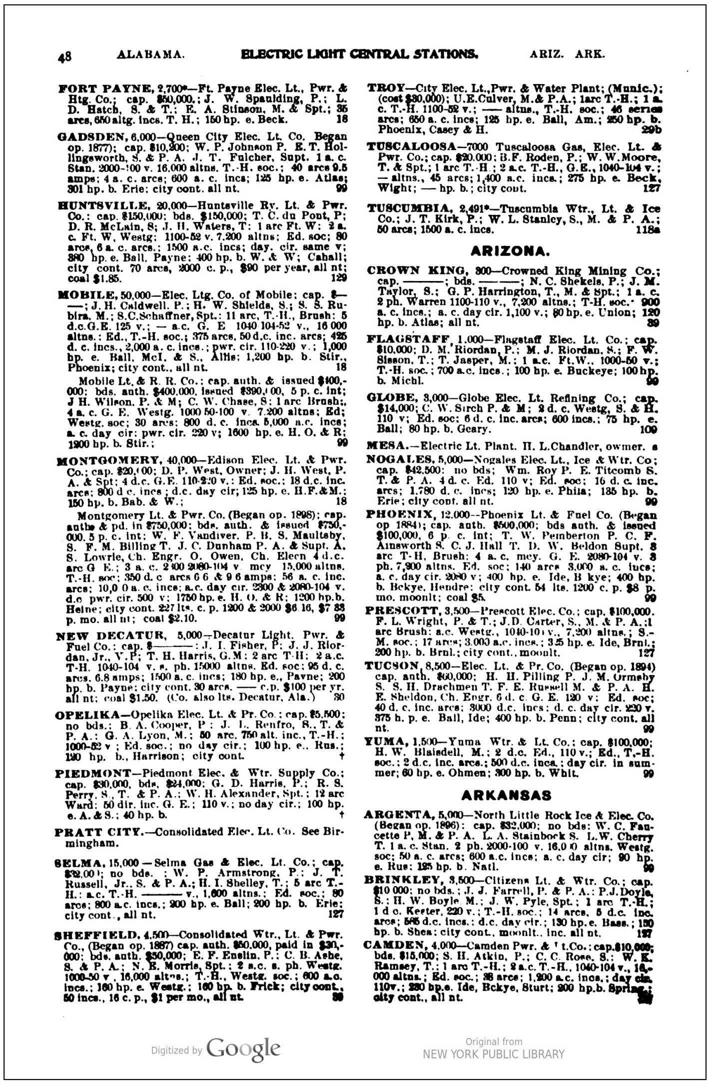
There are a few points worth noticing to understand the process of collecting data from these sources. First, there is a noticeable difference between the quality of the images for different directories. Some documents were perfectly readable (See pane C and D in Figure 2.7), while others were less so (pane B). Additionally, for each document, some pages were more readable than others. Second, there were inconsistencies in how data were collected, which had to be categorized and corrected using an automated procedure. For example, notice that in pane C, the city contract

---

[9] The McGraw directory list is available in Appendix A.2, Figure 5.3. The Powers and the McGraw directory used similar abbreviations, but some were added in later years.

- text enclosed in a green box - counts 459 arcs. The same number appears in the general text, clearly a replication, which needs to be counted only once. This and other similar inconsistencies had to be adjusted through a review of the automated scraping procedure, which is described in detail in the next section.

Figure 2.5: A typical page in the Powers' directory, 1900. Page 58.



48 ALABAMA. **ELECTRIC LIGHT CENTRAL STATIONS.** ARIZ. ARK.

FORT PAYNE, 2,700*—Ft. Payne Elec. Lt., Pwr. & Htg. Co.; cap. $50,000.; J. W. Spaulding, P.; L. D. Hatch, S. & T.; E. A. Stinson, M. & Spt.; 35 arcs, 650 altg. incs. T. H.; 150 hp. e. Beck. 18

GADSDEN, 6,000—Queen City Elec. Lt. Co. Began op. 1877); cap. $10,200; W. P. Johnson P. E. T. Hollingsworth, S. & P. A. J. T. Fulcher, Supt. 1 a. c. Stan. 2000-100 v. 16,000 altns. T.-H. soc.; 40 arcs 9.5 amps; 4 a. c. arcs; 600 a. c. incs; 125 hp. e. Atlas; 301 hp. b. Erie; city cont. all nt. 99

HUNTSVILLE, 20,000—Huntsville Ry. Lt. & Pwr. Co.: cap. $150,000; bds. $150,000; T. C. du Pont, P; D. R. McLain, S; J. H. Waters, T; 1 arc Ft. W; 2 a. c. Ft. W, Westg; 1100-52 v. 7.200 altns; Ed. soc; 80 arcs, 6 a. c. arcs.; 1500 a.c. incs; day. cir. same v; 380 hp. e. Ball, Payne; 400 hp. b. W. & W; Caball; city cont. 70 arcs, 2000 c. p., $90 per year, all nt; coal $1.85. 129

MOBILE, 50,000—Elec. Ltg. Co. of Mobile; cap. $——; J. H. Caldwell. P.; H. W. Shields, S.; S. S. Rubira, M.; S.C.Schaffner, Spt.: 11 arc, T.-H., Brush; 5 d.c.G.E. 125 v.; — a.c. G. E 1040 104-52 v., 16 000 altns.: Ed., T.-H. soc.; 375 arcs, 50 d.c. inc. arcs; 425 d. c. incs., 2,000 a. c. incs.; pwr. cir. 110-220 v.; 1,000 hp. e. Ball, McI. & S., Allis; 1,200 hp. b. Stir., Phoenix; city cont., all nt. 18

Mobile Lt. & R. R. Co.; cap. auth. & issued $400,000; bds. auth. $400,000, issued $390,000, 5 p. c. int; J H. Wilson, P. & M; C. W. Chase, S; 1 arc Brush;, 4 a. c. G. E. Westg. 1000 50-100 v. 7.200 altns; Ed; Westg. soc; 30 arcs; 800 d. c. incs. 5,000 a.c. incs; a. c. day cir; pwr. cir. 220 v; 1600 hp. e. H. O. & R; 1200 hp. b. Stir.; 99

MONTGOMERY, 40,000—Edison Elec. Lt. & Pwr. Co.; cap. $20,000; D. P. West, Owner; J. H. West, P. A. & Spt; 4 d.c. G.E. 110-220 v.: Ed. soc.; 18 d.c. inc. arcs; 800 d c. incs; d.c. day cir; 125 hp. e. H.F.&M.; 150 hp. b. Bab. & W.; 18

Montgomery Lt. & Pwr. Co. (Began op. 1898); cap. auths & pd. in $750,000; bds. auth. & issued $750,000. 5 p. c. int; W. F. Vandiver. P. B. S. Maultsby, S. F. M. Billing T. J. C. Dunham P. A. & Supt. A. S. Lowrie, Ch. Engr. O. Owen, Ch. Elecn 4 d.c. arc G E.; 3 a. c. 2300 2080-104 v. mcy 15,000 altns. T.-H. soc; 350 d. c arcs 6 6 & 9 6 amps; 56 a. c. inc. arcs; 10,000 a. c. incs; a.c. day cir. 2300 & 2080-104 v. d.c pwr. cir. 500 v; 1750 hp. e. H. O. & R; 1200 hp.b. Heine; city cont. 227 lts. c. p. 1200 & 2000 $6.16, $7.83 p. mo. all nt; coal $2.10. 99

NEW DECATUR, 5,000—Decatur Light. Pwr. & Fuel Co.: cap. $———: J. I. Fisher, P; J. J. Riordan, Jr., V.P; T. H. Harris, G.M; 2 arc T-H; 2 a.c. T-H. 1040-104 v. s. ph. 15000 altns. Ed. soc; 95 d. c. arcs. 6.8 amps; 1500 a. c. incs; 180 hp. e., Payne; 200 hp. b. Payne; city cont. 30 arcs. —— c.p. $100 per yr. all nt; coal $1.50. (Co. also lts. Decatur, Ala.) 30

OPELIKA—Opelika Elec. Lt. & Pr. Co.; cap. $5,500; no bds.; B. A. Cooper, P; J. L. Renfro, S., T. & P. A.; G. A. Lyon, M.; 50 arc, 750 alt. inc., T.-H.; 1000-52 v; Ed. soc.; no day cir.; 100 hp. e., Rus.; 120 hp. b., Harrison; city cont. †

PIEDMONT—Piedmont Elec. & Wtr. Supply Co.; cap. $30,000, bds. $24,000; G. D. Harris, P.; R. S. Perry, S., T. & P. A.; W. H. Alexander, Spt.; 12 arc Ward; 50 dir. inc. G. E.; 110 v.; no day cir.; 100 hp. e. A. & S.; 40 hp. b. †

PRATT CITY.—Consolidated Elec. Lt. Co. See Birmingham.

SELMA, 15,000—Selma Gas & Elec. Lt. Co.; cap. $32,000; no bds. ; W. P. Armstrong, P.; J. T. Russell, Jr., S. & P. A.; H. I. Shelley, T.; 5 arc T.-H.; a.c. T.-H. ——— v., 1,600 altns.; Ed. soc.; 80 arcs; 800 a.c. incs.; 200 hp. e. Ball; 200 hp. b. Erie; city cont., all nt. 127

SHEFFIELD, 4,500—Consolidated Wtr., Lt. & Pwr. Co., (Began op. 1887) cap. auth. $50,000, paid in $30,000; bds. auth. $50,000; E. F. Enslin, P.; C. B. Ashe, S. & P. A.; N. E. Morris, Spt.; 2 a.c. s. ph. Westg. 1000-50 v., 16,000 altns.; T.-H., Westg. soc.; 600 a.c. incs.; 160 hp. e. Westg.; 160 hp. b. Frick; city cont., 50 incs., 16 c. p., $1 per mo., all nt. 38

TROY—City Elec. Lt.,Pwr. & Water Plant; (Munic.); (cost $30,000); U.E.Culver, M.& P.A.; 1arc T.-H.; 1 a. c. T.-H. 1100-52 v.; —— altns., T.-H. soc.; 46 series arcs; 650 a. c. incs; 125 hp. e. Ball, Am.; 250 hp. b. Phoenix, Casey & H. 29b

TUSCALOOSA—7000 Tuscaloosa Gas, Elec. Lt. & Pwr. Co.; cap. $20,000; B.F. Roden, P.; W. W. Moore, T. & Spt.; 1 arc T.-H.; 2 a.c. T.-H., G.E., 1040-104 v.; — altns., 45 arcs; 1,400 a.c. incs.; 275 hp. e. Beck, Wight; — hp. b.; city cont. 127

TUSCUMBIA, 2,491*—Tuscumbia Wtr., Lt. & Ice Co.; J. T. Kirk, P.; W. L. Stanley, S., M. & P. A.; 50 arcs; 1500 a. c. incs. 118a

ARIZONA.

CROWN KING, 800—Crowned King Mining Co.; cap. ———; bds. ———; N. C. Shekels, P.; J. M. Taylor, S.; G. P. Harrington, T., M. & Spt.; 1 a. c. 2 ph. Warren 1100-110 v., 7,200 altns.; T-H. soc. 900 a. c. incs.; a. c. day cir. 1,100 v.; 80 hp. e. Union; 120 hp. b. Atlas; all nt. 39

FLAGSTAFF, 1,000—Flagstaff Elec. Lt. Co.; cap. $10,000; D. M. Riordan, P.; M. J. Riordan, S.; F. W. Sisson, T.; T. Jasper, M.; 1 a.c. Ft.W., 1000-50 v.; T.-H. soc.; 700 a.c. incs.; 100 hp. e. Buckeye; 100 hp. b. Michl. 99

GLOBE, 3,000—Globe Elec. Lt. Refining Co.; cap. $14,000; C. W. Sirch P. & M; 2 d. c. Westg, S. & H. 110 v; Ed. soc; 6 d. c. inc. arcs; 600 incs.; 75 hp. e. Ball; 80 hp. b. Geary. 109

MESA.—Electric Lt. Plant. H. L. Chandler, owmer. s

NOGALES, 5,000—Nogales Elec. Lt., Ice & Wtr. Co; cap. $42,500; no bds; Wm. Roy P. E. Titcomb S. T. & P. A. 4 d. c. Ed. 110 v; Ed. soc; 16 d. c. inc. arcs; 1,780 d. c. incs; 120 hp. e. Phila; 185 hp. b. Erie; city cont. all nt. 99

PHOENIX, 12,000—Phoenix Lt. & Fuel Co. (Began op 1884); cap. auth. $500,000; bds auth. & issued $100,000, 6 p. c. int; T. W. Pemberton P. C. F. Ainsworth S. C. J. Hall T. D. W. Beldon Supt. 8 arc T-H, Brush; 4 a. c. mcy. G. E. 2080-104 v. 3 ph. 7,200 altns. Ed. soc; 140 arcs 3,000 a. c. incs; a. c. day cir. 2080 v; 400 hp. e. Ide, B kye; 400 hp. b. Bckye. Hendre; city cont. 54 lts. 1200 c. p. $8 p. mo. moonlt; coal $5. 99

PRESCOTT, 3,500—Prescott Elec. Co.; cap. $100,000. F. L. Wright, P. & T.; J.D. Carter, S., M. & P. A.; 1 arc Brush; a.c. Westg., 1040-104 v., 7,200 altns.; S.-M. soc.; 17 arcs; 3.000 a.c. incs.; 325 hp. e. Ide, Brnl.; 200 hp. b. Brnl.; city cont., moonlt. 127

TUCSON, 8,500—Elec. Lt. & Pr. Co. (Began op. 1894) cap. auth. $60,000; H. H. Pilling P. J. M. Ormsby S. S. H. Drachmen T. F. E. Russell M. & P. A. H. E. Sheldon, Ch. Engr. 6 d. c. G. E. 120 v; Ed. soc; 40 d. c. inc. arcs; 3000 d.c. incs; d. c. day cir. 220 v. 375 h. p. e. Ball, Ide; 400 hp. b. Penn; city cont. all nt. 99

YUMA, 1,500—Yuma Wtr. & Lt. Co.; cap. $100,000; H. W. Blaisdell, M.; 2 d.c. Ed., 110 v.; Ed., T.-H. soc.; 2 d.c. inc. arcs.; 500 d.c. incs.; day cir. in summer; 60 hp. e. Ohmen; 300 hp. b. Whit. 99

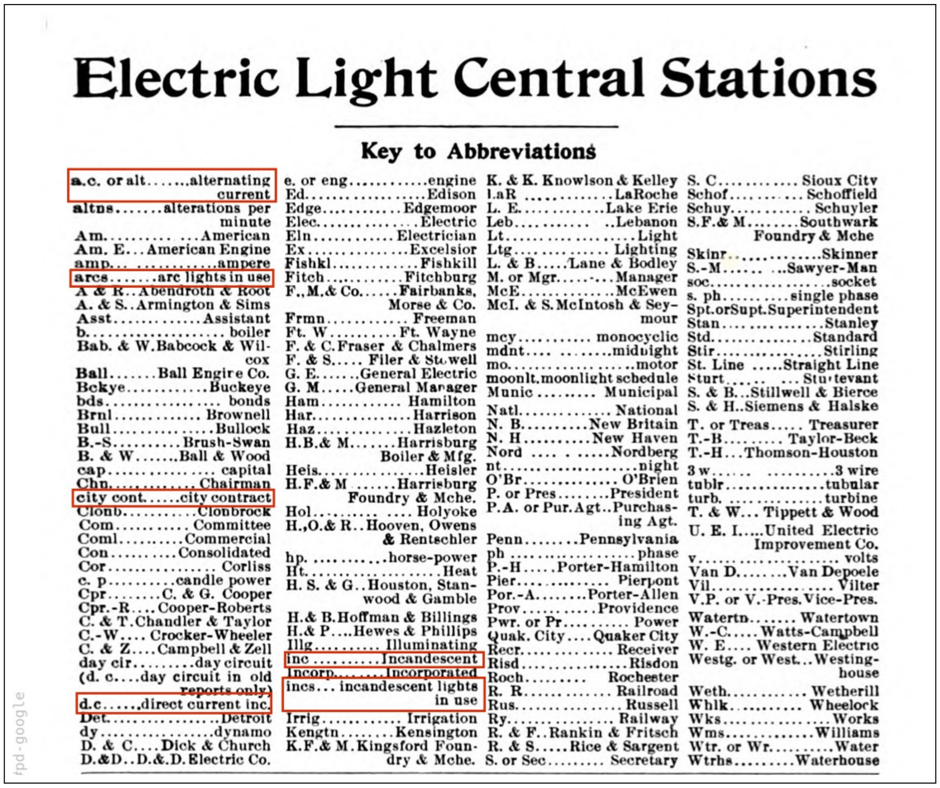
ARKANSAS

ARGENTA, 5,000—North Little Rock Ice & Elec. Co. (Began op. 1896): cap. $32,000; no bds: W. C. Faucette P, M. & P. A. L. A. Stainbock S. L.W. Cherry T. 1 a. c. Stan. 2 ph. 2000-100 v. 16,000 altns. Westg. soc; 50 a. c. arcs; 600 a.c. incs; a. c. day cir; 90 hp. e. Rus; 125 hp. b. Natl. 99

BRINKLEY, 3,500—Citizens Lt. & Wtr. Co.; cap. $10 000; no bds.; J. J. Farrell, P. & P. A.; P.J.Doyle, S.; H. W. Boyle M.; J. W. Pyle, Spt.; 1 arc T.-H.; 1 d c. Keeter, 220 v.; T.-H. soc.; 14 arcs, 5 d.c. inc. arcs; 585 d.c. incs.; d.c. day cir.; 130 hp. e. Bass.; 150 hp. b. Shea; city cont., moonlt., inc. all nt. 127

CAMDEN, 4,000—Camden Pwr. & Lt.Co.; cap.$10,000; bds. $15,000; S. H. Atkin, P.; C. C. Rose, S.; W. K. Ramsey, T.; 1 arc T.-H.; 2 a.c. T.-H., 1040-104 v., 16,000 altns.; Ed. soc.; 38 arcs; 1,200 a.c. incs.; day cir. 110v.; 280 hp.e. Ide, Bckye, Sturt; 200 hp.b. Spring.; city cont., all nt. [illegible]

Figure 2.6: List of Abbreviations in the Powers' directory, 1900. Page 57



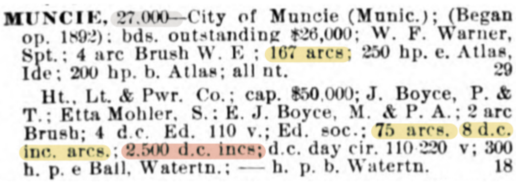
Electric Light Central Stations

Key to Abbreviations

a.c. or alt	alternating current
altns	alterations per minute
Am.	American
Am. E.	American Engine
amp.	ampere
arcs	arc lights in use
A & R.	Abendroth & Root
A. & S.	Armington & Sims
Asst	Assistant
b.	boiler
Bab. & W.	Babcock & Wilcox
Ball	Ball Engire Co.
Bckye	Buckeye
bds	bonds
Brnl	Brownell
Bull	Bullock
B.-S	Brush-Swan
B. & W	Ball & Wood
cap	capital
Chn.	Chairman
city cont.	city contract
Clonb.	Clonbrock
Com	Committee
Coml	Commercial
Con	Consolidated
Cor	Corliss
c. p	candle power
Cpr	C. & G. Cooper
Cpr.-R	Cooper-Roberts
C. & T.	Chandler & Taylor
C.-W	Crocker-Wheeler
C. & Z	Campbell & Zell
day cir	day circuit
(d. c.	day circuit in old reports only)
d.c	direct current inc.
Det.	Detroit
dy	dynamo
D. & C.	Dick & Church
D.&D.	D.&.D. Electric Co.
e. or eng	engine
Ed.	Edison
Edge	Edgemoor
Elec.	Electric
Eln	Electrician
Ex	Excelsior
Fishkl	Fishkill
Fitch	Fitchburg
F., M.& Co.	Fairbanks, Morse & Co.
Frmn	Freeman
Ft. W	Ft. Wayne
F. & C.	Fraser & Chalmers
F. & S.	Filer & Stowell
G. E.	General Electric
G. M	General Manager
Ham	Hamilton
Har.	Harrison
Haz	Hazleton
H.B.& M.	Harrisburg Boiler & Mfg.
Heis.	Heisler
H.F.& M	Harrisburg Foundry & Mche.
Hol	Holyoke
H.,O.& R.	Hooven, Owens & Rentschler
hp.	horse-power
Ht.	Heat
H. S. & G.	Houston, Stanwood & Gamble
H.& B.	Hoffman & Billings
H.& P.	Hewes & Phillips
Illg	Illuminating
inc	Incandescent
Incorp.	Incorporated
incs	incandescent lights in use
Irrig	Irrigation
Kengtn	Kensington
K.F.& M.	Kingsford Foundry & Mche.
K. & K.	Knowlson & Kelley
LaR	LaRoche
L. E.	Lake Erie
Leb	Lebanon
Lt	Light
Ltg	Lighting
L. & B	Lane & Bodley
M. or Mgr.	Manager
McE	McEwen
McI. & S.	McIntosh & Seymour
mcy	monocyclic
mdnt	midnight
mo.	motor
moonlt.	moonlight schedule
Munic	Municipal
Natl.	National
N. B.	New Britain
N. H	New Haven
Nord	Nordberg
nt	night
O'Br	O'Brien
P. or Pres	President
P.A. or Pur. Agt.	Purchasing Agt.
Penn	Pennsylvania
ph	phase
P.-H	Porter-Hamilton
Pier	Pierpont
Por.-A.	Porter-Allen
Prov	Providence
Pwr. or Pr	Power
Quak. City	Quaker City
Recr.	Receiver
Risd	Risdon
Roch	Rochester
R. R	Railroad
Rus.	Russell
Ry.	Railway
R. & F.	Rankin & Fritsch
R. & S.	Rice & Sargent
S. or Sec	Secretary
S. C	Sioux City
Schof	Schoffield
Schuy	Schuyler
S.F.& M	Southwark Foundry & Mche
Skinr	Skinner
S.-M	Sawyer-Man
soc	socket
s. ph	single phase
Spt. or Supt.	Superintendent
Stan	Stanley
Std.	Standard
Stir	Stirling
St. Line	Straight Line
Sturt	Sturtevant
S. & B.	Stillwell & Bierce
S. & H.	Siemens & Halske
T. or Treas	Treasurer
T.-B	Taylor-Beck
T.-H	Thomson-Houston
3 w	3 wire
tublr	tubular
turb.	turbine
T. & W.	Tippett & Wood
U. E. I.	United Electric Improvement Co.
v	volts
Van D.	Van Depoele
Vil.	Vilter
V.P. or V.-Pres.	Vice-Pres.
Watertn	Watertown
W.-C.	Watts-Campbell
W. E.	Western Electric
Westg. or West.	Westinghouse
Weth.	Wetherill
Whlk	Wheelock
Wks	Works
Wms	Williams
Wtr. or Wr.	Water
Wtrhs	Waterhouse

Figure 2.7: Examples of Directories

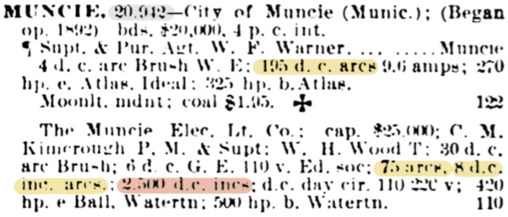
A. Powers Directory, 1900



MUNCIE, 27.000—City of Muncie (Munic.); (Began op. 1892); bds. outstanding $26,000; W. F. Warner, Spt.; 4 arc Brush W. E; 167 arcs; 250 hp. e. Atlas, Ide; 200 hp. b. Atlas; all nt. 29

Ht., Lt. & Pwr. Co.; cap. $50.000; J. Boyce, P. & T.; Etta Mohler, S.; E. J. Boyce, M. & P. A.; 2 arc Brush; 4 d.c. Ed. 110 v.; Ed. soc.; 75 arcs. 8 d.c. inc. arcs.; 2.500 d.c. incs; d.c. day cir. 110-220 v; 300 h. p. e Ball, Watertn.; — h. p. b. Watertn. 18

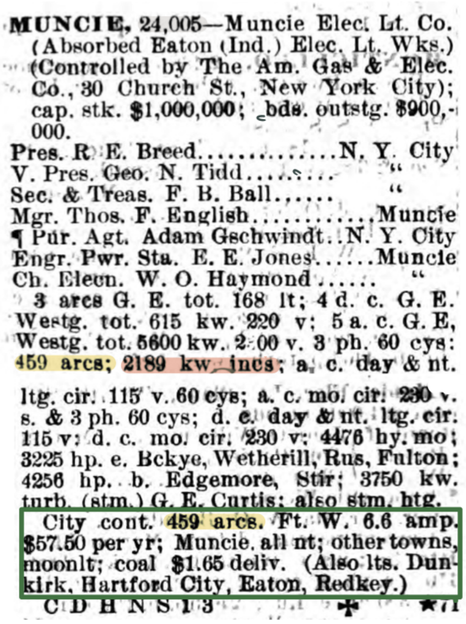
B. Powers Directory, 1903



MUNCIE, 20.942—City of Muncie (Munic.); (Began op. 1892) bds. $20,000, 4 p. c. int.
¶ Supt. & Pur. Agt. W. F. Warner. ... .Muncie
4 d. c. arc Brush W. E; 195 d. c. arcs 9.6 amps; 270 hp. e. Atlas, Ideal; 325 hp. b. Atlas. Moonlt. mdnt; coal $1.95. ✠ 122

The Muncie Elec. Lt. Co.; cap. $25.000; C. M. Kimbrough P. M. & Supt; W. H. Wood T; 30 d. c. arc Brush; 6 d. c. G. E. 110 v. Ed. soc; 75 arcs, 8 d.c. inc. arcs.; 2.500 d.c. incs; d.c. day cir. 110 220 v; 420 hp. e Ball, Watertn; 500 hp. b. Watertn. 110

C. McGraw Directory, 1911

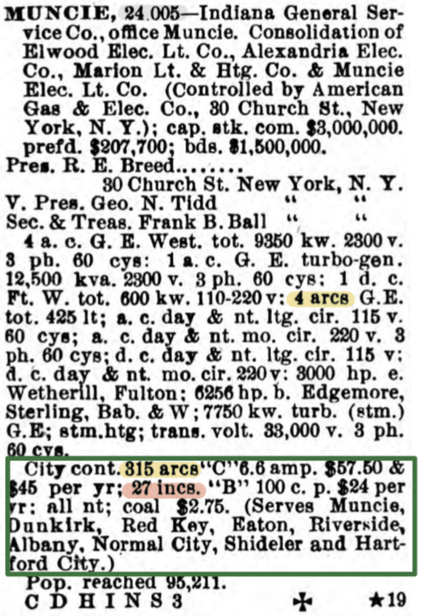


MUNCIE, 24,005—Muncie Elec. Lt. Co. (Absorbed Eaton (Ind.) Elec. Lt. Wks.) (Controlled by The Am. Gas & Elec. Co., 30 Church St., New York City); cap. stk. $1,000,000; bds. outstg. $900,000.
Pres. R. E. Breed..................N. Y. City
V. Pres. Geo. N. Tidd.........   "
Sec. & Treas. F. B. Ball.......   "
Mgr. Thos. F. English.............Muncie
¶ Pur. Agt. Adam Gschwindt. .N. Y. City
Engr. Pwr. Sta. E. E. Jones......Muncie
Ch. Elecn. W. O. Haymond......   "

3 arcs G. E. tot. 168 lt; 4 d. c. G. E. Westg. tot. 615 kw. 220 v; 5 a. c. G. E, Westg. tot. 5600 kw. 2200 v. 3 ph. 60 cys: 459 arcs; 2189 kw. incs; a. c. day & nt. ltg. cir. 115 v. 60 cys; a. c. mo. cir. 230 v. s. & 3 ph. 60 cys; d. c. day & nt. ltg. cir. 115 v; d. c. mo. cir. 230 v; 4476 hy. mo; 3225 hp. e. Bckye, Wetherill, Rus, Fulton; 4256 hp. b. Edgemore, Stir; 3750 kw. turb. (stm.) G. E. Curtis; also stm. htg.

City cont. 459 arcs. Ft. W. 6.6 amp. $57.50 per yr; Muncie. all nt; other towns, moonlt; coal $1.65 deliv. (Also lts. Dunkirk, Hartford City, Eaton, Redkey.)
C D H N S 1 3 ✠ ★71

D. McGraw Directory, 1919



MUNCIE, 24.005—Indiana General Service Co., office Muncie. Consolidation of Elwood Elec. Lt. Co., Alexandria Elec. Co., Marion Lt. & Htg. Co. & Muncie Elec. Lt. Co. (Controlled by American Gas & Elec. Co., 30 Church St., New York, N. Y.); cap. stk. com. $3,000,000. prefd. $207,700; bds. $1,500,000.
Pres. R. E. Breed........
30 Church St. New York, N. Y.
V. Pres. Geo. N. Tidd   "   "
Sec. & Treas. Frank B. Ball   "   "

4 a. c. G. E. West. tot. 9350 kw. 2300 v. 3 ph. 60 cys: 1 a. c. G. E. turbo-gen. 12,500 kva. 2300 v. 3 ph. 60 cys: 1 d. c. Ft. W. tot. 600 kw. 110-220 v: 4 arcs G.E. tot. 425 lt; a. c. day & nt. ltg. cir. 115 v. 60 cys; a. c. day & nt. mo. cir. 220 v. 3 ph. 60 cys; d. c. day & nt. ltg. cir. 115 v; d. c. day & nt. mo. cir. 220 v: 3000 hp. e. Wetherill, Fulton; 6256 hp. b. Edgemore, Sterling, Bab. & W; 7750 kw. turb. (stm.) G.E; stm.htg; trans. volt. 33,000 v. 3 ph. 60 cys.

City cont. 315 arcs "C" 6.6 amp. $57.50 & $45 per yr; 27 incs. "B" 100 c. p. $24 per yr: all nt; coal $2.75. (Serves Muncie, Dunkirk, Red Key, Eaton, Riverside, Albany, Normal City, Shideler and Hartford City.)
Pop. reached 95,211.
C D H I N S 3 ✠ ★19

Notes: Highlighted in grey, the population estimate for that year. In yellow and orange, the data points for arc lights and incandescent lights respectively. The green box shows the data about the city contract. Source: See Text.

2.3.2 OCR practice for historical documents

At the time of writing, only three papers discuss the process of using Optical Character Recognition (OCR) to extract data from historical documents. In Correia and Luck (2023), they focus on best practices for extracting data from historical balance sheet records. Meanwhile, Albers and Kappner (2023) examines the extraction of data from city directories, specifically using an 1880 Berlin directory as an example. Both papers concentrate on creating the database and provide a set of steps for practitioners. Hegghammer (2022), on the other hand, tests the performance of three OCR engines, two commercial and one public, by comparing their results on a diverse range of documents. This section will outline a data extraction pipeline customized for the Powers and McGraw central stations directories, following best practices as described in the previously mentioned papers. The following is an overview of the general procedure, followed by a detailed examination of both datasets. Figure 2.8 summarizes the process in a small graph which I have adapted from Correia and Luck (2023). Note that the Powers directories only required Step 3 onward, while cleaning the McGraw directories required also Step 1 and 2. After obtaining the raw images (.JPG files) by calling the HathiTrust API, the general steps are the following:

1. **Image processing**. To enhance the OCR engine's ability to read scanned images, the first step is to transform them. There are two ways in which I processed the images:

   (a) **Color distortions corrections** This step aims to correct any distortions in the digitized images caused by the scanning process or the wear and tear of historical documents. The scanned pages may have distortions in size, geometry, and color, which can lead to non-uniform shades on the page's surface. To address this issue, the most commonly used methods are binarization and grayscale conversion. Binarization converts images to black and white pixels, while grayscale conversion desaturates images. It is also recommended to combine grayscale and binarization of the image after inverting its colors. This step is particularly important for the business directories used in this research paper, as the scanning process was uneven, resulting in some volumes being better preserved and/or scanned than others.

(b) **Columns segmentation** To extract each column of text separately and ensure proper reading order, page segmentation was necessary. The analyzed pages had two columns, so identifying structural elements to distinguish the columns was important. The python OpenCV package provided useful functions for analyzing page structure. The focus was to identify the vertical line dividing the two columns of text. By using this line as a reference point, each column of text was separated and extracted separately [10]

2. **OCR**. I used Tesseract, an open-source OCR engine, to extract text from the modified images. While Tesseract has weaker performance compared to commercially available OCR engines like Google Document AI and Amazon, as noted by Hegghammer (2022), it is the only free option and its utilization required less time and budget management, which made it the preferred choice for this research paper.

---

[10] The code and a summary of the functions used is available in the online GitHub repository of the author.

3. **Data Extraction**. The procedures adopted for data extraction are highly dependent on the researcher and on the documents themselves. For the Powers and McGraw business directories I used a combination of the following methods:

    (a) **Split cities** Looking at the typical page in the Powers’ directory (Figure 2.5), the reader can see that each city is listed in a separate paragraph. To assign each data value to the correct location, I had to split the text document into city-level paragraphs, and assign the correct State name to each city.

    (b) **Split city contract** In Figure 2.7, each city has a section for the station’s characteristics and a section for the city contract. To differentiate between the two, I identified the beginning of the city contract paragraph by locating the word ’city’ or something similar using fuzzy matching.

    (c) **Regex** Finally, I used a regex pattern to locate data points relating to arc lights and incandescent lights in the text. Regex, short for regular expression, is a sequence of characters that defines a search pattern and is used to match and manipulate text. For example, in order to find all the data points related to arc lights, I used the following expression:

    ```
    ’r’[\s][Aa]r[eac][s]?’
    ```

    Which is able to find variations of the word ’arcs’, including common mistakes by the OCR engine, such as identifying the letter ’c’ as an ’e’ or an ’a’. After locating the correct word, it is easy to catch every digit within the next (or the previous) n characters, and store it in a dataframe.

4. **Validation Checkpoint**. During an OCR process, validation is the act of verifying and fixing errors in the output by comparing it to the ground truth. Ground truth is the known and accurate text that corresponds to the input document. It must be created manually or thoroughly evaluated by a human reviewer. In this step, I compared the output of the extracted data to a manually extracted sample of data points. I selected the total number of cities in the 11 Western States, with the selection of cities covering random page ranges in each document, given that the states were listed alphabetically. By selecting cities from Western states, the processing time for each city was reduced, as I did not need to find randomly selected cities in the original scanned files. For the Powers business directory, the error rate is 14%, with an average difference of 15% from the true value.

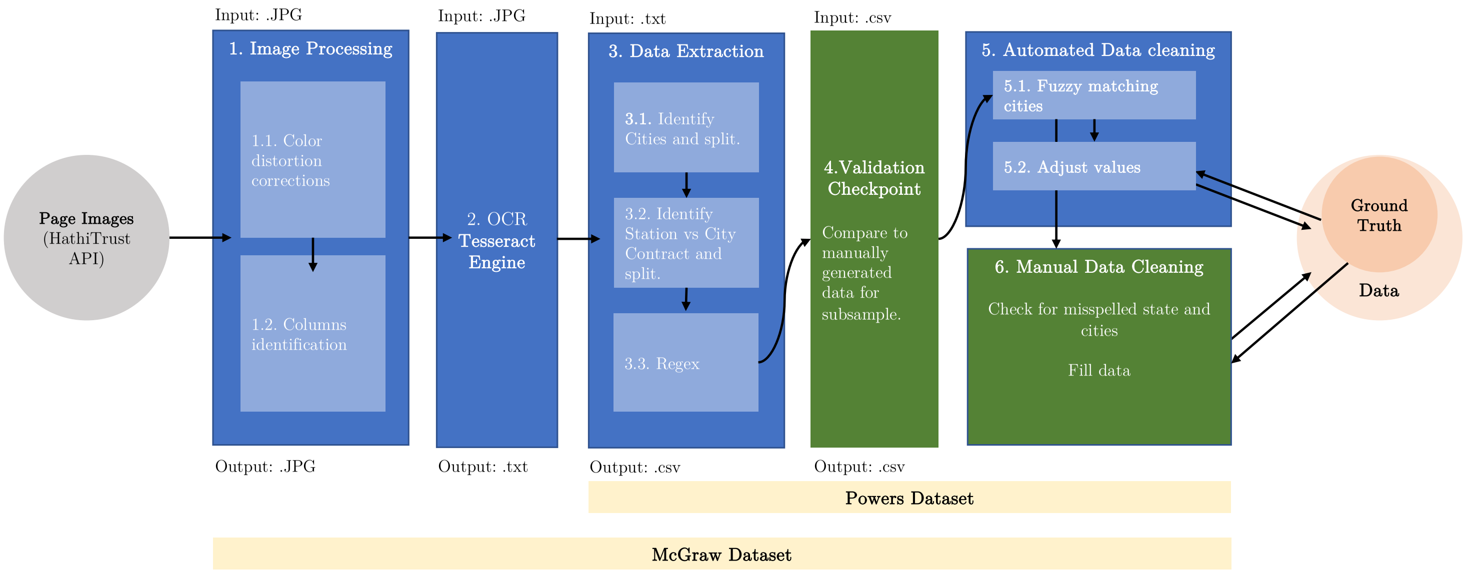
5. **Automated Data Cleaning**.

    (a) **Fuzzy matching cities** Next, the data are passed through an automated data cleaning step which uses a Levenshtein Distance algorithm to compare the city names extracted from the OCR with a list of all cities in the U.S. with more than 500 people. The Levenshtein Distance algorithm is a string metric for measuring the difference between two sequences of characters by counting the minimum number of operations required to transform one sequence into the other. If the distance is smaller than two characters, the matches are correct, and I keep the data points in a separate datset and I proceed to Step 5.2. If the matches are wrong, I move to Step 6 below.

    (b) **Adjust values** Finally, I run any additional cleaning procedure to validate the data further and to produce a clean data frame. In the case of the Powers' dataset, for each year, if there are repeated cities with identical values, I use the mean. Otherwise I sum the values for each city. Cities with two different data points could be the result of one city being split into two different pages or columns (Note that detailed additional steps are described in the section below).

6. **Manual Data Cleaning**. To rectify inaccurately OCR'd city names, I manually enter the correct names and data values. While this task is time-consuming, it cannot be avoided entirely for historical documents, as noted in the literature (Albers and Kappner, 2023). Improving the OCR output can reduce the need for this task, but it is likely an unavoidable necessity.

Figure 2.8: Data extraction pipeline



2.3.3 1900: Powers Data-set

The data for 1900 were scraped from the Powers' Central Station Directory and Buyers' Manual, printed by the E.L. Powers Company and for which I had access to the yearly editions printed between 1900 and 1903. As mentioned above, this kind of directory reported information about power stations for each city in the U.S. and listed them alphabetically by state. To process this directory, I proceeded as described above, and I added some additional manual data cleaning steps, which I will describe here.

First, although I had access to the four consecutive years for each power station, I averaged the observations to add an additional validation checkpoint and to correct for potential scraping mistakes. Therefore, my data points represent incandescent and arc lights at the beginning of the 1900s, rather than in a specific year. Raw data obtained from this first step are displayed as a dot-density map in Figure 2.11 and Figure 2.10.

Secondly, I checked my data against state-level data from the Electrical Census, the only available data source about incandescent light diffusion in the early 20th century, which was described in the section above. Grouping observations at the state level, I concluded that discrepancies between the already aggregated census numbers and my numbers were substantial, so I adjusted the city-level observations using the data I collected as weights to re-distribute the state-level electrical census data. After comparing the Electrical Census data collection method - described in the technical manual chapter of the census itself - with the Powers directory's one, I concluded that a possible reason for discrepancies is the way in which companies were asked to count for incandescent lights in use. While in the Electrical Census all types of lights were converted to 16-candlepower lights and counted, in the Powers' catalogue there was no option to specify whether the incandescent lights were 32- or 16-candlepower, eventually resulting in a lower number[11].

After being weighted, counts of incandescent lights vary across states and divisions, with an average/median mismatch of +76/+60% and a negative mismatch only in the South Atlantic region (-13%). For arc lights, the mismatches are smaller, with an average 18% mismatch, and a negative mismatch only in the South Atlantic region (-28%).

The final step was to convert city-level data to county-level data. The Powers catalogue was collected at the city level, approximating the power station location. I converted this dataset to a county level one by assuming that stations had a 30 miles radius (60 miles in diameter) in which they could serve, and then by calculating the number of incandescent and arc lights provided by all stations in a given county weighted by the area covered by a given station. Figure 2.9 shows a simplified example. Station A in blue and station B in orange are both in County 1. County 1 will get 100% of the lights assigned to Station A, but only 70% of the lights of Station B.

My choice is based on the fact that many cities - including Muncie, in the example above -, served other neighbouring cities and towns which could have been located in a neighbouring county. In practice, I drew a 30 miles radius circle around the centroid for a given city. Then, I intersected that circle with the counties' boundaries. The

---

[11] For the interested reader, more about how the Electrical Census enquired about incandescent lights is available in the "Technical Aspects of the Period" chapter, under "Incandescent lamps". For each year, there is a detailed survey of the technology used during that period U.S. Bureau of the Census (1907)

number of lights assigned to one city from one station is proportional to the area of the circle overlapping the county area. The final number of lights for each county is given by the sum of all portions of ’power station service areas’ within one county’s boundaries.

Figure 2.9 shows a simplified example. Station A in blue and station B in orange are both in County 1. County 1 will get 100% of the lights assigned to Station A, but only 70% of the lights of Station B. The final spatial distribution of arc and incandescent lights for 1900 is displayed in Figure 2.12 and 2.13 and in Table 2.4 and 2.5.

Figure 2.9: Graph of Station to County-Level Calculations

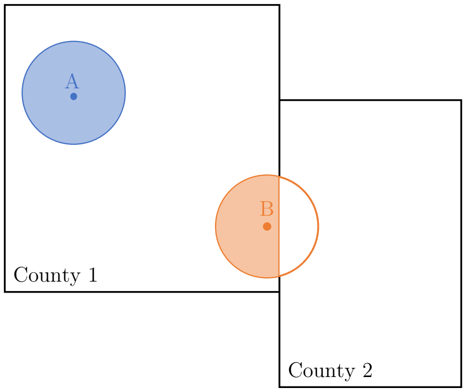
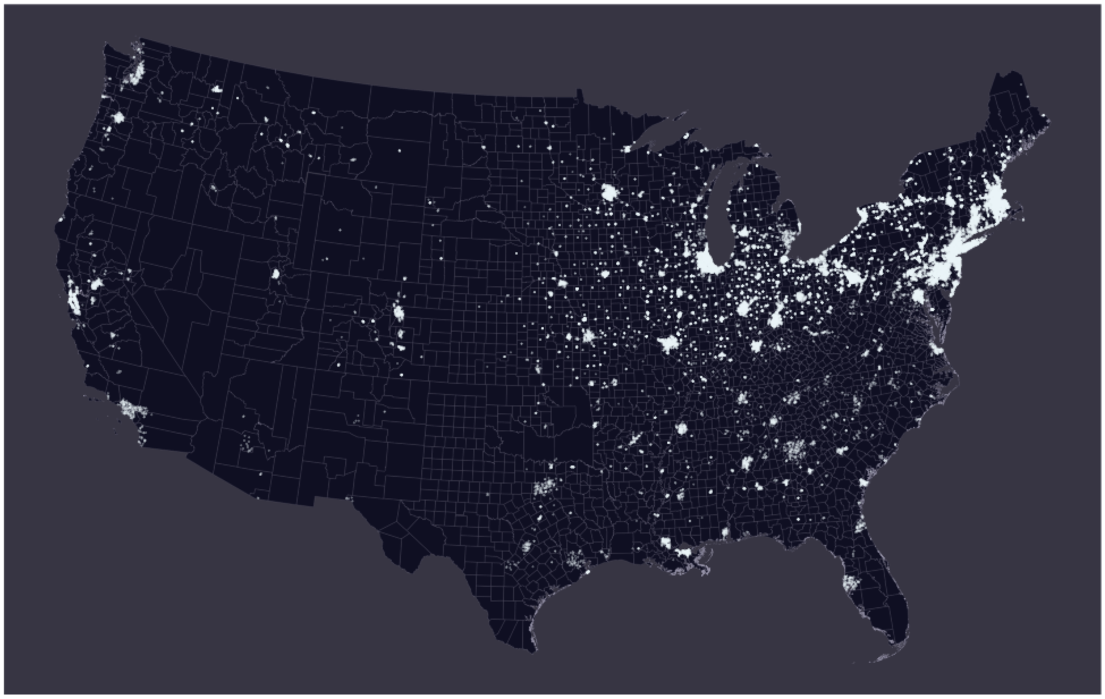
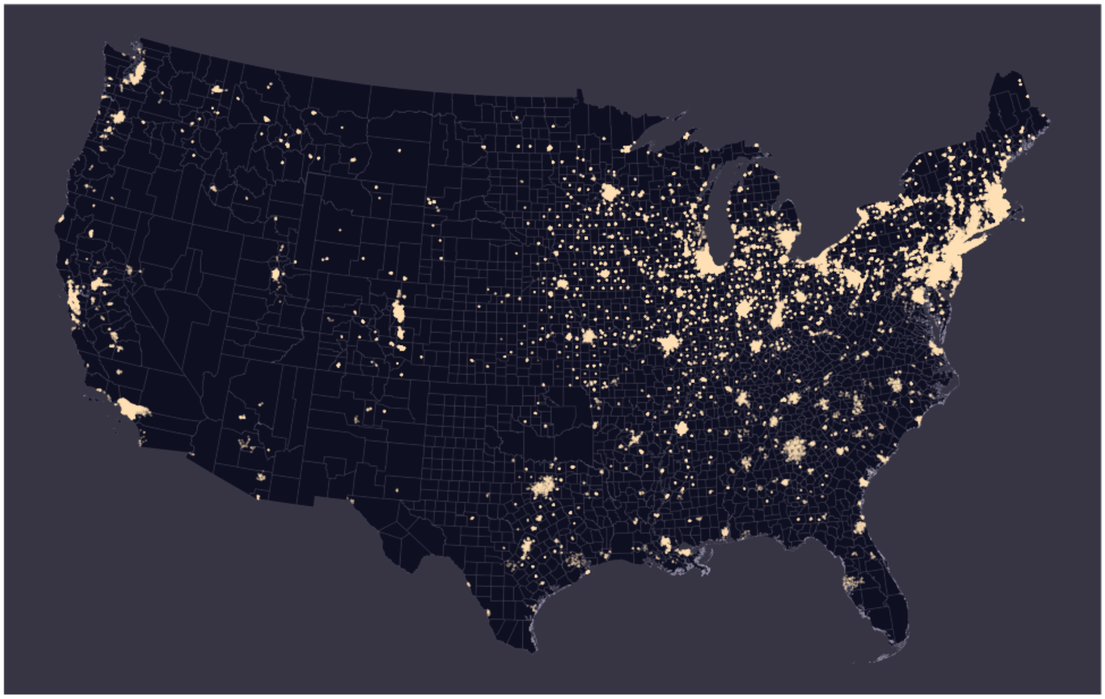


Figure 2.10: City-level Arc Lights, 1900



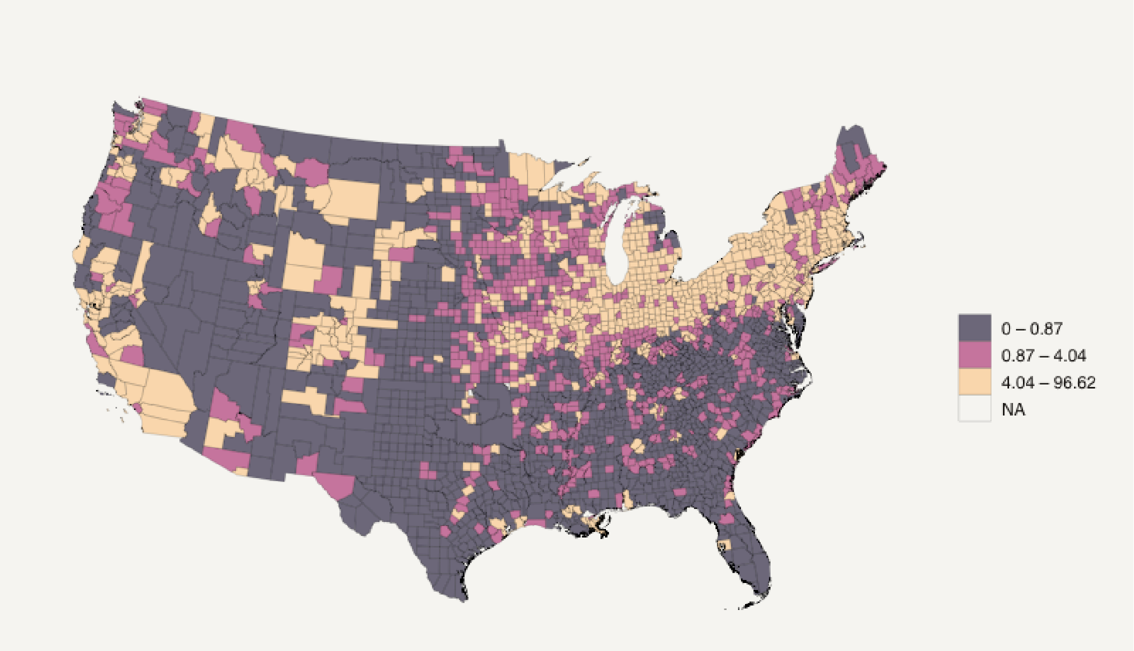
Notes: This is a dot density map which randomly plots a number of dots proportionate to the number of Arc Lights in a given location. Source: See Text.

Figure 2.11: City-level Incandescent Lights, 1900



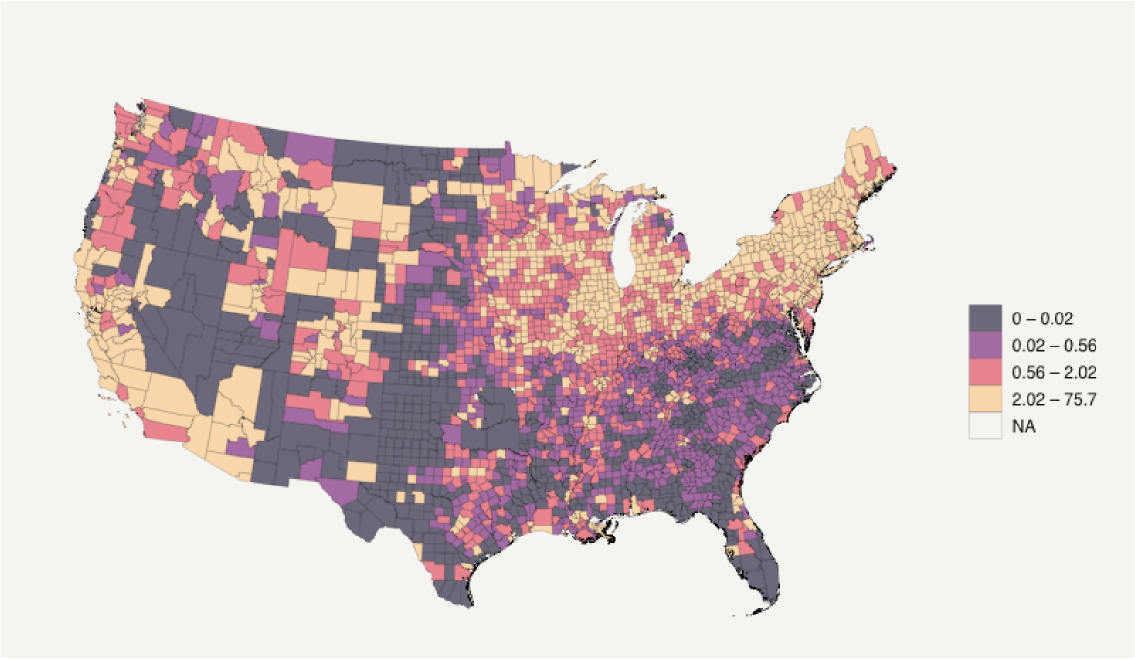
Notes: This is a dot density map which randomly plots a number of dots proportionate to the number of Incandescent Lights in a given location. Source: See Text.

Figure 2.12: Number of Arc Lights p.c., quartiles, 1900



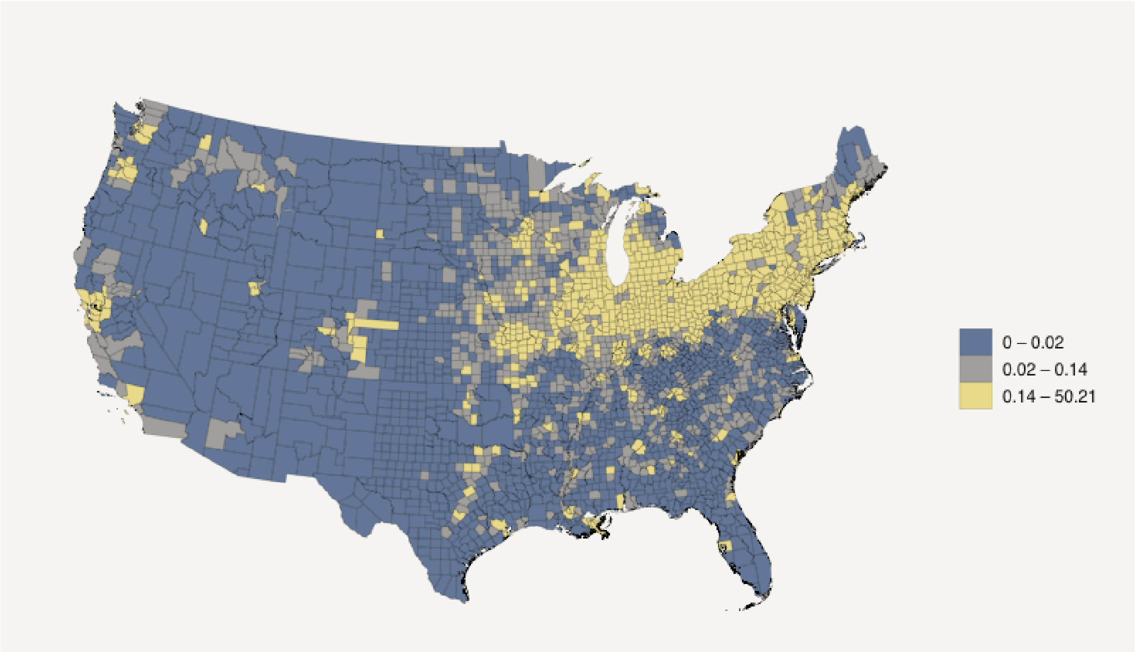
Notes: The first two quartiles are merged because of the skewed distribution of arcs p.c. (the first 25% of data points are equal to 0). Missing data (NAs) are due to missing county-level population from the census. Source: See Text.

Figure 2.13: Number of Inc. Lights p.c. (00s), quartiles, 1900



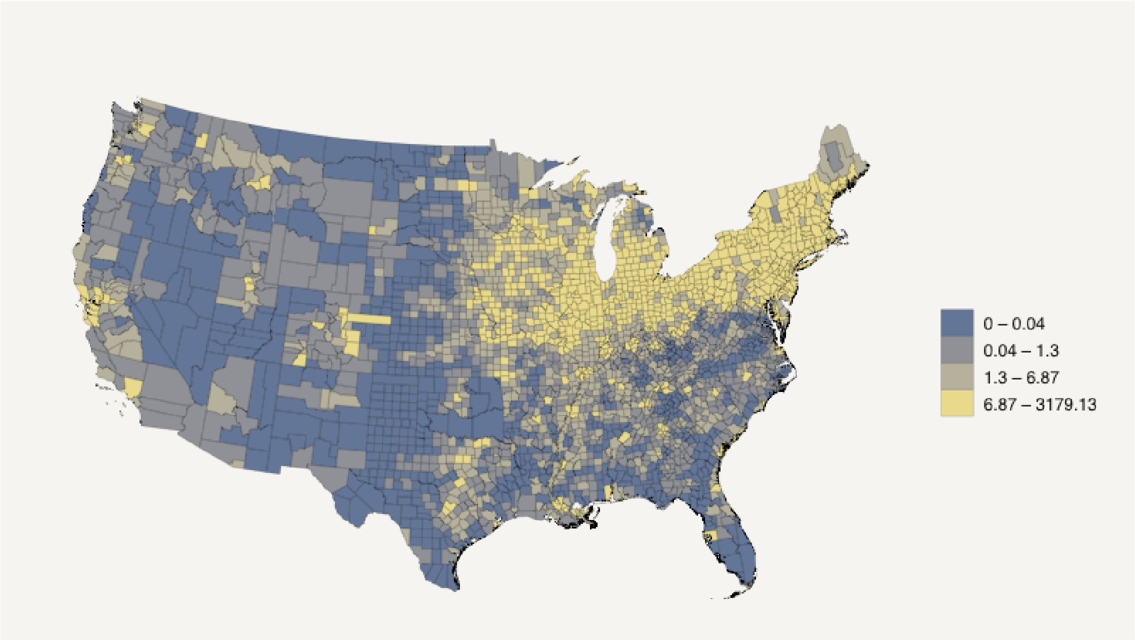
Notes: Data points are in hundreds. Missing data (NAs) are due to missing county-level population from the census.
Source: See Text.

Figure 2.14: Number of Arc Lights p. sq. mile, quartiles, 1900



Notes: The first two quartiles are merged because of the skewed distribution of arcs p.c. (the first 25% of data points are equal to 0). Missing data (NAs) are due to missing county-level population from the census. Source: See Text.

Figure 2.15: Number of Inc. Lights p. sq. mile, quartiles, 1900



Notes: Data points are in hundreds. Missing data (NAs) are due to missing county-level population from the census. Source: See Text.

2.3.4 1910 and 1920: McGraw Data-set

For 1911, I scraped data about the number of incandescent lights and arc lights, as well as the power capacity (kW) at the city level from the McGraw Central Station Directory, published in 1911 by the McGraw Publishing company. To scrape the data, I followed a procedure similar to the one used for the Powers' data-set [12]

In this data set, for some observations, the number of incandescent lights had to be

---

[12] Again, the interested reader can access the original code at the link provided above.

estimated from the amount of kW powering up the total of incandescent lights served by a given station. Using contemporary data about the diffusion of different types of light bulbs, I concluded that the most diffused light bulb had 16 candlepower - equivalent to 200 lumens - and 3.5 watts per candle, making it a 56-watt light bulb. Diving the number of kW by the wattage would give me the estimated number of incandescent lights powered up by a given station. In this data set, the discrepancies between the directories and the census were larger and of higher variance, possibly because for some observations I had to estimate the number of lights. For incandescent lights, the median/mean weighted count difference is 86%/100%, with the Middle Atlantic region being the most under-reported. The arcs discrepancy is also quite high, with a median/mean of 60%/77%. Again, I rely on the weights to have a better portrait of the spatial variation of electric light adoption, but I trust the Electrical census estimates in terms of magnitude.

For 1920, the McGraw Directory did not report the number of lights in operation with consistency. This is because starting in the 1910s, companies equipped more and more customers with reading meters, allowing them to be billed according to the actual usage in kW, rather than by light bulb. Eventually, companies stopped separating kW for lights from other uses of electricity, such as appliances, from their reports. As a result, for 1920 I had to estimate the total number of incandescent lights using the state-level total counts from the electrical census while keeping the weights from the 1910 McGraw Directory. Finally, given that during the 1910s arc lights started to be replaced by incandescent lights, the 1920s census does not report the total number of arc lights, making it impossible to estimate these additional data points.

Table 2.4: Incandescent Lights Data Adjustment.

	1900 Lights (000s)			1910 Lights (000s)			Differences	
Division	Powers	Census	Diff.	McGraw	Census	Diff.	Powers - McGraw	Census
East North Central	2577	4392	0.70	5129	18591	2.62	0.99	3.23
East South Central	286	463	0.62	1498	2147	0.43	4.23	3.63
Middle Atlantic	2498	6136	1.46	5669	22096	2.90	1.27	2.60
Mountain	384	612	0.59	1662	2239	0.35	3.32	2.66
New England	1565	2425	0.55	3662	7956	1.17	1.34	2.28
Pacific	509	1210	1.38	3778	7955	1.11	6.42	5.57
South Atlantic	460	611	0.33	2026	3887	0.92	3.41	5.36
West north Central	1026	1785	0.74	2804	8327	1.97	1.73	3.67
West South Central	250	545	1.18	3022	3286	0.09	11.07	5.03

Notes: Difference (Diff.) is in percentage points. E.g. 0.7 = 70%. Source: See text.

Table 2.5: Arc Lights Data Adjustment.

	1900 Lights (000s)			1910 Lights (000s)			Differences	
Division	Powers	Census	Diff.	McGraw	Census	Diff.	Powers - McGraw	Census
East North Central	96	111	0.155	113	138	0.223	0.18	0.25
East South Central	11	11	0.000	20	18	-0.063	0.73	0.62
Middle Atlantic	91	123	0.344	89	182	1.037	-0.02	0.48
Mountain	7	9	0.341	14	10	-0.294	1.02	0.06
New England	44	47	0.063	150	45	-0.702	2.39	-0.05
Pacific	14	21	0.487	31	31	-0.012	1.24	0.49
South Atlantic	17	17	0.011	33	21	-0.353	0.95	0.24
West north Central	28	35	0.256	34	41	0.212	0.21	0.17
West South Central	9	12	0.283	17	20	0.176	0.82	0.67

Notes: Difference (Diff.) is in percentage points. E.g. 0.7 = 70%. Source: See text.

Table 2.6: Differences, weighted and not, 1900-1920.

	Arc Lights				Inc Lights					
	1900-1910				1900-1910				1900-1920	
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)
East North Central	0.94	0.83	0.74	0.77	17.10	2.15	49.40	5.42	179.32	21.95
East South Central	9.49	6.76	8.06	6.24	447.11	8.24	486.80	11.20	1167.10	34.29
Middle Atlantic	1.57	0.53	3.38	1.23	0.00	1.96	0.01	3.65	0.03	17.44
Mountain	11.99	13.56	7.57	4.98	950.89	23.04	1212.95	14.78	2540.07	37.73
New England	0.28	0.78	0.08	-0.37	0.01	1.20	0.01	1.85	0.03	6.72
Pacific	10.50	2.55	8.95	1.23	605.58	9.36	538.71	8.12	1606.63	24.24
South Atlantic	14.81	11.14	5.90	5.43	449.32	16.17	468.81	23.62	1642.68	81.37
West north Central	3.64	1.17	4.30	1.07	474.23	4.37	894.93	8.09	2602.31	27.76
West South Central	15.53	117.00	19.78	103.73	714.57	49.75	1067.21	35.64	2679.38	95.03

Notes: For 1900-1910: (1) Mean, k (2) Median (3) Weighted mean, k (4) Weighted median. For 1900-1920: (1) Weighted mean, k, (2) Weighted median. Source: See text.

2.4 Analysis

In 1900, half of the counties had at least 62.9 incandescent lights for a thousand people, which increased to 448 and 1262 in 1910 and 1920 respectively. As expected, given that the arc technology was used only for street lights, industrial buildings and the occasional wealthy household, the numbers for arcs are of a different magnitude. In 1900, the median county had less than one arc per a thousand people. Arc lights were installed in city centres, reflected in the low values of arc lights per square mile. In 1900, the median county had access to only 2 arc lights for every hundred square miles. In 1920, the median county had approximately 1200 incandescent lights per thousand people, almost twenty times more than the value in 1900. On the other hand, arc lights diffused only up to 1910, when the amount per capita doubled. After 1910, arc lights started to be replaced by incandescent lights. Indeed, the business directories did not report arc lights anymore, and there are no data for arc lights on the Electrical Census either. One of the main drivers of electricity adoption was whether or not in a given area there was enough demand to either open a new power station or extend the existing power supply. In the next section, I will look at two different types of counties: counties with major city centres and a high population density, and counties with a lower population density, which better represents the average American city in a low-density county in 1900.

Table 2.7: County characteristics, 1900 to 1920

		1900	1910	1920
Inc. Lights p.c.(000s)	mean	171	1034	2975
	sd	364	4248	6725
	median	62.9	448	1262
Inc. Lights p. sq. mi.	mean	16.6	50.9	181
	sd	113	167	659
	median	1.4	12.5	37.2
Arc Lights p.c.(000s)	mean	3.48	5.45	.
	sd	7.19	10.7	.
	median	0.967	2.27	.
Arc Lights p. sq. mi.	mean	0.36	0.354	.
	sd	2.16	1.2	.
	median	0.0207	0.0581	.
Density (p. sq. mi.)	mean	85.6	106	118
	sd	854	1099	1211
	median	30.3	31.8	32.9
Urban (%)	mean	13.5	17	20
	sd	21.2	22.9	24.3
	median	0	1.04	12
White (%)	mean	86.5	87.4	88.4
	sd	21.1	20.3	19
	median	97.7	98	98.2

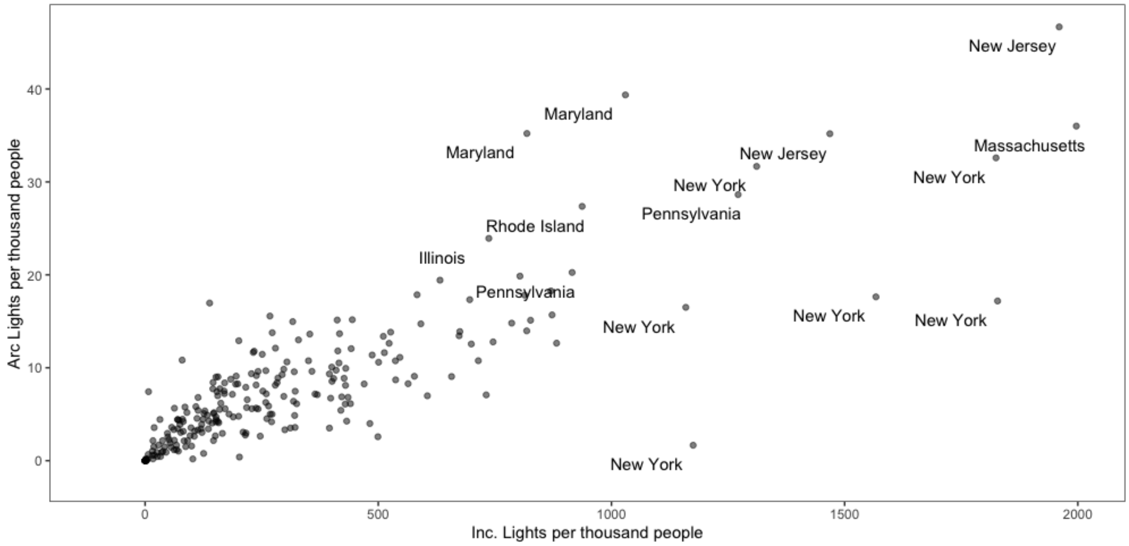
2.4.1 Counties with major urban areas

In 1900, the U.S. had only a few major big urban centres. Table 2.8 shows summary statistics for counties with more than 50 people per square mile, which at that time was considered a county with at least a major town. The number of lights in the average "big city county" vary by division. Unsurprisingly, the early industrial counties of the North East (Middle Atlantic and New England regions) lead the table with 508 and 443 incandescent lights per thousand people and approximately 10 arc lights per thousand people; right after, the Pacific region, with 361 incandescent lights per thousand people. Looking at the relationship between arc lights and incandescent lights per capita, one can observe that within these densely populated centres, there is a relatively linear relationship between the two (see Figure 2.16). Locations in the North East had adopted arc lights early for streets and industrial centres. They also adopted incandescent lights at the same rate.

Table 2.8: Summary Statistics for Big Cities in 1900

	Inc. Lights p.c.		Arc Lights p.c.	
Division	mean	median	mean	median
Middle Atlantic	508	400.3	11.0	9.2
New England	443	320.6	10.0	7.3
Pacific	361	189.1	6.5	4.7
East North Central	290	252.5	8.1	7.1
West South Central	245	171.4	7.3	5.6
West north Central	154	106.5	4.6	4.3
East South Central	114	74.9	4.0	3.9
South Atlantic	86	25.9	2.8	0.8
All	310	202.2	7.6	5.8

Figure 2.16: Arc vs Inc. Lights p.c. in Big Cities in 1900



Notes: Numbers are per thousand. Source: Author, see Text.

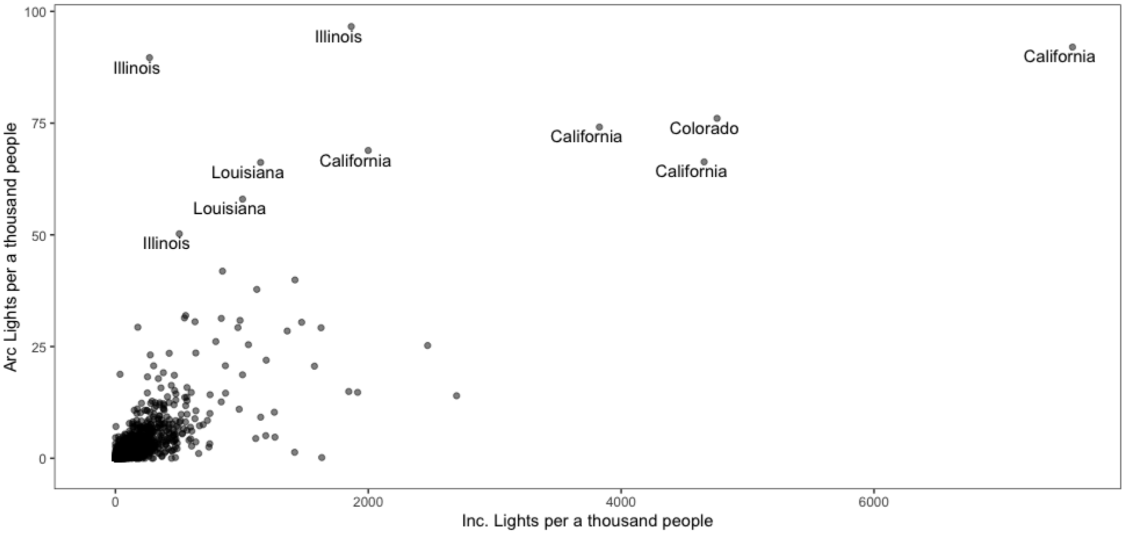
2.4.2 The average county

In 1900, the average county had between 10 and 50 people per square mile. These counties, similar to the "Middletown" Muncie analyzed by Nye (1997), had some incandescent lights installed (a median of 45.6 per thousand people), possibly in some public buildings or in some factories (see Table 2.9). When it comes to having a system of arc lights installed, almost thirty per cent of the counties did not have any. Here, the relationship between arc and incandescent lights is different and highlights some regional differences (see 2.17): counties in California, especially those situated in the valleys between the coast and the mountain, had a very high number of incandescent lights per capita and arc lights per capita; counties in the Midwest, such as in Illinois, had a lower number of incandescent lights per capita, more similar to the densely populated centres, but a very high value for arc lights. This might be explained by the presence of industrial centres, which adopted arc lights for internal illumination.

Table 2.9: Summary Statistics for the Average County in 1900

	Inc. Lights p.c.		Arc Lights p.c.	
Division	mean	p50	mean	p50
Pacific	836	291.3	15.8	7.8
Mountain	703	338.9	10.6	6.3
New England	473	426.0	4.1	2.9
Middle Atlantic	398	349.6	7.7	5.7
East North Central	242	196.3	6.1	3.9
West north Central	161	107.9	2.6	1.5
West South Central	74	27.2	1.8	0.0
East South Central	46	13.5	0.9	0.1
South Atlantic	29	4.8	0.7	0.1
All	141	45.6	2.8	0.7

Figure 2.17: Arc vs Inc. Lights p.c. in the Average County in 1900



Notes: Numbers are per thousand. Source: Author, see Text.

2.4.3 From regional to county-level variation in diffusion

Comparing the previously available state-level and regional-level values with these new county-level data, one can observe a few facts. Table 2.3 indicates that in the New England and Pacific region there were 400 and 600 incandescent lights per 1000 persons. However, the evidence in Figures 2.16 and 2.17 indicates that there was great variation within these regions. Indeed, certain counties in California had more than 4000 incandescent lights per 1000 persons (i.e., 4 lights per person), whereas in certain Massachusetts, New York and New Jersey counties there were over 2000 lights per 1000 capita (i.e., 2 lights per person). In other words, Californian counties were leading the way.

It also suggests that in certain counties, there was more than one light bulb per person. This suggests that by 1900 wealthy families in the top Californian county (with 6 lights per person) had more than 17 light bulbs (assuming a family of five and half the lightbulbs being used by households, and the rest by industry, commercial and public administration buildings). In other words, most rooms in their household would have been lit using electricity. While these light bulbs were generally the equivalent of 16 candlepower or 200 lumens, this indicates that households had the potential to consume more than 15 million lumen hours (= 200 lumens * 17 light bulbs * 12 hours per day (i.e. ignoring daylight hours) * 365 per year) per year using electric lighting.

Meanwhile, the average wealthy household (with 0.836 lights per person) in the Pacific region had the potential to consume 1.8 million lumen hours per year. The average household in the U.S. (with 0.141 lights per person) had the potential to consume 310,000 lumen hours per year using electric lighting. As a comparison, Fouquet and Pearson (2006 p.168) estimated that in the UK the average household consumed 274,000 lumen hours per year – which included gas and electric lighting. Naturally, U.S., Pacific or California households would not have consumed to their maximum potential and would have consumed only a fraction of this electric lighting potential, while also using gas lighting in urban areas and kerosene in rural areas.

2.5 Dataset potential

This new dataset will be publicly available. In this section I suggest three possible uses of the data, focusing on applications in economics and economic history.

2.5.1 Measuring the impact of electricity access in the early 1900s

This dataset can be used to estimate the impact of having access to electricity on social and economic indicators, both at the individual and county or city level. The availability of data on electric lights from the early 1900s extends the time frame for which researchers can access electricity usage data. There has been an increased interest among economists in using historical settings to test hypotheses, and digitization of historical data has made it possible to investigate the impact of historical events and/or their long-term effects on contemporary indicators of economic activity (Abramitzky, 2015).

The process of electrification of the U.S. has seen renewed interest as well. Recent papers include the seminal work of Kline and Moretti (2014), who identified a quasi-experimental scenario by comparing counties within the Tennessee Valley Authority electrification project and those counties which were proposed but not accepted.

While they did not have access to electricity usage data at the county level, they used the fraction of households with access to a radio to assess whether treatment and control counties differed with respect to electricity access.

Kitchens and Fishback (2015) were the first to use the Federal Power Commission National Power Survey map, which shows the location of power plants and the grid as of 1935. In their paper, they included distance to the nearest generation station and distance to the nearest transmission line as a covariate of their main variable of interest, access to rural electrification loans. More recent papers such as Lewis (2018) and Lewis and Severnini (2017) also make use of the Federal Power Commission Map from 1930, together with new data for power plants in 1960. They use distance to the nearest (large) power plants as a proxy for access to electricity at a lower cost. Similarly, Gaggl et al. (2021) digitize two new maps from the U.S. Army Corps of Engineers and build a new variable based on the number of kilometers of high-voltage power lines for each county. Finally, two recent working papers are using data similar to the dataset presented here. Fiszbein et al. (2020) digitize a large volume of data from the Census of Manufacturers, creating variables at the county-industry level for energy intensity. Vidart (2020) extracts information from the McGraw-Hill Central Station directory for 1911 and 1919 but focuses on the generation capacity of each power plant, rather than the endowment of incandescent and arc lights.

Combining this new data with existing datasets about electricity access in the early 20th century will allow new identification procedures, or at least confirm that distance from the nearest power plant in a given year predicts access to electricity for contemporaneous or subsequent years.

2.5.2 Measuring the Spatial Diffusion of Technology

Secondly, this new dataset can help explore is the empirical study of spatial technology diffusion. While differences in technology adoption are highly correlated with cross-country differences in per capita income, both theory and empirical studies show that technology diffuses slowly, both across and within countries. In a 2012 paper by Comin et al. (Comin et al., 2012), the authors examine the diffusion of transportation, communication, and industrial technology in a panel of 161 countries over 140 years. Their hypothesis is that the slow diffusion of technology is also due to the spatial distance between countries. Their results show a robust and significant effect of the impact of a country's neighbors' technology adoption on its own adoption rate. The same theory which guided that paper could be used to explore which factors played a role in the diffusion of electric lights in the U.S. In their model, the technology is diffused through interactions between adopters and non-adopters. These interactions are random but the probability of an interaction increases with proximity. Having access to county or city level data, the interested researcher could empirically look at how specific geographical features of a country, such as access to markets and/or transportation options, being located near the coast or in the hinterland, and/or being located close to a plain or a mountain range, correlate with adoption. The fact that this dataset is panel means that it could be estimated using fixed effects, controlling for potential confounders at the state or county level.

2.5.3 As a Proxy for Local Economic Development

Another way to utilize this dataset is by using lights per capita at the city level as a proxy for local economic development. This is similar to what other studies have done with electric light data obtained from satellite images, which are referred to as luminosity data (Michalopoulos and Papaioannou, 2018). While luminosity data have been used since the early 1990s, new papers in economic geography only started using them a decade later. The literature has used different transformations of the image data, such as the sum of total night light, the share of image pixels considered lit, and the average night light. Henderson et al. (2018), and later Pinkovskiy and Sala-i Martin (2014), showed that light density at night is a proxy for economic activity across and within countries.

Luminosity data generated from satellite imagery has the objective advantage of being a variable that is free from biases generated in the process of data collection itself. Although the dataset presented in this paper does not benefit from this characteristic, it still provides some advantages. Lights per capita (or per square mile) can be used as a proxy for local development, even in rural areas of the U.S. where we currently do not have local GDP or other income measures. In fact, for the early 1900s, data about city-level characteristics were only available for locations with more than 8,000 people ("Financial Statistics of Cities Having a Population of 8,000 to 25,000", 1903). However, this dataset covers every single inhabited (or uninhabited) center with at least one power station, including locations with even fewer than 100 inhabitants.

2.6 Conclusion

In this paper, I described the creation of a new county-level dataset about the diffusion of arc lights and incandescent lights between 1900 and 1920 in the U.S. My primary data sources are business directories, which were originally published to list central stations and their services at the city level. Using image processing and

Optical Character Recognition (OCR), I computationally extracted those data from the scanned business directories; then, I adjusted my observations using official data from the Electrical Census. To my knowledge, this is the first dataset covering three decades with such granularity.

From an initial analysis of the newly available data, I report three main findings. First, different areas of the U.S. observed different rates of diffusion of incandescent and arc lights. The different rate is much more visible at the county than at the state or regional level. Lights were mostly an urban phenomenon and up to 1920, this division between densely populated centres and less populated ones was evident. Second, counties in the Pacific area and especially in California were leading in terms of consumption of lights per capita by a large margin. This pattern is also visible in counties of the western states with a major city centre, like the state capitals of Montana and Utah. Higher consumption of light services in the western U.S. might be explained by the early expansion of hydropower, which reduced costs and justified adoption by entities, such as households, with a smaller load[13]. Third, the diffusion of arc lights is more evident in certain areas than others, and it is not directly correlated with the diffusion of incandescent lights. This is due to the different levels of industrialization, given that arc lights were used in public buildings and factories only.

In closing this chapter, I highlighted three key potential uses of this new, publicly available dataset. First, the dataset can be used to measure the social and economic impact of having access to electricity in the early 1900s, on both individual and community levels. It expands upon existing historical data, allowing us to test

---

[13] This is discussed in length in section 3.2, where I discuss the diffusion of electricity in the western american states.

hypotheses and analyze long-term effects, and could potentially show strong links between access to electricity and economic activity. It also allows us to explore whether proximity to power plants accurately predicts access to electricity during that time period. Second, the dataset can be used to study how technology spread across the U.S., specifically focusing on the spread of electric lighting. It can help us understand the geographical and societal factors that influenced adoption rates and the interactions between those who adopted the technology and those who did not. Lastly, the dataset can act as a measure of local economic development by considering the number of lights per person at the city level. While it may not be as objective as data from satellite images, it provides valuable insight into the development of even the smallest and most rural areas in the early 1900s.

Chapter 3

White Coal: Impact of early rural electrification in the Western U.S., 1900-1930.

3.1 Introduction

As of 2018, approximately 800 million people lack access to electricity worldwide. Of those, 85% live in rural areas (International Energy Agency (IEA) et al., 2020). In the U.S. of the early 1900s, while the majority of urban dwellers had access to electricity within their homes, 90% of rural families were not granted electricity service. This rate is even more surprising if one considers that in Europe, rural electrification was close to saturation. In the 1930s, the U.S. government bridged the gap between urban and rural electrification rates by heavily investing in infrastructure and credit programmes, which resulted in universal electrification at the national level by 1960.

Today, national governments, international development agencies and donors invest heavily in programs that aim to extend electricity access to rural areas [1]. These investments are driven by the positive perception of electrification experiences in now-developed countries, as well as by the solid cross-country macro evidence about the correlation between GDP per capita and electricity access (Stern et al., 2018). On the other hand, evidence at the micro level has been mixed. For instance, increased electricity access is associated with positive effects on female employment in South Africa and Nicaragua (Dinkelman, 2011; Grogan and Sadanand, 2013) and with an increase in educational attainment in Brazil and Vietnam (Lipscomb et al., 2013; Khandker et al., 2009) while other studies report no statistically significant changes in either employment or school enrolment in India and Kenya (Burlig and Preonas, 2016; Lee et al., 2020b). When discussing the historical experience of the United States, however, results so far have been in agreement: the rural electrification programme started in the 1930s was a success and it had a positive effect on employment rates as well as on agricultural productivity (Kline and Moretti, 2014; Kitchens and Fishback, 2015; Lewis and Severnini, 2017; Lewis, 2014; Gaggl et al., 2021).

In fact, most of the empirical studies about the process of rural electrification in the United States present a single narrative: before the 1930s, the majority of cities were electrified, but only a negligible percentage of the rural population had access to electricity. Then, in 1930, rural development programs were created and American farms were eventually able to connect to the grid. Although this narrative is not misguiding, the story behind the process of rural electrification in the U.S. starts

---

[1] For instance, the World Bank committed \$ 6.2 billion over the last five years. Additionally, in the past two years, they committed \$ 1 billion for mini-grids and off-grid systems, mostly in Sub-Saharan Africa.

with an early gradual expansion in the Western States[2], only later combined with a governmental push. West of the 100th parallel, the combination of semi-arid lands with mountainous regions, a lack of sufficient coal resources, and favourable institutions led the Western States - with the primacy of California - to adopt electric power earlier than the rest of the country.

The aim of this paper is twofold: to give a precise narrative of the adoption of electric power in the Western U.S., and to estimate its effects on the rural economy before the government programs of the 1930s. The rest of the paper is organised as follows. In the second section (3.2), I provide an extended historical background and I investigate the historical process through which the Western States invested in electricity access earlier than the Eastern States. By presenting new qualitative and quantitative evidence, I conclude that earlier adoption was driven by geographical (the presence of hydropower potential) and economic factors (the demand for power for irrigation and the high prices of traditional fuels). In the third section (3.3), I present a simple model (Kitchens and Fishback, 2015) which I use to derive the main channels through which electricity might have an impact on the rural economies. My main hypothesis is that rural electrification led to increased land values, with an ambiguous effect on farm profitability. After describing the available data and summary statistics in section 3.4, I present the empirical model, which is based on a continuous difference-in-differences framework with decadal trends and controls for baseline covariates (Wooldridge, 2007; Lewis and Severnini, 2017). There are no available data for electricity access at the micro level for 1900 to 1930, so I opt for the proxy which has already been tested in the literature: distance from the nearest hydroelectric power plant. Results presented in section 3.6 show that electrification between 1900

---

[2] These states are: Arizona, Colorado, California, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming

and 1930 did not have significant effects on the rural economy of the Western States, except for an increase in land values in the decade between 1900 and 1910. The main findings show that a 10 miles reduction in distance to the nearest hydroelectric power plant leads to a 2.7% increase in farmland values. A series of robustness tests in section 3.7 confirms the main results, with some caution due to relatively high Conley standard errors.

I hypothesise that the absence of effects is driven by four causes: first, electrification rates in the West were still too low. Second, the farms that had access to electricity were potentially served by off-grid solutions, making the proxy (distance to the nearest power plant) noisy and unable to capture results. Third, as most of the effects are captured in the earliest decade (1900-1910), these effects are probably the outcome of the adoption of electrical lights, both in households and in local communities. As electrical farm equipment - a complementary technology- was only developed only later, in 1920, we do not observe effects on agricultural productivity. In the final section (3.8) I discuss the results and conclude.

This paper potentially contributes to three strands of literature. By looking at the spatial differences between the Western and the Eastern United States in terms of processes of electrification, this paper contributes to historical studies which portray the long-term processes of technology adoption (Nye, 1997; Hughes, 1993). By extending the analysis of the impact of electrification to the early 1900s, this paper contributes to economic history and labour economics studies about the impacts of electricity in the United States (Kline and Moretti, 2014; Kitchens and Fishback, 2015; Lewis and Severnini, 2017; Lewis, 2018). Finally, by comparing these results with the mixed evidence of developing countries, this paper contributes to impact evaluation studies in development economics (Lee et al., 2020a,b; Dinkelman, 2011; Lipscomb et al., 2013; Burlig and Preonas, 2016).

3.2 Historical Background

3.2.1 Rural Electrification in the U.S.

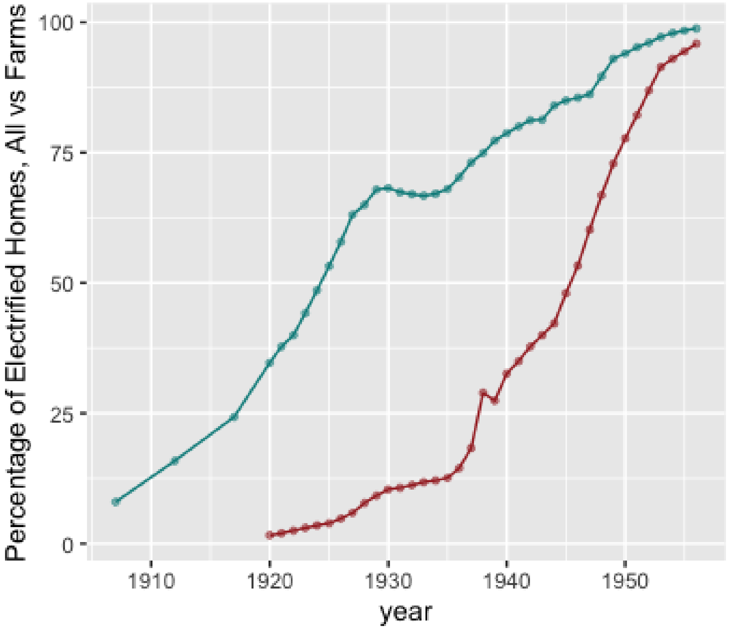
The process of electrification of the U.S. has been recounted as a tale with two phases. The electrification of urban centres started in the late 19th century with the construction of central power stations and the adoption of electric lights in streets and homes. Then, more than thirty years later, rural electrification had to be pushed by federal agencies such as the Rural Electrification Administration (REA, 1936) and the Tennessee Valley Authority (TVA, 1933). This narrative is well documented by the data available in the Historical Statistics of the United States - see figure 3.1 below, showing the different trends in electrification rates for urban homes and rural farms - and it has been the object of study of many economic and social historians[3], their focus being the radical changes that electricity had on everyday life as well as the enormous orchestrated efforts needed to expand the power transmission system.

Similarly, economists focused on quantifying the economic impact that the expansion of electricity access to the countryside had on several indicators of local productivity and growth. For instance, Kitchens and Fishback (2015) showed how low-interest loans issued by the REA between 1935 and 1939 helped the expansion of transmission lines, with a positive effect on crop output and land values. Similarly, Kline and Moretti (2014) evaluated the impact of the TVA between 1930 and 1960, showing how it positively affected agricultural and manufacturing employment. On a more general note, a series of papers by Severnini (2014), Lewis and Severnini (2017), and Lewis (2018) presented empirical evidence about the positive effect of the expansion of the transmission lines in the rural areas of the U.S., with a focus on land values

---

[3] For instance in *Electrifying America* by Nye (1997) or *Networks of Power* by Hughes (1993)

Figure 3.1: % Electrified Homes vs Farms



Notes: The blue line is for homes, while the red one is for farms. *Sources*: Historical Statistics of the United States, section S, p.510.

and agricultural productivity, and child mortality and fertility respectively.

While presenting valid empirical evidence about the process of electrification, all of these studies emphasised the narrative that rural electrification proceeded slowly and uniformly from the urban to the rural areas, starting only around 1930, and thanks to federal support policies. Their conclusion is not erroneous nor inaccurate, but rather incomplete in that it does not consider the period before the 1930s when rural electrification was already expanding significantly in the Western states. Possibly, the main reason behind the lack of observations about electricity access before the 1930s is the fact that data availability is incredibly scarce, and substantial evidence can only be found from 1925, the year in which the Census and the REA surveys started collecting official statistics. Additionally, the majority of the available data are not granular enough - most of them are collected at the national level - to defend this hypothesis if not by using solely anecdotal evidence.

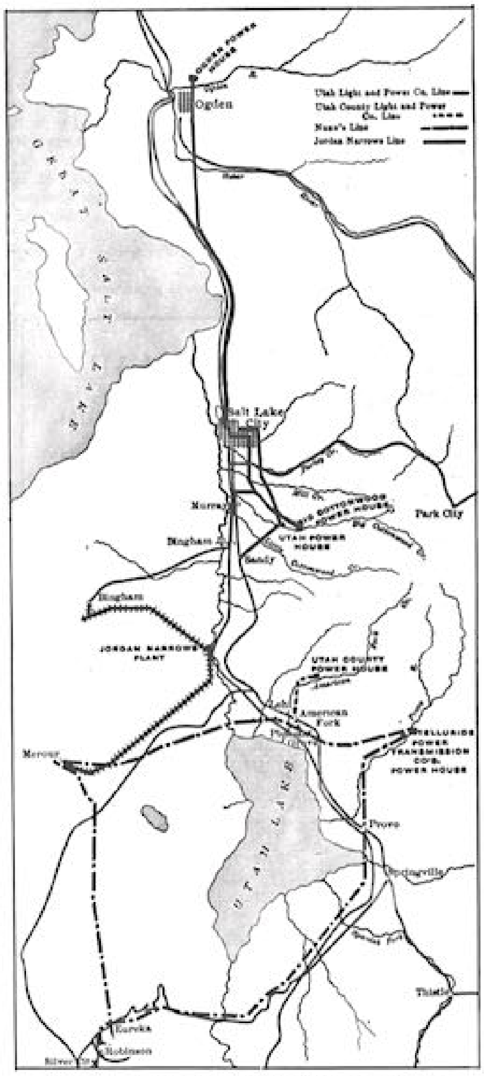
In this paper, I present a new narrative and collect new primary evidence to show that rural electrification was not the product of a single big push, but of an initial gradual expansion driven by private investment in the Western states, only later joined by government assistance. In this first chapter, I first present the differences between electrification in the Western versus the Eastern states; then I present some historical evidence about the effects of rural electrification in the Western states. In the following chapters, this evidence will serve as the basis for the theoretical framework I will use to estimate those impacts.

3.2.2 West vs East

In every single issue of the *Electrical World*, one of the first American engineering publications about "current progress in electricity and its practical applications", there is a lengthy section about the most recent developments of power transmission in the West. This was also true for other publications I have consulted, such as the *The Journal of Electricity and Western Industry*, or *The Journal of Electricity, Power and Gas*. Starting in 1901 with maps of the transmission systems in Utah, Colorado and California, up to 1923 with a detailed map of Western US Transmission Lines - see figure 3.2 and figure 3.3 below -, the attention of the contemporary scientific community was on the states west of the 100th meridian.

The reason behind this copious material is quite evident: transmission of electrical power was developing quickly in the Western states in comparison with the Eastern

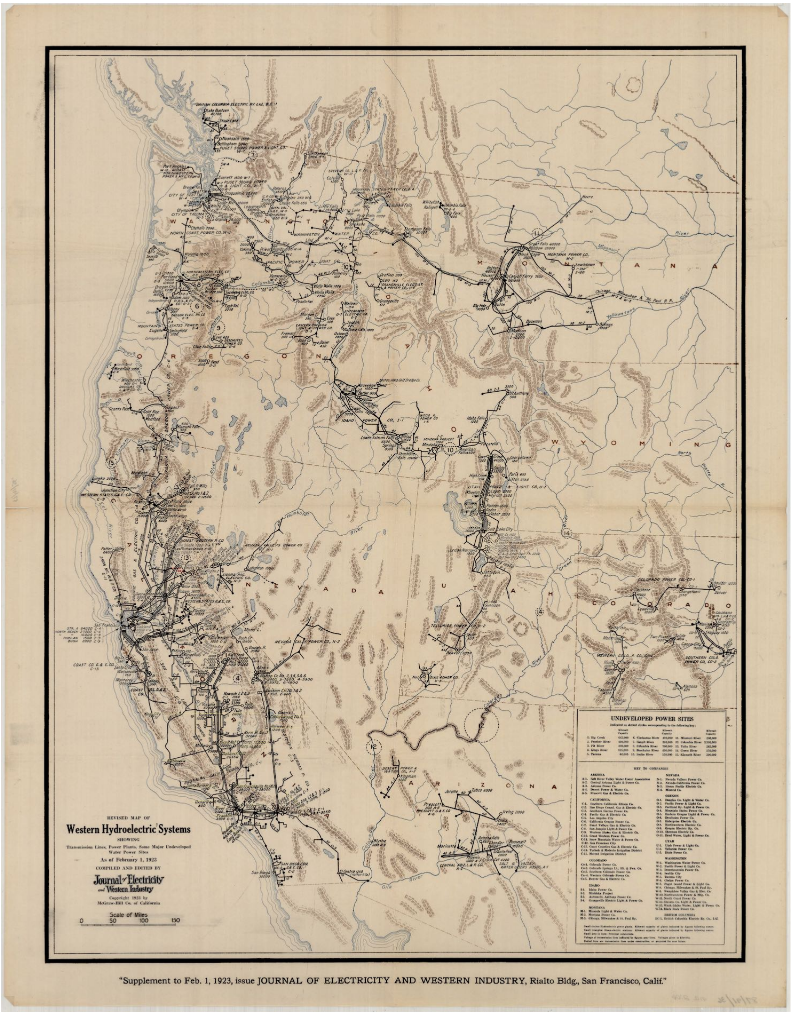
Figure 3.2: Map of Transmission Lines, Utah, 1901.



MAP OF HIGH-TENSION TRANSMISSION LINES IN UTAH.

Source: Electrical World, Volume 37, p. 592, 1901.

Figure 3.3: Map of Western Transmission Lines, 1923.



Source: The Journal of Electricity and Western Industry, 1923.

side. Initially driven by the demand of mining companies located in the Rocky Mountains, Sierra Nevada and Cascade ranges, hydroelectric power developed together with the technology that made long transmission lines possible. Longer transmission lines meant that it was possible to connect the generating stations with the urban centres driving new demand for electric power. California in particular stood out as being at the forefront of innovation in terms of transmission lines [4] In a review of the existing transmission systems operating in 1914, Electrical World shows that most of them (and the longest ones) are in the Western US, especially in California, Utah and Colorado. By looking at the collection of maps, the reports available in the above-mentioned journals, and additional newly collected data from primary sources, we can make two general observations about rural electrification in the Western states:

1. First, California was a true exception to the norm but the other states followed closely, compared to the Eastern States. In 1925, 24% of Californian farms had access to electricity. Most importantly though, Washington (18%) followed closely and Utah (13%) and Idaho (11%) were next, with a percentage rate higher than the more densely populated regions of New England (see figure 3.4). Although the rates might seem negligible in magnitude, they are remarkable when comparing them to the aggregate data for the whole US which records only 4% of farms as electrified in that year.

2. Secondly, electrification rates for the West go up in three jumps (see table 3.1), approximately every 5 to 7 years. One jump before 1930, before 1925 and 1930 (from 12.4 to 26.2%), one between 1930 and 1937 (from 26.2 to 41.7%) and

---

[4] This is true only when considering the U.S. As pointed out by Hughes (1993), if compared with other developed countries (Europe), the US was lagging behind in terms of rural electrification. In Germany for instance, rural electrification was close to saturation in the 1910s. The main reason behind this is the difference in population density.

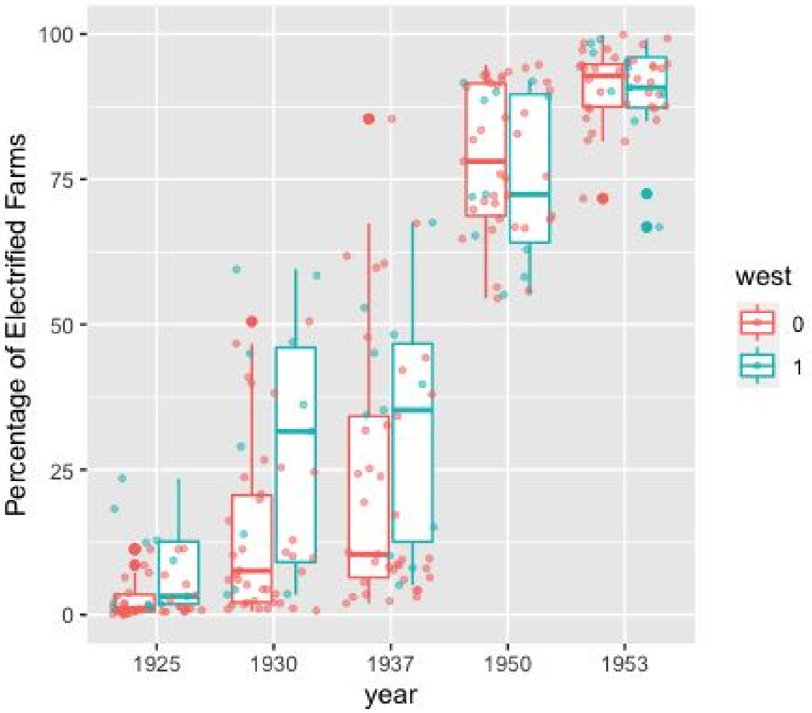
one between 1937 and 1950 (from 41.7 to 83.5%). On the contrary, in the rest of the U.S. there was essentially only one jump in electrification rates between 1937 and 1950 (from 14.5 to 76.6%).

Table 3.1: % Electrified Farms

Year	West	East
1925	12.4	2.3
1930	26.2	9.2
1937	41.7	14.5
1950	83.5	76.6
1957	93.2	90.6

Sources: Electric Light and Power Industry Report, NELA, 1931; Report to the Rural Electrification Administration, 1937; Annual Report of the department of Agriculture, 1953. Author's calculations.

Figure 3.4: Map of Western Transmission Lines, 1923.



Source: see table 3.1

California exceptionalism has received significant attention in the literature. In his history of the development of the American grid, Hughes (1993) dedicated an entire chapter ("California White Coal") to California's highly developed transmission system. A few factors are aligned to make California the leader in hydropower and long-distance transmission. The presence of mountain ranges provided huge water potential to be deployed thanks to water-turbines[5]. At the same time, scarcity of coal[6] resources made steam-generated electricity very expensive and the prices were high enough to ignite interest in developing alternative energy sources. Then, urban demand from the coastal cities spurred investment in developing "point-to-point" transmission lines, rather than transmission networks. That is to say, the initial connection was made between hydroelectric power plants in the mountains and cities on the coast, rather than generating in the cities and then connecting to the periphery (Hughes (1993); Markwart (1927), Markwart (1940))[7].

Within this background of urban hydropower demand and investment, rural electrification followed. In the West, agricultural communities had their own demand for electricity: given the semi-arid climate and the low levels of precipitation, water scarcity was a recurrent problem. There, farmers could use electricity-powered pumps to divert surface water and extract groundwater and between 1890 and 1930, demand for irrigation pumps was almost exclusively recorded for Western states (Selby, 1949). Again, California had the highest demand, driven by a transition to

---

[5] According to Hughes, water-turbines powerful enough developed only around 1900, which was perfect timing if combined with the development of the technology needed for long-distance transmission (Hughes, 1993)

[6] According to the Twelfth and Thirteenth Census of the U.S. (1900 and 1910), the only Western State producing coal was Colorado, which had a few mines of low-quality bituminous coal. Apart from relative scarcity, transportation costs to major urban centres on the Pacific coast made coal prices high compared to the East.

[7] See figure 5.4 in the Appendix for a graphical representation.

a very water-intensive "Mediterranean" type of agriculture based on fruit and nut trees. Water for agriculture was almost exclusively pumped by using electricity: in a paper by Olmstead and Rhode (1988), the authors used the number of irrigation pumps in a county as a proxy for electrification of farms.

In addition to the higher load driven by irrigation needs, the possibility of connecting to the grid partially materialised because of chance. Most of the original transmission lines built between 1900 and 1910 were 60 kilovolts, a voltage suitable for diversions along the transmission line itself. A.H. Markwart, the vice-president of Pacific Gas and Electric Company (the biggest power company in California at the time) described this peculiarity as follows:

> *"Long-distance hydroelectric transmission lines could be tapped en route because this voltage (60 kV) permitted substation costs which bore a proper economic relation to the size of loads to be carried."*

Once farmer communities started connecting to the grid, power companies realised that their irrigation-fuelled load was complementary to the existing load from urban centres, making it profitable to keep serving the rural areas[8]. Looking at data presented in the engineering publications mentioned above, there is sufficient evidence to conclude that a high number of the systems in the Western States were in fact 60 kV, with some exception for transmission systems that served big urban centres such as Los Angeles or San Francisco and that had a higher voltage (Electrical World, Volume 59, 1912, p.1225). That said, I hypothesise that given the available statistics shown above, the rest of the Western States followed California in expanding rural

---

[8] On a secondary note, but still relevant, World War I fuel conservation needs made hydropower from an attractive cheap alternative to a "patriotic necessity" (Williams (1988) citing the Journal of electricity, 1918).

electrical lines to agricultural communities earlier than 1930, albeit with great variance across states with different environmental and therefore agricultural conditions and load. In the West then, access to electricity for rural consumers happened earlier than what the general narrative presents. I will bring additional evidence to support this statement in the rest of the chapter.

Roughly speaking, we consider the states completely west of the 100th meridian of being part of the West. Specifically, these states are Montana, Wyoming, Colorado, New Mexico, Idaho, Utah, Nevada, Arizona, Washington, Oregon and California. The divide between West and East is based more on climate and natural geography than one might think. In 1878, John Wesley Powell identified the "Arid Region" of the U.S. as the land west of the 51-cm-per-year rainfall line, which tracks the actual 100th meridian closely (Powell, 1879). More recently, Seager et al. (2018) collected enough data to confirm that the 100th meridian is in fact a natural divide which can be tracked using an aridity index built on precipitation and evapotranspiration data. Therefore, when hypothesising that the rest of the Western States electrified like California and faster than the Eastern states, I am implying that west of the 100th meridian, climatic conditions called for irrigation to make agriculture viable, creating demand for pumps and therefore electricity. I am also saying that the rugged landscape which presents high variability in the range of elevations, created the optimal environment for developing hydropower, providing the base for the development of cheaper electricity generation.

In 1916, Electrical World collected data about 45 rural distribution lines in a small sample of the U.S (Electrical World, Volume 68, 1916, p.704). 39 of those lines are from the Western States: 26 from Washington, 5 from California, 3 from Idaho, and

5 from Utah. Out of 45 rural lines built in the West, most of them mention access to electrical light as the construction purpose (24), but 17 of them mention irrigation and only 7 mention other power purposes. Additionally, all of the lines are serving hydroelectric plants.

To be clear, development and distribution to rural communities in the West were still not cost-effective, possibly giving more weight to Markwart (1927) hypothesis about lines being "tapped en route". In the same survey of small distribution lines in 1916, Electrical World pointed out that for most of the lines the returns are negative, possibly due to the too high construction costs compared to the rate that they were able to charge. Nonetheless, they also point out that:

> *"More figures were not given for Eastern Territories because few lines serving low-density loads are installed there, this being the practice of many larger companies."*

This can be read as evidence that the Western and Eastern states were following a different pattern in terms of electrification rates, regardless of the returns from investment. Possibly, given the lower construction costs in the West compared to the East, the capital-at-risk in expanding the line was lower. For instance, in Central New York or in Illinois, two states with very high electrification rates, costs per mile for rural lines were more than double the costs in Central California (2750 $ and 2485 $ versus 860 $ per mile, see table 3.2). The costs in other Western states were also lower.

Table 3.2: Distribution Lines

	Washington			**California**					
	East	West	Central	South	Central	**Idaho**	**Utah**	**Illinois**	**NY**
Lines (number)	7	6	13	4	58	3	5	4	2
Length (mi)	100.5	3.8	40.4	150.5	109.4	29.4	52.8	31.3	12.0
Cost ($/mi)	1345	1790	935	1985	860	1400	1000	2485	2750

Source: Electrical World, Volume 59, 1912.

To summarise, historical evidence about the differences in electrification rates in the West versus the East seems to point to three factors:

- **Environmental factors**:
  - **scarcity of coal** meant higher fuel prices and necessity to find an alternative to steam-generated electricity.
  - **higher elevations** meant more hydropower potential and therefore cheaper electricity generation from hydropower.
  - **lower precipitations** meant potential demand for electricity for pumping water for agricultural purposes, hinting at potential higher loads. Higher loads meant that utility companies could justify serving rural customers.
- **Economic factors**: cheaper construction costs possibly due to cheaper rights-of-way allowed for lower-cost investments.
- **Contingent factors**: farmers located underneath transmission lines connecting hydropower to urban centres were able to ”tap en-route”, starting a demand in California which was imitated elsewhere.

Apart from setting the scene for evaluating the effect of electricity access in the West, all of these factors will play a role in the identification strategy. In the next section, I will delve deeper into the evidence about Western rural electrification.

3.2.3 Western Rural Electrification

Having established that the Western states had on average a higher percentage of farms connected to the grid, I can now clarify what are the factors that drove electricity adoption West of the 100th meridian, while also highlighting the differences that there are within the West itself. First, it should be made clear that power generation in the Western states relied on the development of hydropower. Steam-powered central stations were located in relatively big urban centres and in some mining communities. Data about rural electrification rates (in specific, percentage of farms with access to electricity) at the state or county level are only available starting from 1925, so we cannot observe the early uptake of electricity by looking at that metric. Instead, we can make estimations by observing available geographical maps of transmission lines. To do so, I have digitised two sets of maps, one for 1912 and one for 1923.

The collection of maps presented above shows us how transmission lines evolved spatially [9]. In 1912, Electric World emphasised how state-level companies operating separately in different regions were slowly connecting to create a unique transmission system. Some of the most interesting examples were the enterprises which generated hydropower as a by-product of irrigation projects. These projects were located in states with arid climates and very low precipitations. Examples are the Salt River Valley Project in Arizona or the transmission system operated by the Northern

---

[9] See the Appendix B.2 for an example of the original 1912 maps.

Figure 3.5: Transmission Lines and Line Density, 1912.

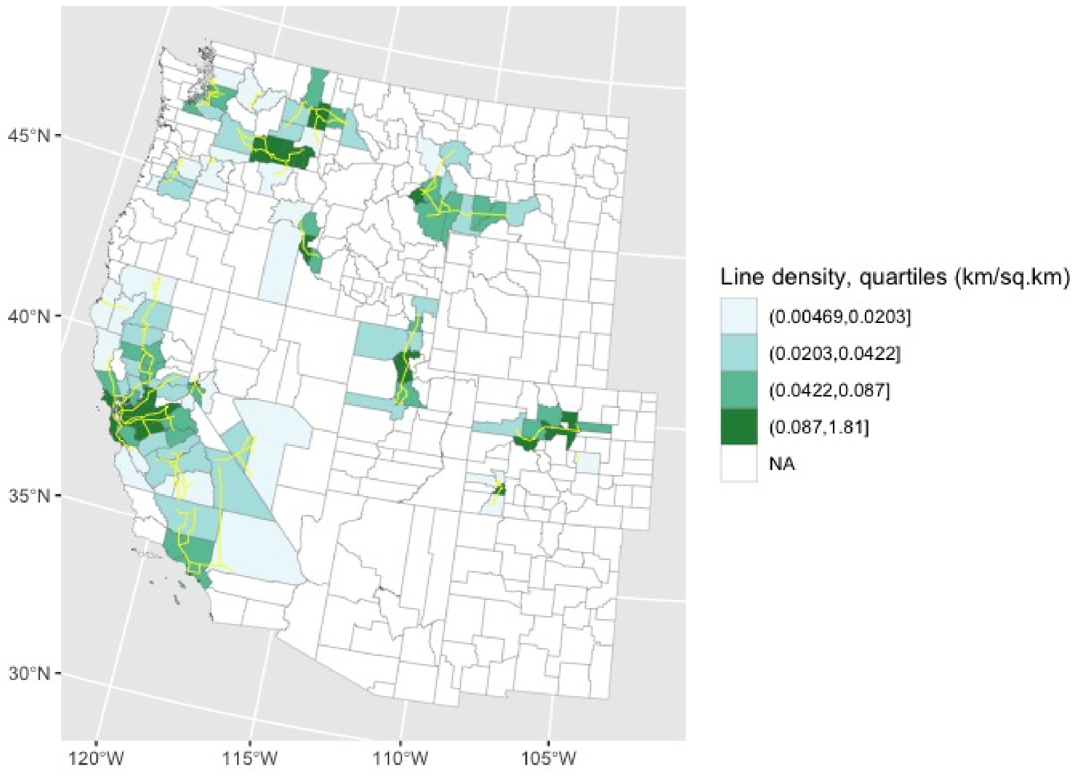
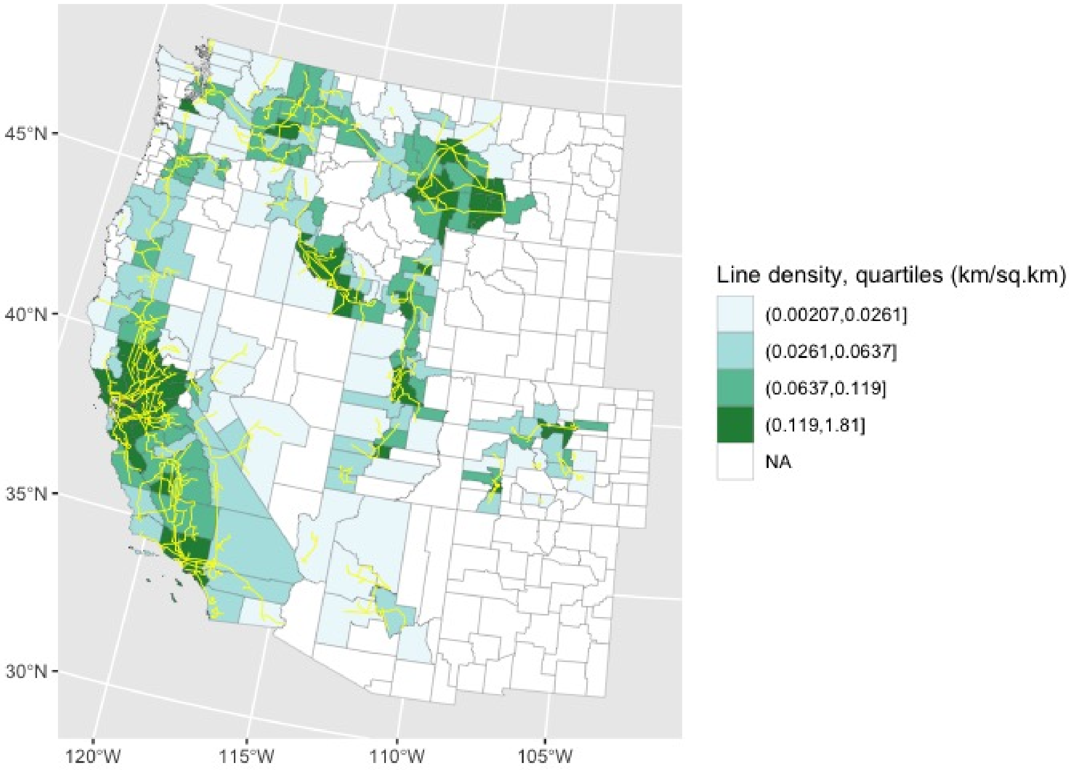


Figure 3.6: Transmission Lines and Line Density, 1923.



Colorado Power Company, serving 1000 sq. miles north of Denver. There were also transmission systems driven by demand from mining communities such as the Bishop Creek development by the Nevada-California Power Company. Long transmission lines across states, such as the connection between Idaho and Utah by the Telluride Company, were still considered exceptional. Finally, states with high precipitation rates such as Washington and Oregon were generally well served because of the high demand from mining, industries and lighting purposes.

By looking at the map for 1923 - see figure 3.3 above for the original map and figure 3.6 for the digitised one-, we can observe two things: transmission lines intensified in all areas, but they intensified more in areas which were already highly served, such as in the Pacific states (California, Washington, Oregon), along the Snake River in Idaho and north of the Great Salt Lake down to Salt Lake City. Areas which were originally served because of irrigation projects (Arizona and Colorado) also developed, but not as much. In Montana, the Montana Power Company developed greatly in order to serve mining companies, where energy was supplied *"at the lowest average rate per kilowatt-hour offered by any company doing a similar business"*. In Washington, lines were extended to several agricultural districts (Electrical World, Volume 75, 1920, p.120), but given the lower loads -higher precipitations meant no need for irrigation agriculture-, these were only distribution lines (with a low voltage, 13 kV) reaching communities that could justify the costs of the extension by expected returns or by contributing to the construction costs.

3.2.4 Use of electricity in Western Rural US and its effects

As it has already been noted, the principal use of electricity in rural areas of the Western states was to pump water for irrigation projects. Solving the problem of water scarcity had been a priority since the first irrigation works during the agricultural booms of the 1860s, which contributed to the construction of ditches and reservoirs. Again, California led the development in water infrastructure during these years (Pisani (2002), Olmstead and Rhode (1988), Selby (1949)). In 1902, the federal government passed the Reclamation Act, which provided federally subsidised funding for major water infrastructure projects. The primary motivation for the Act was to provide agricultural water, but the infrastructure itself provided secondary benefits, including hydropower development. [10]

Apart from granting access to water, electricity served the rural customer in household operations, farm operations and rural community services. Although the initial demand for electricity in the West was driven by demand for irrigation, the demand for electric light in the home followed closely. Access to electric light usually meant a better standard of living and possible improvements in health conditions. It also meant longer hours for household chores and farmyard activities. Already in 1916, the U.C. Journal of Agriculture observed that: *"Electrical manufacturers have been busily at work building electrical household appliances that would give the farmer's wife every comfort and convenience that her city sister enjoys."* In 1925, farms in California had a 68% saturation rate for flatirons, and around 35% for washing machines, vacuum cleaners and toasters (Moses, 1929; Williams, 1988; Nye, 1997).

In terms of farm operations, the application of electricity to farm work was still in

---

[10] The link between access to electricity and agricultural water is an endogenous one: electricity generated by hydroelectric power plants helped extract water for irrigation, but locations that had more access to water were also more likely to get access to electricity, given that all of the electricity was generated by hydropower. Nonetheless, there is sufficient evidence to show that the effects of the Bureau of Reclamation projects were only felt after 1940 (Pisani (2002); Edwards and Smith (2018)), possibly because water extraction technologies were not developed enough. Therefore, this should not affect my identification strategy when trying to estimate the effect of access to electricity.

an experimental stage up to 1920, and its usefulness depended on the type of work considered. For transportation (hauling goods) or fieldwork (tillage or harvesting), electrical power was not superior to other types of mechanical power, but for stationary work (such as bailing, grinding and other farmyard activities, such as churning butter), electricity revealed itself as a very useful source of power. This is why dairy and poultry farmers were early electricity users, partially because they were located close to urban areas (and therefore they had earlier access to transmission lines) and partially because of the type of activities they were involved in, which made use of appliances such as electric churners, milkers and hatchers. Additionally, citrus growers and other orchardists used electricity for frost protection and dehydration of fruits and nuts. Finally, in addition to serving the individual farms, rural electric service was already changing the activities of small rural communities and towns by providing entertainment and the possibility to meet during the evenings thanks to the electric lights (Beal, 1940).

Even though there is evidence of the uses of electricity on the farm and in rural communities, there is no empirical estimate of its effects for the period before 1930. In the next chapter, I will try to derive first intuitively and then theoretically the possible effects of the use of electricity.

3.3 Theoretical Background

3.3.1 Local markets and Rural Electrification

Given the spatial nature of the data, previous literature used variations of Rosen (1974) and Roback (1982) models to build a theoretical framework for the analysis. These general equilibrium models capture shocks to local economies simply and intuitively which is easily adaptable. In its original formulation, the production function is built at the city level. Each city is a competitive economy using labor (L), capital (K), land (T) - or another fixed factor F - and a local "amenity" factor (A), sometime called a "productivity shifter" (Moretti, 2011), which summarises the advantages or disadvantages of a specific location in space.

In studying the impact of electricity access, this framework has been adapted in several different ways. In a paper by Kline and Moretti (2014), the authors model U.S. counties as small open economies with price-taking behaviour on all inputs in which the differences in county-level output result from spatial factors (differences in amenity, locational productivity advantages and agglomeration externalities). Outcomes of interests (Y_{it}) are produced in each county following the typical Cobb-Douglas production function ($Y_{it} = A_{it} K_{it}^{\alpha} F_{i}^{\beta} L_{it}^{1-\alpha-\beta}$).

In this specific structural formulation, A_{it} is a locational advantage which consists of agglomeration and locational advantages as well as the advantage of accessing specific location-based amenities such as electricity, which in Kline and Moretti (2014) equals having access to the Tennessee Valley Authority investment.

In a simplified (and earlier) version of the above-mentioned model, Greenstone et al. (2008) builds a one-sector model with firms -instead of counties- maximising a generalised production function of the type:

$$\mathrm{max}_{\mathrm{L,K,T}}\, \pi = \mathrm{f(A, L, K, T) - wL - rK - qT}$$

This same model has been adapted to accommodate the case of increased access to electricity by Kitchens and Fishback (2015), in which the authors model the effect of REA loans on farms rather than firm productivity. In their version, energy is included explicitly as a separate input (E), given that REA loans directly impacted energy prices. According to their specification, access to REA loans affects the farm's profitability through increased productivity, lower profits through raised wages and increased land prices, and increased profits due to lower energy prices, with an overall ambiguous effect.

Nearly identical Roback-style models have also been used to describe the impact of access to other spatially varying factors affecting productivity. In Hornbeck (2012) the authors formally develop the theoretical framework outlined by Foster and Rosenzweig (2004) in order to explain variation in outcomes due to different access to groundwater. In their model there are two production sectors, agriculture and industry, and they are both modelled with a production function in which firms maximise inputs given the productivity shifter, which includes access to the aquifer.

In a slightly different fashion, Lewis and Severnini (2017) make use of the Rosen-Roback model so that the effects of electricity access result in increased agricultural productivity, but also in improvements in the quality of rural housing. While farm productivity effects are the same, as housing quality enters the utility function of workers, effects on wages are ambiguous. This is because housing quality is also a location-dependent amenity (A) but does not enter the production function.

3.3.2 A Simple Theoretical Framework and Hypotheses

As observed in the historical background, the effects of access to electricity depend on the use of electricity itself. As a generalisation, we can divide these effects in three types and see how they can be included in a theoretical framework:

- **Household effects**: effects derived from the use of technologies *within* the home such as the use of electric light and other appliances. As access to electricity, and therefore the possibility to use electric lights and other appliances, is location specific, these effects will enter the profit function through the productivity shifter, A.

- **Farm effects**: effects derived from the use of specific farm equipment such as irrigation pumps or milking machines. These effects will enter the profit function through the productivity shifter (again, because access to electricity is location specific) and through a reduction in input prices (e.g. reduction in energy prices, which is also location specific, as it depends on electricity access).

- **External effects**: effects at the community level which derive from the use of power for other purposes, such as for industry or entertainment. These effects will not be directly considered in this study as they fall outside the profit function.

That said, we propose a simple theoretical framework which could help us explain empirical results and that incorporates the effects of access to electricity into the amenity (productivity shifter) input and into input prices. Similarly to Hornbeck (2012); Greenstone et al. (2008); Kitchens and Fishback (2015) we have a model where all firms (agricultural firms in this case) in a county use labor L, capital K, land T and energy E to produce an internationally traded good with a fixed and normalised price equal to one. Firms in each county (c) and sector maximise the following expression:

$$\max_{\mathrm{L,K,T,E}} \pi_c = \mathrm{f(A, L, K, T, E) - wL - rK - qT - sE} \tag{3.1}$$

Input prices are w for labor, r for capital, q for land, and s for energy.

To formulate our hypotheses with regards to the effect of electricity access we can write the above function in terms of the optimal level of inputs, and then make these inputs depend on the access to electricity (H, as in hydropower) e.g. for labour $L^*(A(H), w(H), r, q(H), s(H))$. Then, to see how profits change when firms gain access to electricity thanks to hydropower, we can take the total derivative of Π_c, profits, with respect to H, electricity access. After collecting terms and cancelling out inputs minus their marginal product (perfect factor markets) we get:

$$\frac{d\Pi_c}{dH} = \left(\frac{df}{dA}\frac{dA}{dH}\right) - \frac{dw}{dH}L^* - \frac{dq}{dH}T^* - \frac{ds}{dH}E^* \tag{3.2}$$

In which we can see that access to electricity, H, affects Π_c through an effect on the productivity shifter, A, as well as on the input prices. Note that K is not affected as it is traded on the national market so that local conditions do not impact its price. From equation 3.2, it follows that our hypotheses can be derived as:

- $\frac{dA}{dH} > 0$: Access to electricity affects the productivity shifter positively through the use of new appliances in the home as well as on the farm, resulting in higher profits.

- $\frac{dw}{dH} \& \frac{dq}{dH} > 0$: Access to electricity affects input prices. A change in A increases wages and land prices affecting profits negatively.

- $\frac{ds}{dH} < 0$: Access to electricity decreases energy prices affecting profits positively.

Overall, according to this simple model, effects on rural economies are ambiguous as we do not know whether gains from a change in A will be offset by increases in input prices.

3.4 Data and Summary Statistics

3.4.1 Data

To test whether access to electricity had an effect on the rural economy, I compiled a county-level dataset for U.S. Western states between 1900 and 1930. All of the variables are from the Agricultural Census and the general Census Haines et al. (2018); Haines (2018), both collected every 10 years. The main dependent variables are general demographic indicators (population, share of urban and rural population), farm characteristics (farmland value - which is the total value of land and buildings divided by the total number of farm acres -, farm output in $, total number of farms, average farm size in acres), and general agricultural sector indicators (share of wheat and corn crops, livestock per acre of farmland, irrigated acres and share of irrigated acres).

Considering the absence of county-level data pertaining to electricity access from 1900 to 1930, I constructed a proxy variable to represent this factor: the proximity to the nearest hydroelectric power plant (refer to Figure 3.7). This metric has been employed in previous research as an instrumental variable for electricity access (Lewis, 2014; Kitchens and Fishback, 2015) or as a similar proxy (Lewis, 2018; Lewis and Severnini, 2017; Severnini, 2014) in studies examining the rural electrification process in the U.S. post-1930s. In the initial phase of this analysis, the aim is to reproduce the findings of these studies, focusing on the early stages of electrification.

The proxy variable was derived from the EIA-860 survey data provided by the U.S. Energy Information Administration (EIA, 2020)[11]. Utilizing this variable rests on the assumption that counties in closer proximity to a hydroelectric power plant incurred lower costs to access hydroelectric power. In earlier research, the validity of this assumption can be empirically examined by analyzing the percentage of households with electricity access in a specific year and determining how the variation in this access correlates with the changes in proximity to the nearest power plant. In this study, the strength of this relationship is supported by the anecdotal evidence presented in the historical background section: already in early 1910s, after the development of the first federal dams projects, hydroelectric power generation was the cheapest source of energy. In the early 1900s, long distance transmission lines were still in the development phase but in the Western states, especially California, coastal cities drove the demand for lines to be built between the coast and the mountain ranges. There, counties started to access electric power by 'tapping en route'. Fig 3.7 shows a map of power plants' openings between 1900 and 1910 and between 1910 and 1930. In 1900 there were only 4 operating hydro power plants, a number which increased to 44 in 1910, 90 in 1920 and 159 in 1930. In 1900, the average power supplied by a hydroelectric power station was around 5.8 megawatts, thirty years later the average capacity was 17.1 megawatts.

Following previous studies (Lewis, 2018), I address selection biases by limiting the

---

[11] This dataset can be accessed publicly at `https://www.eia.gov/electricity/data/eia860/` and contains information about the location, as well as the opening and closing years of hydroelectric power plants, with records dating back to the late 1880s.

number of counties included in the final dataset. From an initial sample of 413 counties, I dropped 12 counties considered metropolitan areas in 1900 [12]. Then, I trimmed samples to control for endogeneity of treatment inclusion - counties too far away from a power plant are unlikely to get access to electricity -, based on the distance from the nearest hydropower plant in 1900. There are 144 counties less than 200 miles away from a power plant, 235 counties less than 300 miles, and 308 counties less than 400 miles. The main specification is the more conservative, with 144 counties; the rest are used as robustness checks.

Table 3.3: Characteristics of Hydro-electric Power Plants in the Western States

Year	MW (mean)	MW (sd)	Number
1900	5.8	4.3	4
1910	7.9	8.5	44
1920	13.9	19.4	90
1930	17.1	23.6	159

Sources: EIA-860. Author's calculations. Power is in average megawatt for a given year.

3.4.2 Summary Statistics

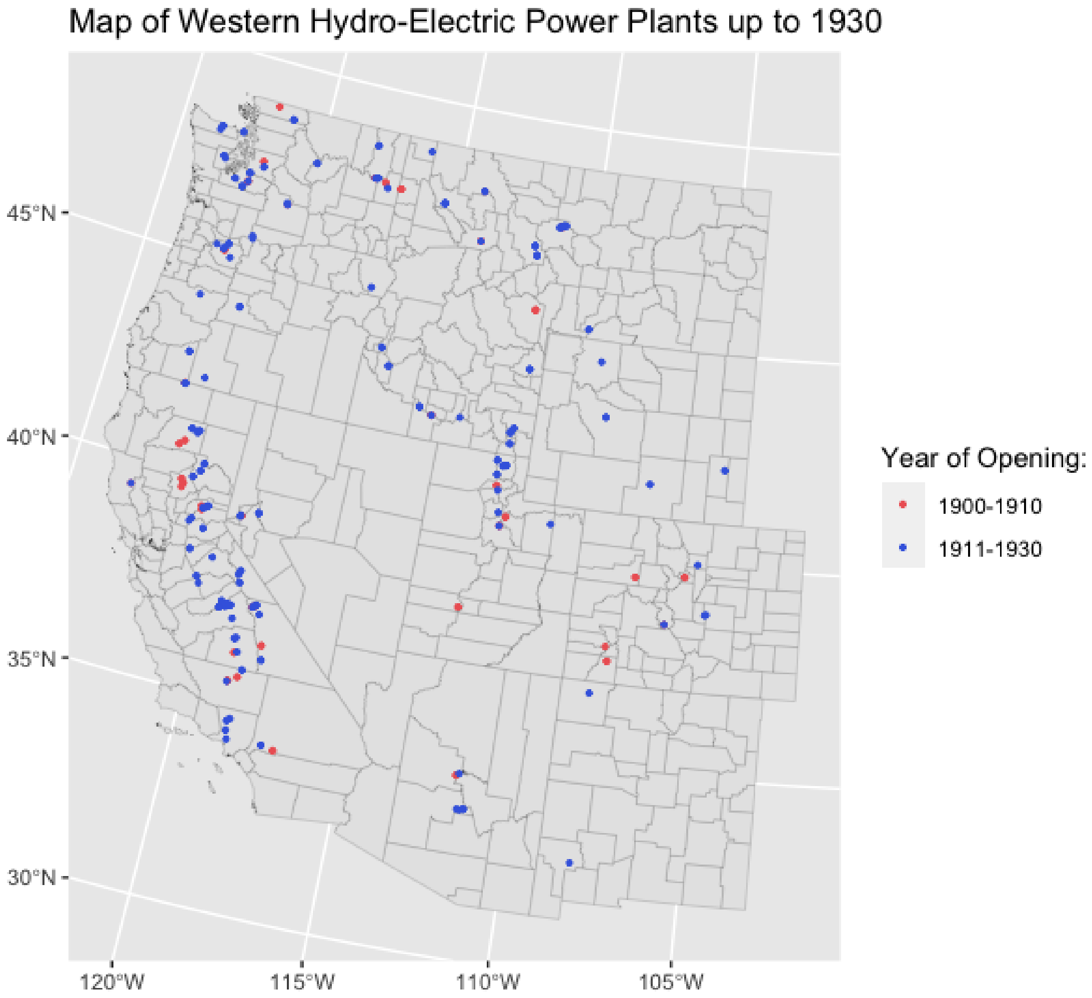
In the summary statistics for the final dataset, several general trends and region-specific trends are observed. The West experienced ongoing growth in farmland area after 1900, while this growth slowed down in the rest of the U.S. This increase in farmland area in the West can be attributed to new settlements and land reclamation projects, which made additional land available for farming.

Between 1900 and 1930, the West was a predominantly white and rural region. On

---

[12] These are, in alphabetical order: Alameda, Arapahoe (now Denver), Los Angeles, King, Multnomah, Pierce, Pueblo, Sacramento, Salt Lake, San Francisco, Silver Bow, Spokane. These were the counties with more than 50% of the total population living in a city with more than 25'000 people. They loosely reflect metropolitan districts in 1930.

Figure 3.7: Western Hydro-Electric Power Plants



Source: see text.

average, over ninety percent of a county's population was white, and the share of people living in small urban centers—with more than two thousand people—never exceeded twenty-three percent throughout the entire period.

The increase in farm output per acre reflects the general trend of agricultural production being particularly favorable up until the First World War. This period, culminating in the 1910-1914 "Golden Age of Agriculture" (Gardner, 2006), saw an increase in agricultural good prices, primarily driven by rapid urbanization and exports. As in any economic boom, land prices rose, up to the end of the First World War, when Europe restarted production and massively decreased the demand for American produce. This is reflected in almost all county-level variables in our dataset. For instance, farm output per farmland acre rose from 4 to 17 dollars in 1920, to then fall to 13 dollars in 1930. Other indicators - corn and wheat acreage and irrigation related variables - were chosen to examine whether electricity impacted farm operations through increased water use. While corn, a very water-intensive crop, has an almost non-existent share throughout the whole period, wheat fluctuated around 2% with no significant changes, but with considerable variation. The same is true for the amount and share of irrigated acreage. Looking at the explanatory variable, the average distance to the nearest hydroelectric power plant for counties was 120 miles in 1900, which decreased threefold to 40 miles by 1930, a meaningful variation.

3.5 Empirical Framework

This model can be interpreted as a Differences-in-Differences approach with multiple time periods, continuous treatment and individual panel data and has been used by Lewis (2018) and Lewis and Severnini (2017) when looking at the effect of access to

Table 3.4: Summary Statistics

	1900	**1910**	**1920**	**1930**	**Mean**
Total Population	9507.7	15064.6	18791.5	24185.4	16912.2
	(8813.4)	(14399.3)	(19700.5)	(30719.6)	(20830.0)
Rural Population	7385.0	10410.3	11898.6	13854.5	10893.6
	(5789.6)	(7912.7)	(9650.8)	(12701.4)	(9654.2)
% Urban, 25k cities	0.00266	0.0149	0.0281	0.0366	0.0206
	(0.0158)	(0.0865)	(0.113)	(0.131)	(0.0978)
% Urban, 2k cities	0.126	0.176	0.216	0.237	0.189
	(0.189)	(0.214)	(0.231)	(0.237)	(0.222)
Density, per sq. mile	7.149	10.88	13.39	17.08	12.14
	(8.794)	(13.29)	(17.57)	(26.38)	(18.12)
% White	0.934	0.951	0.957	0.935	0.944
	(0.0706)	(0.0482)	(0.0569)	(0.0800)	(0.0658)
Farms	777.2	1057.9	1349.3	1509.7	1175.1
	(718.0)	(954.5)	(1269.4)	(1582.3)	(1210.1)
Average Farm Size, acres	408.1	343.6	335.2	441.5	382.6
	(416.6)	(326.4)	(278.7)	(597.8)	(425.3)
Farm area, 000s acres	262.3	282.4	321.9	367.9	309.0
	(284.4)	(286.3)	(294.1)	(351.3)	(307.4)
Farmland value, 000s $	4169.3	11175.7	21413.3	25534.4	15633.4
	(6269.2)	(13772.8)	(28627.9)	(38605.6)	(26628.5)
Farmland value per acre, $	14.22	40.16	67.60	80.17	50.68
	(11.63)	(33.82)	(60.99)	(103.3)	(67.77)
Farm output, 000s $	945.6	1490.2	5216.0	4420.0	3038.9
	(1105.3)	(1789.5)	(6854.8)	(7759.4)	(5610.4)
Farm output per acre	4.186	5.626	17.55	13.36	10.24
	(2.372)	(4.038)	(19.37)	(21.04)	(15.56)
Livestock, 000s	811.1	1477.7	2568.8	848.3	1425.8
	(670.0)	(1260.9)	(2122.2)	(807.0)	(1518.2)
Livestock per acre	4.931	6.789	10.49	2.963	6.287
	(4.305)	(3.903)	(6.124)	(2.297)	(5.188)
% Wheat Acreage	0.0250	0.0151	0.0260	0.0225	0.0223
	(0.0539)	(0.0418)	(0.0485)	(0.0486)	(0.0485)
% Corn Acreage	0.000504	0.000419	0.000992	0.00109	0.000755
	(0.00151)	(0.000872)	(0.00326)	(0.00282)	(0.00235)
Irrigated Acreage	19843.5	33143.6	50982.1	48756.0	38250.3
	(26352.5)	(43295.2)	(67749.0)	(72972.8)	(57301.5)
% Irrigated Acreage	0.140	0.195	0.234	0.179	0.187
	(0.184)	(0.199)	(0.266)	(0.222)	(0.222)
Distance nearest power plant, tens miles	12.24	6.383	5.186	4.406	7.062
	(4.930)	(3.671)	(3.309)	(3.151)	(4.919)

Notes: Data are collected at the county level for counties less than 200 miles from an hydroelectric power plant. There are in total 144 counties and no missing data. Quantities in dollar values are not adjusted for inflation, they will be adjusted in the estimation.

electricity in the US for the 1930-1960 period.

To implement it, I follow the methodology of Wooldridge (2007), which uses a combination of first differencing and a two-way fixed effects estimator. Results are then interpreted as a Differences-in-Differences estimate, where the differences are within units -in our case, counties- and over time. In this specification, unconfoundedness or exogeneity of treatment is conditional on observable and unobservable covariates. Time-constant unobservables are differenced out in the first stage, observables are added as a series of baseline controls, so that parallel trends are conditional on baseline characteristics. As we consider a continuous variable, overlap between groups equals balance in covariates between counties with relatively high and low electricity access levels.

The specification is described by the following equation:

$$y_{ct} = \alpha_c + \lambda_t + \beta distH_{ct} + \gamma X_{c,1900} \cdot t + \varepsilon_{ct} \tag{3.3}$$

where α_c and λ_t denote county-level and time (decade) fixed-effects respectively. $DistH_{ct}$, distance to a hydroelectric power plant, is our main variable of interest, for which β captures the effect of electricity access on a series of economic and social outcomes. Following similar studies (Kitchens and Fishback, 2015; Lewis and Severnini, 2017), we add two series of baseline covariates interacted with decade-time effects, which relax the parallel trend assumption: $X_{c,1900}$, a series of baseline characteristics (population density, share of urban population, manufacturing employment). In different specifications, I also add a series of baseline characteristics for the nearest metropolitan area (+Metro: population density, manufacturing employment, distance to), and additional geographic characteristics (+Geo: average precipitation between 1900-1930, average and range of elevation). The inclusion of these baseline covariates enables the analysis to account for varying trends in outcomes based on the specified characteristics. For example, counties with higher average precipitation levels may exhibit different trajectories, as their benefits from electricity access could be distinct. By controlling for these factors, the study can more accurately isolate the effects of electricity access on the rural economy and better understand the underlying relationships.

Once we take first differences, (3) reduces to:

$$\Delta y_{ct} = \eta_t + \beta \Delta DistH_{ct} + \gamma X_{c,1900} + \Delta \varepsilon_{ct} \tag{3.4}$$

In the final specification, variations in outcomes within a county are associated with shifts in electricity access, as indicated by the proximity to the nearest hydroelectric power plant. This relationship is examined while accounting for baseline characteristics and decade-specific trends that control for unique events during the period, such as the "Golden Years" of agriculture and the onset of the Great Depression in 1929. Our additional key assumption is that the distance to the nearest hydroelectric power plant is exogenous. This assumption is supported by the fact that the location of hydroelectric power plants was determined by exogenous topographical and geographical factors, as outlined in the historical background section. Furthermore, the demand for power was primarily driven by major urban centers, which have been excluded from the dataset.

3.6 Results

All specifications have clustered standard errors at the county level to control for serial correlation [13]. Given the two-step procedure in which it is possible to run first-differences and then add decade fixed effects, all specifications have county fixed effects and time (decade-level) trends. I have also run specifications with state trends and state-time trends, with no significant difference in the results.

3.6.1 Farmland Value and Farm Output

Table 3.5 below reports results for main outcomes related to the rural economy. I report results for the coefficient of interest only, distance to the nearest hydroelectric power plant ($DistH$).

Every specification from column (1) to column (3) has county fixed-effects (implemented by first differences), decade trends and state trends. In each table, column (1) includes baseline controls, column (2) adds controls at the metropolitan area level (demographics controls for the nearest urban area) and column (3) adds geographic level controls. Baseline controls and metropolitan area controls are from the base year, 1900. Geographic controls are time-invariant.

In general, a decrease in 10 miles in distance to the nearest power plant, a proxy for having access to electricity, is correlated with an increase in farmland value per acre. Across all three specifications, the coefficient on $DistH$ is positive and statistically significant at the 1% level. This implies that an increase in distance to the nearest hydroelectric power plant is associated with an increase in the farmland value,

---

[13] Following the literature, I have also run this same specification with Conley Standard Errors. Results for the regressions with Conley Standard Errors are reported in the robustness checks (Chapter 3.7).

holding other factors constant. Specifically, the coefficient on $DistH$ tells us that electrification is associated with a (0.027*100) 2.7% increase in value. The coefficient remains relatively stable across specifications, suggesting that the relationship is robust to the inclusion of additional controls. In 1910, the mean land value per acre was 42$, and the mean farm was approximately 345 acres; a 2.7% increase in value would equal approximately 340$ per farm.

The coefficient on $DistH$ for the log number of farms is positive and statistically significant at the 10% level in the first two specifications, indicating that an increase in distance to the nearest power plant is associated with a higher number of farms, although the relationship is weaker compared to farmland value. Specifically, in column (1), a 1-unit increase in distance is associated with a 1.2% increase in the number of farms (0.012 * 100 = 1.2%). However, the coefficient is not statistically significant in column (3), which might suggest that the relationship between distance to the nearest power plant and the number of farms is sensitive to the inclusion of controls.

The coefficient on $DistH$ for log farm output is positive but not statistically significant across all specifications, suggesting that there is no clear evidence of a relationship between distance to the nearest power plant and farm output when controlling for other factors. The lack of significance might indicate that electricity access is not the primary driver of farm output, or that the relationship between electricity access and farm output is more complex than what the model captures.

The coefficient on $DistH$ for log rural population is positive but not statistically significant across all specifications, which indicates that there is no strong evidence of a relationship between distance to the nearest power plant and rural population when accounting for other factors. This may suggest that other factors, such as employment opportunities or amenities, play a more significant role in determining the size of the rural population than electricity access.

Table 3.5: Main Results

	Coeff. on $\Delta DistH$		
	(1)	(2)	(3)
Controls:	base	+ metro	+ geo
Log Farmland Value	0.027^{***}	0.026^{***}	0.025^{***}
	(0.008)	(0.008)	(0.009)
Log # of Farms	0.012^{*}	0.012^{*}	0.011
	(0.007)	(0.007)	(0.007)
Log Farm Output	0.006	0.003	0.0004
	(0.009)	(0.009)	(0.009)
Log Rural Population	0.003	0.005	0.005
	(0.005)	(0.006)	(0.006)
Observations	444	444	444

Note: *p<0.1; **p<0.05; ***p<0.01

Except for the growth in land values and the quantity of farms, the outcomes of this study markedly differ from those reported by Lewis and Severnini (2017); Kitchens and Fishback (2015); Kline and Moretti (2014). Their research indicated that, between 1930 and 1960, electrification contributed to an expansion in agricultural output, the number of farms, total land dedicated to agriculture, and farm size, which contrasts with the results observed in the current analysis.

These results lead to two observations. First, electricity, as a general-purpose technology, necessitated an extended period - in this instance, more than a few decades - to visibly impact the rural economy. This delay in effects can be attributed to (1) the delay in innovation of and (2) the delay in the adoption of complementary technologies. Between 1900 and the late 1910s, the implementation of electricity for specific agricultural tasks remained in an experimental phase. It was not until 1912-1914 that companies began expressing interest in producing and marketing electric farm equipment, with initial testing conducted in highly productive Californian regions (Spence, 1962). Consequently, despite the West's early access to electricity compared to the rest of rural America, its influence on agricultural production was minimal.

Second, while electricity could be employed for efficient water pumping in irrigation, many farmers were unaware of this application. It was not until after the 1930s that the practice of "dry farming" - cultivating crops without irrigation in areas with limited precipitation - was considered the optimal method for managing land west of the 100th parallel (Gardner, 2006). This belief may have restricted the use of electricity for irrigation purposes, thereby explaining the magnitude of the observed outcomes. Interestingly, these results contrast with the primary sources from that era, which largely conveyed optimism regarding the prospects of investing in the electrification of farming in the West (World, 1912, 1920).

3.6.2 Population and Employment

The results for the general demographics and urbanisation rates portray a similar story. The coefficient on $DistH$ for log rural population is positive but not statistically significant across all specifications, which indicates that there is no strong evidence of a relationship between distance to the nearest power plant and rural population when accounting for other factors.

Rural electrification appears to have had no significant effect on population numbers or urbanization rates between 1910 and 1930. The West did experience an increase in both rural and urban populations during this time, but most of the change occurred within urban centers, which were excluded from the analysis. In comparison with previous research, it seems that access to electricity only began to impact migration patterns after 1930, slowing down the pace of rural depopulation and contributing to an increase in agricultural employment. This may be due to the limited prevalence of electricity-dependent technologies during the study period, which prevented electricity from having a substantial impact on agricultural productivity and labor demand.

Regarding manufacturing employment, data was only available for 1920 and 1930. In first differences, this equates to a single set of observations. There is a 7% effect suggesting that electricity may have had an impact on local manufacturing employment, but the estimates are only statistically significant at the 10% level. This hints at a potential influence of electricity access on manufacturing jobs at the local level but should be interpreted with caution due to the limited data available.

3.6.3 Timing and light

To conclude, the results point to one significant effect: an increase in land values. The variable farmland value per acre is calculated as the value of the land itself, without buildings and/or equipment. To understand why outcomes in land values are affected by the change in proximity to power plants, I explored whether it was the value of the land or the value of the buildings driving this relationship. I have also looked at

Table 3.6: Manufacturing 1920-30

	Manufacturing Employment		
	base	+ metro	+ geo
	(1)	(2)	(3)
$\Delta DistH$	0.076^{*}	0.072^{*}	0.062
	(0.040)	(0.039)	(0.045)
Observations	148	148	148
Adjusted R^2	0.030	0.034	0.033
F Statistic	1.384 (df = 12; 135)	1.322 (df = 16; 131)	1.265 (df = 19; 128)
Note:		*p<0.1; **p<0.05; ***p<0.01	

whether there were differences in effects according to the decade (1900-10, 1910-20 or 1920-30). Looking at Table 3.7, it appears that there are no significant differences in the contribution of land versus buildings; both receive a 3% effect from a 10 miles reduction in distance to the nearest power plant. Nonetheless, when controlling for interactions between the main explanatory variable of interest $DistH$ and the decade trends, it is possible to observe how the effect is driven by the change between 1900 and 1910 only. This is an interesting result which might be explained by two facts. While most farm equipment was still in the exploratory phase between 1900 and 1910, electric lights were widely used in urban households and were increasingly adopted by rural ones. The adoption of electric lights might be what was driving an appreciation of land and farm building values, possibly by increasing the quality of life in a certain location. Secondly, the switch in sign for the decades 1910-1920 and 1920-1930 can be explained by the slow onset of the long agricultural depression which followed the golden years of agriculture, terminating around 1918, at the end of the war. The crisis was driven by overproduction and a subsequent decline in agricultural prices. An improvement in technology such as the possibility to connect to the electric grid was definitely not enough to counter the negative effect of the farm crisis.

Table 3.7: Land and Buildings Values by Decade

Buildings:	Without (1)	With (2)	Only (3)
$\Delta DistH$	0.030^{***}	0.032^{***}	0.025^{***}
	(0.008)	(0.008)	(0.008)
$\Delta DistH$ *** year 1920**	-0.021	-0.025	-0.013
	(0.023)	(0.023)	(0.022)
$\Delta DistH$ *** year 1930**	-0.026	-0.064^{***}	-0.060^{***}
	(0.023)	(0.023)	(0.021)
Observations	444	444	444
R^2	0.672	0.668	0.396
F Statistic (df = 23; 420)	37.446^{***}	36.819^{***}	11.975^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

3.7 Robustness Checks

In this section, I present a series of robustness checks to substantiate the validity of the main findings. First, I employ Conley standard errors as a mean to address the potential issue of spatial autocorrelation in the error term, which can arise when nearby observations exhibit similar unobserved characteristics. This approach allows me to account for potential biases in the estimates arising from spatial dependence. Next, I undertake a placebo test by examining counties situated more than 200 miles from the treatment area. These counties should not have been directly affected by the reduction in distance, thereby serving as a useful control group to test the specificity of the treatment effect. A lack of significant effects in this placebo group would support the validity of the main results. I also implement an additional placebo test by measuring the impact of electricity access on land values in 1900, prior to the change in distance to the nearest hydroelectric power plant. This analysis helps confirm that the results are not driven by pre-existing trends or confounding variables that may have been present before the treatment took place. If the pre-treatment land values exhibit no significant changes, this would further reinforce the causal interpretation of the main findings. Lastly, I explore the dynamic effects of the treatment over time.

3.7.1 Conley Standard Errors

Table 3.8 incorporates Conley standard errors at different cutoffs (50km, 500km, 1000km, and 2000km). For this specification, the regression is run without weights for farmland area, so the resulting coefficient is smaller (0.019 vs 0.027). The coefficients on $DistH$ are positive across all specifications, however, the statistical significance of this relationship varies across the different cutoffs. In the 50km and 1000km cutoff specifications (columns 1 and 3), the coefficients are statistically significant at the 10% level, while in the 2000km cutoff specification (column 4), the coefficient is statistically significant at the 5% level. In contrast, the coefficient in the 500km cutoff specification (column 2) is not statistically significant. The interaction terms are not statistically significant in any case. This suggests that there is no clear evidence of a change in the relationship between distance to the nearest power plant and farmland values in 1920 and 1930 compared to the base year. Overall, the results indicate that distance to the nearest power plant is positively associated with farmland values, although the strength of this relationship varies across different cutoffs.

Table 3.8: Conley Standard Errors for Main Specification, no farm land area weights.

	Log Farmland Value			
	50km/31mi	500km/310mi	1000km/621mi	2000km/1242mi
	(1)	(2)	(3)	(4)
$\Delta DistH$	0.019^{*}	0.019	0.019^{*}	0.019^{**}
	(0.010)	(0.012)	(0.011)	(0.007)
$\Delta DistH$ *** year 1920**	-0.023	-0.023	-0.023	-0.023
	(0.018)	(0.020)	(0.019)	(0.016)
$\Delta DistH$ *** year 1930**	-0.023	-0.023	-0.023	-0.023
	(0.022)	(0.022)	(0.017)	(0.015)

Note: *p<0.1; **p<0.05; ***p<0.01.

The variation in significance across different cutoffs could be attributed to several factors. First, the spatial correlation of the data points might differ at various cutoffs. As the distance between the observations increases, the degree of spatial dependence might decline, impacting the standard errors and the statistical significance of the estimated coefficients. This could result in the observed changes in significance levels across different Conley standard error cutoffs. Second, the number of observations within specified cutoffs could vary, leading to fluctuations in the power of the test and the chances of detecting a statistically significant relationship. For instance, when the cutoff is larger, there might be more observations included in the analysis, which could increase the power of the test and improve the chances of detecting a significant relationship. Third, there could be heterogeneous effects of distance to the nearest power plant on farmland values at different spatial scales. The relationship between these variables might not be uniform across space, resulting in different effect sizes for different regions. This heterogeneity could contribute to the varying significance levels across the examined cutoffs. Fourth, there might be threshold effects in the relationship between distance to the nearest power plant and farmland values. For example, the impact of distance on farmland values might be more pronounced beyond a certain distance threshold, or the effect might be non-linear. These threshold effects could contribute to the observed variations in significance levels across different Conley standard error cutoffs. Lastly, the degree of noise captured by the model could change as the cutoff changes, influencing the statistical significance of the relationship. As the cutoff increases, the model might capture more noise or confounding factors, which could affect the precision of the estimates and the significance levels.

3.7.2 Placebo Tests

In the placebo test using counties more than 200 miles away from a power plant, presented in table 3.9, the results provide mixed evidence for the effect of a reduction in distance to the nearest power plant on farmland values, number of farms, farm output, and rural population. The coefficients for $DistH$ in Model (1) for Log Farmland Value and Model (3) for Log Farm Output are not statistically significant, which suggests that there is no strong evidence to support a causal relationship between the reduction in distance to power plants and farmland values or farm output for the placebo group. This is in line with the expectations for a placebo test, as there should be no effects observed in this group.

However, the coefficients for $DistH$ in Model (2) for Log Number of Farms and Model (4) for Log of Rural Population are statistically significant at a 1% level. This indicates that the reduction in distance to the nearest power plant has a positive and significant effect on the number of farms and rural population in the placebo group. It is possible that the placebo test fails in these cases due to other confounding factors that were not accounted for in the analysis or because of the spillover effects of power plants' influence that extend beyond the 200-mile threshold.

Table 3.10 presents the results of an alternative placebo test that examines the impact of a reduction in distance to the nearest power plant on earlier farmland values, which theoretically should not be influenced by it. The coefficients for $DistH$ across all three models (base, +msa, +geo) are statistically insignificant, with p-values exceeding 0.1. This suggests that the reduction in distance to the nearest power plant has no detectable effect on earlier farmland values. These findings are consistent with the expectations of a placebo test, as no causal relationship should be observed when assessing the impact of future changes on past outcomes.

Table 3.9: Main Specification on Counties >200 miles from a Plant

	Log Farmland Value (1)	Log Num. Farms (2)	Log Farm Output (3)	Log Rural Pop. (4)
$\Delta DistH$	0.001 (0.005)	0.026^{***} (0.006)	-0.009^{*} (0.006)	0.012^{***} (0.004)
$\Delta DistH$ *** year 1920**	0.010 (0.016)	-0.022 (0.015)	0.011 (0.022)	-0.026^{*} (0.014)
$\Delta DistH$ *** year 1930**	0.008 (0.009)	-0.018^{*} (0.010)	0.040^{***} (0.013)	-0.009 (0.008)
Observations	782	782	782	780
Adjusted R^2	0.689	0.302	0.430	0.253

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 3.10: Effect of t+1 changes on t values

	Log Farmland Value		
	base	+ msa	+ geo
	(1)	(2)	(3)
$\Delta DistH$	0.013	0.010	0.015
	(0.022)	(0.023)	(0.023)
$\Delta DistH$ *** year 1920**	−0.002	0.001	−0.008
	(0.037)	(0.037)	(0.037)
Observations	284	284	284
Adjusted R^2	0.226	0.249	0.250
Note:	*p<0.1; **p<0.05; ***p<0.01		

3.7.3 Dynamic

Table 3.11 presents the results of a dynamic panel data analysis that explores the heterogeneous effects of a reduction in distance to the nearest power plant on farmland values over time. The table comprises four models (lead 0, lead 1, lead 2, and lead 3), which represent the contemporaneous effects as well as the lagged effects of the reduction in distance on farmland values at different time horizons: the same decade, one decade after, two decades after, and three decades after the reduction, respectively.

In the contemporaneous model (lead 0), the coefficient for $DistH$ is 0.030 and is statistically significant at the 1% level, indicating that a reduction in distance to the nearest power plant has a positive and significant effect on farmland values within the same decade. However, in the lead 1 model, the coefficient for $DistH$ is 0.004 and is statistically insignificant ($p > 0.1$), suggesting that there is no discernible effect on farmland values one decade after the reduction in distance. In the lead 2 model, the coefficient for $DistH$ is -0.017 and is statistically significant at the 5% level, implying that there is a negative and significant effect on farmland values two decades after the reduction in distance to the nearest power plant. Lastly, in the lead 3 model, the coefficient for $DistH$ is 0.014 and is statistically significant at the 5% level, indicating a positive and significant effect on farmland values three decades after the reduction in distance.

The coefficient for the 1920 year dummy is negative in both the lead 0 and lead 1 models but is only statistically significant in the lead 2 model (0.039 at the 5% level), suggesting that the effect of the reduction in distance between 1910 and 1920 has a positive and significant impact on farmland values two decades later. In contrast, the coefficient for the 1930 year dummy is negative in the lead 0 model and statistically insignificant in the lead 1 model, implying that the effect of the reduction in distance between 1920 and 1930 is not discernible one decade later. The dynamic panel data analysis demonstrates that the impact of a reduction in distance to the nearest power plant on farmland values exhibits temporal variation, with both positive and negative significant effects observed across different time horizons. The interaction terms reveal that the effect of the reduction in distance has evolved across the decades, with significant positive effects on farmland values observed two decades after the reduction in distance between 1910 and 1920.

Table 3.11: Dynamic Effects

	lead 0 (1)	**lead 1 (+10)** (2)	**lead 2 (+20)** (3)	**lead 3 (+30)** (4)
$\Delta DistH$ (1900-1910)	0.030^{***}	0.004	-0.017^{**}	0.014^{**}
	(0.011)	(0.007)	(0.007)	(0.007)
$\Delta DistH$ * **year 1920** (1910-1920)	-0.021	-0.027	0.039^{**}	
	(0.021)	(0.019)	(0.017)	
$\Delta DistH$ * **year 1930** (1920-1930)	-0.026	-0.005		
	(0.020)	(0.017)		
Observations	444	444	296	148
Adjusted R^2	0.654	0.602	0.179	0.189

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Summarizing the previous results, this study suggests that rural electrification in the American West between 1900 and 1930 had a limited impact on the rural economy, except for an increase in farmland values (+2.5%), and that effects varied across decades (+3% for 1900 to 1910 but not significant for the other decades). The results contrast with earlier studies on U.S. rural electrification between 1930 and 1960, potentially due to the limited adoption of electricity-dependent technologies during the study period and the ongoing agricultural depression. The increase in land and farm building values might be driven by the adoption of electric lights, improving the quality of life in rural areas.

Robustness tests present mixed results. Conley standard errors show that the positive relationship between distance to the nearest power plant and farmland values has varying statistical significance across cut-offs. The placebo outsample test yields mixed evidence, with no strong evidence supporting a causal relationship between the reduction in distance to power plants and farmland values or farm output, but a positive and significant effect on the number of farms and rural population in the placebo group. The placebo lead test aligns with expectations, showing no detectable effect of the reduction in distance to the nearest power plant on earlier farmland values. The dynamic panel data analysis reveals that the impact of the reduction in distance on farmland values varies temporally, with significant positive effects observed two decades after the reduction in distance between 1910 and 1920.

3.7.4 Incandescent lights per capita

Given the mixed results obtained in previous chapters, I reevaluated the validity of using $DistH$ as a proxy for access to electricity. In the earlier analysis, I found that the relationship between $DistH$ and various economic performance indicators, such as farmland values, were inconsistent, thus prompting further investigation.

In the absence of direct measures of households' access to electricity prior to 1930, I introduced a new dataset in Chapter 2. This dataset details the number of incandescent and arc lights per capita at both city and county levels for the years 1900, 1910, and 1920. The purpose of this chapter is to examine the relationship between the number of incandescent lights per capita and $DistH$ in greater depth. This section aims to determine if this relationship can clarify the mixed results previously observed.

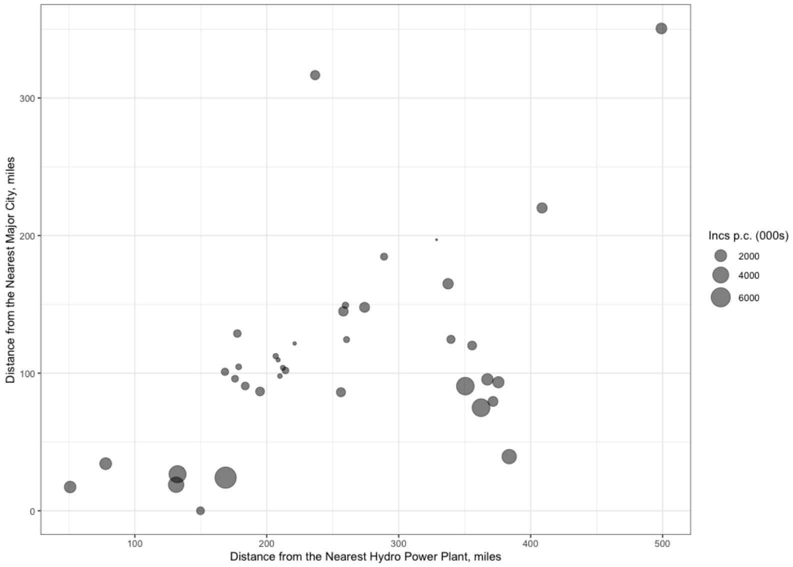
I begin by excluding counties with exceptionally high numbers of incandescent lights per capita[14], along with major metropolitan areas, as these were excluded in the main analysis. This exclusion is justified as these areas may introduce outliers.

---

[14] Five counties in California, two counties in Colorado and Jefferson County in Montana

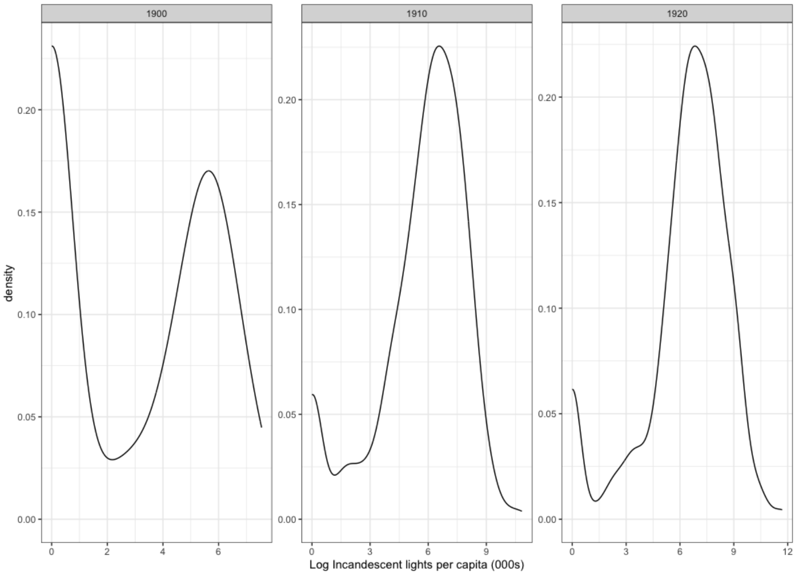
In theory, the variation in incandescent lights per capita could be explained by a county's distance from the nearest metropolitan area, where isolated power plants were located, and by the distance from the nearest hydroelectric power plant, which could predict the cost of electricity. Section 3.2.2 previously discussed the phenomenon in the Western States, especially California, where farmers could 'tap en-route'. This term refers to the practice of directly accessing power from transmission lines. As such, the likelihood of having access to electricity depended not only on proximity to a major city or a hydroelectric power plant but also on the proximity to the transmission line itself. This is reflected in the distribution of incandescent lights per capita, and in the non-linear relationship observed between incandescent lights per capita, $DistH$ and distance from the nearest metropolitan area.

Figure 3.8: Non-linear relationship between lights per capita, $DistH$ and distance from a major city ($DistMSA$), bins.



Notes: The dots are binned observations. I first categorized counties based on incandescent lights per capita, then I plotted the average distance from the nearest power plant and the average distance from the nearest large city for each bin. The size represents the number of incandescent lights per capita for each bi. *Source*: see text.

Figure 3.9: Distribution of Log Incandescent Lights per Capita by Year



Source: see text.

Figure 3.8 illustrates how the number of incandescent lights per capita varies with distance from a power plant and distance from the nearest metropolitan area. Clusters of counties with high numbers of lights per capita are typically found within 100 miles of major urban areas. However, the proximity of these counties to hydroelectric power plants varies considerably. Some lie within a 200-mile radius, while others are located at a distance exceeding 300 miles from the nearest hydroelectric power plant.

Figure 3.9 shows the shape of the distribution of lights per capita by decade. During 1900, many counties (around 25%), including those near hydroelectric power plants, didn't use incandescent lights. Despite a decrease in such counties in 1910 and 1920, some areas still lacked access to incandescent lights (around 5%).

Based on these observations, a comprehensive understanding of the relationship between per capita incandescent light usage and $DistH$ requires a two-fold approach. This involves separately examining the extensive margin - the switch from not having access to electricity and having it - for the year 1900, and the intensive margin - the increase in lights per capita - for the year 1910 and 1920.

I implement this approach by conducting several linear regressions. Firstly, an Ordinary Least Squares (OLS) regression of the logarithm of incandescent lights per capita on $DistH$ and $DistMSA$, incorporating State fixed effects and clustered errors, is run separately for all counties and specifically for those included in the main regressions (i.e., counties located less than 200 miles from a power plant) for each year. Secondly, for the year 1900 only, an OLS regression on access to incandescent lights (a dummy variable) is run with the same specifications as above, given that only a small fraction of counties lacked access in 1910 and 1920 (approximately 5%)—this represents the extensive margin. Finally, to capture the intensive margin, an OLS regression on incandescent lights per capita is run separately for all years but only for counties that had access to incandescent lights per capita in 1900 (dummy = 1), with the same specifications as above.

I will discuss the results of the analysis on a year-by-year basis. Panels 1 and 4 in Table 3.12 depict the outcomes of a simple linear regression of the logarithm of incandescent lights per capita on $DistH$ and $DistMSA$ for the year 1900, using counties located less than 200 miles from a hydroelectric power plant (counties in the sample) and the entire sample, respectively. All panels incorporate state fixed effects. The results are both positive and significant, with a 10-mile increase in $DistH$ corresponding to a 12%/7% reduction in lights per capita, indicating the vital role of proximity to power plants in determining access to electricity.

Table 3.13 displays the results for the extensive margin alone, corresponding to a linear regression of $DistH$ and $DistMSA$ on a dummy variable that signifies whether or not access to lights was available in 1900. A 10-mile increase in $DistH$ reduces the probability of having access to lights by approximately 2.6%. For the entire sample, the decrease in magnitude is smaller, standing at 1.3%. Both these results are significant and underscore the influence of geographical proximity on access to electricity.

However, when examining the results for the intensive margin, as shown in Table 3.14, Panel 1, there is no discernible relationship between the distance from a power plant and lights per capita. This suggests that the location of a county only determined the potential for access to electricity, while other factors—such as local infrastructure, economic status, and population density—likely influenced the actual adoption rate.

In the subsequent years of 1910 and 1920, the impact of $DistH$ on incandescent lights per capita is significant only when examining all counties. For those located close to a power plant, the coefficients are not significant. This observation, shown in Table 3.12, panels 2 and 4 for 1910 and panels 3 and 6 for 1920, is also reflected in the intensive margin in Table 3.14.

$DistMSA$, which denotes the distance from the nearest metropolitan area, predicts the number of lights per capita in the logarithmic model. However, it provides mixed results in the intensive margin model, and overall with a low R^2, suggesting that other factors, potentially socio-economic, also played a crucial role in the diffusion of electricity in the years 1910 and 1920.

In sum, while geographical proximity to power sources significantly impacted access to electricity, other factors determined the extent of electricity adoption. Specifically, proximity to power plants was crucial in the initial stages of electricity access, as reflected in the year 1900 data. However, as time progressed, the influence of distance on per capita incandescent light usage declined, particularly in counties closer to power plants, as evidenced by the 1910 and 1920 data. Instead, the distance from the nearest metropolitan area, and possibly other socio-economic factors, assumed a greater role in the diffusion and uptake of electricity. This suggests a complex interplay of geographical, infrastructural, and socio-economic factors in shaping the electricity adoption patterns during the early 20th century.

Table 3.12: Determinants of Log Incandescent Lights per Capita

	Counties in Sample			**All Counties**		
	1900	1910	1920	1900	1910	1920
	(1)	(2)	(3)	(4)	(5)	(6)
DistH	-0.120^{**}	-0.017	0.027	-0.071^{***}	-0.087^{***}	-0.090^{***}
	(0.048)	(0.047)	(0.055)	(0.018)	(0.020)	(0.022)
DistMSA	-0.170^{***}	-0.132^{***}	-0.148^{***}	-0.083^{***}	-0.024	-0.039^{**}
	(0.043)	(0.034)	(0.036)	(0.018)	(0.016)	(0.017)
Observations	144	144	144	405	405	405
Adjusted R^2	0.323	0.212	0.164	0.248	0.230	0.245

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

3.8 Discussion

The study's findings present a new perspective on the rural electrification narrative in the American West from 1900 to 1930. While the data does indicate that the rural counties west of the 100th parallel experienced higher transmission lines and

Table 3.13: Extensive margin: Determinants to Access to Incandescent Lights, 1900

	Counties in Sample	**All Counties**
	(1)	(2)
DistH	-0.0259^{***}	-0.0132^{***}
	(0.0083)	(0.0031)
DistMSA	-0.0261^{***}	-0.0141^{***}
	(0.0075)	(0.0031)
Observations	144	405
Adjusted R^2	0.3665	0.2749
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 3.14: Intensive margin: Determinants of Incandescent lights per Capita.

	All Counties			**Counties in Sample**	
	1900	1910	1920	1910	1920
	(1)	(2)	(3)	(4)	(5)
DistH	0.78	-165.40^{***}	-388.28^{***}	-14.09	-94.08
	(4.09)	(36.48)	(100.83)	(40.68)	(151.13)
DistMSA	-1.02	78.07^{***}	112.11	-83.37^{***}	-283.38^{***}
	(4.16)	(29.09)	(77.75)	(29.06)	(98.61)
Observations	230	376	376	140	140
Adjusted R^2	0.01	0.11	0.07	0.20	0.12
Note:					*p<0.1; **p<0.05; ***p<0.01

electricity rates, the overarching economic implications of electrification on these rural areas seem modest, except for a 2.5% increase in farmland values. On the other hand, there are no significant changes in farm output, the number of farms, or the rural population. The robustness of these findings was assessed with a series of tests, including the addition of Conley standard errors, a placebo test using out-sample data,a placebo test using lead data, and a dynamic effects analysis, which showed a more nuanced relationship between electricity access and the local economy. Moreover, I tested the strength of the variable $DistH$ as a proxy for electricity access. While this relationship was initially strong, this correlation weakened over time, hinting at other factors at play. This discussion chapter will delve deeper into these findings.

3.8.1 Noisy treatment and off-grid connections

The robustness check looking at the efficacy of $DistH$ as a proxy for electricity adoption concluded that while this variable strongly predicted electricity access, as demonstrated by access to incandescent lights in 1900, it did not provide significant explanatory power for the subsequent years. Despite incandescent lights not being a direct measure of electricity adoption, the results from the previous tests may imply that distance to the nearest hydroelectric power plant might be an excessively noisy measure for the considered period.

There is substantial evidence that this measure is in fact correlated with electricity access from recent literature (Lewis, 2018; Lewis and Severnini, 2017; Kitchens and Fishback, 2015; Kline and Moretti, 2014; Gaggl et al., 2021) but all of these studies consider an extended period of time, up to 1960. Only one study (Gaggl et al., 2021) considers the period between 1910 and 1940, but the outcomes were only measured in 1940. Although indirectly, this could suggest that before 1930 being located near a transmission line that was not synonym with higher access to electricity. The higher electrification rates I presented in the historical background might be driven by other causes.

According to Nye (1997), between 1910-1930, rural customers in the Western US were aware of the possibilities they had in terms of adopting new technologies but still could not cover the costs of attaching to the main transmission line systems. As an alternative, they installed their own generating systems, a type of electrification that the current development literature would call "off-grid" (Lee et al., 2020a). Around 1910, there were several alternatives for the farmer in terms of generating energy. The cheapest option was to use a small generator to convert water energy into mechanical energy, also called a "water motor". A similar solution, also requiring water pressure, was to use a water dynamo. Farmers who had access to a stream could build their own small dams. There were then solutions that did not involve the use of water. The most flexible solution was to use a gasoline or kerosene-driven system, the "Delco-Light", originally produced in Dayton, Ohio. In 1910, the engine was only run to charge batteries (Nye, 1997), but in 1920, Delco-Light started selling "Farm-lighting sets" (Electrical World, Volume 75, 1920), clearly addressing what was driving electricity demand in the rural counties of the U.S. Perhaps surprisingly, windmills were an even more common solution. This could explain the non-linear relationship between proximity to a power plant and access to electricity: farmers showing up as having access to electric lights in 1900-1930 could have been located anywhere as long as they had their own generators. On a different note, off-grid connections had a much lower wattage than a connection to the grid, making it unsuitable to power machinery and limiting the effect of electrification, a parallel situation to developing countries today. For example, in a paper by Lee et al. (2020b) the author highlights that in Kenya, off-grid home solar systems do not satisfy household energy needs as appliances need to have a higher wattage.

3.8.2 Heterogeneous treatment effects and complementary technology

Secondly, electricity access, or the potential access to electricity in rural areas, might have taken a considerable amount of time to have a visible effect on the rural economy. This delay might be due to the complementary nature of other technologies, such as household devices or machinery used in agricultural and rural production. Although there is a lack of granular data to investigate this analysis for the U.S. in the early 1900s, we can make inferences by examining the most recent evidence about rural electrification in developing countries.

In a recent study by Lee et al. (2020a), the authors find that mere access to electricity is not sufficient to bring about significant economic outcomes. By reviewing the most important results of modern literature, they compared earlier studies of electrification impacts in high and middle-income countries like South Africa and Brazil and later results from electrification in contemporary low-income Kenya and India. In their paper, they argued that the fact that recent electrification programs have ambiguous results is driven by the fact that households cannot afford complementary inputs to electrification (cit.). This result is also supported by their most recent experimental paper, which shows that it is the wealthier adopters who derive benefits from having access to electricity.

Applying these findings to our study, it could be inferred that the lack of significant effects might be due to rural households being, on average, either too poor to afford the complementary technology or that the technology was not sufficiently advanced to justify the purchase. This perspective offers a possible explanation for the discordance in the results between the rise in farmland values and the lack of significant changes in other economic indicators in the period under study.

3.9 Conclusion and Further Research

Looking again at the 1923 map of long-distance transmission lines in the Western U.S. (refer to Figure 3.3), it is apparent that there was considerable anticipation surrounding the distribution of electricity over vast distances. Engineering journals optimistically touted investments in new hydroelectric projects, and companies extended their electric lines to reach rural agricultural communities. Despite these advances, early rural electrification in the Western States from 1900 to 1930 had only a modest impact on the economy. This paper suggests two potential explanations: the presence of a noisy treatment and the absence of complementary technologies.

A closer examination of the data reveals that the most significant effects of a reduction in distance to the nearest power plant occurred between 1900 and 1910, potentially driven by the adoption of electric lights. However, the substantial acceleration in electrification rates mostly took place between 1920 and 1930, as presented in table 3.1. This seemingly contradictory finding might be explained by hypothesizing that the impact on farmland values was indirect, driven by the adoption of electric lights at the community level rather than direct effects at the household level, hinting at possible economic spillover effects (as suggested by Lipscomb et al. (2013)). This idea resonates with the findings of Burlig and Preonas (2016), who demonstrated that while the rural electrification program in India did boost electricity availability and consumption, it did not significantly affect the economy.

Moreover, this study illustrates that the process of rural electrification in the U.S. was more intricate than depicted by studies focusing on the successful period between 1930 and 1960. It suggests that the success of those “Big Push” policies(Murphy et al., 1989; Rosenstein-Rodan, 1961) might have been facilitated by specific program features: the program was structured to assist farmers via loans to agricultural cooperatives, thereby relieving individual farmers of the burden to connect to the grid. It was launched in conjunction with other New Deal programs and supplemented with awareness and educational programs for farmers - supports that were absent during the early 1900s. The absence of these supports is reflected in the non-significant effects on farm output and other extensive and intensive production measures.

This understanding of the early phase of rural electrification provides a new context for interpreting the economic impacts of electrification policies during this period. Furthermore, these findings draw compelling parallels with contemporary experiences of electrification. Much like the early 20th century U.S., many countries are still grappling with the economic implications of expanding electrical access to rural areas. This study, therefore, offers valuable insights that can help inform and shape electrification strategies in these regions, emphasizing the importance of complementary policies and supports to maximize the economic benefits of electrification.

This study suggests two compelling directions for future research in the field of rural electrification. Firstly, an in-depth investigation into the adoption and role of off-grid technologies, such as water motors and Delco-Light systems, in the broader electrification process could provide deeper insights into the electrification narrative.

Such an exploration could highlight the ways in which these alternative solutions bridged the gap in rural electrification during periods of limited grid connectivity. An analysis of the variables that drove the uptake of these technologies, their efficacy in fulfilling energy requirements, and the catalysts behind their ultimate replacement with grid electricity could also inform current policy and strategy in regions where off-grid solutions are being leveraged to extend electricity access. As of now, the adoption of off-grid solutions during the early electrification period remains largely unexplored, save for the historical accounts provided by Hirsh (2018). Given the rich potential of this topic, an economic analysis would complement these historical narratives.

Secondly, this study's findings suggest the need for a broader examination of the social impacts of rural electrification in the early 1900s. While the research demonstrates a significant correlation between electrification and increased farmland values, the other economic effects are less defined. However, the real value of electrification may have unfolded in more nuanced, societal ways, such as enhancements in education, health outcomes, and shifts in social structures and gender roles, potentially driven by the time saved through electricity reducing manual tasks. A dedicated study quantifying these social impacts could provide a better understanding of the full spectrum of benefits derived from early electrification. Such a study could also offer insights for modern electrification initiatives in developing nations, highlighting the broad social advantages that can accompany these efforts. An example of this kind of work is the study by Lewis (2018), who looks at the impact of 1930-1960 rural electrification on infant health and women's fertility. A more recent example is the research on women's participation to the workforce by Vidart (2020). New research should use similar methods but focus on the early years of adoption.

Chapter 4

The Effects of Weather Shocks on Labour Markets: the Corn Belt in the 1930s

4.1 Introduction

Climate change is predicted to increase the frequency and intensity of extreme weather events during this century(IPCC, 2021) and a growing body of literature shows that weather fluctuations have a negative impact on agricultural productivity and crop yields (Schlenker and Roberts, 2009; Deschênes and Greenstone, 2007). Additionally, the latest projections using a new generation of crop and climate models show even more pessimistic results: mean end-of-century corn productivity is shifted from a range of +5% - +1% to a range of -6% - -24% (Jagermeyr et al., 2021). Given these potentially severe impacts, it is essential to learn about the way in which society might adapt. Because of the rarity of these events, it is not easy to find case studies for which the given weather extreme has affected a large area and for which data are available for a large number of people and for an extended period of time.

This is where historical insight might become essential. Between 1930 and 1939, the U.S. experienced the worst series of droughts in U.S. history (Seager et al., 2008; Cook et al., 2014). The Dust Bowl - which refers to areas eroded by topsoil blown off overplowed farmland - was part of that environmental shock, but it affected only a small number of counties in the Plains area. The droughts impacted a much larger area, extending from the Plains to the farmlands in Minnesota and Montana (see Figure 4.2). Although the Dust Bowl erosion was exceptional, the droughts were also an exogenous weather shock which caused substantial losses to crops (Sutch, 2011), and caused people to migrate (Sichko, 2020; Gutmann et al., 2016). This paper aims to study the impact of the 1930s droughts on the labour market, with the purpose of learning about mid to long-term adaptation strategies of individuals hit by weather shocks.

Studies have highlighted different possible responses to climate change extremes: farmers can change consumption and saving patterns as 'temporary responses' (Gbetibouo et al., 2010; Di Falco et al., 2011; Hisali et al., 2011); they can also change input use and agricultural practices (e.g. an increase in fertilizer use, change in crop) (Aragón et al., 2021; Sutch, 2011; Griliches, 1957); finally, additional studies looking at the changes in trade and migration patterns (Feng et al., 2015; Sichko, 2020). However, to this date, there is less evidence about another potential impact of weather extremes: sectoral reallocation. The agricultural risk literature shows that farming households have numerous *ex-post* and *ex-ante* adaptation strategies to cope with weather shocks (Rose, 2001) and the movement of labour between sectors is one of those strategies. Assuming fixed prices and free movement of labour between sectors and regions, a reduction in yields reduces agricultural wages, causing movement between regions - migration - and movement between sectors - labour reallocation. While we have substantial evidence about migration (Branco and Féres, 2021), there is less evidence about sectoral reallocation (Jessoe et al., 2018; Branco and Féres, 2021; Hill et al., 2021). This is especially true when looking at the direct effect of weather on labour supply (Hill et al., 2021).

In this paper, I focus on the Corn Belt area of the U.S., which I define as the following states of the Midwest: Indiana, Illinois, Iowa, Missouri, Nebraska, Kansas and South Dakota (see Figure 4.2) One of the reasons why I chose this area is the great extent to which it was affected by droughts. For example, after the 1934 and the 1936 drought, in Kansas, only 20% of the planted acreage was harvested. In Illinois, 80% of the planted land produced useful crops. This makes it ideal for testing variation in the impacts of a weather shock. Secondly, the Corn Belt, as the name suggests, was an almost single-crop economy: crop production centred on corn and livestock, which in turn depended on lower-quality corn for feeding. Additionally, the single-crop economy allows me to focus on the specific range of temperatures which badly affected corn. Thirdly, data for this period and this area are relatively good. Data about employment status come from the full linked census (Ruggles et al., 2020): while only 6% of the full census can be linked, 30% of the male population in those states can be linked from one year to the other - that is, approximately 2.1 million individuals. Data about crop yields are available yearly between 1925 and 1950 at the county level, and there is a homogeneous amount of weather stations from which to extract temperature and precipitation data points.

My approach uses weather variation at the county level in the form of 'Harmful Degree Days', and measures its impacts on labour markets by looking at its effects on employment and wages. Like Feng et al. (2015) I first measure the impact of the droughts on corn yields between 1930 and 1940 using county-level data. Then, I use full census-linked data from 1930 and 1940 to measure the aggregate impact of the 1930s drought on the probability of being employed in 1940 and for wages. I estimate results separated for individuals living on a farm and not, and for individuals who migrated - to another state - and not.

The rest of this chapter is organized as follows. In section 4.2, I look at the contribution of previous literature, as well as the general historical context. In section 4.3 I review the theory about shocks to agricultural production and use it to draw a series of testable hypotheses. In section 4.4 I go over the data creation process. The empirical strategy is in section 4.5 and the analysis is in section 4.6. I conclude by summarizing the results, addressing limitations and proposing additional steps forwards in 4.7 and 4.8.

4.2 Literature Review

This paper builds on two strands of literature: studies from development economics which look at the relationship between productivity shocks and local labour markets, and studies that look at the specific case of the 1930 U.S. and the impact of droughts on the local economy.

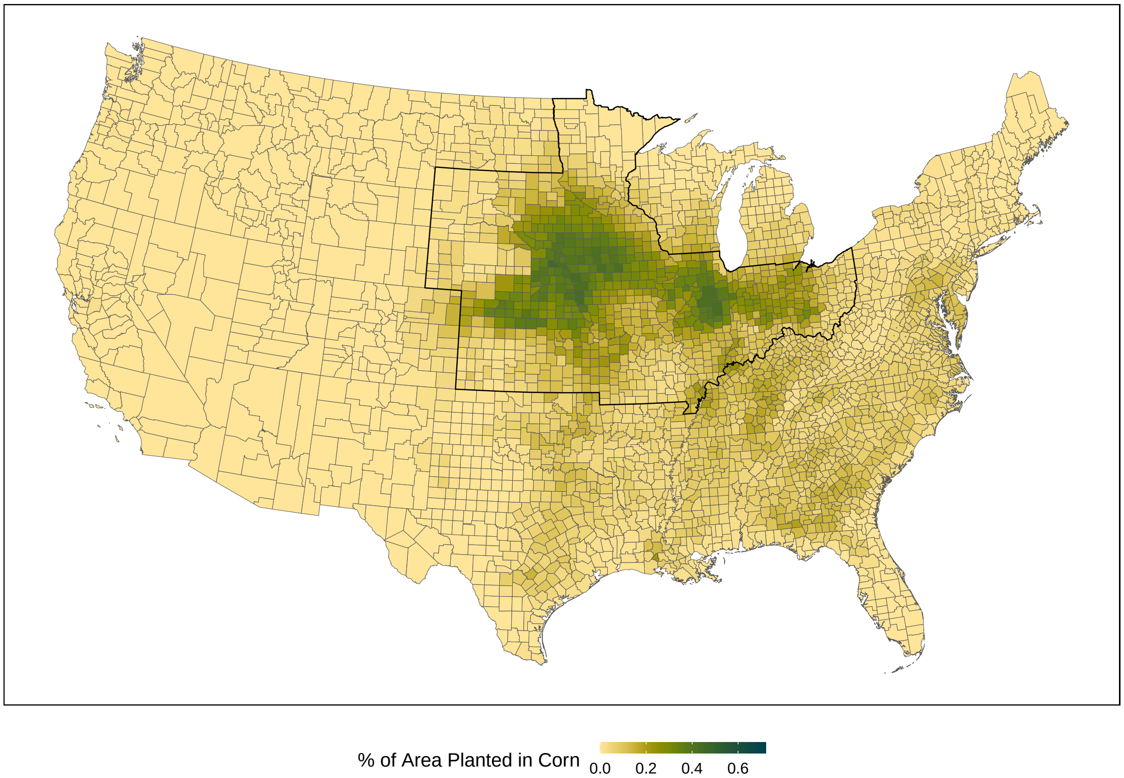
Figure 4.1: North Dakota Farmer and Stunted Wheat

Source: Picture by A. Rothstein, 1936 available for free at archives.gov

4.2.1 Rainfall as an instrument and other Weather Shocks

Initially, literature in development economics used weather shocks as an instrument for agricultural income or economic growth. Miguel et al. (2004) and Jayachandran (2006) use rainfall variation in a cross-country panel and cross-district panel respectively. In these studies, rainfall is a strong predictor of yield and in the case of countries with a broad agricultural sector, precipitation levels are also correlated with economic growth.

Figure 4.2: Area(%) Planted in Corn, 1930.



Notes: 0.1 means 10% *Source*: See text.

More recent studies have looked at the impact of weather shocks on labour directly. Bastos et al. (2013) uses employment data from local areas in Brazil and drought indices built from rainfall data points. They find that a higher frequency of droughts during the previous decade reduces local value-added, employment and adjusted wages in the agricultural sector. District and individual-level data from India are used in Emerick (2018). In this study, the author uses deviation in precipitation as an instrument for agricultural productivity and finds that exogenous increases in agricultural output cause modest increases in the non-agricultural labour share in rural districts. In Jessoe et al. (2018), the authors use weather shocks - expressed as harmful and growing degree days and precipitation - to look directly at how this negatively impacts individual labour opportunities in Mexico.

4.2.2 The case of 1930s U.S. Droughts

Nowadays, weather shocks cause substantial stress to developing economies which heavily rely on the agricultural sector. In the 1930s, approximately 21% of the U.S. population was still working in agriculture. Although markets were developed and the typical farmer heavily relied on commercialization and integration with the rest of the economy, the typical farmer did not have the same mechanisms - policy and insurance - to cope with extreme risk. In the early 1930s, large areas of the U.S. were afflicted by intense droughts and many farmers struggled to continue their businesses. Among the most affected were the Great Plains, which were also the scenario for the worst drought in U.S. history in 1934 (Cook et al., 2014). Already the year after, in 1935, agricultural economists reported how this was the worst drought recorded in the country. It affected over 75% of the area (27 States) by reducing the yields of food grains and cotton and forcing a reduction in livestock numbers. The 1935 'Yearbook of Agriculture' reported that local supplies of certain food products were short in many affected areas and that farmers suffered a decline in income (Eisenhower, 1935). Some areas in the Great Plains region were also affected by the Dust Bowl, a period of erosion due to strong winds blowing off the topsoil and moving it across the area. The causes of the drought were the anomalous temperatures and the low precipitation, but overworking the soil exacerbated the severity of the erosion (Schubert et al., 2004).

Homesteaders were originally attracted to the Great Plains region because of unusually high precipitation levels in the early 20th century. Additionally, World War I generated high demand for agricultural products, raising prices and creating a further incentive for settlement in the area. Removing the grassland and planting crops created dry topsoil which was then lifted off by the strong winds, causing dangerous dust storms (Hornbeck, 2012). In the same period, the U.S. was going through the Great Depression, which spurred a new period of economic interventions and stimulus programs. For the farmers affected by droughts condition, the Agricultural Adjustment Act(AAA) was supposed to alleviate some of the adjustment costs. However, AAA payments were given to farm owners and large farms, as opposed to sharecroppers and tenants (Fishback et al., 2005).

There is a growing number of studies that looks at weather shocks' effects on the local economy in the 1930s. Hornbeck (2012) and Hornbeck (2020) looked at the impacts of the Dust Bowl on county-level outcomes. Both studies concluded that the Dust Bowl reduced land values and that the local economy mainly recovered through migration (as opposed to sectoral reallocation). Long and Siu (2008) explores the differences between Dust-Bowl and Non-Dust-Bowl migrants. They find positive labour market effects for migrants (such as an increase in wages) but that at the same time these migrants were less likely to move, possibly due to income constraints. The study most similar to this paper is by Sichko (2020). In his research, the author looks at the effect of drought - measured through the Palmer Drought Index - and how this interacted with migration. He finds that there was positive selection into migration, confirming the income constraint and that people moved from a rural to another rural location.

While the development economics literature studies the relationship between weather shocks and labour markets at the regional or district level, the historical literature uses individual-level data, but mostly to observe effects on migration. This paper attempts to combine the insights from the development economics literature with the case of the 1930s U.S. droughts to study the effect that weather shocks had on employment and wages at the individual level. Formal hypotheses to guide the empirical research are outlined in the next section.

4.3 Theoretical Background

4.3.1 Direct and indirect effects of a shock on agricultural production

Previous empirical literature considers the weather shocks experienced in the U.S. between 1930 and 1940 in a short-term scenario, especially when looking at the episode of the Dust Bowl. In the simple production function model by Hornbeck (2012), a farmer can choose between technologies to gain rent from its land assets. In the short term, she allocates land to be used with these specific technologies. For instance, she may allocate half of her land to corn, and the other half to hay. When an external shock hits the farmer's production, it might decrease the profitability of one technology, for instance, corn production. In the short-run, it would be hard for the farmer to move its land assets from corn to hay, so production experiences losses, because the land is 'constrained' to its original allocation. In the long run, the farmer can slowly convert to a different land allocation, using a different combination of technology. In this case, she might move two-thirds of production to hay. This adjustment might take time to be implemented because adjustment costs are high initially and decline over time due to learning by doing and/or technological adoption (Griliches, 1957; Foster and Rosenzweig, 1995; Sutch, 2011). To summarize, in the short-term an exogenous shock causes a reduction in agricultural revenues, a reduction in land values, and a reduction in demand for land. In the long-term, the farmer recoups agricultural revenues proportionally to how much adjustment in land allocation she was able to do.

While the production function approach only looks at the effects of a given shock within a sector, additional general equilibrium effects can be easily summarized using a spatial equilibrium approach (Roback, 1982; Rosen, 1974). Within this framework, one can observe how a local negative shock affects different regions: it will reduce agricultural productivity in the county where it happened (county A) which in turn will reduce land values and output, as well as lead to an increase in prices. Open markets will increase prices in the other area (county B) as well, and higher demand for land in the non-affected area will raise land values. Additionally, a decrease in agricultural labour productivity in county A will cause a decline in agricultural wages, inducing a movement to county B - migration - or a reallocation to the non-agricultural sector.

4.3.2 Effects of a shock on farming households

The agricultural household model described by Singh et al. (1986a) serves as a key reference for studying changes in agricultural labor supply. This framework and its subsequent adaptations have one distinguishing characteristic: households are simultaneously consumers and producers: they produce partly for sale and partly for their own consumption, and they both provide and hire labour inputs. Theoretically, this means that labour supply is endogenously determined by the production function via the income constraint: households adjust the amount of time they allocate working on the farm also based on how farm profits are going [1].

Formally, labour supply is defined as [2]:

$$F^{*}\left(w, p_a, p_{na}, \mathbf{A}, \mathbf{Z}, T, \bar{y}\right) \tag{4.1}$$

which is determined by:

- w: wages.
- p: prices of agricultural and non-agricultural goods.
- $\mathbf{A}$: fixed inputs such as capital and land.
- $\mathbf{Z}$: household characteristics
- T: Time allocation
- $\bar{y}$: additional exogenous income

In which prices are exogenous, but inputs can vary spatially and temporally, affecting the final labour supply. This simple framework has been used to model the behaviour of farming households in rural regions of the contemporary developing world, but it can also describe how farmers participated in the market in the U.S. in the 1930s. In 1935-36, approximately 8.9% of farm families were receiving relief; the rest (non-relief families) had an income that ranged between \$500 and more than \$10,000, with a bit more than one-half falling within the \$500-\$1500 bracket. In the intermediate income

---

[1] Note though that most models assume separability/recursiveness: the household can make production decisions independently of consumption and labour-supply decisions. However, the opposite does not hold, because both depend on income which is determined by the farms' profits Singh et al. (1986b).

[2] I follow the notation of Hill et al. (2021).

group, between \$1000 and \$1250, farming families earned 55% in cash and the rest in kind, including food and wood fuel produced on the farm itself [3]. The author specifies that on average farm families produced two-thirds of their food supplies, but they still needed to purchase the rest from the market, highlighting the mixed producer-consumer role of the households (Monroe, 1940). Additional research by Leonard et al. (2011) also confirms that household characteristics, such as the age of the household head, have an effect on farm operations, confirming the role of household characteristics ($\mathbf{Z}$) in determining labour supply.

In the Singh et al. (1986b) model, farm production is an input in the farming household consumption; that is, households are partially self-sufficient. Having the possibility to rely on income sources outside of the market is one of the ways in which farmers insure against risk, including the effects of bad weather spells. Following the model of Strauss (1986) as cited in Singh et al. (1986a), Rose (2001) develops a simple agricultural household model when weather risk enters the labour supply function. In her model, supplying labour to the market - outside of the agricultural sector - is both an ex-ante and ex-post strategy to cope with uncertain weather. A similar model has been developed by Jessoe et al. (2018) after the work of Ravallion (1988). In this model, a negative weather shock decreases agricultural labour demand, reducing demand for hired labour. Additionally, a contraction in demand for non-agricultural goods also decreases the demand for non-agricultural labour. Finally, a negative weather shock increases labour migration.

In light of the theoretical insights from the previous literature, the historical context, and the introduction of weather risk in the household production function, I put forth

---

[3] In Monroe (1940), the author writes: “All families of farm operators are entrepreneurs; the farm enterprise is a family undertaking, the concern of all members old enough to work.”

a set of hypotheses to guide my investigation:

1. First, a negative weather shock, such as severe drought or storms, would likely decrease agricultural productivity - corn yields in this case. In response, farmers might attempt to adapt through optimizing input allocation, such as adjusting the use of labor, land, and capital resources. The demand for agricultural labor in the impacted county would decrease, reducing the number of laborers hired on farms. This is a logical conclusion derived from the production function model.

2. Second, the weather shock-induced reduction in agricultural production would also lower the farmer family's income. This income loss, in turn, would lead to a decrease in the consumption of non-agricultural goods, consequently reducing the demand for non-agricultural labor in the affected area. In the short-term, this might result in a decrease in employment in these areas. This hypothesis aligns with the insights from both the producer-consumer household model and the Roback spatial equilibrium model.

3. Finally, According to both the Roback spatial equilibrium model and the producer-consumer household model, such a weather shock is also expected to spur an increase in migration. This could occur due to a decrease in local wages following the shock and the availability of higher wages in neighboring locations. Alternatively, a decrease in the demand for hired labor within farming households could result in a surplus of labor, also driving migration. This hypothesis has found support in numerous studies, particularly for counties in the Great Plains that suffered significantly from both drought and erosion (Long and Siu, 2008; Hornbeck, 2020). In this study, I aim to test this hy-

pothesis in areas that experienced less extreme, yet still significant, weather shocks.

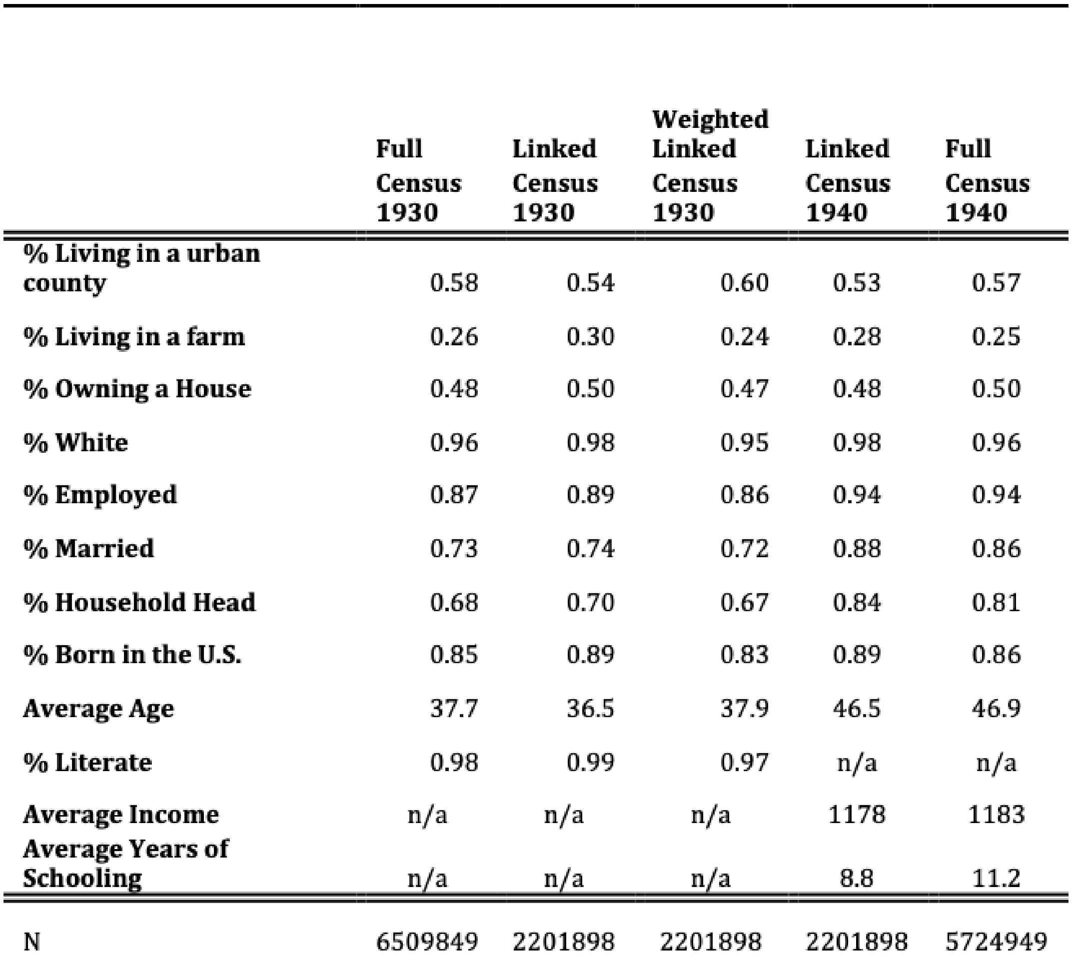
4.4 Data and Summary Statistics

Data for this paper have been collected from different sources. Data at the individual level are from the 1940 and 1930 censuses. Data at the county level are collapsed information from the censuses, and county-level data from IPUMS USA (Ruggles et al., 2021). In the following sections, I will go into more detail about the specific data used and present some summary statistics. I will also test the relationship between my weather risk variables and measures of agricultural productivity.

4.4.1 Linked Census Data

Employment and demographics data at the individual level come from the full census for 1930 and 1940. In order to have data about the transition of people from one occupational sector to the other, I needed to observe the same individual at two different points in time. To do so, I linked individuals using a fully anonymous crosswalk between individual IDs created by the Census Linking Project (Abramitzky et al., 2019). To focus on working-age men, I restricted my sample to men between 20 and 60 years old. I opted for a high minimum age to exclude men from having the option to attend school. Second, I kept only men living in 8 Corn Belt States in 1930 (Indiana, Illinois, Iowa, Minnesota, Missouri, Nebraska, Kansas and South Dakota), which reduced my sample to 6.5 million (from a total of 65 million men living in the U.S.). Out of the total 18 million links available in the Census Linking Project (between 1930 and 1940), 2.2 million are in our dataset, approximately 33% of the total population of the Corn States. Unavoidably, sampling data creates selection biases. In my 1930 linked census, for instance, there are 30% of people living on a farm, while in the full census only 26% reside in one (see Table 4.1). Similarly, there are some discrepancies in the percentage of people who are household heads (68% vs 70%) and how many are employed (87% vs 89%).

Table 4.1: Employment data, Linked vs Full census



	Full Census 1930	**Linked Census 1930**	**Weighted Linked Census 1930**	**Linked Census 1940**	**Full Census 1940**
% Living in a urban county	0.58	0.54	0.60	0.53	0.57
% Living in a farm	0.26	0.30	0.24	0.28	0.25
% Owning a House	0.48	0.50	0.47	0.48	0.50
% White	0.96	0.98	0.95	0.98	0.96
% Employed	0.87	0.89	0.86	0.94	0.94
% Married	0.73	0.74	0.72	0.88	0.86
% Household Head	0.68	0.70	0.67	0.84	0.81
% Born in the U.S.	0.85	0.89	0.83	0.89	0.86
Average Age	37.7	36.5	37.9	46.5	46.9
% Literate	0.98	0.99	0.97	n/a	n/a
Average Income	n/a	n/a	n/a	1178	1183
Average Years of Schooling	n/a	n/a	n/a	8.8	11.2
N	6509849	2201898	2201898	2201898	5724949

Source: Ruggles et al. (2021) and authors calculations.

To adjust for these discrepancies, I run a probit model (Abramitzky et al., 2019; Sichko, 2020) to calculate the probability of being linked based on a series of variables I am interested in: age, living in an urban centre, living on a farm, owning a house, race, employment status, married or not, nativity and whether literate or not, all calculated in 1930. Probabilities are then used in inverse probability weights to adjust for the selection biases. Finally, in order to focus on individuals who were not still attending school, and therefore were not yet in the labour force full time, we exclude them (-2.6%). Summary statistics are available in in table 4.2 below. Note that the statistics are for the non-adjusted dataset; weights will be used in the estimation.

Table 4.2: Summary Statistics for 1930 and 1940 data, Linked Census.

Baseline characteristics in 1930		
	mean	sd
Age	36.8	10.9
White (%)	97.6	15.2
Married (%)	75.5	43
Head (%)	71.4	45.2
Born abroad (%)	11	31.3
Urban (%)	53.2	49.9
Living in a Farm (%)	29.9	45.8
Owner (%)	49.3	50
Employed in Agriculture (%)	28.2	45
Wage-earner (%)	55.4	49.7
Unemployed (%)	6.88	25.3
Occscore (median total income in hundreds of 1950$)	19.14	13.3
Outcomes in 1940		
	mean	sd
Employed in Agriculture (%)	24.7	43.1
Employed in Non-Agriculture (%)	62	48.5
Wage-earner (%)	58.6	49.3
Unemployed (%)	5.18	22.2
Not in Labour Force (%)	9.07	28.7
Farmer, worse off (%)	1.9	13.6
Migrant, State (%)	18.6	38.9
Migrant, County (%)	13.8	34.5
Wage ($)	1171	1480
Log Wage	3.16	5.57
Occscore (median total income in hundreds of 1950$)	22.8	12.5

Table 4.2 summarizes baseline demographic and economic characteristics for the linked census data in 1930 and outcomes for that same sample in 1940. In 1930, the average age was approximately 36.8 years with a standard deviation of 10.9 years. The population was predominantly White (97.6%), and most were married (75.5%)

and identified as the head of their household (71.4%). 11% were born abroad, slightly over half lived in urban areas (53.2%), and about 30% lived on a farm. Home ownership was split fairly evenly with 49.3% owning their property, and 28.2% were employed in agriculture. 55.4% were working for wages, with a small percentage of unemployment at 6.88%. The median total income in hundreds of 1950 dollars (Occscore) was 19.14. In 1940, the number of people employed in agriculture decreased to 24.7%, while those in non-agriculture employment increased to 62%. Wage earners slightly increased to 58.6%, and unemployment decreased to 5.18%. Notably, 9.07% were not part of the labour force. A small percentage (1.9%) of farmers were worse off, while a significant percentage migrated at the state (18.6%) and county (13.8%) levels. The average wage increased to $1171, but the log wage value suggests significant wage disparity. The median total income in hundreds of 1950 dollars increased to 22.8.

4.4.2 Weather Data

In order to create a variable which summarizes the weather shocks experienced by counties in the Corn Belt area, I had to aggregate daily-level data to a decade-level measure. In this paragraph, I will discuss the data sources as well as the estimation process. Next, I will also look at the relationship between the weather risk variable and measures of agricultural productivity.

Estimation of county-level weather data

Daily weather data are from approximately one thousand stations' observations located in the contiguous United States between 1930 and 1940. The data were collected by the Global Historical Climate Network Daily (GHCN), an integrated database of daily climate summaries from land surface stations across the globe (NOAA, 2021) [4]. In terms of data quality, between 1930 and 1940, there are in average 356 observations per year per each weather station, making it a relatively good data set in terms of variation over time, especially comparing it with the previous years (see Figure 4.3 below). Therefore, given that there are no substantial missing observations from weather stations, I do not need to interpolate the values for the missing dates (a step which has been suggested by the literature (Auffhammer et al., 2013).

In terms of variation over space, the density of stations varies across the contiguous U.S. but there is a high density of stations in the Corn Belt area and numbers do not vary substantially over the years. In the map presented in Figure 4.4 the core Corn Belt area is highlighted in green and the Corn Belt states (the sample area) in orange. There are 255 stations within the boundaries of the Corn Belt states, more than 25% of the total.

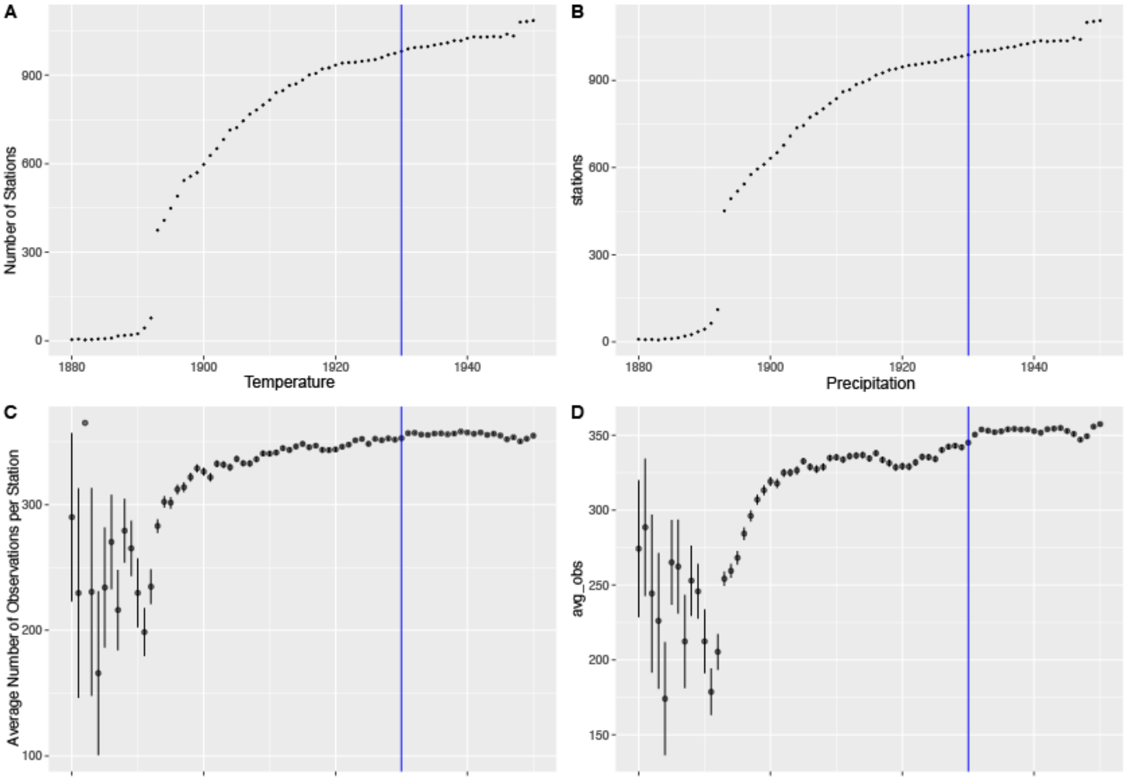
The collected data include daily maximum and minimum temperatures and total precipitation. To create a daily observation for a weather variable at the county level, I utilised a spatial averaging technique called kriging [5]. Kriging, much like the more commonly used inverse distance weighting (IDW), uses observations at proximate points to estimate values at unknown locations. In IDW, weights assigned to each point are determined by an inverse distance matrix, whereas in kriging, weights are derived from a variogram, a model that quantifies the spatial autocorrelation between observations. The specific variant of kriging I employed, universal krig-

---

[4] Data are published by the US National Oceanic and Atmospheric Administration (NOAA) and accessible online at this link

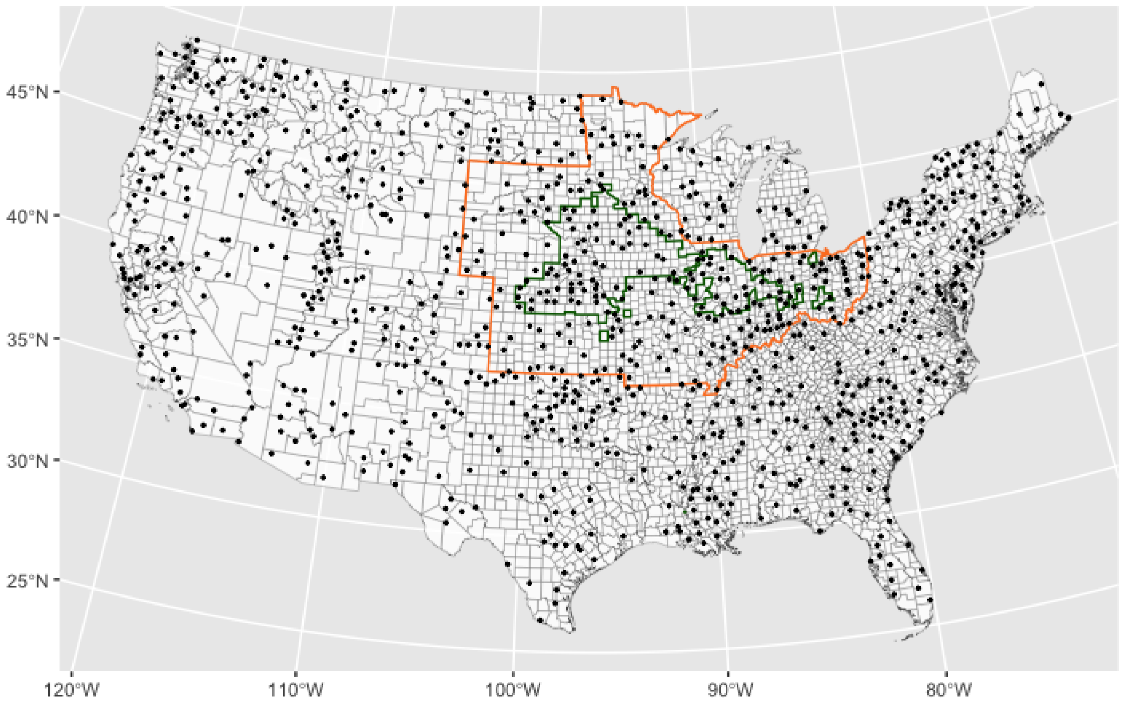
[5] Specifically, I employed a universal kriging algorithm using the gstat and automap packages in R.

Figure 4.3: Variation in observations per station over time.



Source: NOAA (2021), author's calculations.

Figure 4.4: Variation in observations per station over space.



Notes: The area in green is the "core" Corn Belt area. The are in orange are the Corn Belt states (states with at least one county in the Corn Belt Area) *Source*: NOAA (2021); Ruggles et al. (2021).

ing, accounts for east-west and north-south temperature trends in the estimation of weights. Universal kriging was applied for each day and each month of the year, with final values estimated at county centroids [6]. This kriging process was conducted under the assumption that the weather patterns exhibit spatial autocorrelation and stationarity. Spatial autocorrelation refers to the phenomenon where geographically closer values are more likely to be similar than those further apart. Stationarity is the assumption that these spatial processes do not change over time.

One should note that this method has its limitations, which include potential bias if these assumptions are not strictly met in the data. For instance, if significant changes in the weather patterns occur over time (non-stationarity), the predictions made from earlier observations might not accurately reflect the current state. Another limitation is that kriging might not work as well with sparse or irregularly distributed data, because it relies on having nearby points to generate predictions.

Growing Degree Days

While the weather estimates are at the day level, employment data are at the decade level. To link the two datasets, I need to prepare weather variables that summarize variation over each year and then over a decade. In the theory section, I explained how the relationship I want to look at is the one between weather and employment, specifically through the channel of agricultural productivity and wages. That means I need to use a weather variable which has its main impact on agricultural production and as a consequence, agricultural wages. Given that agricultural production in the Corn Belt area is the homonymous crop, I chose to build my main explanatory

---

[6] This methodology follows closely to that used by Bleakley and Hong (2017), while Jessoe et al. (2018) used IDW.

variable based on the variable that better explains variation in corn yields: growing degree days above 29 degrees Celsius, which I call Harmful Degree Days (HDD).

According to the most recent results about the effect of weather on corn yields, temperature is the best predictor for corn yields in the contiguous U.S. (Schlenker and Roberts, 2009; Auffhammer et al., 2013; Ortiz-Bobea, 2021). In specific, the aforementioned and additional agronomic studies found that there is a nonlinear relationship between weather and yields: heat has a beneficial effect up to a threshold, after which the relationship is inverted and higher temperatures are detrimental to crop growth. Technically speaking, when air temperature is above the high threshold or below a crop-specific base temperature, crop development stops because plants cannot thermoregulate. To measure this concept of heat accumulation, agronomists and farmers use Growing Degree days(GDD), which are expressed as:

$$T\left(\bar{h}, \underline{h}, t_0, t_1\right) = \int_{t_0}^{t_1} GDD(t)dt \quad \text{where } GDD(t) = \begin{cases} \bar{h} - \underline{h} & \text{if } h(t) > \bar{h} \\ h(t) - \underline{h} & \text{if } h(t) \in]\underline{h}; \bar{h}] \\ 0 & \text{if } h(t) \leqslant \underline{h} \end{cases} \tag{4.2}$$

In the case of corn, the lower threshold $\underline{h}$ is 10°C, while the higher threshold $\bar{h}$ is 29°C. Due to computational constraints, I employed a simplification of the GDD calculation for this paper, which is frequently used in the existing literature (Jessoe et al., 2018; Roberts and Schlenker, 2011). I calculated GDD piecewise using daily mean temperatures, derived by averaging the maximum and minimum temperatures provided in the raw data. The simplified equations for GDD is:

$$\text{GDD}(T) = \min\left[\frac{(T_{\max} - T_{\min})}{2} - 10, 19\right] \text{ if } 10\text{C} < T \leq 29C \tag{4.3}$$

To capture the effect of harmful temperatures above 29°C and therefore account for the nonlinearity, I also calculate an additional measure, harmful degree days (HDD) as follows:

$$\text{HDD}(T) = \frac{(T_{\max} - T_{\min})}{2} - 29 \text{ if } T \geq 29C \tag{4.4}$$

The crop's heat absorption only happens during a specific growing season, from planting to harvesting. Therefore, to calculate the yearly GDDs and HDDs, I sum over the growing season, which is assumed to be between March 1st and August 31st. Additionally, I calculate yearly precipitation levels by summing over all days of the growing season.

The GDDs/HDDs-Corn Yield relationship

Given that employment data are at the decade level, and I am working with a cross-section, I needed to find a way to summarize the information within the GDDs/HDDs and corn yield relationship into one time-invariant variable. To do so, I first explored whether there is evidence of a nonlinear relationship between GDDs/HDDs and corn yields between 1930 and 1940 in the Corn Belt area.

In the original work by Schlenker and Roberts (2009), the authors look at the relationship between the yields of three major U.S. crops (corn, soy and wheat) in the rain-fed agricultural area of the U.S. (East minus Florida) for the years 1950 to

2000. In this study, they confirm that cumulative exposure to heat above the crop-specific threshold (HDDs) is the strongest predictor of yield outcomes. That means that for corn, a day at 40°C instead of a day at 29°C (the temperature at which growth is maximized), leads to a -7% reduction in average county yield. Additional takeaways from this study are also helpful in highlighting the importance of HDDs in independently explaining corn yields: first, the nonlinear relationship stays stable in all regions of the U.S., second, different growing seasons do not alter the results, and finally, the relationship does not depend on precipitation. In the following work by the same authors Roberts and Schlenker (2011), they extend the study of the GDDs/HDDs relationship to the 1910s for one State in the U.S. corn belt, Indiana. To do so, they use the results from the previous paper and use a relatively simple piecewise-linear specification of the yield-GDDs relationship with newly estimated weather data. The results confirm that corn yields reacted similarly to high temperatures throughout the century. Relevant to our study is the fact that the relationship between extreme heat (expressed in HDDs) and corn yield has been stable between 1930 and 1940, with the only reduction in the steepness of the slope after 1950 [7].

In an effort to understand this relationship better for my sample, I modelled corn yield as a function of GDDs, HDDs and precipitation. What I wanted to find was a confirmation of HDDs as the strongest predictor of corn yields, so that I could use it as a weather indicator that captures the effect of high temperatures on the labour market. To do so, I have collected agricultural data at the county-year level for seven states of the Corn Belt area from the U.S. Department of Agriculture public data (NASS) [8]. My data are a balanced panel of 652 counties in the Corn Belt states

---

[7] There is evidence to show that this is the result of adopting a new variety of drought-resistant corn (the double cross hybrid corn Sutch (2011))

[8] Data are publicly available through a searchable database at this link

(Illinois, Indiana, Iowa, Minnesota, Missouri, Nebraska and South Dakota) observed each year between 1930 and 1940 so that I have in total 6520 observations. Among the states in the Corn Belt, only Kansas had no available data between 1930 and 1940. Looking at the state-level data from the NASS database, Kansas was one of the most affected states[9], so its exclusion can only bias results downwards. Table 4.3 reports variable averages. Between 1930 and 1940, the average HDDs was 17; there were on average 2411 GDDs and approximately 18 inches (465 mm) of precipitation during the growing season. For reference, Figure 4.5 and Figure 4.6 show the spatial variation of temperatures, precipitation and yields as long-term averages. Figure 4.7 shows the variation over time of the same variables.

Table 4.3: Summary statistics

	Mean	SD
Degree Days > 29°C	17.38	27.45
Degree Days 10°C - 29°C	2410.80	117.12
Precipitation (mm)	465.09	147.29
Log Yield	3.02	0.90

---

[9] In 1934 and 1936, the two major droughts years, only 20% of acreage planted was harvested in Kansas. In Iowa, it was around 80%. (Sutch, 2011)

Figure 4.5: Spatial variation in temperature: GDDs and HDDs.

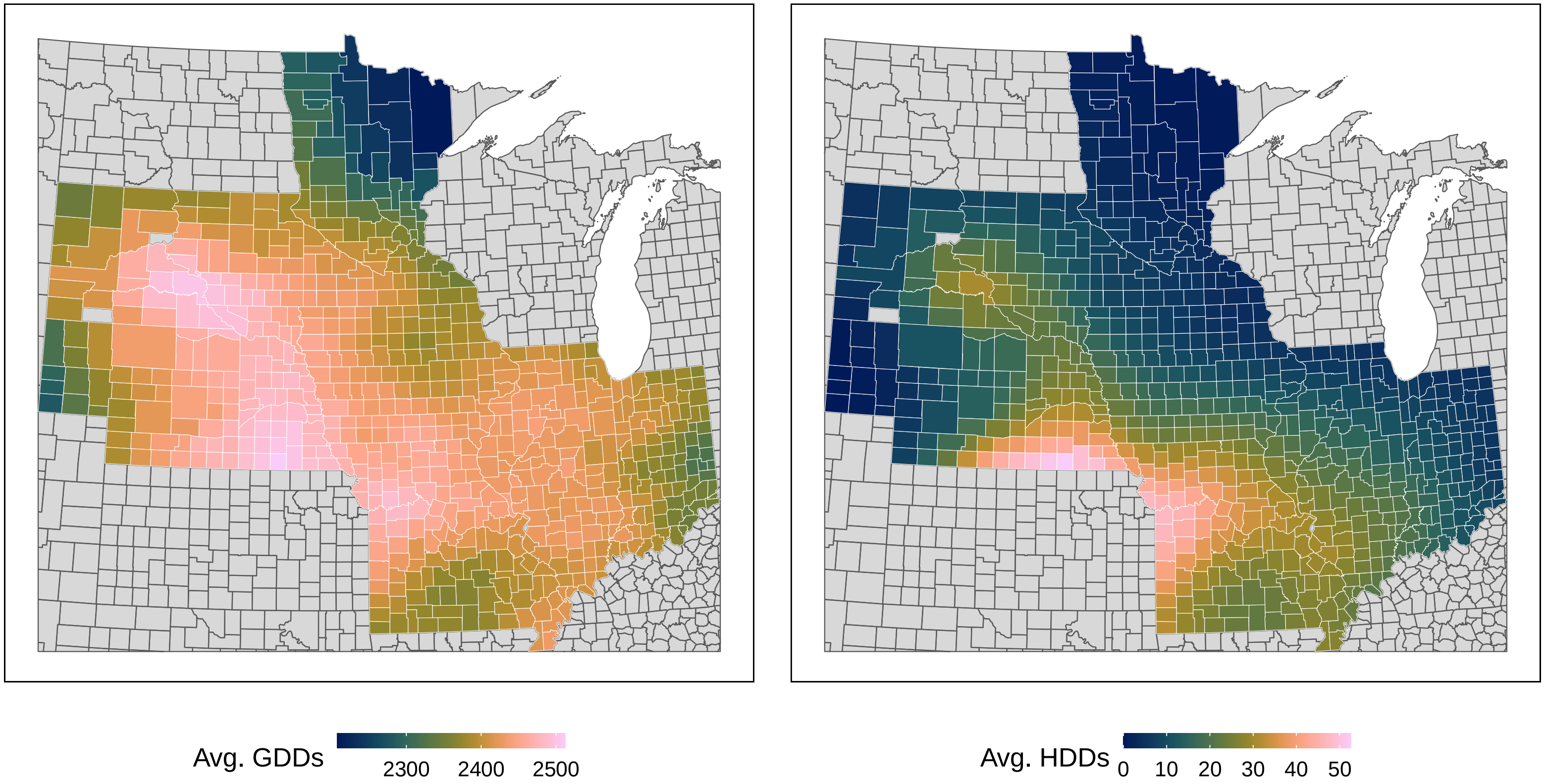


Figure 4.6: Spatial Variation in Precipitation (mm) and Log Yield

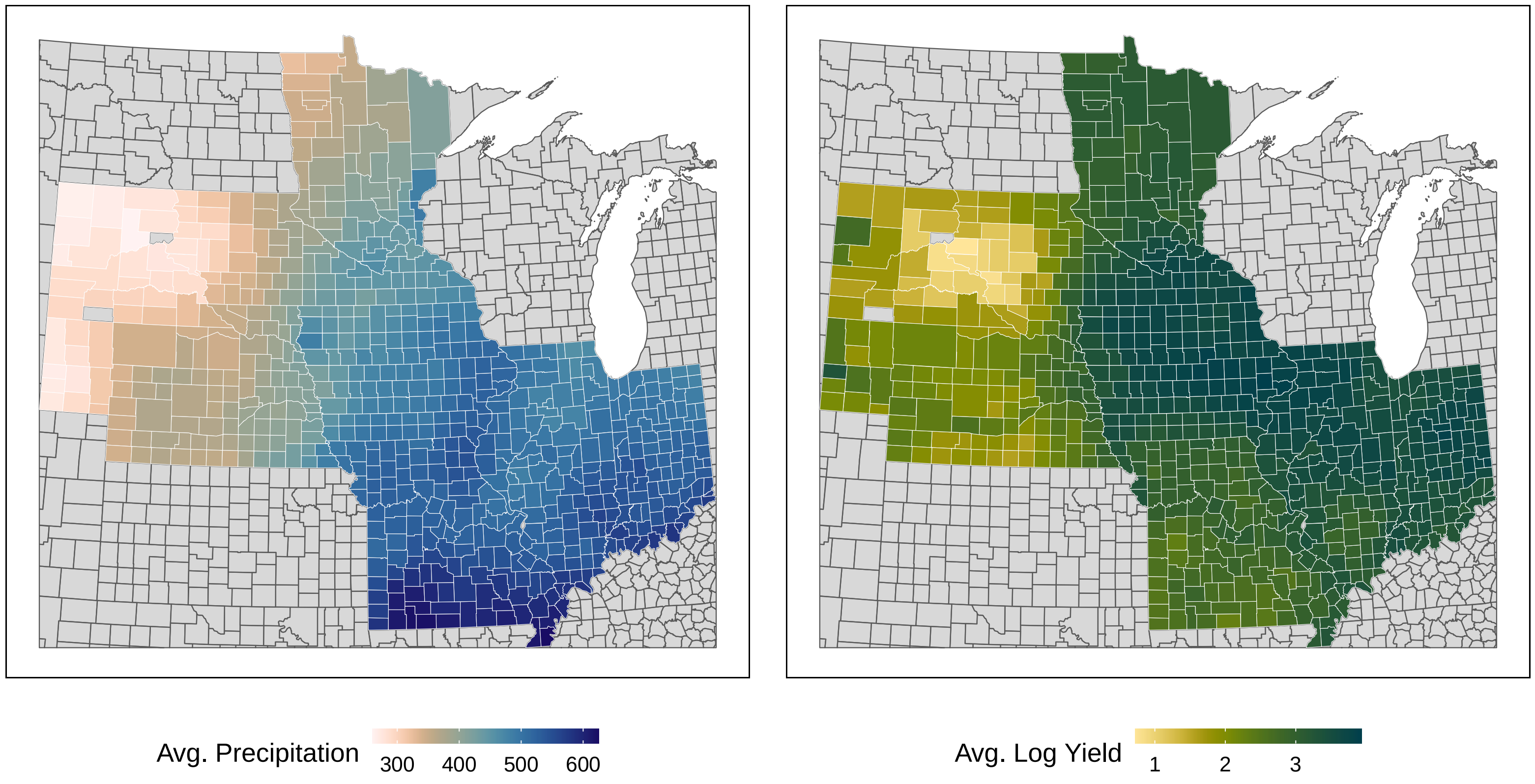


Figure 4.7: Variation over time

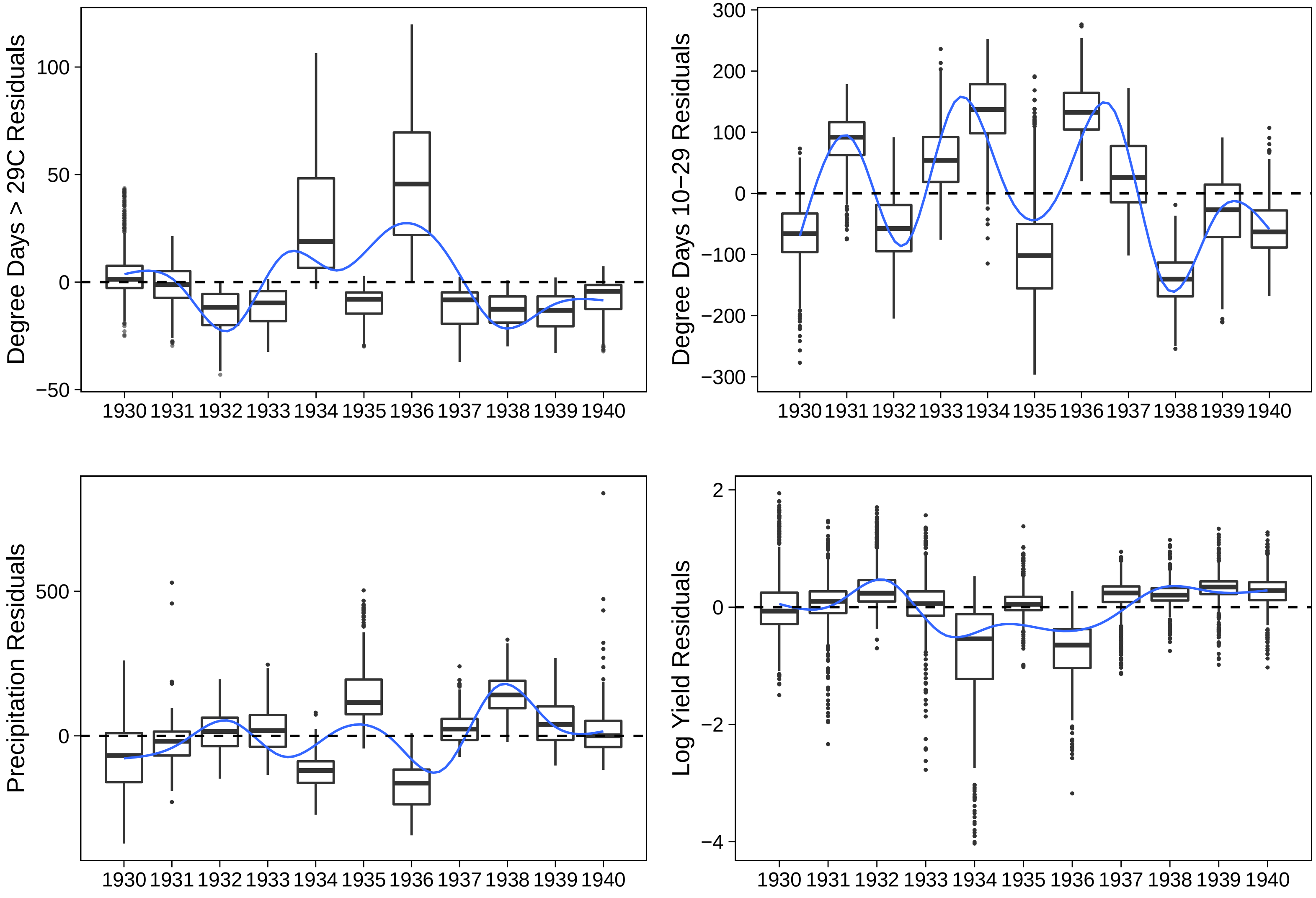


Table 4.4: Nonlinear relationship between Log Yield and Weather

Dependent Variable:	Log(Yield), Bushels per Acre			
Model:	(1)	(2)	(3)	(4)
Variables				
Degree Days > 29°C	-0.0084***	-0.0072**	-0.0085**	-0.0078***
	(0.0021)	(0.0024)	(0.0024)	(0.0018)
Degree Days 10°C - 29°C	-0.0007	-0.0003	-0.0007	-0.0004
	(0.0004)	(0.0003)	(0.0004)	(0.0004)
Precipitation (mm)	13.53**	18.18**	11.24*	16.16**
	(4.496)	(5.910)	(5.208)	(5.431)
Precipitation^2 (mm)	-15.96**	-15.12**	-16.18**	-19.04**
	(5.733)	(5.653)	(6.339)	(6.379)
Fixed-effects				
county	Yes	Yes	Yes	Yes
state				Yes
Time trends				
linear	Yes			
quadratic		Yes		
quadratic (county)			Yes	
quadratic (state)				Yes
Fit statistics				
Observations	6,520	6,520	6,520	6,520
R^2	0.79763	0.80761	0.82263	0.71439
Within R^2	0.52162	0.54521	0.53340	0.44873

Clustered (state) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Similarly to Feng et al. (2015), I test four different model specifications using the following form:

$$Y_{ct} = \beta_1 HDD_{ct} + \beta_2 GDD_{ct} + \beta_3 PRCP_{ct}^2 + \beta_4 PRCP_{ct} + g(t) + \gamma_c + \varepsilon_{ct} \quad (4.5)$$

In which Y_{ct} is the log of average corn yield in county c, HDD_{ct} are harmful degree days - days above 29°C -, HDD_{ct} are growing degree days - days between 10°C and 29°C and $PRCP_{ct}$ is precipitation. Although I have reason to argue that between 1930-1940 there was no substantial change in terms of agricultural technology, time trends are added to control for technological change. Errors are clustered at the state level, also following the existing leading literature (Roberts and Schlenker, 2011; Berry et al., 2014; Feng et al., 2015).

In panel (1) of Table 4.4, the results demonstrate a significant inverse relationship between HDDs and yield: an increase of one harmful degree day reduces yield by approximately 0.7-0.8 percent, assuming all other variables remain constant. In contrast, the relationship between GDDs and yields is not statistically significant and, if existent, is one order of magnitude smaller. These findings align with the work of Feng et al. (2015), except for the insignificance of the GDDs' effect. The negative effect of HDDs on yield is substantial; for instance, a season with ten additional harmful degree days would imply a reduction in yield of about 7-8%, potentially affecting the agricultural income and, by extension, the local economy. The non-significant effect of GDDs could be due to the fact that temperatures within the growing range (10°C - 29°C) do not adversely affect corn yields as much as temperatures above the optimal range. Ultimately, these results confirm a significant relationship between yields and HDDs, and they support the use of HDDs as a variable that can summarize the impact of weather shocks over the decade

A decade of weather shocks

So far I have reduced daily observations of temperatures (min and max) into yearly ones (HDDs and GDDs), and I have established that HDDs are the variable with the strongest effect on yields, which I hypothesize to be a determinant of labour supply. Now, I need to transform my variable of choice, HDDs, so that I can reconcile it to the full census data for 1940. Two papers in the existing literature attempt to do something similar. In a study by Ramcharan (2010) the author uses measures of weather variation over time as an instrument for land concentration. This is because weather is a major determinant of risk in agricultural production. His logic rests on the idea that weather patterns are large sources of spatially co-variant risk. The variable the author names "weather risk" is eventually measured as the standard deviation of precipitation and growing degree days. In a different fashion, Boustan et al. (2010) use proxies for extreme weather conditions to predict outflows from major U.S. cities. This variable is then used to measure the effect of out-migration flows on local labour markets. In their paper, the variables used are 'Months of Extreme Wetness between 1935-40' and 'Average Temperature (1935-1940)'. In this paper, I opted to test two different methods. First I test the relationship between labour market outcomes and the non-linear weather specification - therefore average HDD, average GDD, average precipitation and precipitation squared- as in Aragón et al. (2021). Second, I test a method similar to the one adopted by Ramcharan (2010) and summarize the variation in HDDs as the Log of the standard deviation in HDDs for each county between 1930 and 1940. Within-county variation in HDDs is the strongest predictor of yields so variation in HDDs should summarize how much a county was affected by weather anomalies which were non-beneficial to crop yields between 1930 and 1940. Additionally, although I am assuming that weather affects labour supply through farm profitability (and wages), I am not trying to measure the impact of a reduction in farm profitability between 1930 and 1940, but rather the impact of increased weather variability on sectoral labour reallocation. If a farmer's reaction to weather risk is the result of ex-ante and ex-post adjustments, then the variability of weather over a period should be the trigger of those adjustments.

Table 4.5: Summary statistics for Weather Variables

	1920-1930		1930-1940	
	Mean	SD	Mean	SD
HDDs mean	2.75	2.50	17.38	11.54
HDDs St.Dev.	6.54	5.49	21.98	14.14
GDDs mean	2266.98	32.28	2410.80	49.11
GDDs St.Dev.	109.17	16.84	110.60	14.30
Temperature (°C) mean	16.17	2.05	16.93	1.99
Temperature (°C) St.Dev.	1.03	0.09	0.82	0.13
Precipitation (mm) mean	515.39	99.95	465.09	82.31
Precipitation (mm) St. Dev.	104.49	44.70	122.55	37.46

Table 4.5 shows summary statistics for weather variables for two different periods: 1920-1930 and 1930-1940. HDDs standard deviation for the 1930-1940 decade, the main explanatory variable I am considering is almost two times the standard deviation in the previous decade. On the other hand, GDDs, temperature variation and precipitation variation are more similar in the two decades. This confirms my choice of HDDs Standard Deviation as the second potential proxy for Weather Shocks in the Corn Belt area.

4.5 Empirical Model

In this section, I go through two types of analysis: first I describe the occupational data for my sample using transition matrices (between 1930 and 1940 occupations), and then I use these results together with the hypotheses I outlined in the theoretical background to test the relationship between weather shocks, unemployment and wages.

4.5.1 Changing occupations between 1930 and 1940 in the Corn Belt

To restate, the objective of this paper is to understand whether people living in areas impacted by high variation in HDDs during the 1930-1940 period were more likely to be unemployed or to have changed jobs versus people living in areas less impacted, in the same state and with similar baseline (pre-1930) characteristics. However, for many farmers, the ideal ex-post adaptation strategy was to migrate to a different county. To control for differences between migrants and non-migrants I look at these two samples separately. First I look at the differences between migrants/non-migrants and farmers/non-farmers as they transition between jobs using a transition matrix. Without being able to control for endogenous factors - such as county characteristics, these results are only suggestive of what the relationship might be. Secondly, I will look at the difference across migrant and non-migrant status as well as people living on farms and not in areas with lower versus higher average HDDs (see Tables 4.6 and 4.7).

Table 4.6: Non-migrants

	farm laborers	farmers	laborers/operatives	not employed	professionals/clerks	sales/crafts/service
HDD >p50 (from counties with HDD above median)						
farm laborers	17.92%	40.06%	14.13%	6.68%	4.92%	16.30%
farmers	4.60%	73.41%	5.29%	5.48%	3.90%	7.32%
laborers/operatives	3.54%	7.59%	17.46%	8.16%	16.73%	46.52%
not employed	5.87%	19.09%	7.93%	39.07%	9.82%	18.22%
professionals/clerks	0.99%	4.90%	3.18%	5.89%	64.02%	21.01%
sales/crafts/service	1.58%	4.90%	6.88%	7.14%	16.34%	63.17%
HDD <p50 (from counties with HDD below median)						
farm laborers	20.06%	42.23%	10.76%	6.39%	4.59%	15.97%
farmers	4.29%	74.68%	4.11%	6.08%	3.66%	7.19%
laborers/operatives	2.12%	4.50%	14.62%	7.54%	19.23%	51.99%
not employed	4.70%	15.78%	7.28%	36.29%	12.59%	23.36%
professionals/clerks	0.61%	2.58%	2.92%	5.45%	65.77%	22.67%
sales/crafts/service	0.91%	2.87%	5.73%	6.95%	14.95%	68.59%

Table 4.7: Migrants

	farm laborers	farmers	laborers/operatives	not employed	professionals/clerks	sales/crafts/service
HDD >p50 (from counties with HDD above median)						
farm laborers	13.87%	16.62%	14.19%	7.51%	12.21%	35.60%
farmers	9.51%	23.81%	11.27%	11.50%	13.34%	30.56%
laborers/operatives	5.14%	10.55%	11.12%	9.75%	20.02%	43.41%
not employed	6.12%	11.99%	8.58%	20.21%	18.68%	34.43%
professionals/clerks	2.39%	7.58%	5.12%	8.95%	44.72%	31.23%
sales/crafts/service	3.39%	8.63%	7.53%	9.15%	20.33%	50.98%
HDD <p50 (from counties with HDD below median)						
farm laborers	12.73%	17.64%	12.96%	7.60%	12.93%	36.14%
farmers	6.71%	26.24%	10.10%	11.34%	14.22%	31.39%
laborers/operatives	4.19%	9.94%	11.15%	9.79%	20.80%	44.13%
not employed	3.81%	10.45%	8.74%	21.18%	20.47%	35.34%
professionals/clerks	2.07%	6.76%	5.33%	9.30%	42.77%	33.79%
sales/crafts/service	3.01%	8.40%	7.99%	9.65%	20.29%	50.66%

Most farmers were employed or in the labour force in 1940 across all panels. To be specific, among individuals who stayed in the same state, 6.6% were either non-employed or exited the labour force, while for people who migrated, the number was slightly higher (around 7.5 to 7.6%). Among the farmers who stayed, 40% and 18% of farm labourers either became farmers (owning a farm) or stayed farm labourers. On the other hand, among migrants 35%/30% of farm labourers/farmers switched to the services/craft professions. In areas with average HDDs above the median (top panel in both tables), slightly more farmers decided to leave agriculture, regardless of being a migrant or not. Additionally, in these areas slightly more migrant farm labourers (14.19 vs 12.96%) and farmers (11.27 vs 10.1%) decided to transition to become manufacturing workers in 1940. Substantially more farm labourers who decided to stay in high HDDs areas decided to transition to factory labour (14.13 vs 10.76%). Among the migrants, both in high- and low- HDD areas more people found a job.

Overall, these tables suggest that areas with low and high HDD were different, but also that migration played a very important role in determining outcomes. Among other factors, this is one of the confounding variables I need to control when trying to identify the relationship between weather shocks and employment.

4.5.2 The effect of Weather Risk on Employment

Next, I want to look at the relationship between harmful degree days - HDD - and employment outcomes using a cross-section of individuals with baseline characteristics in 1930 and outcomes in 1940. My first independent variable of interest is built similarly to (Feng et al., 2015) and used in my model as in Aragón et al. (2021) and Jessoe et al. (2018). In this first version I estimate:

$$Y_{iacs1940} = g\left(\beta, W_{c,1930-1940}\right) + \gamma X_{ic1930} + \rho_s + \rho_a + \epsilon_{ics} \tag{4.6}$$

$$g\left(\beta, W_c\right) = \beta_0 DD_c + \beta_1 HDD_c + \beta_2 PP_c + \beta_3 PP_c^2 \tag{4.7}$$

where $Y_{iacs1940}$ is either a binary variable of an individual's employment status in the year 1940 or the logged wage in 1940. Employment status is measured in three different ways: whether an individual is employed for a wage, both in the agricultural and in the non-agricultural sector, whether an individual is employed in the agricultural sector, and whether an individual is employed in a non-agricultural sector. In identifying this relationship I need to consider other channels through which my independent variable of interest - weather - might be related to the outcomes, as well as other variables which might interact with it. In order to control for time-invariant non-observable characteristics such as specific state-level policies and access to labour markets, I run every specification with state fixed-effects (ρ_s). Additional baseline characteristics in 1930 which might have affected outcomes in 1940 are also added through a vector of control variables both at the individual (race, years of education, whether living in an urban centre or not and other demographics) and at the county level - average characteristics such as New Deal spending (Fishback et al., 2005), Erosion (Hornbeck, 2012), soil quality and other topographic characteristics (Fishback et al., 2005), average employment, urbanization etc. ($X_{ic,1930}$). Given that household characteristics are one of the determinants of labour supply, I run my specification with age-cohort specific effects (ρ_a), controlling for the fact that individuals aged 45+ in 1930 might exit the labour force in 1940. As stated in the hypotheses, migration is one of the likely outcomes of a negative weather shock, therefore I run the above specification for migrants and non-migrants and compare the results.

In a second specification, I use a different measure for weather risk, the standard deviation in HDD in the period between 1930 and 1940, logged to adjust for its distribution. As described in section IV, I designed this variable in the manner of Ramcharan (2010) and Boustan et al. (2010). For this specification, I estimate equation 4.6 with logged standard deviation instead of the HDD averages for 1930 to 1940.

4.6 Results

4.6.1 Impacts on Local Employment and Wages

I begin by estimating the effects of harmful degree days - HDDs - and the other weather variables on individual employment outcomes for men who did not migrate. In this case, we consider migration moving to a different State. For each table, results are reported with State F.E., then with State and Age-Cohort F.E., then with additional county-level baseline controls and finally with individual level controls. In Table 4.8, after adding individual controls, a 10 units increase in Average HDDs leads to a 0.0076*100 = 0.76% decrease in employment in the agricultural sector. Moving from an average of 5 to an average of 30 (+25) would lead to a 0.00076*25*100 = 1.9% decrease in employment in the agricultural sector.

Conversely, an increase in GDDs - beneficial degree days - increases employment in the agricultural sector, a relationship also seen with crop yields (see section 4.4.2). The effect of GDDs remains positive and significant across all specifications, reinforcing the evidence of a beneficial role.

Turning to non-agricultural employment in Table 4.9, an increase in Average HDDs again leads to an increase in employment: moving from 5 to 30 Average HDDs leads to 0.00085*25*100 = 2.1% increase. In this case, beneficial GDDs lead to a decrease, confirming the positive effect of GDDs on agricultural employment. Third,

Table 4.10 shows how a negative weather shock would increase the number of people working for wages, either in the agricultural or in the non-agricultural sector by 0.97%, confirming the hypothesis that individuals seek to reduce risk by moving to the less affected sector of the economy or, if staying within the agricultural sector, by getting employed as hired labour in a separate farm. Note that this movement to a different farm/location is possible because I am only excluding individuals who migrated to a different state.

Finally, in Table 4.15, the relationship between Average HDDs and wages is also negative, but the results are less significant. A 10 HDDs increase leads to a 2.3% decrease in wages, which nonetheless confirms the hypotheses. It's important to note that the statistical significance of the relationship between HDDs and wages is weaker here, which might suggest other factors, not captured in the model, that influence wages.

Except for the model about log wages, the R^2 is relatively high, with explained variation between 23% and 32%. Model fit increases with the addition of fixed effects and additional controls (county and individual level).

Table 4.8: Employed in Agricultural Sector, Non-Migrant

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	0.002 27	0.002 24	−0.000 63*	−0.000 76**
	(0.001 87)	(0.001 87)	(0.000 27)	(0.000 24)
Avg. GDD	0.001 08**	0.001 08**	0.000 25***	0.000 26***
	(0.000 34)	(0.000 34)	(0.000 05)	(0.000 05)
Avg. Prec.	−0.013 05***	−0.013 05***	0.001 48***	0.001 31***
	(0.002 11)	(0.002 11)	(0.000 39)	(0.000 36)
Avg. Prec^2	0.000 01***	0.000 01***	0.000 00***	0.000 00***
	(0.000 00)	(0.000 00)	(0.000 00)	(0.000 00)
Num.Obs.	1 743 683	1 743 683	1 743 683	1 743 683
R2	0.082	0.082	0.213	0.328

Table 4.9: Employed in Non-Agricultural Sector, Non-Migrant

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	−0.002 36	−0.002 09	0.000 62*	0.000 85**
	(0.001 78)	(0.001 73)	(0.000 29)	(0.000 31)
Avg. GDD	−0.000 93**	−0.000 95**	−0.000 21**	−0.000 26**
	(0.000 31)	(0.000 30)	(0.000 07)	(0.000 08)
Avg. Prec.	0.011 96***	0.011 86***	−0.001 33**	−0.001 19**
	(0.001 95)	(0.001 90)	(0.000 44)	(0.000 43)
Avg. Prec^2	−0.000 01***	−0.000 01***	0.000 00***	0.000 00**
	(0.000 00)	(0.000 00)	(0.000 00)	(0.000 00)
Num.Obs.	1 743 676	1 743 676	1 743 676	1 743 676
R2	0.057	0.084	0.174	0.280

Table 4.10: Working for wages, Non-Migrant

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	−0.002 07	−0.001 79	0.000 86***	0.000 97***
	(0.001 61)	(0.001 57)	(0.000 26)	(0.000 24)
Avg. GDD	−0.000 98**	−0.001 01**	−0.000 33***	−0.000 31***
	(0.000 31)	(0.000 31)	(0.000 07)	(0.000 07)
Avg. Prec.	0.010 44***	0.010 41***	−0.001 43**	−0.001 06*
	(0.001 79)	(0.001 73)	(0.000 45)	(0.000 42)
Avg. Prec^2	−0.000 01***	−0.000 01***	0.000 00***	0.000 00**
	(0.000 00)	(0.000 00)	(0.000 00)	(0.000 00)
Num.Obs.	1 743 676	1 743 676	1 743 676	1 743 676
R2	0.051	0.085	0.155	0.234

Table 4.11: Log(Wage), Non-Migrant

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	−0.006 67+	−0.006 10+	−0.003 04+	−0.002 33+
	(0.003 61)	(0.003 55)	(0.001 55)	(0.001 34)
Avg. GDD	−0.000 58	−0.000 64	0.000 36	0.000 08
	(0.000 96)	(0.000 96)	(0.000 60)	(0.000 53)
Avg. Prec.	0.032 44***	0.032 38***	0.001 12	0.001 32
	(0.005 07)	(0.005 02)	(0.002 61)	(0.002 18)
Avg. Prec^2	−0.000 03***	−0.000 03***	0.000 00	0.000 00
	(0.000 01)	(0.000 01)	(0.000 00)	(0.000 00)
Num.Obs.	990 948	990 948	990 948	990 948
R2	0.008	0.015	0.028	0.063

4.6.2 Impacts for Migrants

Next, I explore the effects of weather variables on individuals who migrated to a different state after 1935, thus experiencing at least one of the major droughts in their lifetime. The direction of the relationship is largely consistent with what we found for non-migrants, however, the results are less significant and the magnitude of the impacts are smaller.

As shown in Table 4.12, the results indicate a negative relationship between Average HDDs and agricultural employment. Specifically, a 10-unit increase in Average HDDs leads to a decrease in agricultural employment of about 0.34%. This suggests that adverse weather conditions have a less severe impact on migrants' employment in the agricultural sector compared to those who did not migrate.

Next, the impact of Average HDDs on employment in the non-agricultural sector for migrants is positive, albeit small (an increase of about 0.23%, as shown in Table 4.13). This might be due to migrants moving towards regions with less exposure to harsh weather conditions and thus, more stable non-agricultural job opportunities.

With regards to wage work, the results (Table 4.14) are not significant, suggesting that weather variables do not play a meaningful role in determining wage work among migrants. However, it's worth noting that the relationship between Average HDDs and log wages for migrants (Table 4.15) is positive and statistically significant (+2.85%). This observation could be due to migrants relocating from heavily affected areas to regions with better conditions and thus, higher wages.

In terms of R^2 values, they increase as we add more controls across all tables, indicating an improvement in model fit. Yet, as the R^2 values are relatively low, it suggests that while weather conditions have an influence, many other factors not captured in these models also play significant roles in determining employment outcomes for migrants.

In sum, these findings confirm our initial results for non-migrants, showing that weather shocks such as HDDs have discernible effects on employment in both sectors. However, the impacts seem less severe for individuals who migrated after experiencing significant weather events. This underscores the importance of geographic mobility as a potential adaptation strategy in response to changing weather patterns.

Table 4.12: Employed in Agricultural Sector, Migrant

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	−0.000 35	−0.000 37	−0.000 37+	−0.000 34+
	(0.000 40)	(0.000 40)	(0.000 22)	(0.000 21)
Avg. GDD	0.000 41***	0.000 42***	0.000 14*	0.000 15**
	(0.000 09)	(0.000 09)	(0.000 06)	(0.000 06)
Avg. Prec.	−0.004 35***	−0.004 37***	−0.000 60*	−0.000 35
	(0.000 52)	(0.000 52)	(0.000 28)	(0.000 29)
Avg. Prec^2	0.000 00***	0.000 00***	0.000 00+	0.000 00
	(0.000 00)	(0.000 00)	(0.000 00)	(0.000 00)
Num.Obs.	399 275	399 275	399 275	399 275
R2	0.013	0.016	0.026	0.064

Table 4.13: Employed in Non-Agricultural Sector, Migrant

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	0.000 16	0.000 32	0.000 24	0.000 23+
	(0.000 40)	(0.000 38)	(0.000 18)	(0.000 14)
Avg. GDD	−0.000 20*	−0.000 26**	−0.000 02	−0.000 06
	(0.000 09)	(0.000 09)	(0.000 05)	(0.000 04)
Avg. Prec.	0.003 78***	0.003 79***	0.000 54*	0.000 17
	(0.000 48)	(0.000 47)	(0.000 26)	(0.000 25)
Avg. Prec^2	0.000 00***	0.000 00***	0.000 00+	0.000 00
	(0.000 00)	(0.000 00)	(0.000 00)	(0.000 00)
Num.Obs.	399 258	399 258	399 258	399 258
R2	0.005	0.056	0.061	0.113

Table 4.14: Working for wages, Migrant

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	0.000 03	0.000 15	0.000 20	0.000 19
	(0.000 25)	(0.000 24)	(0.000 18)	(0.000 18)
Avg. GDD	−0.000 16*	−0.000 22**	−0.000 05	−0.000 06
	(0.000 07)	(0.000 08)	(0.000 05)	(0.000 05)
Avg. Prec.	0.001 63***	0.001 73***	−0.000 31	−0.000 53+
	(0.000 32)	(0.000 32)	(0.000 27)	(0.000 28)
Avg. Prec^2	0.000 00***	0.000 00***	0.000 00	0.000 00*
	(0.000 00)	(0.000 00)	(0.000 00)	(0.000 00)
Num.Obs.	399 258	399 258	399 258	399 258
R2	0.001	0.057	0.059	0.070

Table 4.15: Log(Wage), Migrant

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	0.004 72**	0.004 81**	0.002 83*	0.002 85**
	(0.001 57)	(0.001 48)	(0.001 15)	(0.001 01)
Avg. GDD	−0.000 72	−0.000 86+	−0.000 36	−0.000 52+
	(0.000 48)	(0.000 45)	(0.000 35)	(0.000 31)
Avg. Prec.	0.009 56***	0.009 88***	0.003 98*	0.002 70+
	(0.001 93)	(0.001 84)	(0.001 79)	(0.001 46)
Avg. Prec^2	−0.000 01***	−0.000 01***	0.000 00*	0.000 00+
	(0.000 00)	(0.000 00)	(0.000 00)	(0.000 00)
Num.Obs.	265 258	265 258	265 258	265 258
R2	0.001	0.018	0.019	0.041

4.6.3 Impacts on Migration

In this last section, I check for the impact of weather shocks on one effect beyond the labour markets: migration. In this specification, there are two binary outcome variables: whether or not an individual migrated to a different county within the same state, and whether or not this individual migrated to a different state. The 1940 census data, which contains information about a person's location five years prior to the census, is used for this purpose. This data ensures that any individual who migrated had spent a minimum of five years in the original county, experiencing at least one major drought year, 1934.

Table 4.16 presents the results for the likelihood of migrating to a different state. Although the simplest specification shows significance, the effect of Average HDDs is negative and decreases when county and individual controls are added. Possibly, counties with higher exposure to weather shocks also have negatively selected individuals who are constrained from moving.

Table 4.17 represents the likelihood of individuals migrating within the same state. Unlike the previous case, the relationship between the Average HDD and migration is positive here, even if the results are not statistically significant. The correlation between Average GDD and migration changes from positive (though small) to negative once county and individual controls are considered, and it is statistically significant. Individuals located in counties with higher GDDs had a slightly higher probability to migrate to a different state and a smaller probability to migrate to a different county.

Like the other models for migrant individuals though, the R^2 is very low, with a range of 2.6% to 4% of variation explained.

Table 4.16: Migrated to a different state

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	−0.001 74*	−0.001 73*	−0.000 21	−0.000 12
	(0.000 81)	(0.000 81)	(0.000 42)	(0.000 39)
Avg. GDD	0.000 12	0.000 12	0.000 25*	0.000 25*
	(0.000 16)	(0.000 16)	(0.000 11)	(0.000 10)
Avg. Prec.	−0.000 79	−0.000 78	−0.003 98***	−0.003 91***
	(0.000 61)	(0.000 60)	(0.000 58)	(0.000 54)
Avg. Prec^2	0.000 00	0.000 00	0.000 00***	0.000 00***
	(0.000 00)	(0.000 00)	(0.000 00)	(0.000 00)
Num.Obs.	2 142 958	2 142 958	2 142 958	2 142 958
R2	0.011	0.011	0.022	0.043

Table 4.17: Migrated to a different county

	State F.E.	State and Age F.E.	+County Controls	+Individual Controls
Avg. HDD	0.001 71*	0.001 76*	0.000 14	0.000 17
	(0.000 79)	(0.000 79)	(0.000 28)	(0.000 28)
Avg. GDD	0.000 08	0.000 07	−0.000 16**	−0.000 17**
	(0.000 12)	(0.000 12)	(0.000 06)	(0.000 06)
Avg. Prec.	−0.002 65***	−0.002 63***	0.001 35***	0.001 33***
	(0.000 69)	(0.000 70)	(0.000 38)	(0.000 38)
Avg. Prec^2	0.000 00***	0.000 00***	0.000 00***	0.000 00***
	(0.000 00)	(0.000 00)	(0.000 00)	(0.000 00)
Num.Obs.	2 142 958	2 142 958	2 142 958	2 142 958
R2	0.006	0.011	0.023	0.026

4.6.4 Alternative specification

In this section, I explore a different approach to encapsulate the concept of weather risk, replacing the previously used measure of Harmful Degree Days. This new approach examines a different metric - the Logarithm of Standard Deviation of HDDs - aiming to examine the relationship of this metric with the variables of interest. This analysis only considers individuals who didn't migrate and only takes into account employment-related outcomes.

Table 4.18 presents the results with all controls, including State and Age Fixed Effects. The coefficients for the Logarithm of Standard Deviation of HDDs are significant and exhibit the same direction as the HDDs coefficients in the previous specification, confirming the robustness of the relationship between weather risks and employment outcomes.

In the agricultural sector, an increase in the Logarithm of Standard Deviation of HDDs corresponds with a decrease in employment. A one unit increase in the variability of harmful weather, agricultural employment decreases by about 1.63%. This suggests that greater weather variability is associated with less employment in this sector. The opposite is true for the non-agricultural sector and wage work, where the positive coefficients suggest that greater weather variability corresponds with increased employment in these sectors. A one unit increase in weather variability could lead to approximately a 1.97% increase in non-agricultural employment and about a 1.71% increase in wage work employment.

The difference in effects between the agricultural and non-agricultural sectors is intuitive given the direct exposure of agricultural activities to weather conditions. The increased weather variability may drive individuals from more volatile agricultural employment to more stable non-agricultural jobs, or push them towards wage work.

It's important to note that these effects confirm those from the main specification. Firstly, looking at employment in the agricultural sector, the coefficient for Log Standard Deviation of HDDs is negative and highly significant, as in the case with Average

HDDs. This supports the previous findings that an increase in weather variability, as indicated by either measure, leads to a reduction in agricultural employment.

Secondly, the coefficients for employment in the non-agricultural sector and wage work are positive and significant. This also confirms the findings of the main model using Average HDDs.

Table 4.18: Different Specification, Log Standard Deviation HDDs

	Employed, Agr. Sector	Employed, Non-Agr. Sector	Employed, wage work
Log. St. Dev. HDD	−0.016 32***	0.019 66***	0.017 10**
	(0.004 47)	(0.004 78)	(0.005 53)
Num.Obs.	1 743 683	1 743 676	1 743 676
R2	0.327	0.280	0.190

4.7 Discussion

The findings presented in the previous section display a consistent picture of how weather shocks impact different facets of the labour market. Table 4.19 and table 4.20 summarize the results for the main coefficient of interest, Harmful Degree Days (HDDs), and for Growing Degree Days (GDDs). The tables show the coefficient on HDD and GDD for the model with all controls included, including state and cohort fixed effects. The coefficient is multiplied by 10, displaying the impact of a 10-units increase. In parentheses, I show the R^2.

The key takeaway is that an increase in Average HDDs leads to a decrease in employment in the agricultural sector, while simultaneously prompting an increase in non-agricultural employment (see table 4.19). This suggests a labour reallocation response to weather shocks, with workers transitioning from the agricultural sector to the non-agricultural sector. This movement might be fueled by agricultural hardships caused by negative weather shocks, leading workers to seek employment in sectors less directly impacted by weather. A similar pattern is observed for wage work, where a negative weather shock leads to an increase in the number of people employed for wages, which may be due to workers resorting to wage work as a fallback option when weather-induced disruptions occur in their regular line of work. Interestingly, the negative relationship between HDDs and wages indicates that weather shocks might depress wage levels, possibly because of an increased supply of labour (from displaced agricultural workers) or reduced demand for labour (due to the negative impact of the shocks on productivity).

Conversely, an increase in growing degree days (GDDs), which generally enhance crop productivity, leads to increased employment in the agricultural sector for both non-migrants and migrants, as predicted by our theory (see table 4.20). Interestingly, this increase in agricultural employment seems to be accompanied by a decrease in non-agricultural employment, which might be a result of labor moving from non-agricultural sectors to agriculture to take advantage of the improved conditions.

Table 4.19: Summary of results of the impacts of HDDs

HDD +10 (R^2)	**Non-Migrants**	**Migrants**
Employed, Agriculture	$-0.76\%^{**}$ (0.32)	$-0.34\%^{+}$ (0.06)
Employed, Not Agriculture	$+0.85\%^{**}$ (0.28)	$+0.23\%^{+}$ (0.11)
Working for Wages	$+0.97\%^{***}$ (0.23)	Not Significant
Log Wage	$-2.3\%^{+}$ (0.06)	$+2.85\%^{**}$ (0.04)
Migrated State		Not Significant
Migrated County		Not Significant

Table 4.20: Summary of results of the impacts of GDDs

GDD +10 (R^2)	**Non-Migrants**	**Migrants**
Employed, Agriculture	$+0.26\%^{***}$ (0.32)	$+0.15\%^{**}$ (0.06)
Employed, Not Agriculture	$-0.26\%^{**}$ (0.28)	Not Significant
Working for Wages	$-0.31\%^{***}$ (0.23)	Not Significant
Log Wage	Not Significant	$-0.52\%^{+}$ (0.04)
Migrated State		$+0.25\%^{*}$ (0.04)
Migrated County		$-0.17\%^{**}$ (0.026)

Comparing my results with existing literature, I have found both similarities and differences. One critical aspect in which my findings align with Jessoe et al. (2018) is that an increase in HDDs leads to a minor reduction in local employment. However, a divergence arises when considering the effects on non-agricultural labor and wage work. This difference could be attributed to the different geographical contexts (Mexico in their study versus Corn Belt states in mine) and temporal contexts (1980-2007 versus 1930s).

Regarding the impact on non-agricultural labor, my results share similarities with those of Branco and Féres (2021) and Albert et al. (2022). Based on their studies in Brazil between 1996 and 2014, both found that areas severly affected by droughts experienced an increase in non-agricultural jobs. This trend is also apparent in my findings.

My results further contribute to the discourse initiated by Hornbeck (2012) and Hornbeck (2020). While their research on the Dust Bowl suggested migration as the primary response to severe weather shocks, my data from the Corn Belt paints a slightly different picture. Instead of moving, farmers may have adapted to the circumstances by switching to non-agricultural sectors. This highlights that coping mechanisms can vary based on the regional context.

Yet, when assessing the relationship between HDDs and migration, my results diverge from those of Sichko (2021). Their study found 'positive selection' into migration during the late 1930s drought, with the more educated individuals choosing to move away from drought-affected regions. In contrast, my analysis does not identify a significant link between HDDs and migration. Again, this discrepancy may arise from my focus on the Corn Belt states, which were generally more developed than the areas studied by Sichko (2021). Consequently, the population in my study area may have had a broader set of coping strategies at their disposal.

4.8 Conclusions

The aim of this chapter was to investigate the effects of the droughts in the U.S. Corn Belt during the 1930s on the labour market, with the purpose of understanding mid to long-term adaptation strategies of individuals hit by weather shocks. After all, climate change is anticipated to generate greater frequency and intensity of extreme weather events with negative impacts on agricultural productivity and crop yields. The rarity of prolonged weather shocks means that the historical evidence will help identify the broader effects.

This paper starts by developing a set of testable hypotheses by including weather risk in the production function of the household and then maximizing over the household's preferences: First, a negative weather shock decreased agricultural productivity – here, corn yields - and farmers sought adaptation through better input allocation. Demand for agricultural labour in the impacted county decreased, reducing the number of labourers hired on farms. Second, reduced production also reduced income for the farmer family, which led to a reduction in non-agricultural goods consumed and, therefore, a reduction in non-agricultural labour demanded in the area where the shock happened. In the short-term, reduction in employment in those areas was an effect. Finally, it was expected that there was an increase in migration, either because of a reduction in wages in the location affected by the shock and relatively higher wages in a neighbouring location or because of decreased demand for hired labour in farming households with a resulting surplus of labour.

Having been tested in previous studies for counties which were highly affected by both drought and erosion (e.g., Hornbeck (2020)), this hypothesis was tested in the current paper for areas that experienced less extreme weather shocks. Employment and demographics data at the individual level come from the complete census for 1930 and 1940. In order to have data about the transition of people from one occupational sector to the other, the study linked 2.2 million individuals (approximately 33% of the total population of the Corn States) over the two censuses. In 1930, most people in the sample were white, born in the U.S. and 29% of them lived on a farm.

To create the weather shocks variable, daily-level data from approximately one thousand stations' observations in the contiguous United States between 1930 and 1940 were aggregated to a decade-level indicator. There are, on average, 356 observations per year per weather station, making it a relatively good data set in terms of variation over time, with a high density of weather stations in the Corn Belt area with slight variation during the 1930s. Given that there is a clear nonlinear relationship between weather and yields for corn, growing degree days (GDD) above 29 degrees Celsius, called Harmful Degree Days (HDD), was used as the key explanatory variable.

The results indicate, first, that a 10 unit increase in average HDDs leads to a 0.76%

decrease in employment in the agricultural sector; moving from an average of 5 to an average of 30 (+25) would lead to a 1.9% reduction in employment in the agricultural sector. Second, an increase in average HDDs leads to an increase in employment in the non-agricultural sector - moving from 5 to 30 average HDDs leads to a 2.1% increase. Third, a negative weather shock would increase the number of people working for wages, either in the agricultural or in the non-agricultural sector, by 2.4%, confirming the hypothesis that individuals seek to reduce risk by moving to the less affected sector of the economy or, if staying within the agricultural sector, by getting employed as hired labour on a separate farm. A 10 HDDs increase leads to a 2.3% decrease in wages, also confirming the hypotheses.

For individuals who migrated to a different state after 1935 (i.e., after at least one of the major droughts), the results are also negative but less significant and smaller (-0.8% for employment in agriculture and +0.5% for employment in the non-agricultural sector), indirectly confirming the results for the individuals who stayed. Interestingly, average HDDs have a positive effect on the wage of migrants, possibly due to the destination of migrants and suggesting that resilient adaptation has the potential to lead to better outcomes in the long run.

Ultimately, this paper confirms the hypotheses in the literature that weather shocks are harmful to agricultural productivity and farmers. However, it does indicate that positive long-run outcomes are possible for those that can and choose to migrate out of agricultural activities suffering due to the impacts of climate change. This conclusion reinforces the importance of policies aiding individuals to adapt to the harmful impacts of climate change.

Chapter 5

Conclusion

In this thesis, themes in economic history were examined using methodologies from environmental, spatial, and development economics. The 20th century U.S. context was the backdrop for investigating the roles of electricity access and climate shocks in local economic dynamics. This concluding chapter outlines the findings, implications, and avenues for further exploration for each of the papers presented.

Chapter 2 introduces a unique county-level dataset detailing the diffusion of arc and incandescent lights in the U.S. from 1900 to 1920. Data was computationally extracted from business directories via Optical Character Recognition (OCR) and refined with Electrical Census data, resulting in granular insights spanning three decades. Initial analysis revealed regionally varied rates of light diffusion, with specific areas like the Pacific region and notably California leading in per capita light consumption. Moreover, distinct diffusion patterns between arc and incandescent lights highlighted the impact of industrialization levels. Beyond its primary findings, the dataset has versatile applications. It will allow for the examination of the socio-economic effects of early 1900s electricity access, and can aid in understanding spatial technology diffusion. Furthermore, it can provide a granular measure of local economic development, particularly valuable for less populous areas. Overall, this dataset constitutes a new resource for exploring the dynamics of technology adoption and its economic implications during the early 20th century.

In this study, the use of OCR was central to extracting data from historical documents. This work involved setting up a specific data extraction pipeline for two business directories - the Powers and McGraw directories - , which integrated best practices from existing literature. Key steps in the pipeline included image processing for color correction, along with automated text parsing and extraction through tools like Tesseract. A significant step was the comparison of OCR output with a manually processed sample to validate accuracy. An automated data cleaning procedure was designed using the Levenshtein Distance algorithm for fuzzy matching, but manual cleaning was necessary to correct inaccuracies, particularly in city names. The undertaking of this work served to highlight the challenges associated with using OCR for historical documents and it showed the importance of rigorous, multi-stage data processing and extraction procedures to ensure dataset accuracy. Therefore, this pipeline provides a useful guide for future research employing OCR for historical document analysis.

Future research can prioritize refining and expanding the derived dataset, while also exploring advanced methods for data extraction from historical documents. Given that the process of OCR and manual data cleaning may introduce errors and inconsistencies, researchers might consider leveraging advancements in Natural Language

Processing (NLP) and language models. In particular, models like GPT-4 or future versions have shown remarkable capability in understanding context and extracting information from unstructured text, potentially improving the accuracy and efficiency of data extraction from historical documents. This would not only reduce the time and effort required for manual data cleaning but could also enhance the overall reliability of the data. This advancement in methodology would enable researchers to handle larger and more complex text-based datasets, opening up new possibilities for historical economic analysis.

Chapter 3 shows that the early process of rural electrification in the American West from 1900 to 1930 had a modest impact on the rural economy, albeit with a 2.5% increase in farmland values. The results show variance across decades, with a notable 3% increase from 1900 to 1910, a period when the most significant effects of reduced distance to the nearest power plant were seen. This is attributed largely to the adoption of electric lights, which improved rural living conditions and likely led to economic spillover effects at the community level. This period of electrification contrasts with the subsequent more successful phase led by the Rural Electrification Administration, potentially due to the limited availability and use of electricity-dependent technologies and ongoing agricultural depression in the early 20th century. Robustness tests yielded mixed results, with varied statistical significance for the positive correlation between proximity to power plants and farmland values, and fluctuating impacts over time. These findings underscored the complexity of rural electrification in the U.S., going beyond narratives centered around the more successful period of 1930 to 1960.

In this study, I have employed the variable of distance to a hydroelectric power plant ($DistH$) as a substitute for the potentially endogenous variable of electricity access. In most instances, $DistH$ is utilized as an instrumental variable in a typical two-stage least squares (2SLS) analysis. However, in historical studies like this one, $DistH$ often serves not as an instrumental variable but as a direct proxy for electricity access, a measure usually not available for the early phases of electrification. This departure from the 2SLS approach bypasses the step that validates the correlation between the instrumental variable and the endogenous variable — a process critical to confirming the instrument's validity. Indeed, without a significant first stage, the instrumental variable is a weak instrument and the 2SLS results might be unreliable.

In section 3.7, I tested the relationship between $DistH$ and a new proxy for electricity access - Incandescent lights per capita. The latter, while a valuable indicator, could not be directly incorporated in the regression due to its endogenous nature. The examination revealed that while a reduction in distance to a hydroelectric power plant did influence the extensive margin, its effect on the intensive margin was insignificant. This finding underlines potential non-linearities with the instrumental variable - or in this case, the proxy - emphasizing the necessity for researchers to rigorously verify the relationship between the instrumental variable and the endogenous variable.

The methodology deployed in this study relies on a Difference-in-Differences (DiD) identification strategy, which hinges on the parallel trends assumption. This assumption posits that, absent the treatment ($DistH$), the treatment and control groups - counties closer to and further from a hydroelectric plant, respectively - would have followed similar trends over time, once conditioned on baseline characteristics. Thus, the identification strategy depends on the comprehensive inclusion of all observable baseline variables. In historical data contexts such as this study, it is the anecdotal evidence of the time and setting that ensures the inclusiveness of all pertinent baseline characteristics. Despite the ample empirical evidence provided to substantiate this context, additional support could further strengthen the study, such as more historical research confirming the connection between local economic development and the placement and growth of hydroelectric power plants.

This study identifies two key areas for future research on rural electrification. Firstly, it highlights the need for a detailed investigation into the adoption and role of off-grid technologies, such as water motors and Delco-Light systems, in the overall electrification process. This research could clarify how these alternative solutions bridged the rural electrification gap during times of limited grid connectivity, and provide valuable insights for current regions leveraging off-grid solutions to expand electricity access. Secondly, the study points to the necessity for a wider examination of the social impacts of early rural electrification. Despite the significant correlation between electrification and increased farmland values, other economic effects are less clear. Exploring the societal changes, like improvements in education, health outcomes, and shifts in social structures and gender roles potentially resulting from time savings through electricity, could enrich our understanding of the full range of benefits from early electrification.

Chapter 4 potentially advances our understanding of the labor market impacts and economic responses to weather shocks. Utilizing individual-level data from the 1930s U.S. Corn Belt states, the study shows that a 10-unit increase in average Harmful Degree Days (HDDs) resulted in a 0.76% decrease in agricultural employment, suggesting a direct impact of weather shocks on agricultural productivity. However, the same increase in HDDs led to a 2.1% rise in non-agricultural employment, demonstrating adaptive responses in the labor market. Furthermore, a negative weather shock increased wage employment in both sectors by 2.4%, but it also resulted in a 2.3% decrease in wages. Among those who migrated following major drought-induced weather shocks, the employment effects were less pronounced but still notable, with a 0.8% decrease in agricultural employment and a 0.5% increase in non-agricultural employment. Notably, HDDs had a positive effect on the wages of migrants, suggesting the potential for long-term positive economic outcomes via adaptation.

This research contributes to two fields: historical U.S. studies examining the impacts of droughts and climate change adaptation studies in developing countries. Historically, the 1930s U.S. droughts, while affecting a far broader area than the Dust Bowl, led to substantial crop losses and human migration. This paper builds on this historical context to look at labor market impacts of less extreme - but more likely - weather shocks (droughts) and in areas with a developed agricultural economy. This work also sheds light on sectoral reallocation as an under-explored climate change adaptation strategy. It focuses on the movement of labor between agricultural and non-agricultural sectors and migration between regions in response to weather shocks. By doing so, it reinforces the necessity for policy interventions that can facilitate such adaptive behaviours.

This chapter made use of a measure of weather shocks which is derived from the non-linear relationship between heat and agricultural yields, Average Harmful Degree Days (HDDs). I first established the relationship between HDDs and corn yields for the counties in my sample, a 1 day increase in HDDs reduces corn yields by at least 0.7%. Then I used the average of HDDs experienced in a decade (1930-1940) as a summarizing measure of weather shocks experienced in a county. My outcome of interest was the employment status of agricultural and non-agricultural workers in the Corn Belt in 1940. In this empirical framework, a primary challenge was identifying a variable to encapsulate the intensity and temporal fluctuations of Heating Degree Days (HDDs) over ten years. This requirement arises from data constraints, with wage data available only for 1940, and individual employment data confined to 1930 and 1940. Other research, such as Jessoe et al. (2018), utilized HDDs and GDDs, facilitated by their access to year-level panel data, which made summarizing HDDs unnecessary. Similarly, Liu et al. (2021) used average temperatures over the growing season, benefiting again from year-level data access. However, to capture the intensity of droughts within specific periods, such differential measures would not have sufficed for this study. Future iterations of this paper will aim to either secure year-level data for a population subset or estimate impacts at the year and county level, thereby accommodating the variation in HDDs.

There are additional potential addition to this research. Firstly, the use of county-level temperatures with individual data may limit the available variation in temperatures. By geolocating individuals in the linked census sample, we could apply location-specific temperatures for more accurate results, as suggested by Berkes et al. (2023). Secondly, the recent release of the 1950 census data offers another expansion opportunity, enabling the examination of weather shocks' impacts up to 1950. Lastly, a significant constraint in our study is the inability to observe heterogenous effects on individuals As suggested by Liu et al. (2021), factors such as liquidity constraints could play a crucial role and explain variation in setoral reallocation. Further research is needed to understand the implications of these factors, making it another valuable direction for future investigation.

To conclude, I would like to summarize the contribution that chapter 3 and 4 thesis might bring to the understanding of development dynamics in current day low- and middle-income countries.

As recently re-stated by Lee et al. (2020a), the success of the Rural Electrification Administration (REA) was, and still is, an example for now-developing countries wanting to increase electricity access in rural areas. When the REA was enacted, in 1935, around 10% of rural households on average had access to electricity, in 1960, this jumped to 96%. Similar rates have been observed more recently. For example, between 1960 and 2000, rural electrification jumped from less than 10% to 75% in Brazil and from less that 10% to 69% in Bangladesh respectively. Both programs were enacted as part of broader projects. The Eletrobras Power Distribution Project I and II in Brazil has been understood to be the most comparable to the U.S. case (Lee et al., 2020a; Lipscomb et al., 2013). It provided investments for expanding the transmission network, increasing generation capacity, but also public policies to handle hyperinflation. Its effects were calculated at the regional level, so that positive effects are the result of rural households accessing electricity, together with rural businesses and institutions.

In this thesis, I looked at the effects of electrification before 1935, when they were not accompanied by big push support policies, but they were the effort of smaller private investments. In the western U.S., electricity was extended to households which were close enough to the grid to justify costs, en-route to bigger cities with higher energy demands. In this case, the effects observed were smaller and with a high variation depending on the decade. Low- and middle- income countries which are still experimenting with programs aimed at extending access to electricity might find these results insightful: extending access to electricity is not guaranteed to have an effect on the economy because its impacts depend on how energy is used and if there are the capabilities and the technologies to do so. In the Western U.S. of the early 1900s, households could only access lightbulbs and minor agricultural technologies still in development. Although previous studies reported substantial effects at the household level for female labour supply, for instance in the seminal study by Dinkelman (2011) in Kenya, more recent findings (Lee et al., 2020b) suggest that effects are heterogeneous and heavily depend on whether the household, or even the individual, has access to credit to purchase appliances, but also on whether the individual is aware of the available appliances.

While chapter 2 and chapter 3 contributed to the broader topic of how technology diffusion impacts rural households, 4 looks at how the natural environment, in the form of climate, has an impact on the rural economy. Agricultural production is the sector of the economy which is the most impacted by weather shocks. Agriculture is also central to employment in developing countries, and it was to the Corn Belt states in the 1930s.

In the 1930s U.S., farming households were still operating like consumer-producer entities (Monroe, 1940). That is to say, both consumption and production occurred within the household (Singh et al., 1986b). This pattern is also prevalent in many developing countries today, especially in rural areas. Therefore, the study's findings on the labor market impacts of weather shocks and the adaptation strategies employed by individuals and households could have significant implications for developing country contexts. Within development economics, there is new interest in understanding sectoral reallocation due to climate change, with several empirical applications in the cases of India and South Africa. In India, weather shocks have been shown to be associated with structural change and movement from the agricultural to the non-agricultural sector (Colmer, 2021), but that long-term impacts are of an overall reduction in non-agricultural employment due to decreased demand for non-agricultural goods (Liu et al., 2021). In South Africa on the other hand, the tertiary and informal sectors are the most negatively impacted (Gray et al., 2022). Although crops impacted, temperatures, and context vary, the mechanisms of sectoral reallocation are similar. Both the historical and the contemporary case can be conceptualized by looking at risk adaptation strategies (Rose, 2001) and reflecting on the impact of shocks on local labour markets (Moretti, 2011). Policymakers could leverage the findings to develop targeted interventions that support adaptive labor reallocation in the face of weather shocks, facilitating the movement of workers between sectors as a means of mitigating the negative impacts of climate change on livelihoods. Additionally, the study's emphasis on the role of migration in response to weather shocks could inform policies that support internal migration as a way to increase economic resilience in rural areas.

Appendices

A Appendix A

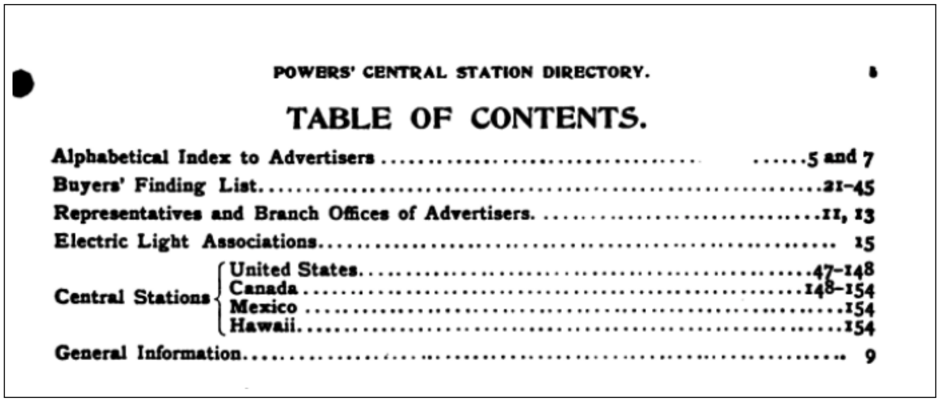
A.1 HathiTrust Catalog References

Table 5.1: Catalog references. Note that the references can be used with the HathiTrust API service to download already OCR'd text and/or page images.

Item	Year	Reference
Powers Central Station Directory	1900	nyp.33433062722495
Powers Central Station Directory	1901	nyp.33433062722487
Powers Central Station Directory	1902	nyp.33433062722479
Powers Central Station Directory	1903	nyp.33433062722461
McGraw Central Station Directory	1911	uiug.30112111811193
McGraw Central Station Directory	1919	uiug.30112039906133

A.2 Pages' examples

Figure 5.1: The Table of Contents of the Powers Central Station Directory for 1900.

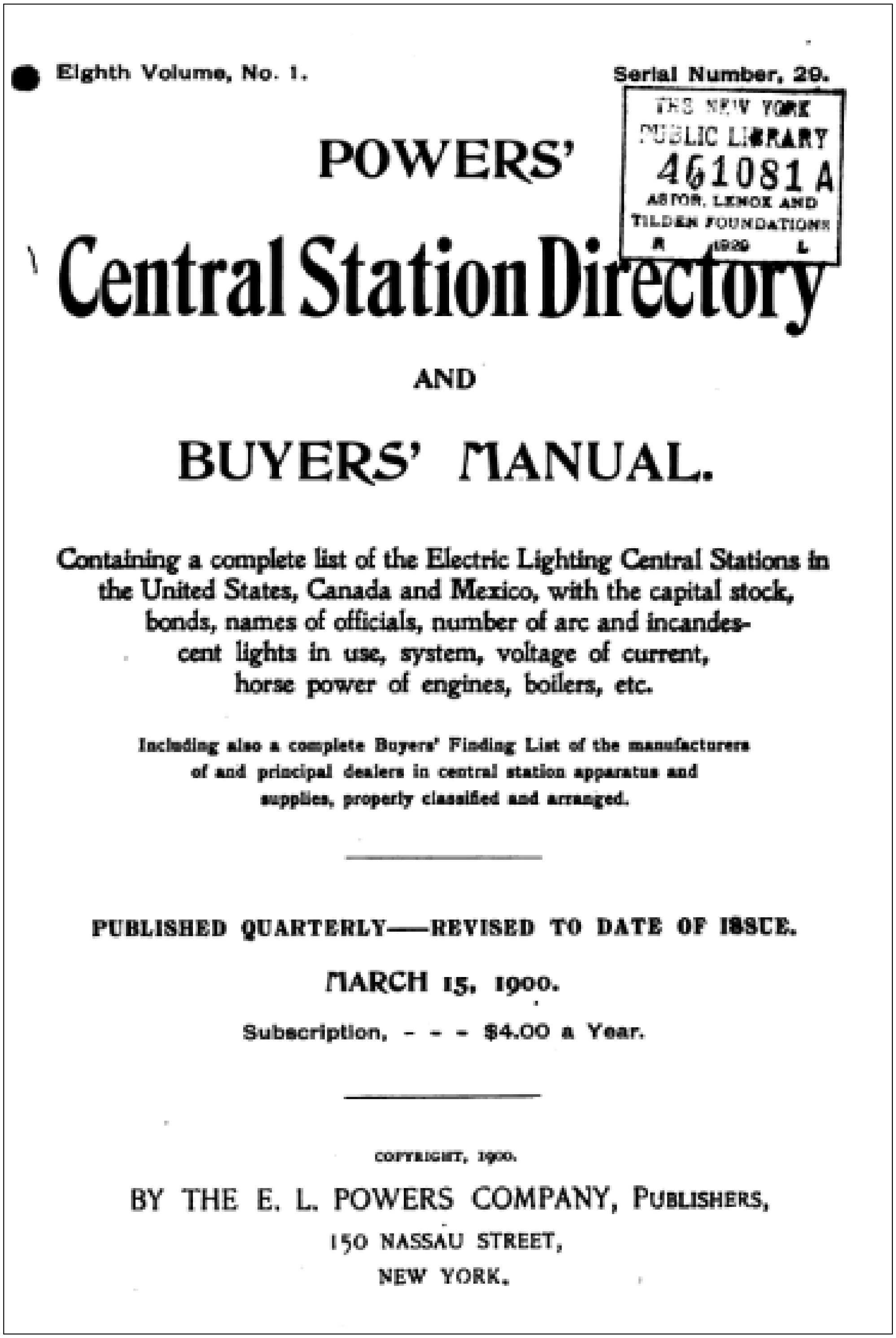


POWERS' CENTRAL STATION DIRECTORY.

TABLE OF CONTENTS.

Alphabetical Index to Advertisers .... 5 and 7
Buyers' Finding List 21–45
Representatives and Branch Offices of Advertisers 11, 13
Electric Light Associations 15
Central Stations
  United States 47–148
  Canada 148–154
  Mexico 154
  Hawaii 154
General Information 9

Figure 5.2: The first page of the Powers Central Station Directory for 1900.



Eighth Volume, No. 1. Serial Number, 29.

POWERS'

Central Station Directory

AND

BUYERS' MANUAL.

Containing a complete list of the Electric Lighting Central Stations in the United States, Canada and Mexico, with the capital stock, bonds, names of officials, number of arc and incandescent lights in use, system, voltage of current, horse power of engines, boilers, etc.

Including also a complete Buyers' Finding List of the manufacturers of and principal dealers in central station apparatus and supplies, properly classified and arranged.

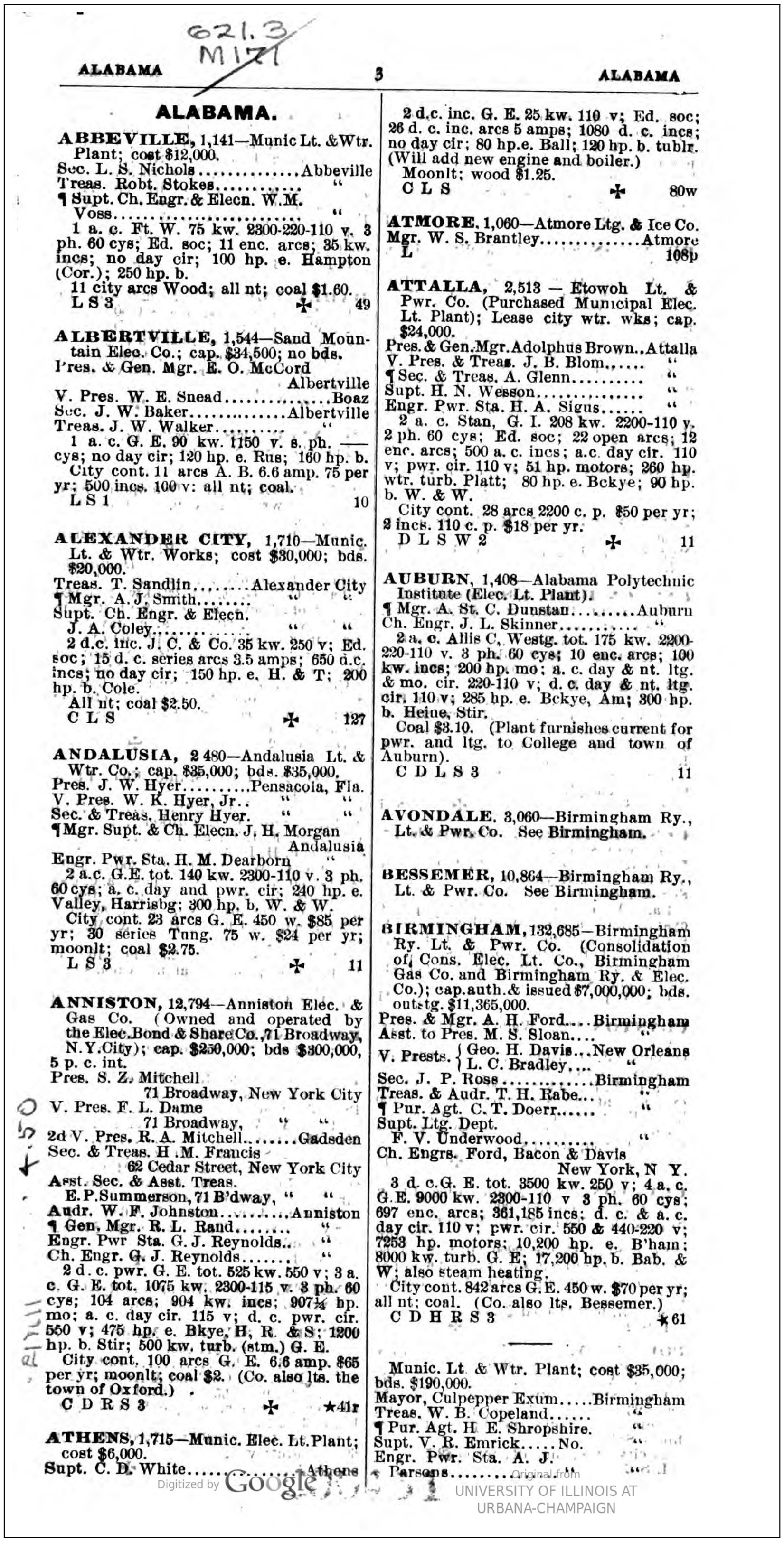
PUBLISHED QUARTERLY——REVISED TO DATE OF ISSUE.

MARCH 15, 1900.

Subscription, - - - $4.00 a Year.

BY THE E. L. POWERS COMPANY, PUBLISHERS,
150 NASSAU STREET,
NEW YORK.

Figure 5.3: List of Abbreviations in the McGraw' directory, 1911. Page 215



ALABAMA.

ABBEVILLE, 1,141—Munic Lt. &Wtr. Plant; cost $12,000.
Sec. L. S. Nichols..............Abbeville
Treas. Robt. Stokes............ "
¶ Supt. Ch. Engr. & Elecn. W. M. Voss......................... "

1 a. c. Ft. W. 75 kw. 2300-220-110 v. 3 ph. 60 cys; Ed. soc; 11 enc. arcs; 35 kw. incs; no day cir; 100 hp. e. Hampton (Cor.); 250 hp. b.

11 city arcs Wood; all nt; coal $1.60.
L S 3 ✠ 49

ALBERTVILLE, 1,544—Sand Mountain Elec. Co.; cap. $34,500; no bds.
Pres. & Gen. Mgr. E. O. McCord Albertville
V. Pres. W. E. Snead...............Boaz
Sec. J. W. Baker..............Albertville
Treas. J. W. Walker.......... "

1 a. c. G. E. 90 kw. 1150 v. s. ph. —— cys; no day cir; 120 hp. e. Rus; 160 hp. b.

City cont. 11 arcs A. B. 6.6 amp. 75 per yr; 500 incs. 100 v: all nt; coal.
L S 1 10

ALEXANDER CITY, 1,710—Munic. Lt. & Wtr. Works; cost $30,000; bds. $20,000.
Treas. T. Sandlin........Alexander City
¶ Mgr. A. J. Smith........ " "
Supt. Ch. Engr. & Elecn. J. A. Coley.............. " "

2 d.c. inc. J. C. & Co. 35 kw. 250 v; Ed. soc; 15 d. c. series arcs 3.5 amps; 650 d.c. incs; no day cir; 150 hp. e. H. & T; 200 hp. b. Cole.

All nt; coal $2.50.
C L S ✠ 127

ANDALUSIA, 2 480—Andalusia Lt. & Wtr. Co.; cap. $35,000; bds. $35,000.
Pres. J. W. Hyer.........Pensacola, Fla.
V. Pres. W. K. Hyer, Jr.. " "
Sec. & Treas. Henry Hyer. " "
¶ Mgr. Supt. & Ch. Elecn. J. H. Morgan Andalusia
Engr. Pwr. Sta. H. M. Dearborn "

2 a.c. G.E. tot. 140 kw. 2300-110 v. 3 ph. 60 cys; a. c. day and pwr. cir; 240 hp. e. Valley, Harrisbg; 300 hp. b. W. & W.

City cont. 23 arcs G. E. 450 w. $85 per yr; 30 series Tung. 75 w. $24 per yr; moonlt; coal $2.75.
L S 3 ✠ 11

ANNISTON, 12,794—Anniston Elec. & Gas Co. (Owned and operated by the Elec. Bond & Share Co., 71 Broadway, N.Y. City); cap. $250,000; bds $300,000, 5 p. c. int.
Pres. S. Z. Mitchell 71 Broadway, New York City
V. Pres. F. L. Dame 71 Broadway, " "
2d V. Pres. R. A. Mitchell........Gadsden
Sec. & Treas. H. M. Francis 62 Cedar Street, New York City
Asst. Sec. & Asst. Treas. E. P. Summerson, 71 B'dway, " "
Audr. W. F. Johnston..........Anniston
¶ Gen. Mgr. R. L. Rand........ "
Engr. Pwr Sta. G. J. Reynolds.. "
Ch. Engr. G. J. Reynolds....... "

2 d. c. pwr. G. E. tot. 525 kw. 550 v; 3 a. c. G. E. tot. 1075 kw. 2300-115 v. 3 ph. 60 cys; 104 arcs; 904 kw. incs; 907¼ hp. mo; a. c. day cir. 115 v; d. c. pwr. cir. 550 v; 475 hp. e. Bkye, H, R. & S; 1200 hp. b. Stir; 500 kw. turb. (stm.) G. E.

City cont. 100 arcs G. E. 6.6 amp. $65 per yr; moonlt; coal $2. (Co. also lts. the town of Oxford.)
C D R S 3 ✠ ★41r

ATHENS, 1,715—Munic. Elec. Lt. Plant; cost $6,000.
Supt. C. D. White.................Athens

2 d.c. inc. G. E. 25 kw. 110 v; Ed. soc; 26 d. c. inc. arcs 5 amps; 1080 d. c. incs; no day cir; 80 hp. e. Ball; 120 hp. b. tublr. (Will add new engine and boiler.)

Moonlt; wood $1.25.
C L S ✠ 80w

ATMORE, 1,060—Atmore Ltg. & Ice Co.
Mgr. W. S. Brantley..............Atmore
L 108p

ATTALLA, 2,513 — Etowoh Lt. & Pwr. Co. (Purchased Municipal Elec. Lt. Plant); Lease city wtr. wks; cap. $24,000.
Pres. & Gen. Mgr. Adolphus Brown..Attalla
V. Pres. & Treas. J. B. Blom...... "
¶ Sec. & Treas. A. Glenn.......... "
Supt. H. N. Wesson............... "
Engr. Pwr. Sta. H. A. Sigus...... "

2 a. c. Stan, G. I. 208 kw. 2200-110 v. 2 ph. 60 cys; Ed. soc; 22 open arcs; 12 enc. arcs; 500 a. c. incs; a.c. day cir. 110 v; pwr. cir. 110 v; 51 hp. motors; 260 hp. wtr. turb. Platt; 80 hp. e. Bckye; 90 hp. b. W. & W.

City cont. 28 arcs 2200 c. p. $50 per yr; 2 incs. 110 c. p. $18 per yr.
D L S W 2 ✠ 11

AUBURN, 1,408—Alabama Polytechnic Institute (Elec. Lt. Plant).
¶ Mgr. A. St. C. Dunstan.........Auburn
Ch. Engr. J. L. Skinner........... "

2 a. c. Allis C, Westg. tot. 175 kw. 2200-220-110 v. 3 ph. 60 cys; 10 enc. arcs; 100 kw. incs; 200 hp. mo; a. c. day & nt. ltg. & mo. cir. 220-110 v; d. c. day & nt. ltg. cir. 110 v; 285 hp. e. Bckye, Am; 300 hp. b. Heine, Stir.

Coal $3.10. (Plant furnishes current for pwr. and ltg. to College and town of Auburn).
C D L S 3 11

AVONDALE, 3,060—Birmingham Ry., Lt. & Pwr. Co. See **Birmingham.**

BESSEMER, 10,864—Birmingham Ry., Lt. & Pwr. Co. See **Birmingham.**

BIRMINGHAM, 132,685—Birmingham Ry. Lt. & Pwr. Co. (Consolidation of Cons. Elec. Lt. Co., Birmingham Gas Co. and Birmingham Ry. & Elec. Co.); cap. auth. & issued $7,000,000; bds. outstg. $11,365,000.
Pres. & Mgr. A. H. Ford....Birmingham
Asst. to Pres. M. S. Sloan.... "
V. Prests. { Geo. H. Davis...New Orleans / L. C. Bradley.... "
Sec. J. P. Ross.............Birmingham
Treas. & Audr. T. H. Rabe... "
¶ Pur. Agt. C. T. Doerr...... "
Supt. Ltg. Dept. F. V. Underwood.......... "
Ch. Engrs. Ford, Bacon & Davis New York, N Y.

3 d. c. G. E. tot. 3500 kw. 250 v; 4 a. c. G. E. 9000 kw. 2300-110 v 3 ph. 60 cys; 697 enc. arcs; 361,185 incs; d. c. & a. c. day cir. 110 v; pwr. cir. 550 & 440-220 v; 7253 hp. motors; 10,200 hp. e. B'ham; 8000 kw. turb. G. E; 17,200 hp. b. Bab. & W; also steam heating.

City cont. 842 arcs G. E. 450 w. $70 per yr; all nt; coal. (Co. also lts. Bessemer.)
C D H R S 3 ★61

Munic. Lt & Wtr. Plant; cost $35,000; bds. $190,000.
Mayor, Culpepper Exum.....Birmingham
Treas. W. B. Copeland...... "
¶ Pur. Agt. H. E. Shropshire. "
Supt. V. R. Emrick.....No. "
Engr. Pwr. Sta. A. J. Parsons.......... " "

B Appendix B

B.1 Additional results

Table 5.2: Additional Variables - All controls

	Log Tot Pop all	Log Rural Pop all	% Urbanization all
	(1)	(2)	(3)
DistH	0.007	0.005	0.001
	(0.006)	(0.006)	(0.003)
1920	-0.231^{***}	-0.210^{***}	-0.008
	(0.055)	(0.047)	(0.022)
1930	-0.241^{***}	-0.189^{***}	-0.034^{*}
	(0.054)	(0.046)	(0.019)
Observations	444	428	428
Adjusted R^2	0.324	0.234	0.081
F Statistic	11.108^{***} (df = 21; 422)	7.218^{***} (df = 21; 406)	2.801^{***} (df = 21; 406)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5.3: Farmland Value - Year Interaction

	Log Farmland Value per Acre		
	base	+ msa	+ geo
	(1)	(2)	(3)
DistHp	0.033^{***}	0.031^{***}	0.030^{***}
	(0.010)	(0.010)	(0.011)
1920	-0.279^{***}	-0.292^{***}	-0.294^{***}
	(0.102)	(0.102)	(0.105)
1930	-1.025^{***}	-1.045^{***}	-1.046^{***}
	(0.104)	(0.105)	(0.107)
DistH * year 1920	-0.024	-0.021	-0.021
	(0.021)	(0.020)	(0.021)
DistH * year 1930	-0.039^{*}	-0.025	-0.026
	(0.021)	(0.020)	(0.020)
Observations	444	444	444
Adjusted R^2	0.641	0.655	0.654
F Statistic	50.497^{***} (df = 16; 427)	43.120^{***} (df = 20; 423)	37.442^{***} (df = 23; 420)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5.4: Main Results - Interactions extension

	Log Farms all (1)	Log Farm Output all (2)	Log Rural Pop all (3)
DistH	0.014	0.011	0.007
	(0.011)	(0.017)	(0.009)
DistH * year 1920	−0.016	−0.024	0.002
	(0.018)	(0.029)	(0.017)
DistH * year 1930	−0.024	−0.018	−0.016
	(0.018)	(0.027)	(0.019)
Observations	444	444	428
Adjusted R^2	0.183	0.658	0.213
F Statistic	5.322^{***} (df = 23; 420)	38.005^{***} (df = 23; 420)	6.036^{***} (df = 23; 404)

Note: *p<0.1; **p<0.05; ***p<0.01

B.2 Additional Figures

Figure 5.4: West vs East Transmission Types

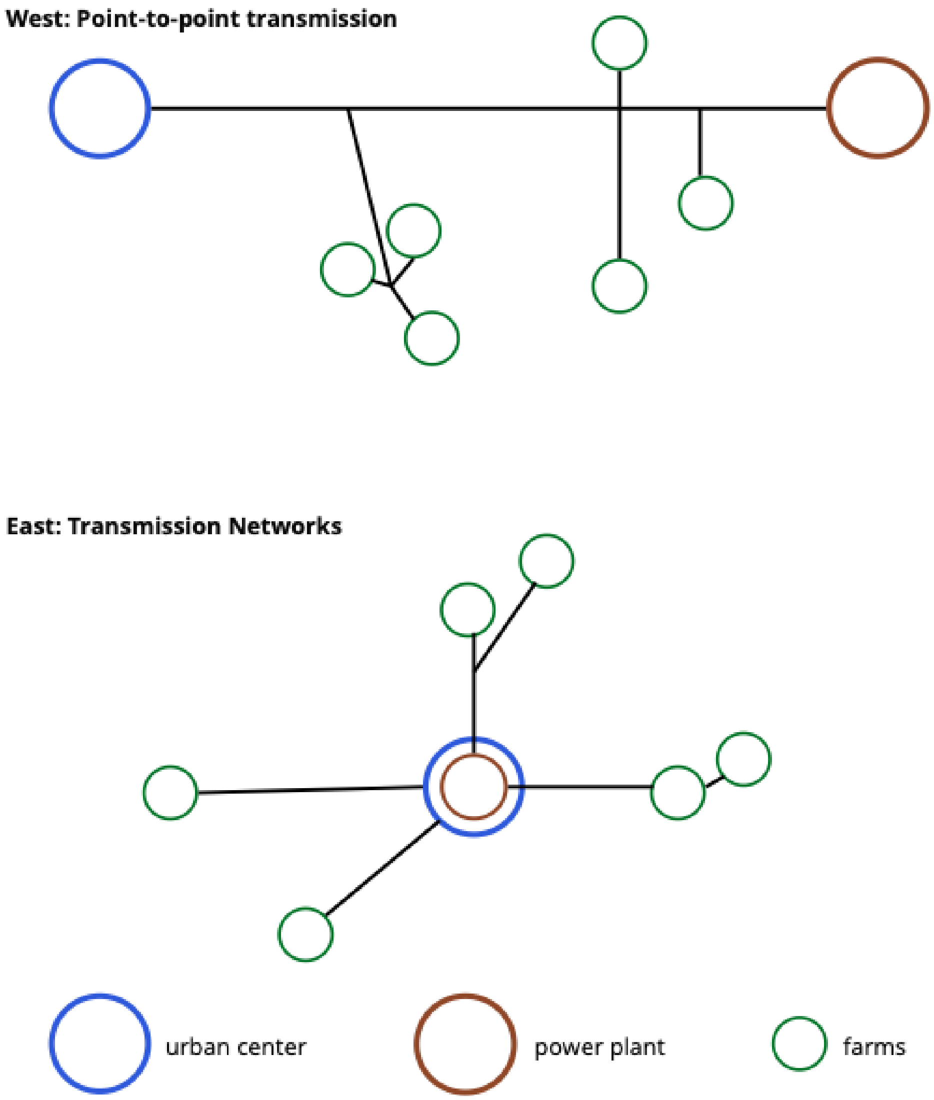


Figure 5.5: Transmission Lines, California and Nevada, 1912.

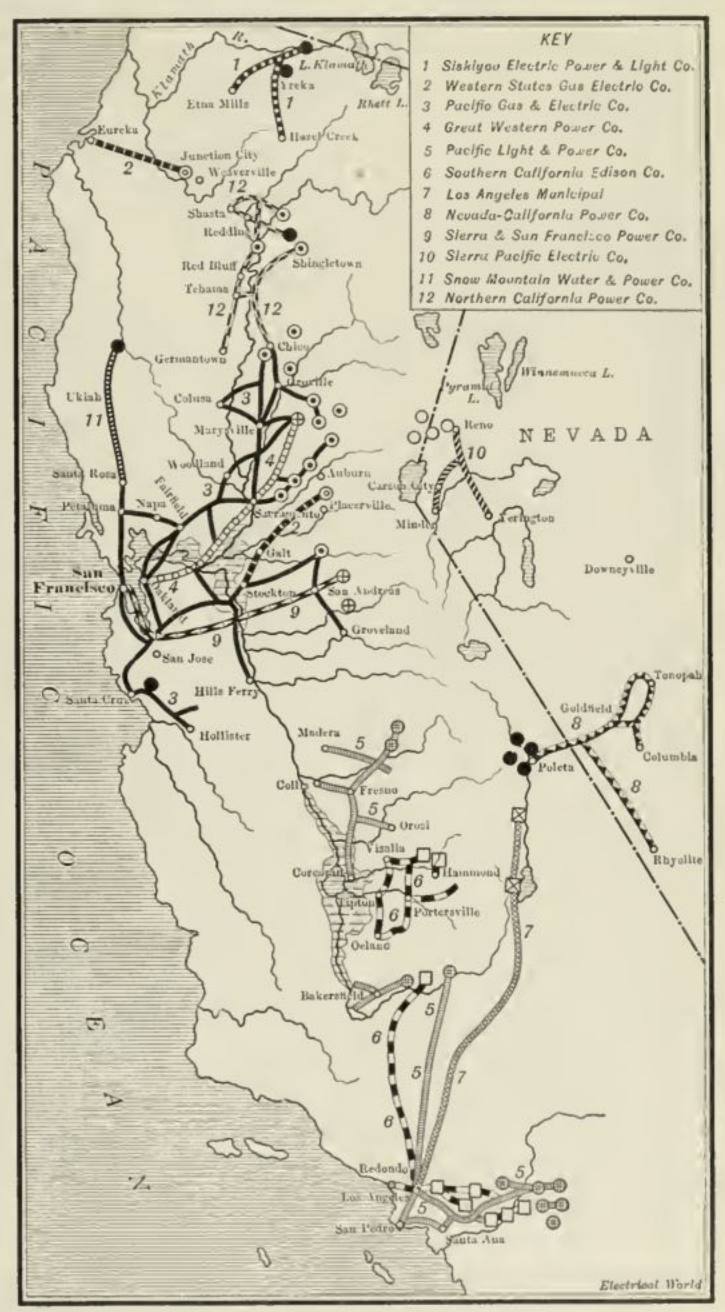
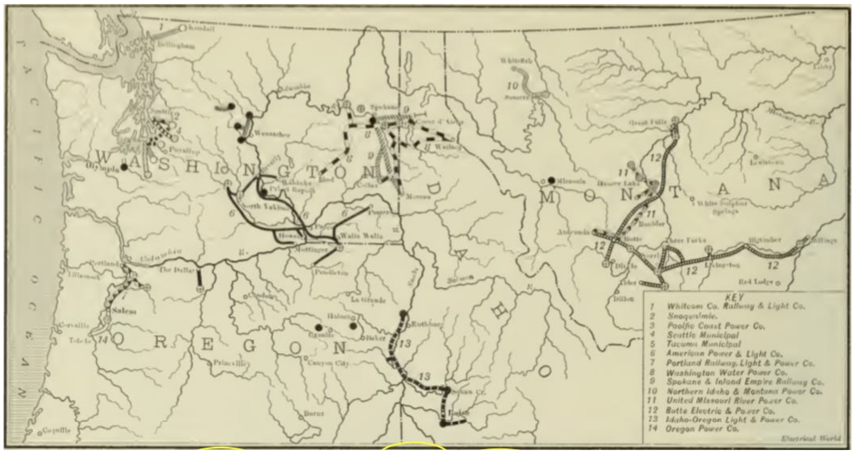


Fig. 2—Transmission Systems in California and Nevada.

Figure 5.6: Transmission Lines, North-West, 1912.



Electrical World

Bibliography

R. Abramitzky. Economics and the Modern Economic HIstorian. *NBER Working Paper Series*, 2015. 58

R. Abramitzky, L. Boustan, K. Eriksson, J. Feigenbaum, and S. Pérez. Automated Linking of Historical Data. *Journal of Economic Literature*, 59(3):865–918, 2019. 138, 140

T. N. H. Albers and K. Kappner. Perks and pitfalls of city directories as a micro-geographic data source. *Explorations in Economic History*, 87:101476, 2023. 35, 39

C. Albert, P. Bustos, and J. Ponticelli. The Effects of Climate Change on Labor and Capital Reallocation. Unpublished Working Paper, 2022. 174

F. M. Aragón, F. Oteiza, and J. P. Rud. Climate Change and Agriculture: Subsistence Farmers' Response to Extreme Heat. *American Economic Journal: Economic Policy*, 13(1), 2021. 16, 126, 156, 160

J. Atack, R. A. Margo, and P. W. Rhode. "Mechanization Takes Command?": Powered Machinery and Production Times in Late Nineteenth-Century American

Manufacturing. *The Journal of Economic History*, 82(3), 2022. 13

M. Auffhammer, S. M. Hsiangy, W. Schlenker, and A. Sobelz. Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, 7(2):181–198, 2013. 143, 147

R. B. Baker, J. Blanchette, and K. Eriksson. Long-run Impacts of Agricultural Shocks on Educational Attainment: Evidence from the Boll Weevil, Dec. 2018. 14

P. Bastos, M. Busso, and S. Miller. Adapting to Climate Change: Long-Term Effects of Drought on Local Labor Markets. *SSRN Electronic Journal*, 2013. 130

L. Bell. *The Art of Illumination.* McGraw-Hill Book Co., 1912. 21

E. Berkes, E. Karger, and P. Nencka. The census place project: A method for geolocating unstructured place names. *Explorations in Economic History*, 87, 2023. 184

S. T. Berry, M. J. Roberts, and W. Schlenker. Corn production shocks in 2012 and beyond: Implications for harvest volatility. In J. Chavas, D. Hummels, and B. D. Wright, editors, *The economics of food price volatility*, pages 59–81. University of Chicago Press, 2014. Section: 2. 155

H. Bleakley and S. C. Hong. Adapting to the weather: Lessons from U.S. history. *Journal of Economic History*, 77(3):756–795, 2017. 146

L. P. Boustan, P. V. Fishback, and S. Kantor. The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression. *Journal of Labor Economics*, 28(4):719–746, Oct. 2010. 156, 162

D. Branco and J. Féres. Weather Shocks and Labor Allocation: Evidence from Rural Brazil. *American Journal of Agricultural Economics*, 103(4):1359–1377, 2021. 16, 127, 174

A. A. Bright. *The Electric Lamp Industry: Technological Change and Economic Development from 1800 to 1947.* The Macmillan Company, 1949. 17, 23, 24

M. Burke and K. Emerick. Adaptation to Climate Change: Evidence from US Agriculture. *American Economic Journal: Economic Policy*, 8(3):106–140, Aug. 2016. 16

F. Burlig and L. Preonas. Out of the darkness into the light? Development effects of rural electrification. Publisher: Energy Institute ad Haas, 2016. 65, 67, 122

F. Caselli and W. J. Coleman II. The U.S. Structural Transformation and Regional Convergence: A Reinterpretation. *Journal of Political Economy*, 109(3):584–616, 2001. Publisher: The University of Chicago Press. 28

J. Colmer. Temperature, labor reallocation, and industrial production: Evidence from india. *American Economic Journal: Applied Economics*, 13(4), 2021. 187

D. Comin, M. Dmitriev, and E. Rossi-Hansberg. The Spatial Diffusion of Technology. Technical Report w18534, National Bureau of Economic Research, Cambridge, MA, Nov. 2012. 60

B. I. Cook, R. Seager, and J. E. Smerdon. The worst North American drought year of the last millennium: 1934. *Geophysical Research Letters*, 41(20):7298–7305, 2014. ISSN 19448007. 126, 131

S. Correia and S. Luck. Digitizing historical balance sheet data: A practitioner's guide. *Explorations in Economic History*, 87, Jan. 2023. 35

R. S. Cowan. *More Work for Mother*. Basic Books, 1983. 18, 25

W. Cronon. *Nature's Metropolis: Chicago and the Great West*. W. W. Norton & Company, 1991. 11

P. A. David. The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox. *The American Economic Review*, 80(2):355–361, 1990. 14

O. Deschênes and M. Greenstone. The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather: Reply. *American Economic Review*, 102(7):3761–3773, 2007. 125

S. Di Falco, M. Veronesi, and M. Yesuf. Does Adaptation to Climate Change Provide Food Security? A Micro-Perspective from Ethiopia. *American Journal of Agricultural Economics*, 93(3):829–846, 2011. 16, 126

T. Dinkelman. The Effects of Rural Electrification on Employment: New Evidence from South Africa. *American Economic Review*, 101(7):3078–3108, Dec. 2011. 65, 67, 186

D. Donaldson and R. Hornbeck. Railroads and economic growth: a "Market Access" approach. *The Quarterly Journal of Economics*, 131(2):799–858, 2016. 13

E. C. Edwards and S. M. Smith. The Role of Irrigation in the Development of Agriculture in the United States. *The Journal of Economic History*, 78(4), Dec.

2018. 84

M. S. Eisenhower. Yearbook of Agriculture. Technical report, United States Department of Agriculture, 1935. 131

K. Emerick. Agricultural productivity and the sectoral reallocation of labor in rural India. *Journal of Development Economics*, 135, Nov. 2018. 130

S. Feng, M. Oppenheimer, and W. Schlenker. Weather anomalies, crop yields, and migration in the US corn belt. Number: March, 2015. 16, 126, 128, 154, 155, 160

P. V. Fishback, W. C. Horrace, and S. Kantor. Did new deal grant programs stimulate local economies? A study of federal grants and retail sales during the great depression. *Journal of Economic History*, 65(1):36–71, 2005. 15, 132, 161

M. Fiszbein, J. Lafortune, E. Lewis, and J. Tessada. Powering Up Productivity: The Effects of Electrification on U.S. Manufacturing. Technical Report w28076, National Bureau of Economic Research, Nov. 2020. 16, 18, 59

A. D. Foster and M. R. Rosenzweig. Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, 103(6):1176–1209, 1995. 133

A. D. Foster and M. R. Rosenzweig. Agricultural productivity growth , rural economic diversity, and economic reforms: India 1970-2000. *Economic Development and Cultural Change*, 52(3):509–542, 2004. 87

R. Fouquet and P. J. Pearson. Seven Centuries of Energy Services: The Price and Use of Light in the United Kingdom (1300-2000). *The Energy Journal*, 27(1), 2006.

17

P. Gaggl, R. Gray, I. Marinescu, and M. Morin. Does Electricity Drive Structural Transformation? Evidence from the United States. *Labour Economics*, 68, 2021. 15, 59, 65, 119

B. L. Gardner. *American agriculture in the twentieth century.* Harvard University Press, 2006. 94, 101

G. A. Gbetibouo, R. M. Hassan, and C. Ringler. Modelling farmers' adaptation strategies for climate change and variability: The case of the Limpopo Basin, South Africa. *Agrekon*, 49(2):217–234, June 2010. 126

H. B. Gray, V. Taraz, and S. D. Halliday. The impact of weather shocks on employment outcomes: evidence from south africa. *Environment and Development Economics*, 28(3), 2022. 187

M. Greenstone, R. Hornbeck, and E. Moretti. Identifying agglomeration spillovers: evidence from million dollar plants. Publisher: NBER Series: Working paper series, 2008. 86, 88

Z. Griliches. Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica*, 25(4):501–522, 1957. 126, 133

L. Grogan and A. Sadanand. Rural electrification and employment in poor countries: Evidence from nicaragua. *World Development*, 43:252–265, 2013. 65

M. P. Gutmann, D. Brown, A. R. Cunningham, J. Dykes, S. H. Leonard, J. Little, J. Mikecz, P. W. Rhode, S. Spielman, and K. M. Sylvester. Migration in the 1930s:

Beyond the dust bowl. *Social Science History*, 40(4):707–740, 2016. 126

M. Haines. Historical, demographic, economic, and social data: The united states, 1790-2002; ICPSR 2896, 2018. 90

M. Haines, P. Fishback, and P. Rhode. United states agriculture data , 1840 - 2010; ICPSR 35206., 2018. 90

T. Hegghammer. OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment. *Journal of Computational Social Science*, 5(1): 861–882, May 2022. 35, 36

J. V. Henderson, T. Squires, A. Storeygard, and D. Weil. The Global Distribution of Economic Activity: Nature, History, and the Role of Trade1. *The Quarterly Journal of Economics*, 133(1):357–406, Feb. 2018. 61

A. E. Hill, I. Ornelas, and J. E. Taylor. Agricultural Labor Supply. *Annual Review of Resource Economics*, 13(1):39–64, Oct. 2021. 16, 127, 135

R. Hirsh. Shedding New Light on Rural Electrification: The Neglected Story of Successful Efforts to Power Up Farms in the 1920s and 1930s. *Agricultural History*, 92(3):296–327, July 2018. 124

E. Hisali, P. Birungi, and F. Buyinza. Adaptation to climate change in Uganda: Evidence from micro level data. *Global Environmental Change*, 21(4):1245–1261, Oct. 2011. 126

R. Hornbeck. The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe. *American Economic Review*, 102(4):1477–1507, 2012. 13, 14, 16, 87, 88, 132, 133, 161, 174

R. Hornbeck. Dust Bowl Migrants: Environmental Refugees and Economic Adaptation. Unpublished Working Paper, 2020. 132, 137, 174, 176

T. P. Hughes. *Networks of Power. Electrification in Western Society 1880-1930.* The John Hopkins University Press, 1993. 10, 67, 68, 73, 75

International Energy Agency (IEA), International Renewable Energy Agency (IRENA), United Nations Statistics Division, The World Bank, and World Health Organization. Tracking SDG7: The energy progress report 2020. Technical report, International Energy Agency, 2020. 64

IPCC. Assessment report 6 climate change 2021: The physical science basis, 2021. 125

J. Jagermeyr, C. Muller, A. C. Ruane, and J. Elliott. Climate change signal in global agriculture emerges earlier in new generation of climate and crop models, Aug. 2021. 125

S. Jayachandran. Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries. *Journal of Political Economy*, 114(3):538–575, June 2006. 129

K. Jessoe, D. T. Manning, and J. E. Taylor. Climate Change and Labour Allocation in Rural Mexico: Evidence from Annual Fluctuations in Weather. *The Economic Journal*, 128(608):230–261, Feb. 2018. 16, 127, 130, 136, 146, 147, 160, 174, 184

S. R. Khandker, D. F. Barnes, H. Samad, and N. H. Minh. Welfare impacts of rural electrification. Evidence from vietnam. Number: September Publication title: World bank impact evaluation series, 2009. 65

C. Kitchens and P. Fishback. Flip the switch: The impact of the rural electrification administration 1935-1940. *Journal of Economic History*, 75(4):1161–1195, 2015. 15, 18, 59, 65, 66, 67, 68, 86, 88, 90, 96, 100, 119

P. Kline and E. Moretti. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *Quarterly journal of economics*, 129(1):275–331, 2014. 15, 58, 65, 67, 68, 86, 100, 119

J. Lafortune, E. Lewis, and J. Tessada. People and Machines: A Look at the Evolving Relationship between Capital and Skill in Manufacturing, 1860–1930, Using Immigration Shocks. *The Review of Economics and Statistics*, 101(1):30–43, Mar. 2019. 14

S. Lebergott. *The American Economy: Income, Wealth and Want*. Princeton Legacy Library. Princeton University Press, 1976. 23

K. Lee, E. Miguel, and C. Wolfram. Does Household Electrification Supercharge Economic Development? *Journal of Economic Perspectives*, 34(1):122–144, Feb. 2020a. 67, 120, 121, 185

K. Lee, E. Miguel, and C. Wolfram. Experimental evidence on the economics of rural electrification. *Journal of Political Economy*, 128(4):1523–1565, 2020b. 65, 67, 121, 186

S. H. Leonard, G. D. Deane, and M. P. Gutmann. Household and farm transitions in environmental context. *Population and Environment*, 32(4):287–317, 2011. doi:

10. 136

J. Lewis. *The impact of technological change within the home*. PhD thesis, University of Toronto, 2014. 65, 90

J. Lewis. Infant Health, Women's Fertility, and Rural Electrification in the United States, 1930-1960. *The Journal of Economic History*, 78(1):118–154, Mar. 2018. 18, 59, 67, 68, 90, 91, 94, 119, 124

J. Lewis and E. Severnini. Short- and long-run impacts of rural electrification: Evidence from the historical rollout of the U.S . Power grid. Number: 11243 Publisher: IZA Institute of Labor Economics, 2017. 15, 18, 59, 65, 66, 67, 68, 87, 90, 94, 96, 100, 119

M. Lipscomb, A. Mushfiq Mobarak, and T. Barham. Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil. *American Economic Journal: Applied Economics*, 5(2):200–231, 2013. 65, 67, 122, 185

M. Liu, V. Taraz, and Y. Shamdasani. Climate change and labor reallocation: Evidence from six decades of the indian census. 2021. 184, 187

J. Long and H. E. Siu. Refugees From Dust and Shrinking Land: Tracking the Dust Bowl Migrants. *The Journal of Economic History*, 78(4):51, 2008. 16, 132, 137

R. S. Lynd and H. M. Lynd. *Middletown: A Study in Modern American Culture*. Mariner Books, 1929. 21

A. Markwart. Power in California. *Journal of the Franklin Institute*, 204(2):145–192, Aug. 1927. 75, 78

A. H. Markwart. Mingled hydro and steam power production in california - past, present, and future. *Electrical Engineering*, 59, 1940. Publisher: American Institute of Electrical Engineers. 75

S. Michalopoulos and E. Papaioannou. Spatial Patterns of Development: A Meso Approach. *Annual Review of Economics*, 10(1):383–410, 2018. 61

E. Miguel, S. Satyanath, and E. Sergenti. Economic Shocks and Civil Conflict: An Instrumental Variables Approach. *Journal of Political Economy*, 112(4):725–753, 2004. 129

D. Monroe. Patterns of living of farm families. Technical report, U.S. Deptartment of Agriculture, 1940. 136, 186

E. Moretti. Local labor markets. In *Handbook of labour economics*, volume 4b, pages 1237–1313. Elsevier, 2011. 86, 187

B. Moses. Electrical statistics for california farms. Publisher: Universoty of California, College of Agriculture, Agricultural Experiment Station, 1929. 84

N. Z. Muller. On the divergence between fuel and service prices: The importance of technological change and diffusion in an American frontier economy. *Explorations in Economic History*, 60:93–111, Apr. 2016. 17

K. M. Murphy, A. Shleifer, and R. W. Vishny. Industrialization and the big push. *Journal of Political Economy*, 97(5), 1989. 123

N. C. f. E. I. NOAA. Global Historical Climatology Network daily (GHCNd), 2021.

143, 144, 145

D. E. Nye. *Electrifying America; Social Meanings of a New Technology, 1880-1940*. MIT Press, Cambridge, 1997. 20, 22, 26, 55, 67, 68, 84, 120

U. B. of the Census. Historical Statistics of the united States Colonial Times to 1970 - Volume 1, 1976. 23, 24

A. L. Olmstead and P. Rhode. An overview of california agricultural mechanization, 1870-1930. *Agricultural History*, 62(3):86–112, 1988. 76, 84

A. Ortiz-Bobea. Climate, agriculture and food. In *Handbook of agricultural economics*. Elsevier, 2021. 147

H. C. Passer. *The electrical manufacturers, 1875-1900 : a study in competition, entrepreneurship, technical change, and economic growth.* Harvard University Press, 1953. 20, 21

M. Pinkovskiy and X. Sala-i Martin. Lights, Camera,... Income!: Estimating Poverty Using National Accounts, Survey Means, and Lights. Working Paper 19831, National Bureau of Economic Research, Jan. 2014. 61

D. J. Pisani. *Water and American Government*. University of California Press, 2002. 84

J. Powell. Report on the lands of the arid region of the united states. Technical report, U.S. Geographical and Geological Survey of the Rocky Mountain Region, 1879. 77

R. Ramcharan. Inequality and Redistribution: Evidence from U.S. Counties and

States, 1890–1930. *Review of Economics and Statistics*, 92(4):729–744, Nov. 2010. 156, 162

M. Ravallion. Expected Poverty Under Risk-Induced Welfare Variability. *The Economic Journal*, 98(393):1171, Dec. 1988. 136

J. Roback. Wages, Rents, and the Quality of Life. *The Journal of Political Economy*, 90(6):1257–1278, 1982. 85, 134

M. J. Roberts and W. Schlenker. The evolution of heat tolerance of corn: Implications for climate change. In *The economics of climate change: Adaptations past and present*. University of Chicago Press, 2011. 147, 149, 155

E. Rose. Ex ante and ex post labor supply response to risk in a low-income area. *Journal of Development Economics*, 64(2):371–388, Apr. 2001. 16, 127, 136, 187

S. Rosen. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–35, 1974. 85, 134

P. N. Rosenstein-Rodan. Notes on the theory of the big push. In H. S. Ellis and H. C. Wallich, editors, *Economic development for latin america*, volume 44. Stockton Press, 1961. 123

S. Ruggles, S. Flood, R. Goeken, E. Meyer, J. Pascas, and M. Sobek. IPUMS USA: Version 10.0 [dataset], 2020. 127

S. Ruggles, S. Flood, S. Foster, R. Goeken, J. Pacas, M. Schouweiler, and M. Sobek. IPUMS USA: Version 11.0 [dataset], 2021. Place: Minneapolis, MN. 138, 139, 145

W. Schlenker and M. J. Roberts. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 106(37):15594–15598, 2009. 125, 147, 148

S. D. Schubert, M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister. On the Cause of the 1930s Dust Bowl. *Science*, 303(5665):1855–1859, Mar. 2004. 131

R. Seager, Y. Kushnir, M. Ting, M. Cane, N. Naik, and J. Miller. Would Advance Knowledge of 1930s SSTs Have Allowed Prediction of the Dust Bowl Drought?*. *Journal of Climate*, 21(13):3261–3281, July 2008. 126

R. Seager, N. Lis, J. Feldman, M. Ting, A. Park Williams, J. Nakamura, H. Liu, and N. Henderson. Whither the 100th meridian? The once and future physical and human geography of America's arid–humid divide. Part i: The story so far. *Earth Interactions*, 22(5), 2018. 77

H. E. Selby. The importance of irrigation in the economy of the west. *Journal of Farm Economics*, 31(4):955–964, 1949. 75, 84

E. Severnini. The power of hydroelectric dams: Agglomeration spillovers. Number: 8082 Pages: 1–70 Publisher: IZA Series: Discussion paper series, 2014. 18, 68, 90

C. Sichko. Migrant Selection and Sorting during the Great American Drought. Unpublished Working Paper, 2020. 126, 132, 140

C. T. Sichko. *Migration During the great American Drought*. PhD thesis, Vanderbilt University, 2021. 175

I. Singh, L. Squire, and J. Strauss. A Survey of Agricultural Household Models:

Recent Findings and Policy Implications. *The World Bank Economic Review*, 1 (1):149–179, 1986a. 134, 136

I. Singh, L. Squire, J. Strauss, and World Bank. *Agricultural household models : extensions, applications, and policy*. The World Bank, 1986b. 135, 136, 186

C. C. Spence. Early uses of electricity in american agriculture. *Technology and Culture*, 3(2):142–160, 1962. 101

D. I. Stern, P. J. Burke, and S. B. Bruns. The impact of electricity on economic development : A macroeconomic perspective. *International Review of Environmental and Resource Economics*, 12(1):85–127, 2018. 65

R. Sutch. The Impact of the 1936 Corn Belt Drought on American Farmers' Adoption of Hybrid Corn. In *The Economics of Climate Change: Adaptations Past and Present*. University of Chicago Press, 2011. 126, 133, 149, 150

D. o. C. U.S. Bureau of the Census. Census of Electrical Industries, 1907, 1907. 18, 29, 42

D. o. C. U.S. Bureau of the Census. Census of Electrical Industries, 1912, 1912. 18, 29

D. o. C. U.S. Bureau of the Census. Census of Electrical Industries, 1917, 1917. 18, 29

D. Vidart. Human Capital, Female Employment, and Electricity: Evidence from the Early 20th Century United States. Unpublished Working Paper, 2020. 18, 59, 124

J. C. Williams. Otherwise a mere clod: California rural electrification. *IEEE Tech-*

nology and Society Magazine, 7(4):13–19, 1988. 76, 84

J. M. Wooldridge. Difference-in-differences estimation, 2007. Pages: 1–19. 66, 96

E. World. Electrical World - Vol. 37, 1901. 71

E. World. Electrical World - Vol. 59, 1912. 18, 76, 79, 101

E. World. Electrical World - Vol. 68, 1916. 77

E. World. Electrical World - Vol. 75, 1920. 18, 83, 101, 120