Public Data Infrastructure and Agricultural Development: Historical Evidence from Soil Mapping

Yiyu Xing*

October 15, 2025

Abstract

This paper investigates how public data infrastructure shapes local economic development, focusing on a large-scale soil mapping program in US history. Since 1899, the National Cooperative Soil Survey (NCSS) was implemented to offer county-level soil maps and classification data. Leveraging newly digitized dataset and the staggered rollout of these maps across counties, I find a significant, persistent increase in local agricultural productivity by 5%, as well as a rising level of monoculture farming. I further provide evidence for a primary channel: by reducing information barriers, soil maps incentivized farmers to adopt fertilizers and machinery, improve the precision of crop choices, expand self-owned farmland, and increase skilled labor input. I also rule out alternative explanations such as local innovation spillovers or aggregate market-based signals induced by these maps. My results thus highlight the tradeoff between productivity growth and sustainability associated with access to uniform technical knowledge.

*Xing: Integrated Substance Use and Addiction Programs (ISAP), University of California, Los Angeles. E-mail: sarrusxing10@g.ucla.edu. I am indebted to my advisor Nicolas L. Ziebarth for always providing incredible feedback, inspiring discussions, and generous support. I am also grateful to Martha Bailey, Hoyt Bleakley, Dora Costa, Michela Giorcelli, and Chris Vickers for sharing thoughtful comments and suggestions. I also thank the participants and discussants from the UCLA Economic History Workshop and the 2025 Economic History Association conference. All errors are my own.

"Each soil has had its own history. Like a river, a mountain, a forest, or any natural thing, its present condition is due to the influences of many things and events of the past..."

— Charles Kellogg, *The Soils That Support Us*, 1956

1 Introduction

How does public data infrastructure shape local economic development and production? Throughout the history, a key function of government is the collection and dissemination of data,[1] in order to modernize national statistical system and promote evidence-based decision making.[2] Such data infrastructure, defined as "*the informational system bedrock that supports the nation*," typically consists of a large network of statistical agencies, units, and programs that collaborate to gather, share, and report information about the economy. Moreover, it is likely a critical adaptive tool in response to economic, public health, or environmental crises.[3] Despite broad recognition of its importance, empirical evidence on the long-run economic consequences of public data infrastructure remains limited.

To addresses this gap, I study the effects of a historical, large-scale soil mapping program in the United States, the *National Cooperative Soil Survey* (NCSS). For the first time, the federal government coordinated with state and local agencies to make detailed soil quality information widely accessible to farmers, researchers, and policymakers, aiming to promote the productive use of soil for agricultural development. This government-led effort provides a unique empirical setting: although economists have extensively studied the causes of agricultural development since the 19^{th} century (e.g., Olmstead and Rhode, 2001; Hornbeck, 2010; Kitchens and Fishback, 2015; Mattheis and Raz, 2019; Smith, 2019; Chan, 2022), the role of *access to information and knowledge* has remained largely unexplored.

Using newly compiled data on the timing and location of NCSS soil surveys, I exploit the staggered roll-out of county-level NCSS mapping programs to identify their causal effects. Specifically, I adopt a stylized staggered difference-in-differences (DD) framework and compile a rich set of data sources to test various hypotheses. I begin by evaluating the direct

[1] This is a responsibility claimed in *Article 1, Section 2 of the U.S. Constitution* (Fox, 2018). In 1787, the Framers of the U.S. Constitution mandated a decennial census of the population to apportion the number of seats for each state in the House of Representatives. Located in *Article I, Section 2* of the Constitution, the Census Clause 3 directs Congress to create laws to facilitate an actual enumeration (Fox, 2018).

[2] See https://fas.org/publication/increasing-responsible-data-sharing-capacity-throughout-government/, last accessed October 15, 2025.

[3] See https://www.amstat.org/policy-and-advocacy/US-Data-Infrastructure, last accessed October 15, 2025.

impact of NCSS mapping on local agricultural productivity, measured by total crop revenue per acre of farmland at the county level. Using detailed county-by-year data from the U.S. Census of Agriculture (1890–1930), I find that access to NCSS mapping programs leads to a statistically significant increase in crop revenue per acre by at least 4%. Notably, this productivity gain is not transitory but persists for at least 15 years following the initial mapping. These results are robust to a range of specifications, including focusing on the adjacent county pairs, incorporating other county-level confounding factors (especially local market access), etc.

In the meantime, I examine the potential costs associated with the provision of soil quality information, including whether such information may induce the concentration of highly suitable crops within a county and, in turn, increase the long-run risk of soil erosion. In the first exercise, I construct two measures of crop concentration: the Shannon-Wiener Index (SWI) and the Herfindahl–Hirschman Index (HHI). I document evidence that access to NCSS mapping programs is associated with a statistically significant in both indexes, suggesting that farmers increasingly specialize in crops most suited to their mapped soil types at the expense of reducing crop diversity. Furthermore, I also check the correlation between the NCSS mapping and a county's erosion status during the 1930s. Using a cross-sectional regression, I find that counties with prior exposure to soil mapping were more likely to experience soil erosion and, on average, had a higher share of eroded land. Both of these findings point to a potential unintended consequence of soil mapping program: while the provision of soil quality information may improve productivity in the short run, it can also generate long-term, negative environmental spillovers by encouraging monocultural farming.

I then turn to empirically test the potential channels driving the observed changes in agricultural development. First, I argue that the NCSS mapping program facilitated the diffusion of more precise and localized soil quality information, thereby reducing uncertainty faced by farmers in making land-use and input decisions. Specifically, I document evidence that the mapping program significantly improved local crop selection, encouraged the adoption of agricultural machinery and fertilizer, increased the share of self-owned farmland, and led to a rise in the density of local farmers.

I also supplement both qualitative and quantitative evidence for the learning effects through this soil mapping program. First, historical reports suggest that the NCSS maps were widely disseminated and popular among key agricultural stakeholders. Second, I show that the effects of the program were much stronger in counties with lower levels of pre-existing agricultural knowledge, specifically, those with fewer local farmers and those lacking agricultural societies or county fairs. These patterns are consistent with a mechanism in which better informa-

tion enabled farmers to make more informed and efficient production decisions, ultimately reshaping the local agricultural landscape.

The second mechanism I examine is whether local innovation spillovers induced by the mapping program primarily drove the productivity gain. By providing a standardized, consistent dataset on local soil conditions, the program could facilitate more effective collaboration among farmers, extension agents, experiment stations, and local planners. This, in turn, may enable better-aligned production decisions, targeted research agendas, and more coherent agricultural development strategies at the local level. To test this channel, I digitize historical annual reports from state agricultural experiment stations and construct a state-year panel tracking the number of publications focused on soil quality and fertilizer topics. Following the implementation of soil mapping, there was a significant increase in the intensity of soil-relevant research. However, such induced, new knowledge did not diffuse substantially among farmers. Thus, the observed productivity gains were not primarily driven by the creation of novel agricultural technologies or discoveries, but rather by farmers' improved ability to apply *existing* knowledge once soil information became standardized and locally accessible.

Another alternative mechanism that could explain the observed changes in local agricultural development is the NCSS mapping program triggered broader market-wide fluctuations (particularly in crop prices), which may have indirectly influenced farmers' production decisions. Specifically, the widespread dissemination of detailed soil quality information could alter regional or national supply patterns, thereby affecting crop prices. These price effects, in turn, would influence local agricultural choices independently of any direct learning or planning benefits associated with the mapping. To investigate this market-based channel, I follow the approach of Kitchens and Rodgers (2023) and construct a county-level measure of exposure to national price fluctuations. After including this price proxy in the specification, I find no evidence that my main results largely changed, suggesting that aggregate market dynamics are unlikely to be the primary driver of the observed effects.

Eventually, I discuss whether the diffusion of soil information acts through co-ethnic networks, particularly among immigrant communities. These networks might have played a key role in facilitating the interpretation and adoption of technical soil data, especially in areas where language barriers or the lack of pre-existing specialization might otherwise hinder learning (Raz, 2025). To test this channel, I use the share of immigrant farmers at the county level as a proxy for the presence of co-ethnic agricultural networks. However, I do not find evidence that the effects of NCSS mapping were significantly stronger in counties with higher immigrant concentrations, suggesting that co-ethnic diffusion may not have been a dominant mechanism in this context.

I further investigate how data infrastructure interacted with physical infrastructure (e.g., railroads) during this period. The provision of localized soil information may have been especially productive when complemented by access to transportation networks, which lowered trade costs and enabled farmers to better respond to new soil quality information. To test this interaction, I employ a triple difference-in-differences specification that exploits variation in both mapping exposure and proximity to local markets. The results indicate that the impacts of NCSS mapping were primarily concentrated in counties with better access to transportation infrastructure, consistent with the view that data and physical infrastructure are complementary in shaping agricultural development.

I employ a range of strategies to establish the credibility and robustness of my baseline identification. Specifically, I conduct a series of balancing tests, implement a synthetic control approach, restrict the sample to adjacent county pairs, exclude never-treated counties, and apply a double lasso selection procedure to address potential selection into NCSS mapping. In addition, I control for a variety of confounding factors, including proximity to agricultural experiment stations and weather stations, as well as proxies for local natural endowments and environmental conditions. To address concerns related to the staggered difference-in-differences design, I adopt the two-way fixed effects estimator proposed by Callaway and Sant'Anna (2021) and demonstrate that my main findings remain robust under this alternative specification.

This paper makes contributions to several strands of literature. First, my work directly speaks to the literature discussing the effects of information access on productivity, especially in the context of agriculture.[4] Most relevantly, Kantor and Whalley (2019) emphasize the local nature of spillovers from agricultural experiment stations on agricultural productivity in the late nineteenth century. Their findings also suggest a reduction in the value of information diffusion through interactions in close proximity between farmers and researchers, as new technologies and other institutions emerged in the twentieth century. Meanwhile, Novaes de Amorim (2023) uses variation in exposure to radio due to topography, showing that the broadcasting of local farm programming led to an increase in the productivity of land used in agriculture that persisted for at least two decades. Nagaraj and Tranchero (2023) also provide new evidence showing the impact of access to confidential microdata through the progressive geographic expansion of the U.S. Census Bureau's Federal Statistical Research Data Centers (FSRDCs).

In particular, my work provides a key distinct contribution compared to prior literature:

[4]For a summary of relevant paper in the agricultural sector, see Suri and Udry (2022).

Kantor and Whalley (2019) highlight the role of proximity to *innovation* in driving up local agricultural productivity.[5] In contrast, I focus on exploring how reducing information barriers in soil quality (via providing soil maps) helps spur local productivity. My findings show that the primary channel is that local farmers can learn *existing* knowledge, not *new* technological advancement due to the mapping. My paper adds novel evidence to this literature by demonstrating that public access to detailed soil quality information through a government-led data sharing initiative also played a critical role in promoting agricultural productivity.

Second, my work also add value to a growing body of papers addressing the economic impacts of maps (Nagaraj and Stern, 2020). Recent flurry of work across disparate subfields as begun to investigate the impact of satellite mapping (Casaburi and Troiano, 2015; Donaldson and Storeygard, 2016; Katona et al., 2025; Nagaraj, 2022), local business maps (Luca, Nagaraj and Subramani, 2022), subway maps (Larcom, Rauch and Willems, 2017), redlining maps (Aaronson, Hartley and Mazumder, 2021; Fishback et al., 2024; Hynsjo and Perdoni, 2024), flood insurance maps (Michel-Kerjan, 2010), and medical innovation (Williams, 2013; Kao, 2024; Kang, 2025). Leveraging the historical NCSS program, my paper offers some of the first causal evidence that mapping efforts can also impose long-run benefits and costs in the agricultural sector by guiding production decisions.

To my knowledge, empirical evidence shedding light on mapping in the agricultural sector is quite sparse. The only closest paper to mine is from Pradhan (2009), documenting the net benefits of the information provided by the soil survey program to the production of selected crops (i.e., corn) after WWII. There are several substantial distinctions between Pradhan (2009) and mine: first, in contrast to Pradhan (2009), I focus on the initial rollout of the National Cooperative Soil Survey (NCSS), which provides a cleaner empirical setting for identifying the program's causal effects. Second, while Pradhan (2009) examines correlations between the presence of NCSS mapping and local corn yields, I employ a variety of identification strategies to estimate causal impacts more rigorously. Finally, and most importantly, Pradhan (2009) does not investigate the underlying economic mechanisms driving the observed benefits, nor does it consider potential unintended consequences or negative effects associated with soil mapping.

Further, this paper broadly contributes to the discussion of how local institutions affect productivity, innovation activities, and economic development. Prior research heavily focuses on

[5]Note Kantor and Whalley (2019) show the positive and persistent total factor production (TFP) effect associated with agricultural experiment stations. That being said, even controlling for key input decisions (i.e., equipment, fertilizer, land, and labor), one would expect a productivity gain stems from being close to agricultural innovation conducted at local agricultural experiment stations.

the role of public infrastructure investment. For example, Hornbeck (2010) examines the impact on agricultural development of the introduction of barbed wire fencing to the American Plains in the late nineteenth century. Chan (2022) examine how market access affects farming output and whether changes to output were driven by increasing concentration in production. Edwards and Thurman (2025) document the importance of a key institutional innovation, the drainage management district, which increased the land value of naturally wet eastern U.S. counties by 20–37 percent. Furman and MacGarvie (2007), Kantor and Whalley (2014), Hausman (2022), and Andrews (2023) all consistently find that with the establishment of a college, local innovative activity increases, although they tend to disagree on the channels through which this effect operates, as well as its magnitude. Furman, Nagler and Watzinger (2021), Karger (2021), Berkes and Nencka (2024), and Gilpin, Karger and Nencka (2024) estimate the effects of public libraries on educational attainment, occupational mobility, and innovation.

My work extends this line of inquiry by focusing on a different form of publicly-provided infrastructure: open data for key inputs. Rather than constructing new physical infrastructure or institutions, the NCSS disseminated detailed soil information to local communities, thereby lowering barriers to learning. It highlights how access to technical knowledge can substitute for or complement other forms of institutional investment in driving long-run productivity gains.

The rest of this paper proceeds as follows: section 2 provides the institutional background. Section 3 presents a simple conceptual framework to guide empirical analysis. Section 4 specified the primary empirical strategies. Section 5 describes a number of datasets enabling to me to implement the empirical analyses; section 6 presents the main results. Section 7 describes and tests relevant mechanisms. Section 8 shows how soil mapping program interacted with physical infrastructure. Section 9 shows how I leverage various strategies to demonstrate the credibility of my empirical strategies. Section 10 concludes and sheds light on some future work.

2 Historical Background

In this section, I briefly discuss the institutional details regarding the National Cooperative Soil Survey (NCSS) and other relevant soil service programs in the early 20th century. Section 2.1 summaries prior state-level soil and geological surveys, Section 2.2 describes the development of the NCSS programs, Section 2.3 discusses the specific contents of NCSS publications, and Section 2.4 shows who would typically use the publication of NCSS programs.

2.1 Prior Soil and Geological Maps

State soil mapping programs and soil surveys were first published in the US in the early 1880s prior to the development of the NCSS (Bockheim and Gennadiyev, 2015). Specially, 11 of the eventual 48 states prepared general soil maps between 1880 and 1927.[6] In particular, six of them were based on physiographic provinces (IN, MS, MO, OR, and TN), two were based on physical soil properties, especially texture (CA, WI), and three showed soil types (DE, MO, and SD). Yet, state soil maps traditionally have been considered to be less important that the larger-scale county maps. Whereas state soil maps have lacked a unified methodology, county maps have had a relatively consistent methodology and a higher level of agricultural value.

Prior to the launch of the NCSS mapping program, 35 states had already conducted their own geological surveys (Figure A38). Unlike the NCSS soil maps and surveys, which emphasized the suitability of soils for different types of crops, these earlier geological surveys primarily focused on the physical and chemical properties of soil (see Figure A40). As such, they offered more descriptive assessments of soil composition rather than actionable guidance for agricultural production or land use decisions. Therefore, they are not as informative as NCSS maps in terms of guiding local farmers.

2.2 Why Scientific Soil Knowledge is Sparse?

It is crucial to mention why compared to other forms of agricultural knowledge (e.g., weather, natural disasters, specialization of crops, etc.), acquiring scientific soil knowledge remains challenging before the launch of these mapping programs. In particular, the barrier to understanding soil quality was substantially higher than that for more observable agricultural information such as temperature, precipitation, or water conditions.

First of all, unlike other types of agricultural knowledge that could be learned through direct observation or accumulated experience, scientific soil knowledge required systematic investigation. Much of it derived from orderly and refined laboratory analyses, the development and testing of complex theoretical frameworks, and the adaptation of principles and methods from related scientific disciplines Hartemink (2021). Moreover, it depends critically on extensive field surveys and precise classification efforts across diverse landscapes. While farmers might gain some intuitive understanding of soils through experience, most substantive advances in soil knowledge could only be achieved through coordinated, scientific studies.

[6]These eleven state-level maps include Louisiana (1880), Wisconsin (1882), California (1883), Tennessee (1896), North Dakota (1902), Delaware (1903), Missouri (1904, 1918), Indiana (1909), South Dakota (1909, 1911), Mississippi (1916), and Oregon (1920). This data comes from Bockheim and Gennadiyev (2015).

Second, tracing the historical development of soil science in the United States reveals that, prior to the nationwide mapping effort, even the basic terminology for describing soil properties was underdeveloped, and only a handful of researchers were engaged in formal study (Hartemink, 2021). By the late nineteenth century, soil surveys had experienced more redirections, scientific missteps, and dogmatic detours than perhaps any other field of agricultural investigation. This scientific immaturity underscores how exceptional the subsequent institutionalization of systematic soil mapping was in transforming a fragmented body of tacit observations into a coherent scientific discipline.

Third, descriptive evidence further suggests that although humans had long learned to use soils empirically, they had not yet managed to comprehend their underlying patterns and complexity (Hartemink, 2021). In other words, practical familiarity with soils preceded scientific understanding: farmers could recognize fertile land or observe crop responses, but the mechanisms governing soil formation, structure, and chemical composition remained largely obscure prior to systematic investigation.

2.3 Development of National Cooperative Soil Survey

Origin of NCSS Mapping. Among the many basic and applied sciences to agriculture is soil science, especially those relating to the genesis, morphology, classification, behavior, and mapping of soils. Collectively, these several fields may be termed, simply "soil survey." The scientific and technical evolution of these fields during the nineteenth century is the subject matter of the first portion of the present study, and thus constitutes a history, not of soil science, but of the origins of "soil survey" in the United States.

Parallel to the general growth of basic and applied sciences, has been the expanded role of science in governmental programs dealing with the problems of agriculture. From the pioneering work in a few states and in the Patent Office of the federal government a century ago, have emerged the many present-day activities of federal and state agencies involving the application of science to agriculture. Publicly supported activities in the collective field of "soil survey", have been carried on principally through the program of the "National Cooperative Soil Survey," a joint federal and state undertaking begun in 1899.

Key administrative steps leading up to the actual establishment of the NCSS as a functioning cooperative program followed a related course beginning at Washington, D. C. in August of 1891, for the fifth annual convention of the Association of American Agricultural Colleges and Experiment Stations. Mark W. Harrington, newly appointed Chief of the U.S. Weather Bureau (Transferred to Department of Agriculture from Department of War on July 1, 1891),

recommended that the colleges

> "*...add some special meteorological investigations such, for example, as the relation of forests to climate, reactions of the soil and climate, the influence of climate upon vegetation, ...and similar topics.*"

In his annual report for 1891, Harrington also reported that

> "*This new field of investigation thus outlined appears so important and so full of promise of results of great practical value to agriculture, if the work is systematically arranged and carried out that I think it would be well to establish a division in this Bureau devoted to the subject of meteorology in its relations to agricultural soils (111):585.*"

Legislative authorization was provided by the fifty-second Congress in the Agricultural Appropriation Act of 1894 (approved March 3, 1893). Under the subappropriation of the Weather Bureau, this Act contained the authority as "*... investigations on the relations of climate to organic life ...*" In February of 1894, the Division of Division of Agricultural Soils was established. In the first few years, the role of the Division was primarily administrative. It is the Agriculture Appropriation Act for 1897 that makes the unit become the Division of Soils and start to systematically conduct nationwide soil surveys.

In May 1899, an annual appropriation of $16,000 was granted for an investigation of the relation of soils to climate and organic life, as well as an investigation of the texture and composition of soils. It marked the official establishment of the National Cooperative Soil Survey. The area mapped in 1899 totaled 720,000 acres of four separated soil surveys in Cecil County, Maryland, the Connecticut Valley, the Salt Lake Valley of Utah, and the Pecos Valley, New Mexico. The statement of policy, which formed the basis of the cooperative structure of the soil survey program, was claimed as

> *Wherever possible it has been the policy of the division to work in close cooperation with the state experiment stations, the state geological surveys, boards of agriculture, or other local institutions. This cooperation, so far as it has been carried, has proved mutually satisfactory and beneficial to the local institutions and to this division. In a general way it relieves the local institutions of all responsibility of the direction of the work and the preparation of the material, while the expense to them is very much less in every case than if they had to employ competent soil experts, even if such experts could be obtained, which is questionable.*

Throughout the early twentieth century, the NCSS program has aroused considerable popular-

ity and public interest. During the 1900s and 1910s, cooperation between the federal Bureau and agencies of the participating states was established from the start, and nearly all soil surveys were carried on cooperatively. Systematic procedures, principles, and techniques of soil survey interpretation, including the application of soil survey data to general soil management and to soil erosion control, were gradually evolved and applied. In particular, these early soil mapping programs developed *soil classifications*,[7] technical refinements in soil characterization and soil mapping,[8] improvements in the quality and detail of the interpretive information on the basis of agricultural value.[9], and soil management tools.

During the 1920s and the early 1930s, the NCSS program experienced several changes in terms of political setting, financial status, economic environment, and many others. First, this period coincided with a succession of Republican administrations. The demand for a return to "normalcy" led to a relaxation of public responsibilities in policy relating to natural resources. Second, the national agricultural economy was beset by an extended depression. Due to the demands of WWI, the agricultural sector experienced fluctuations, especially the price of wheat. Along with other factors (e.g., weather conditions, nonfarm price levels, etc), a sharp and prolonged break in farm priced when worldwide demand diminished after 1920. Third (and maybe most remarkably), the primary goal of soil surveys shifted from "agricultural development" to "soil erosion control." The deepening political and economic crises after 1929 brought forth and unprecedented demand for soil survey data on erosion. The emergence of the *Soil Conservation Service* as significant factors in policy and planning occurred after 1933.

In the two decades of the New Deal and Fair Deal,the Soil Survey was profoundly affected by the public attitudes, the political philosophy, and the federal administrative structures which emerged from the national election of 1932. In 1952, the National Cooperative Soil Survey and the Soil Conservation Service were integrated together. Among the vase political changes, by the New Deal were those involving public attitudes toward the nature and extent of the federal government's responsibility to individual farmers. The earlier "Conservation Movemen" and the national land policies which developed before 1933 dealt primarily with nonagricultural lands, and emphasized public ownership as the remedy for misuse of land.

It is worth noting that there are no traceable records detailing how counties or areas were

[7]For example, the soil types recognized in the first wave of NCSS include Jordan sandy loam (from Utah), Pecos sand (from New Mexico), and Podunk fine sandy loam (from the Connecticut Valley).

[8]

[9]As claimed by Whitney, "... the Bureau has never attempted to classify the soils according to their current rank as factors in our agricultural output, but rather to weigh them on the basis of their potentiality, no matter whether they were being used to their full capacity, were lying idle, or were still in their virgin state."

selected for inclusion in the mapping program. Consequently, it is infeasible to exploit administrative selection rules or predetermined eligibility criteria as in other studies on place-based policies (see Kline and Moretti, 2013; Karger, 2021; Andrews, 2023; Berkes and Nencka, 2024). Instead, I implement a series of complementary identification strategies to establish the robustness and credibility of the baseline difference-in-differences design. Details will be discussed in Section 9.

Types of NCSS Mapping. There were three main categories of NCSS mapping programs during the early 20th century. The first was the detailed soil survey conducted at a scale of 1 inch to 1 mile, which remained the core output of the NCSS during this period. The second type, initiated around 1906, involved very large-scale mapping of small areas where field experimental work was either ongoing or planned. These specialized surveys are further discussed under "Soil Survey Interpretation." One such example appears in Appendix Table A41. The third category, known as the reconnaissance soil survey, represented the opposite extreme in terms of detail and scale. These surveys were designed to provide general information on soils and agricultural potential over broad areas. They were conducted rapidly and published at coarse scales ranging from 1 1/6 miles to 1 1/2 miles. Reconnaissance surveys began in 1908 and eventually became a permanent and important component of the NCSS program.

In terms of identification, a key limitation of the latter two types is that they were conducted on areas rather than full counties. As a result, the county border often does not align with the boundaries of these surveys, complicating efforts to merge the data with other county-level historical records. This mismatch introduces challenges in seeking to construct consistent county-year panels, since it is frequently unclear which counties were fully or partially covered by a given area survey. Thus in the sample of analysis, I focus primarily on the standard 1:1 mile county-level surveys, which offer the most consistent geographic coverage and compatibility with other data.[10]

2.4 Contents of National Cooperative Soil Survey

A typical NCSS mapping program consisted of three key components, providing rich information on local soil characteristics and other relevant agricultural knowledge:

Soil maps: as indicated by its name, an NCSS mapping program always produced a detailed map of the county or area covered. These maps classified and visualized distinct soil types

[10]In a robustness check, I also use the alternative coding method to address the issue of partially mapped counties/areas.

across the landscape, using a standardized system of color-coding, symbols, and legends. These maps enabled users to identify variation in soil quality, composition, and distribution for different types of agricultural activities. They often incorporated topographical features and were produced at proper scales to guide agricultural decision-making. As shown in Figure 1, this is an example NCSS map taken in Wood County, Wisconsin, 1917. Such maps were included as *advanced sheets* and could be purchased separately since 1904.

Soil Descriptions: Along with the maps, there were usually comprehensive narrative descriptions of each soil type identified. These profiles include the physical and chemical properties of each soil series (e.g., texture, structure, drainage capacity, fertility, depth to bedrock, and pH levels), as well as . The descriptive content was informed by extensive field sampling and laboratory analysis. Such information was critical for understanding the agronomic potential and limitations of each soil unit. Figure 2 and Figure 3 provide examples for such descriptions regarding soil proportions and recommended crops, respectively.

Supplementary Agricultural Information: Each mapping program also provided practical guidance on land use and management tailored to local soil conditions. These recommendations covered optimal crop types, irrigation needs, fertilization strategies, and erosion control measures. In some cases, they also offered advice on pasture management, forestry use, and construction suitability. This component reflected the broader goal of the NCSS to promote sustainable land stewardship through science-based interventions.

2.5 Reaction to National Cooperative Soil Mapping

As mentioned by Gardner (1998), the general acceptance of the Soil Survey by local communities in the early days was "*implicit in the increasing volume of local requests for surveys, and in the gradual expansion of the program.*" The practical value of the Soil Survey was established and the demand for surveys increased steadily. By 1908, the Bureau of Soils had on file requests, over and above its capacity to meet, for soil surveys of six hundred and fifty-four additional areas, covering nearly three-quarters of a million square miles. This aggregate area for which soil surveys were requested amounted to almost four times the total area covered by the Soil Survey

Reflecting this growing demand, the Bureau revised its publication practices. In the earliest years, all Soil Survey materials, including maps, narrative reports, and summaries, were compiled annually into a single bound volume titled Field Operations of the Bureau of Soils. Starting in 1903, however, the Bureau introduced "advance sheets": individually issued map-report separates, distributed more flexibly via Members of Congress and Bureau officials. This

shift significantly improved the accessibility and timeliness of survey information (315):1153.

Widespread use of the NCSS maps and reports quickly emerged, with three distinct groups of consumers actively engaging with the materials:

Individual Farmers. The soil survey is meeting increasingly gratifying evidence of the extent to which the soil maps are being used by the various interests for whose benefit they are made. An inspection of this list covering the areas reported upon to July 1, 1905, shows that a great variety of interests are concerned with the results of the survey work. These demands are distributed remarkably uniform all over the United States. Large numbers of requests for the soil surveys have been distributed in response to direct applications. Especially in the case of tobacco, these maps directly went into the hands of individual farmers, who are anxious to learn whether any of the soil types adapted to the production of new crops are to be found upon their own farms.

New Settlers. The NCSS also played a key role in guiding new settlers and prospective land buyers, especially in areas recently opened to agricultural development. These users relied on soil surveys to understand local climate conditions, soil characteristics, and crop suitability. For many, the reports served as an essential planning tool to identify which kinds of farming would be most profitable in the region.

Investment Companies. Another important use of the NCSS reports is that made by investment companies, who desire information not only regarding particular plots of ground, but who also desire an account in general terms which will give them an insight into the agricultural characteristics of entire counties. The proper investment of the funds intrusted to these companies demand a knowledge not only of soil and climate, but also of the broader features of transportation, market facilities, and the general prosperity of the agricultural community.

3 Conceptual Framework

In this section, I follow Kantor and Whalley (2019) and use a simple growth model to capture the effects of accessing soil quality information on agricultural productivity. In contrast to Kantor and Whalley (2019), I highlight how the standardization and dissemination of existing knowledge *per se* can generate productivity gains.

In this setting, a region has numerous farmers that produce final output using an intermediate good (i.e., crop seeds) as an input. Having better information regarding the match between soil and seeds would improve the suitability of this intermediate good.

Production. There is a sequence of time periods $t = 1, 2, ...$ indexing the vintage of agricultural technology. People consume a single composite agricultural commodity with price normalized to one, produced by farmers endowed with a single acre of farmland. Farmers use intermediate goods to produce agricultural output Y according to the following production function:

$$Y = Ax^{\alpha}F^{\mu}$$

where A is the productivity of the seed used, x is the quantity of other factors of production used (i.e., labor, equipment, etc.) with $0 < \alpha < 1$, $0 < \mu < 0$, and $0 < \alpha + \mu < 1$. Having additional information on soil quality can scale up productivity in production as $A_n = \gamma A_o$, where $\gamma > 1$ is the size of this increase.

Knowledge. The technology farmers used in production depends on their type, whether they have high or low level of knowledge regarding the match between soil and seeds. For simplification, we assume there are two types: (1) farmers with high-level of soil-crop knowledge, $A^H = A_n$, and (2) farmers with low-level of soil-crop knowledge, $A^L = A_o$. Farmers with an initial low level of soil-crop knowledge need to acquire at least K units of know-how to be a high type.

In this context, farmers would acquire know-how from the NCSS mapping program. *Learning* requires the availability and level of knowledge carried by NCSS maps. For simplicity, we assume each NCSS map would carry a fixed amount of soil-crop information s within the corresponding county.[11] Yet, the extent of utilizing such information depends on a farmer's pre-existing agricultural knowledge, legacy, or social networking. Thus, a farmer can eventually receive $\frac{s}{c}$ units of know-how from the mapping program. c ranges from 0 to C. Intuitively, this indicates that when a farmer has already acquired sufficient amount of agricultural knowledge about local soil, the marginal value she/he can derive from the soil mapping program would be relatively low.

A local farmer's stock of know-how is then defined as

$$E(s) = [\sum_{0}^{t}(\frac{1}{1-\rho})^{t}\frac{s}{c}]^{\phi}$$

where s and c denote that within each mapped county, a local farmer could receive $\frac{s}{c}$ units of soil-crop knowledge from each NCSS mapping program. The know-how can depreciate at rate ρ with marginal effect of soil information of ϕ.

[11] We can assume this level of information depends on the supply-side factors (e.g., the expenditure and effort of field work conduced locally).

A threshold $c(s)^*$ determine the type of local farmers. When the amount of know-how carried by the map is below this threshold ($c < c(s)^*$), farmers would be low type; when the amount of soil information is beyond this threshold, local farmers are considered as high type. This threshold is

$$c(s)^* = \frac{\sum_0^t (\frac{1}{1-\rho})^t s}{K^{\frac{1}{\phi}}}$$

Within each county, the distribution of c is considered as an uniform one, between 0 and C, i.e., $s \sim U[0, C]$. As a result, the share of farmers with high knowledge is $\frac{c(s)^*}{C}$ and with low knowledge is $[1 - \frac{c(s)^*}{C}]$.

Expected Land Productivity. When the NCSS map and report are available in a county, both high know-how type and low know-how type of farmers can derive new soil-crop information. Here I did not allow for the new technology emerging as a result of mapping programs as Kantor and Whalley (2019). This is simply because as shown in Section 7, there is little empirical evidence that (1) county-level soil maps spurred local agricultural innovation, and (2) counties closer to local agricultural experiment stations would benefit more from such soil mapping programs, in terms of productivity gains.

As a result, the expected land productivity for each county is given by

$$\begin{aligned} Y &= \frac{c(s)^*}{C} Y_H + (1 - \frac{c(s)^*}{C}) Y_L \\ &= \frac{c(s)^*}{C} A_n x^\alpha F^\mu + (1 - \frac{c(s)^*}{C}) A_o x^\alpha F^\mu \\ &= [1 - (1-\gamma)\frac{c(s)^*}{C}] A_o x^\alpha F^\mu \end{aligned}$$

We can thus capture the *learning effect* of NCSS mapping program by obtaining the productivity differential ΔY_C. This is basically a comparative statistic with respect to prior agricultural knowledge of farmers:

$$\Delta Y_C = (1-\gamma)\frac{c(s)^*}{C^2} A_o x^\alpha F^\mu < 0$$

This result indicates with the launch of the NCSS mapping program, counties with less prior knowledge about soil quality would get more productivity gains. In particular, the resulting productivity gains do not stem from the adoption of newly available technologies, but from a better utilization of existing factors of production (i.e., equipment, fertilizer, and labor) and the intermediate goods (i.e., seeds).

4 Empirical Strategy

In this section, I describe two primary identification strategies used in the empirical analysis: the baseline difference-in-differences (DID) design and the adjacent-county difference-in-differences (DID) design. I also list the advantages and limitations of each.

4.1 Baseline Difference-in-Differences Design

The primary identification strategy in this analysis exploits the staggered rollout of county-level NCSS (National Cooperative Soil Survey) mapping programs over time. By comparing counties with similar agro-ecological conditions and initial endowments that are subject to the same prevailing national market prices, any differential trends in agricultural outcomes can plausibly be attributed to variation in exposure to the NCSS mapping efforts. The specification is shown as follows:

$$y_{cst} = \alpha + \beta NCSS_{cst} + X_{cst}^{'}\gamma + \rho_t + \eta_c + \epsilon_{cst} \tag{1}$$

where c denotes counties, s represents states, and t indexes years. y_{cst} is the key outcome of interest: the total crop revenue per acre (of farmland) in county c located in state s in year t. $NCSS_{cst}$ is an indicator whether there is a NCSS-based map in county c in state s in year t. X_{cst} is a set of county-specific controls. I include county fixed effects (η_c) and ρ_t as year fixed effects, respectively. Therefore, β captures the average treatment effect of access to NCSS mapping on agricultural productivity, under the assumption of parallel trends in the absence of treatment.

A key threat to identification with this baseline difference-in-differences specification is the potential violation of the parallel trends assumption. If counties that received NCSS mapping earlier were already on a different trajectory in terms of agricultural productivity (due to unobserved investments, institutional capacity, or other time-varying county-specific factors), then the estimated effect of NCSS access may conflate the true treatment effect with these underlying trends.

For example, if early-treated counties were systematically more likely to receive complementary agricultural extension services or had higher baseline administrative capacity to adopt new farming technologies, then the identifying assumption would be compromised. Additionally, selective timing of rollout would introduce *endogeneity* into treatment assignment, further biasing the estimates. To directly test th parallel trends assumption, I estimate a

stylized event study model as follows in equation 2:

$$y_{cst} = \alpha + \sum_{\tau} \beta_{\tau}(Treat_{cs} \times \textit{Year since mapping} = \tau) + \rho_{cs} + \eta_t + \epsilon_{cst} \tag{2}$$

where τ represents the year since policy change, grouped by 5-year intervals. In this specification, I include county fixed effects and year fixed effects, respectively. the estimate associated with one year prior to the treatment is normalized to zero. This way, $\beta_P t + i$ also captures the dynamic effects of restricting PA.

Note there is a growing strand of work emphasizes the sensitivity of a staggered difference-in-differences specification: the traditional staggered difference-in-differences estimation cannot fully capture the dynamic treatment effects with staggered adoption and heterogeneous treatment effects (Goodman-Bacon, 2021; Sun and Abraham, 2021).

4.2 Adjacent County Difference-in-Differences Design

One challenge to the baseline difference-in-differences (DD) specification is that, even after controlling for time-invariant characteristics via county fixed effects, counties that received NCSS maps may still differ systematically from those that never participated in the program, due to various unobserved time-varying factors.

To address this concern, I leverage the fact that adjacent counties tend to be more comparable in terms of natural conditions, economic structure, political institutions, and social environments, relative to the broader cross-county sample. Therefore, I construct a restricted sample that only includes neighboring counties on either side of the NCSS mapping rollout boundary. This approach strengthens the identifying assumption by ensuring that the treated and untreated groups are more locally comparable, thereby reducing bias from omitted variables that may confound the relationship between mapping and agricultural outcomes.

Specifically, this sample is constructed as follows:

- I first identify all counties that were ever mapped by the NCSS program between 1899 and 1930.
- I then construct a county-pair sample based on two criteria: (1) within each pair, two counties must share a county border, and (2) within each pair, at least one county must have been mapped during the sample period.
- I pick each unique individual county included in this county-pair sample. The resulting adjacent-county sample thus includes all mapped counties and their directly neighboring

unmapped counterparts.

- Note the final sample of interest is still composed of individual counties. In the robustness check, I directly use the adjacent-county pair sample.

This adjacent-sample thus gives me a *balanced* panel of 1826 counties, spanning 1890 to 1930. The alternative identification strategy is implemented using the same specification as Equation 1 and the only difference lies in the sample restriction. The coefficient of interest, β, continues to capture the causal effect of NCSS mapping on local agricultural outcomes but now we are comparing treated and untreated counties that are more plausibly similar in terms of unobserved local characteristics.

Of course, a more restrictive version of this adjacent-county DD design would further include county-pair fixed effects (Huang, 2024), which directly exploits variation induced by the NCSS mapping within each county pair. In other words, this specification identifies the effect of mapping by comparing treated and untreated counties that are directly adjacent to one another, while controlling for all pair-specific, time-invariant factors. This approach sharpens identification by narrowing the source of variation to within-pair differences in timing and treatment status. The corresponding specification is:

$$y_{cpt} = \alpha + \beta NCSS_{ct} + X_{ct}^{'}\gamma + \rho_t + \eta_{cp} + \epsilon_{cpt} \tag{3}$$

where c denotes counties, p represents states, and t indexes years. The model includes year fixed effects and county-pair fixed effects to account for both temporal shocks and stable pair-specific confounders. In the final sample, there are 2111 county pairs included (compose of 1826 counties). Given this structure, it is not feasible to further include county-by-year fixed effects, which would absorb nearly all the identifying variation associated with the staggered rollout of NCSS mapping programs. Results based on this adjacent county pair design are presented in Table A2 and Figures A1-A5.

5 Data

In this section, I summarize the primary data sources I used throughout the empirical analysis: (1) the timing and location records of the universe of NCSS mapping programs (1899-1930), (2) the decennial Census of Agriculture (1890-1930), (3) the decennial Census of Population (1880, 1900, 1910, 1920, and 1930), (4) the Global Agro-Ecological Zones (GAEZ) project, and (5) the historical records of research conducted by local agricultural experiment stations.

5.1 Mapping Progress of National Cooperative Soil Survey

I construct a novel dataset of all NCSS mapping programs using the newly digitized historical records collected by the Natural Resources Conservation Service (NRCS).[12] I compiled this list from the original NCSS program correspondence from 1899 to 1930. I assign each mapping program to the county in which it was implemented and use the publication years as "treatment" years since they are available for the county. The county-level NCSS has several distinctive features. First, an NCSS survey may cover an *area*, defined as a unit that can include multiple counties or portions of counties (as shown in Appendix). In such cases, it becomes difficult to delineate clear boundaries between treated and control counties, since a single survey area may straddle both. Second, multiple surveys within the same county occurred in approximately 1 percent of recipient places. In multi-NCSS counties that get mapped at different points in time, I assign the county-level mapped year to the first time that an NCSS program was conducted.

From the full universe of NCSS programs, I construct a balanced panel of counties serving as the foundation of the empirical analysis. To ensure consistency and comparability, I implement sample restrictions as follows: First, I exclude NCSS programs that span multiple counties, as noted earlier in Section 2.2. These multi-county surveys were most common in western states such as California, Oregon, and Washington and present challenges for aligning soil survey boundaries with county-level data. Imposing this restriction leaves 81% of mapped counties in the sample.

Second, I restrict my analysis to programs published prior to 1930. This cutoff is motivated by both economic and political features. As discussed in Section 2, the agricultural sector experienced significant fluctuations during and after World War I, which influenced both the demand for and the interpretation of soil information. In addition, the onset of the New Deal in the 1930s introduced sweeping policy and administrative reforms that fundamentally reshaped the institutional context of American agriculture. Importantly, the focus of soil surveys also shifted in this period: beginning in the late 1920s and continuing into the 1930s, NCSS programs increasingly emphasized conservation planning and erosion control, rather than the facilitation of agricultural development. These later surveys are thus *less comparable* to the early mapping efforts that are the focus of this study.

In Figures 4-6, I plot the mapping progress of NCSS from 1910 to 1930. As of 1930, approximately 40% of US counties got their first NCSS maps in place. By 1930, approximately 40%

[12] See https://www.nrcs.usda.gov/sites/default/files/2022-09/Links-to-Scanned-Manuscripts-in-BOX.pdf, last accessed October 15, 2025.

of U.S. counties had received their first NCSS maps.

5.2 Census of Agriculture

The primary data used to estimate the impact of the National Cooperative Soil Survey (NCSS) programs on local agricultural development between 1890 and 1930 are drawn from the Census of Agriculture (Haines, Fishback and Rhode, 2018). I compile county-level information on key agricultural variables, including total crop revenue, crop-specific farm acreage, the value of farm equipment, and expenditures on fertilizers, covering the full sample period. The main outcome variable of interest is crop revenue per acre, constructed as the total value of production for nine principal crops: wheat, barley, corn, oat, potato, sweet potato, cotton, buckwheat, and rye. For each crop, production is calculated as the quantity harvested multiplied by its corresponding national-level price, following the methodology of Kantor and Whalley (2019).

There are several advantages to using this revenue-based measure of crop-specific, land productivity. First, as emphasized by Costinot, Donaldson and Smith (2016), if price variation across locations reflects differences in market power or access, then changes in market integration may be conflated with changes in land productivity per se. In this case, relying on self-reported total crop revenue per acre, which is directly recorded in the Census of Agriculture, may introduce bias. Second, self-reported revenues typically exclude the value of crops retained for on-farm use, particularly those fed to livestock, as the data are based on realized market transactions. In contrast, my preferred measure is constructed from the quantities of crops grown rather than sold, offering a more comprehensive, land-specific indicator of agricultural output. Third, the method of reporting crop revenue in the Census of Agriculture itself changed over time, complicating comparisons across years. Notably, the 1920 Census transitioned from collecting revenue figures directly from farmers to calculating them by multiplying reported crop quantities by regional prices obtained from external sources (Kantor and Whalley, 2019). This inconsistency further supports the use of a consistently constructed, quantity-based revenue measure over time.

By merging the NCSS mapping data to the Census of Agriculture data, my full, balanced sample includes 2,148 counties, of which 920 had been mapped by 1930. To have consistent units of observation over time in spite of changes in county boundaries, I adjust all data to conform to 1930 boundaries.

5.3 Census of Population

Another primary data source I rely on in this paper is the U.S. Census of Population (1880, 1900–1930), which provides decennial, nationally representative demographic and occupational information at the county level. Using this dataset, I construct measures of both the total number of farmers and the number of immigrant farmers in each county. These variables are central to my analysis, as they allow me to explore the composition of the agricultural labor force and to examine how immigrant presence may have shaped or interacted with the impacts of soil information and other local development policies. The richness of the Census data also enables me to control for broader population trends and labor force characteristics that might otherwise confound the results. I further use the Census of Population to construct county-level density of immigrants and immigrant farmers, respectively.

5.4 Global Agro-Ecological Zones Project

A key dimension of input adjustment farmers can take is the choice of crop cultivated locally. To assess if NCSS mapping induced any response along this line, I compile information from the FAO's Global Agro-Ecological Zones (GAEZ) project, which provides measures of the suitability for cultivating individual crops at a detailed disaggregated geographic level. Using the most updated version, I construct a county-crop-specific suitability class using data from the *Suitability and Attainable Yield* section. These measures of crop-specific potential productivity are based on expert knowledge of climatic features affecting agricultural production processes; they do not rely on a statistical analysis of production patterns observed across the world. In addition, though based on climatic records for 1961-1990, they provide good proxies for historical conditions (see Nunn and Qian (2011)).

Technically, GAEZ data includes the classification of a normalized suitability index SI:

$$SI = 100 \times (90 \times VS + 70 \times S + 50 \times MS + 30 \times mS + 15 \times VmS + 0 \times NS) \times 0.9 \quad (4)$$

with values stretching over the interval of 0 to 10000 and where *VS, S, ... , NS* are the shares of area extents of different suitability classes in a 5 arc-minute grid cell.[13]

[13] VS: Very suitable (VS); Prime land offering best conditions for economic crop production.
S: Suitable; Good land for economic production.
MS: Moderately suitable; Moderate land with substantial constraints. Economic production requires high product prices for profitability.
mS: Marginally suitable; Commercial production not viable. Land could be used for subsistence production when no other land is available.
VmS: Very marginally suitable; Production not feasible.
NS: Not suitable; Production not possible.

The normalized suitability index has been calculated for 53 crops, by level of inputs/management and separately for rain-fed and irrigated conditions. Results relate to either an entire grid cell or to a fraction of the grid cell indicated by the land cover data as cropland. For the map compilations regarding cropland it is assumed that farmers have used the better part of the suitability distribution in a grid cell. Results of calculated SI values for each grid cell are grouped into 8 aggregate classes, termed as follows: (1) Very high, when SI>8500; (2) High, when $7000 < \text{SI} \leq 8500$; (3) Good, when $5500 < \text{SI} \leq 7000$; (4) Medium, when $4000 < \text{SI} \leq 5500$; (5) Moderate, when $2500 < \text{SI} \leq 4000$; (6) Marginal, when $1000 < \text{SI} \leq 2500$; (7) Very marginal, when $0 < \text{SI} \leq 1000$; and (8) Not suitable, when $\text{SI} = 0$.

Based on these classifications, I construct a crop-county level indicator that identifies whether a crop is the *least suitable* among all considered crop types for that county. This indicator captures the degree of misalignment between actual crop choice and underlying agroecological potential, and serves as a proxy for inefficient production decisions in the absence of detailed soil quality information.

5.5 Records from Agricultural Experiment Stations

Experiment stations were expected to publish the results of their investigations and research activities. Following Kantor and Whalley (2019), I leverage soil-relevant published findings to construct a measure of local research density at the agricultural experiment stations. To do so, I compile the historical annual reports of state agricultural experiment stations since 1889.[14] As shown in Figure 7, such reports list station's publications addressing fertilizer and soil issues. For example, in 1907, Kentucky's experiment station, there were analyses and valuations of commercial fertilizers collected during the preceding year and included in their report. Similarly, in New Hampshire, their station's publication included the results of analyses of 114 sample of fertilizer materials inspected during 1906.

Specifically, I define a research activity directly related to soil as an event in which experiment station listed one publication that addresses soil composition, soil treatment, fertilizer use, soil productivity, etc. This may includes chemical analyses of soil or fertilizer samples, field trials evaluating fertilizer effectiveness, and studies on soil amendments or erosion. This enables me to construct a time-varying measure of local soil research activity, which serves as a proxy for the intensity of institution-based research. It is worth noting I construct this panel at the state-year level, instead of the county-year level. This is simply because agricultural experiment stations typically serve the entire state, and most states had only one primary

[14] These reports are available here: https://catalog.hathitrust.org/Record/100114549, last accessed October 15, 2025.

station during the period.

6 Main Results

In this section, I present my main results for the impacts of NCSS mapping programs on agricultural productivity and ecological risk, respectively. I also discuss how to interpret these estimates.

6.1 Effects on Agricultural Productivity

Table 1 presents the baseline estimates of the effects of NCSS soil mapping programs on agricultural productivity, measured as total crop revenue per acre of cropland (as defined in Section 5).[15] Columns (1)–(5) report results using the full sample of 2148 counties, while Columns (6)–(10) focus on a more restrictive specification exploiting within-pair variation across borders of adjacent counties, comprising 1826 counties in total (see Equation 2). In Columns (4)-(6) and (8)-(10), I include county-level market access controls, as connecting markets had a substantial effect on local farming outputs (Chan, 2022). As a result, the sample period is restricted to 1890–1920 due to the unavailability of market access data for 1930.[16] Columns (2) (4) (5) (7) (9) and (10) include year-region fixed effects, controlling for region-specific time-varying shocks, such as localized climate events or natural disasters, regional exposure to immigrants, political, industrial, or technological shocks.[17]

Across both sets of specifications, the estimated effects of NCSS mapping are consistently positive and in most cases, statistically significant. In Columns (1) and (2), I show that when taking into account region-specific, time-varying shocks, local crop revenue per acre increased by 2.8 log points at least. In Column (3), controlling for market access and using the full sample, I find that an NCSS program increases total crop revenue per acre by 5.34%. Relative to the 1890 pre-treatment mean of \$3.422, this corresponds to an increase of approximately \$0.182 in 1890 dollars. The magnitude of my estimate remains stable in Columns (4) and (5) when introducing year-region fixed effects and weighing by the acre of land for each county in 1900, suggesting that regionally correlated shocks are unlikely to be driving the

[15] These estimates are robust using "total crop revenue per acre of farmland (instead of cropland) as well." In the appendix, I also show that the ratio between cropland and farmland per county remains stable following the NCSS mapping.

[16] Market access measures are drawn from Donaldson and Hornbeck (2016), which provides consistent county-level data through 1920.

[17] For example, as mentioned in Fiszbein (2022), the period of 1870–1920, sometimes called the *Second Industrial Revolution*, was characterized by a series of massive transformations in technologies, spanning electricity, chemicals, steel, transportation, communications, management, and marketing.

results. Notably, the magnitude of these effects declined when including the period of the 1920s, suggesting productivity gains from NCSS mapping were most pronounced in the earlier decades. Possible explanations could be that during the 1920s, there were several substantial breakthroughs in the agricultural sector.

I also implement the adjacent-county design. As reported in Columns (8) (9) and (10), the estimated effects remain positive and statistically significant, with an average increase of 5.76% in crop revenue per acre. These results reinforce the conclusion that the observed productivity gains are unlikely to be driven by selection bias or unobserved spatial heterogeneity. The modest reduction in the magnitude (5.6 vs. 5.1) relative to the main specifications is consistent with the expectation that neighboring counties share similar environmental characteristics and other features, thereby producing a more conservative estimate. In Section 9, I also show that when introducing county-pair fixed effects, which imposing even more restrictive controls, the estimated increase in crop revenue per acre is 4.1 log points, further mitigating concerns about selection.

To explore dynamic effects, I supplement the baseline results with event-study specifications, presented in Figures 8 and 9, corresponding to Equation 3. Figure 8 shows that productivity started to rise immediately following the implementation of an NCSS mapping program, with an estimated increase of roughly 7% within the first five years. This effect is persistent, with gains sustained for at least a decade after treatment. Figure 9, which leverages the adjacent-county design, reveals a similarly robust pattern: treated counties experience a lasting increase in crop revenue per acre of approximately 4% to 5%, relative to their untreated neighboring counterparts. These results underscore the durability of NCSS-induced productivity gains and reinforce the credibility of the identification strategy.

Taken together, the estimates from both the TWFE models and the event-study framework both provide consistent and compelling evidence that the dissemination of soil quality information through NCSS programs contributed meaningfully to local agricultural development. As a benchmark, Kantor and Whalley (2019) finds that reallocating agricultural research to maximize geographic proximity would have increased national total factor productivity (TFP) by 5% in 1910, with effects persisting for two decades. In line with this, my results suggest that access to locally tailored soil information raised crop revenue per acre by approximately a similar magnitude. These findings highlight the long-run value of spatially targeted knowledge diffusion and suggest that such information/data infrastructure may play a complementary and critical role in promoting agricultural productivity.

6.2 Effects on Agricultural Risk

In this section, I turn to discuss the potential agricultural risk associated with the soil mapping program. In Section 5.2.1, I assess whether the provision of additional soil quality information would encourage crop specialization and reduce local crop diversity, resulting in monoculture farming. In Section 5.2.2, I further examine if the soil mapping status prior to the Dust Bowl is correlated with local soil erosion status later.

6.2.1 Crop Diversity

Beyond its effects on agricultural productivity, I also examine whether NCSS soil mapping influenced the level of monoculture farming and crop biodiversity within a county. This analysis is motivated by the concern that improved soil information, while intended to promote more efficient land use, may have unexpectedly encouraged crop specialization, which could delete the soil of nutrients (making the soil less productive over time), reduce organic matter in soil, cause significant erosion and finally decline in crop productivity (Power and Follett, 1987; Belete and Yadete, 2023). By revealing which crops were most suitable and/or profitable for a given soil type, this soil mapping program could have steered farmers toward a narrower set of options and concentrate their production. In this sense, while NCSS mapping aimed to improve productivity, it may have also introduced trade-offs at the expense of imposing negative environmental spillovers.

To test this idea, I construct two relevant measures of local crop diversity. The first one is a widely used index called *Shannon-Weiner Index* (SWI) (see for example, Song et al., 2010; Mahecha-Vásquez, Sierra and Posada, 2017), specified as follows:

$$\text{SWI}_{ct} = \sum_{i} \textit{share of cropland acreage}_{ict} \times log(\textit{share of cropland acreage}_{ict}) \tag{5}$$

where i refers to crop type (i.e., wheat, corn, oat, barley, potato, rye, buck, and sweet potato), c refers to the county, and t represents the calendar year. $\textit{share of cropland acreage}_{ict}$ is the share of crop-specific farmland acreage for crop i in county c, year t. By construction, The higher the value of SWI, the lower the diversity of crops in a particular county.[18]

The second measure of local crop diversity if the Herfindahl-Hirschman Index (HHI), defined as shown in the following equation in a similar way as Chan (2022):

[18]For ease of interpretation and comparability with the Herfindahl-Hirschman Index (HHI), which increases as concentration rises, I take the inverse of the SWI in the empirical analysis. Thus, a higher value of the transformed measure reflects lower crop diversity.

$$\text{HHI}_{ct} = \sum_{i} \textit{crop-specific revenue}_{ict}^{2} \tag{6}$$

where i refers to crop type (i.e., wheat, corn, oat, barley, potato, rye, buck, and sweet potato), c refers to the county, and t represents the calendar year. *crop-specific revenue*$_{ict}$ is the total revenue per acre for crop i in county c, year t. Similarly, the higher the value of HHI, the lower the diversity of crops within a county.

Table 2 presents the results. The first four columns report estimates for the full sample of counties (2,148 in total), while the last four restrict the analysis to adjacent counties sharing borders. Even-numbered columns control for county-level market access information, allowing me to assess whether the effects of soil mapping on concentration are mediated by access to transportation infrastructure.

Using the full county sample, I find that implementing an NCSS mapping program is associated with a statistically significant increase of 0.007 in the inverse-transformed Shannon-Wiener Index (SWI) (significant at the 10% level). This is equivalent to a 1.41% increase relative to the 1890 sample mean (–0.498), suggesting a modest rise in input-based crop concentration. Similarly, the Herfindahl-Hirschman Index (HHI) increases by 0.012 units (significant at the 5% level), which corresponds to a 2.99% increase relative to the 1890 average of 0.401. Taken together, these results indicate that NCSS mapping programs triggered a shift toward more specialized cropping practice.

In Columns (5)–(8), I repeat the analysis using the adjacent-county sample. The results are consistent and a little larger in terms the magnitude. In this restricted sample, NCSS mapping is associated with a 0.007 increase in the SWI (Column 6), and a 0.029–0.036 increase in the HHI (Columns 7 and 8), equivalent to a 10.2% increase over the 1890 mean HHI of 0.284. These stronger effects suggest that even when limiting the comparison to neighboring counties with similar agro-climatic conditions, soil mapping meaningfully increased crop concentration.

Most interestingly, in Figures 10-13, I show that the impact of NCSS mapping on crop diversity started to be statistically significant after at least a decade relative to the program. This delayed response suggests that farmers may have gradually adjusted their cropping decisions as they internalized and acted upon the soil information, leading to more pronounced shifts toward specialization over time. This gradual growth in crop concentration over time thus highlights the long-run influence of spreading information on production, as well as the potential for a key trade-off between efficiency and ecological sustainability.

6.2.2 Relationship with Soil Erosion

The second exercise I did to examine the potential long-run environmental costs of NCSS mapping is to explore whether *early exposure* to soil mapping programs is associated with greater soil erosion risk after 1930. Specifically, I investigate the correlation between a county's mapping status prior to 1930 and the severity of soil erosion reported during the 1930s. To do so, I construct a county-level indicator of erosion severity based on archival data compiled by the U.S. Department of Agriculture and the Soil Conservation Service, which began systematic assessments of erosion damage during that period in response to the Dust Bowl and related environmental crises. I employ the following cross-sectional regression to estimate this correlation:

$$y_{c,1930s} = \alpha + \beta NCSS_{c,1899-1930} + \gamma X_{c,1930} + \epsilon_{ct} \tag{7}$$

where $y_{c,1930s}$ is the outcome of interest, representing the presence of any eroded soil within a county c. $NCSS_{c,1899-1930}$ represents the status of being mapped till the calendar year of 1930. $X_{c,1930}$ is a vector of variables at the county-level in 1930.[19] Key coefficient β thus captures the correlation we are interested in: whether a mapped county would experience more severe soil erosion later. Soil erosion data are drawn from Hornbeck (2012) and are merged accordingly by each county.

The point estimates based on this equation are shown in Table 3. In the first two columns, I focus on the extensive margin of soil erosion: whether there was any presence of soil erosion during the 1930s. Columns (3) and (4) turn to focus on the intensive margin of soil erosion, the natural logarithm of specific share of soil erosion for each county. Throughout the analysis, I include all counties of interest. In even-numbered columns, I weight the regression by its total land area in 1930 so that these estimates capture the effects for a typical acre of land rather than the entire county.

Based on the regression results, we can tell that along both the extensive and intensive margins, there is always a statistically significant correlation between soil mapping and soil erosion. Specifically, getting mapped by the NCSS program is associated with an increase in the likelihood of having any soil erosion by 5.3 percentage points, as well as a risk in the share of soil eroded by at least 9.8 percentage points. These findings suggest that while soil mapping may have enabled more intensive agricultural development, it was also associated with increased, long-term vulnerability to soil health.

[19] The only exception is market access, which the most recent one is as of 1920.

7 Mechanisms

In this section, I empirically test various potential mechanisms at play to explain changed in agricultural productivity and agricultural risks shown in Section 6. Specifically, I discuss the role of farmers' input adjust (crop choice, capital, and human capital) through learning, the coordination across government agencies, the aggregate market shocks/responses, and knowledge diffusion through social network, respectively.

7.1 Farmers' Input Decisions

The first and primary channel I argue is that access to soil quality information significantly reduces uncertainty in agricultural production, enabling farmers to make more informed decisions regarding crop selection, technology adoption, the use of land, and labor input. In the absence of such information, farmers have to rely on costly trial-and-error learning or inherited knowledge, which may not align with agronomic realities. Soil surveys thus represent an important channel through which information can improve the efficiency of production. In the following subsections, I assess farmers' adjustment in a variety dimensions of inputs and also show supporting evidence for whether and how they learn soil information via NCSS.

7.1.1 Farmers' Choice of Crops

I begin by investigating whether access to soil quality information enabled farmers to make more informed decisions regarding their crop selection and improve the match between local soil and the corresponding crop choice. Specifically, once the soil mapping program is available, soil quality information becomes publicly accessible to farmers, reducing information frictions and enabling more accurate matching between crops and land, especially in regions where prior knowledge of soil characteristics was limited. Farmers can adjust their crop selection to better align with the agronomic potential of their land, shifting away from less suitable crops toward those that are better adapted to local conditions. To measure the suitability of crop choice, I need to identify whether a crop is (theoretically) suitable for each county conditional on the natural endowment.

Following Nunn and Qian (2011), I use data from the Food and Agriculture Organization (FAO)'s Global AgroEcological Zones (GAEZ) 2022 database. These data measure the suitability for cultivating individual crops at a detailed disaggregated geographic level. Using the most updated version, I extract information on crop-specific suitability class from the *Suitability and Attainable Yield* section. This section presents results of the final step in the GAEZ crop suitability and productivity assessment, combining agro-climatic potential yields

with soil/terrain evaluation results (i.e., yield reduction factors due to the constraints induced by soil limitations and prevailing terrain-slope conditions).

To empirically test this idea, I exploit a triple DD specification as follows:

$$y_{ict} = \alpha + \beta_1 \textit{least suitable}_{ic} \times \text{NCSS}_{ct} + \beta_2 \text{NCSS}_{ct} + X_{ict} + \gamma_{it} + \pi_{ic} + \mu_i + \pi_c + \phi_t + \epsilon_{ict} \quad (8)$$

where i stands for crop type, c refers to county, and t means year. The dependent variable is the logarithm of crop-specific acreage per farm. $\textit{least suitable}_{ic}$ is an indicator for whether a crop type i is the least suitable type for a county c and NCSS_{ct} is an indicator for whether there was an NCSS mapping program in county c, year t. Accordingly, I include crop-year fixed effects, crop-county fixed effects, crop fixed effects, county fixed effects, and year fixed effects.

The coefficient of interest, β_1, thus captures conditional on getting mapped by a county, whether local farmers are less likely to choose the least suitable crop. A negative and statistically significant estimate would suggest that farmers indeed responded to improved soil information by shifting land use away from crops poorly matched to local soil conditions. Based on the qualitative evidence, this is indeed the case:

> *About 400 different types of soil have been encountered by the Soil Survey since the beginning of its field work in 1899 ... the study of soils through the field work is coming to have two primary objects. The first of these, the discovery of the best crops for each soil, is known as the soil fertility or the soil management problem ... it is easily seen that the farmers are not securing the best possible results from their agricultural pursuits through a lack of knowledge of the intimate relation between crop and soil which is comprehended in the term "crop adaptation."*

Table 4 presents the main results, showing how access to NCSS soil mapping affects crop selection. The dependent variable is defined as the logarithm of crop-specific acreage per farm. Columns 1–4 use the full sample of 2,148 counties, while Columns 5-8 restrict the analysis to 1,826 counties sharing county borders over time to address concerns about selection and dissimilar conditions. In Columns (3) (4) (7) and (8), I replace the indicator for the least suitable crop with crop-county and crop-year fixed effects, offering a more conservative identification strategy that accounts for persistent differences in crop suitability and aggregate trends in acreage.

Across all specifications, we can observe that the coefficient on the interaction term ($\textit{least suitable}_{ic} \times NCSS_{ct}$) is negative and statistically significant at the 1% level. In the full county

sample (Columns 1–4), estimates range from –0.064 to –0.084, indicating that following the introduction of soil mapping, farmers reduce their planting of crops classified as least suitable by approximately 6.2% to 8.1% relative to other crops and non-mapped counties. The neighboring-county sample gives consistent results (–0.061 to –0.088), further supporting the robustness of this finding.

Together, these results provide compelling evidence that access to detailed soil information encouraged farmers to adjust away from agronomically mismatched crops. These effects are both statistically and economically meaningful, consistent with the hypothesis that public information infrastructure improves input allocation in agricultural production. By enabling farmers to better match land with crop-specific ecological potential, the dissemination of soil information plays a role analogous to other forms of productivity-enhancing infrastructure, such as weather forecasting or irrigation networks.

7.1.2 Farmers' Investment in Technology and Land

In response to better soil quality information, farmers can also adjust other production inputs in response to improved information. I test whether capital utilization is a dimension of input adjustment by examining changes in (1) the value of agricultural machinery and equipment per acre, and (2) the likelihood of owning farmland, following the implementation of NCSS mapping.

Main results are presented in Table 5. Across both the full sample and the neighboring counties specifications, the introduction of NCSS mapping is associated with a statistically significant increase in capital investment. In the all-counties specification (Columns 1 and 2), the value of agricultural machinery per acre increases by 0.452 units ($p < 0.01$), and the cost of fertilizers rises by 0.207 units ($p < 0.01$). Similarly, the likelihood of self-owned land increases by 1.2 percentage points ($p < 0.01$), indicating that farmers may respond to more precise soil information by expanding or consolidating ownership.

These patterns hold when restricting the sample to adjacent counties only (Columns 4–6), which account for spatial heterogeneity in soil, endowments, and local market conditions. The magnitude of the effects is slightly smaller but remains statistically significant: machinery value increases by 0.331, fertilizer costs by 0.171, and the likelihood of owning farmland by 1.1 percentage points. These findings suggest that improved soil quality information enables farmers to optimize input usage and make more informed investment decisions in both capital and land markets.

To further explore the dynamics of input adjustments, Figures 14-16 present event-study estimates using adjacent counties, allowing us to visualize the evolution of treatment effects. Consistent with the baseline DD estimates, all three event studies show minimal pre-trends and a clear upward shift in outcomes following treatment. The machinery investment effect emerges quickly and remains stable over time (Figure 14), suggesting prompt capital reallocation once more reliable soil information becomes available. Fertilizer use also increases gradually post-treatment (Figure 15). Figure 16 shows a steady rise in farmland ownership over time, indicating that land-use adjustments may respond more slowly but persistently to the intervention. Overall, these dynamic estimates reinforce the interpretation that NCSS mapping helps reduce production uncertainty and facilitates re-optimization of both capital and land input decisions at the county level.

7.1.3 Input in Human Capital

Another key dimension of input choice is about human capital. If soil mapping reduces uncertainty about land productivity, farmers may respond by adjusting their labor inputs, becoming more willing to increase total farm labor or reallocate tasks toward skilled or better-educated workers. In this case, access to improved soil information not only encourage greater labor investment but also shift the composition of labor toward individuals better equipped to interpret technical information.

Table 6 presents evidence on this margin. I examine two outcomes: (1) the number of farmers per 100,000 population, and (2) the number of literate farmers per 100,000 population. The results show a significant increase in both measures. In the full sample (Panel A), NCSS mapping is associated with an increase of 10 farmers per 100,000 people, who were primarily literate farmers (Columns (3) and (4)). These estimates remain robust using the neighboring counties sample (Panel B). These results suggest that improved soil quality information indeed spurs not only an expansion in farm labor supply but also a shift toward more educated labor. One interpretation is that improved soil information enables more complex, skill-intensive farming practices (such as tailored input decisions, crop rotation, or soil management strategies), which require more literate farmers.

7.1.4 Learning Effect through Soil Mapping

I further present both descriptive and quantitative evidence about people learn from the early use and dissemination of historical NCSS soil maps.

Descriptive Evidence. First, records from the Department of Agriculture suggest that these soil surveys were not merely bureaucratic exercises but were actively requested and used by

local stakeholders, especially individual farmers.

> *"The soil survey is meeting increasingly gratifying evidence of the extent to which the soil maps are being used by the various interests for whose benefit they are made. A careful check list is kept of the demand for advanced sheets of each area. An inspection of this list covering the areas reported upon to July 1, 1905, shows that a great variety of interests are concerned with the results of the survey work. The number of requests is steadily increasing, and 40,000 separates have been distributed from the Bureau's office alone to meet the demand for the report made in 1902 and 1903. The sources of these demands are distributed with remarkable uniformity all over the US."*

In the meantime, investment companies also firmly believed that people read the soil survey reports and form perceptions towards the value of local land.

> *In addition to the reports sent to the individuals, which usually convey information to a single person or to a small group, a large number of reports are annually asked for by real estate and investment companies, who use these reports to furnish accurate and unprejudiced information of the regions in which their farming lands may be located.*

The above descriptions thus not only highlight the demand for soil survey outputs was not subtle, but also suggest that information generated was rapidly disseminated and valued by both individuals and private sectors. The uniformity in geographic demand also supports the idea that soil information was considered broadly across various agricultural contexts.

NCSS surveys also played a complementary role in translating experimental findings from agricultural colleges and experiment stations into more accessible knowledge at the local level. While experiment stations were often geographically limited in scope, the availability of localized soil surveys allowed farmers to apply experimental results to their own land and farming practice.

> *"The people of the United States are provided in each state and territory with at least one agricultural college and experiment station where scientific research along agricultural lines is carried on, and where information is imparted to the younger members of the community. A vast amount of scientific agricultural work has been accomplished by these colleges and experiment stations...In order that the vast amount of valuable information which has been secured by these experiment stations may be made of the most direct application to each farming community*

within the State served, it is necessary for the farmer to know not only the details of any particular experiment, but also for him to be able to judge where his own farm contains the type of soil upon which the experiment has been carried on, and whether it lies within the same general climatic belt as the experiment station itself... The Soil Survey thus renders the result of experiments conducted at the various experiment stations applicable throughout a wide range of territory not restricted by State lines."

To empirically test for the presence of a *learning* effect (i.e., whether farmers acquired soil information and related agricultural knowledge directly through soil mapping reports), I investigate whether the impact of NCSS mapping varies by the level of *pre-existing local agricultural knowledge* within a county. The underlying hypothesis is that if soil mapping primarily served as a learning tool, then its marginal value should be greater in areas where prior knowledge was more limited. In other words, we would expect stronger effects in counties with weaker prior exposure to scientific farming practices, technical assistance, or institutional research.

Heterogeneity by Agricultural Societies. I collect information on the number of county-level agricultural societies and farmers' clubs (as of 1876) as the first proxy for local agronomic knowledge. These institutions played a key role in facilitating the exchange of farming practices and disseminating technical advice among rural producers. Their presence indicates the extent to which local communities had already developed mechanisms for diffusing agricultural knowledge prior to the arrival of soil mapping programs. For example, according to *List of Agricultural Societies and Farmers' Clubs (July 4, 1876)*

The necessity for these societies is becoming more and more apparent, as, in order to make agriculture profitable, farmers must ... devote their energies to increase the productive capacity of the soil, and thereby secure a fair return for the capital invested and labor bestowed, enabling them to adopt those rules and regulations of business life which yield such satisfactory results in commercial and manufacturing pursuits .

In Figure 17, I present an example of these records, which include the name, location, year of organization, number of members, and whether there was a fair for each institution. Based on these data, I construct two indicators of local agricultural knowledge accumulation: (1) an indicator for whether a county had at least one agricultural society or farmers' club, and (2) an indicator for whether a county hosted local agricultural fairs through these institutions. The first measure captures the basic presence of formal knowledge-sharing organizations, while the second reflects a more active form of community-level agricultural engagement and public

dissemination, *county fairs*.

Table 7 reports the results. Across multiple outcomes (revenue per acre, machinery investment, fertilizer usage, and farm ownership), the estimated effects of NCSS mapping are consistently larger in counties without prior agricultural societies or fairs. For example, the increase in log(revenue per acre) is 0.038 in counties with societies (Panel A), but nearly twice as large (0.071) in counties without them (Panel B). Similarly, fertilizer expenditure rises by 0.376 units in counties lacking societies/farmers' clubs, compared to a statistically insignificant 0.070 in counties with them.

Panels C and D show a similar pattern when checking the heterogeneity by the presence of county fairs: counties without agricultural fairs exhibit larger gains in revenue, the use of machinery, and fertilizer. The interaction pattern is particularly strong for fertilizer investment and HHI. These results support the interpretation that soil mapping programs facilitated learning, especially in settings where prior institutional or social infrastructure for agronomic knowledge diffusion was limited. The muted effects in counties with more developed pre-mapping knowledge systems also suggest the declining marginal returns to extra information that these communities had already accumulated.

I also use the following specification to test this heterogeneous effects by interacting the term $NCSS_{ct}$ and an indicator for $Agr\ Society_{c,1876}$, and adding the $Agr\ Society_{c,1876}$-year fixed effects. The key coefficient β_1 captures the differential effects of soil mapping.

$$y_{ct} = \alpha + \beta_1 NCSS_{ct} \times Agr\ Society_{c,1876} + \beta_2 NCSS_{ct} + X_{ct} + Agr\ Society_{c,1876} \times \theta_t + \mu_c + \epsilon_{ct} \quad (9)$$

Point estimates are presented in Table 8. Consistent with the results shown in Table 7 and the event-study results, we can see the coefficients associated with the interaction term are negative (except for the SWI) and some are statistically significant (the adoption of machinery and fertilizer). Meanwhile, the baseline estimates for $NCSS_{ct}$ are significant, positive, and similar in the magnitude relative to my main results.

These findings show further support to the hypothesis that one primary channel through which soil mapping influenced agricultural outcomes was the diffusion of technical knowledge. In counties that lacked formal agricultural societies in the late 19th century, NCSS mapping appears to have delivered new information and guidance that spurred changes in farming practice, especially the adoption of technologies. In contrast, counties with already established agricultural societies/county fairs may have already gotten access to agronomic expertise, printed bulletins, demonstration farms, or informal farmer networks, thereby diminishing the

marginal informational value of NCSS reports.

Heterogeneity by Farmers' Experience. I also examine heterogeneity in treatment effects by the local density of farmers in 1880, which is the second proxy for local agricultural knowledge accumulation. The intuition is that in areas with fewer (experienced) farmers, especially skilled, educated farmers, information constraints were more likely to bind (i.e., without word-to-mouth communication). Thus the marginal value of soil surveys should be higher in such areas by filling this knowledge gap.

Table 9 presents these results. Panel A and B compare counties above vs. below the median number of total farmers per county in 1880. In Panel C and D, I compare counties by the number of literate farmers. This measure better captures the local capacity to absorb knowledge about local soil quality. In both dimensions, I find the results are consistent with the learning effect.

Specifically, the increase in *log(value of machinery)* is 0.101 in counties with fewer farmers in 1880, compared to just 0.025 in counties with more farmers. Likewise, fertilizer expenditures rise by 0.377 points in below-median farming counties, but only 0.150 points in above-median ones. These differences are even more pronounced when focusing on farmers' literacy: low-literacy counties experience larger gains in all three investment margins (machinery, fertilizer, and land ownership). The revenue per acre effect is positive and significant across all groups, but its magnitude is consistently larger in counties with fewer or less literate farmers. This again suggests that even where productivity was increasing across the board, the informational value of NCSS maps was particularly important in more informationally disadvantaged contexts.

In Table 10, I also show the results based on a similar specification in Equation 9. Theses estimates again confirm that among counties with fewer farmers prior to the soil mapping, the impacts are stronger for agricultural productivity, the use of machinery and fertilizer, and the share of self-own farmland. Taken together, these results reinforce the interpretation of soil mapping as a technology of learning. The larger effects in counties with fewer (less literate) farmers point to the role of NCSS in reducing informational frictions and facilitating the translation of technical knowledge into improved input use, as well as productivity.

Heterogeneity by Specialized Soil Knowledge. Lastly, I examine whether this learning effect is stronger in counties with less specialized soil knowledge prior to the mapping. If standardized information indeed filled existing knowledge gaps in soil quality, these areas are more likely benefited from mapping. In particular, I construct a proxy for local soil knowledge: the share of improved farmland per county. This is simply motivated by a statement from

Stoll (2003):

> *The fertility of soil became a source of worry for a group of North American planters and farmers in the decades following the War of 1812 ... Most of all, they recognized a link between an enduring agriculture and an enduring society in the long-settled places and picked up an old word to name their efforts: improvement ... improvement meant the changes that enabled land to be cultivated in the most prosperous possible over the longest possible time...*

Leveraging information on improved farmland acres in the 1890 Census of Agriculture, I use the following specification to test the idea.

$$y_{ct} = \alpha + \beta_1 NCSS_{ct} \times share_{c,1890} + \beta_2 NCSS_{ct} + \beta_3 share_{c,1890} + \theta_c + \phi_t + \epsilon_{ct} \tag{10}$$

In this specification, β_1 captures the heterogeneous treatment effect of soil mapping conditional on prior soil knowledge: a negative β_1 would imply that counties with less specialized soil knowledge (i.e., lower shares of improved farmland) gained more from standardized soil information. Intuitively, the NCSS filled larger informational gaps in these areas, helping farmers correct misperceptions about soil fertility and adopt more appropriate cultivation practices.

I report these estimates in Tables 11 and 12. Consistent with the hypothesis, both the split-sample and continuous-interaction it shows that there are stronger gains from mapping in counties that initially had less specialized soil knowledge. As shown in Table 11, the estimated treatment effects are roughly twice as large for below-median areas, particularly for revenues, machinery use, and fertilizer adoption. Table 12 reinforces this pattern: the negative coefficients on the interaction term imply that the productivity and technology-adoption benefits of mapping decline with higher pre-existing improvement activity. In short, standardized soil information primarily benefited places that had more limited prior understanding of soil quality, consistent with an information-gap or "catch-up" interpretation rather than simple complementarity.

7.1.5 Significance of Learning Effect

To assess the importance of producers' input adjustments in explaining the productivity gains from soil mapping, I conduct an additional exercise following the approach of Kantor and Whalley (2019) and estimate the total factor productivity (TFP) effect. The basic idea is to capture to what extent the learning channel could explain my findings in productivity and

crop concentration.

To do so, I estimate the following specification, which is a modified version of the baseline DD model not including variables in per acre terms but rather county totals. Meanwhile, I directly control for rural population, farm equipment, fertilizer expenditure, farmland acres, and market access information. I also include the interaction term between farm equipment and fertilizer expenditure. This allows me to directly test whether observed changes in input use account for the effects of NCSS mapping on farm productivity.

$$\begin{aligned} y_{ct} =& \alpha + \theta NCSS_{ct} + \beta_1 log(machinery)_{ct} + \beta_2 log(fertilizer)_{ct} + \beta_3 log(machinery) \\ & \times log(fertilizer) + log(farmlandacre)_{ct} + log(rural\ population)_{ct} + \\ & log(market\ access)_{ct} + \phi_c + \mu_t + \epsilon_{ct} \end{aligned}$$

where the outcomes of interest include both the total crop revenue of each county and the crop concentration indexes. If access to soil quality information prompted farmers to reallocate or intensify input application efficiently, then controlling for these inputs should attenuate most of the estimated effects from the NCSS treatment indicator (θ). A substantial reduction in the magnitude and statistical significance of θ would thus support the hypothesis that input adjustment is the primary channel.

I report these estimates in Table 13 and Table 14. Table 13 presents the results for total crop revenue. Across all counties (Columns 1–3), the estimated coefficient on the NCSS mapping indicator is small and statistically insignificant (or significant at the 10% level) when inputs are controlled for. This finding persists even when adding region-by-year fixed effects and applying historical acreage weights (Column 3). Focusing on the sample of adjacent counties (Columns 4-6), a similar pattern persists: the baseline estimates still show insignificant effects. Table 14 also presents estimates for crop concentration after controlling for the full set of input variables. Consistent with the revenue results, the coefficients are substantially attenuated, reinforcing the interpretation that the NCSS mapping program primarily influenced productivity and specialization through changes in farmers' input decisions, rather than through channels unrelated to production behavior.

These patterns suggest that once producers' input choices are fully incorporated into the analysis, the estimated effect of NCSS mapping on total crop revenue becomes statistically insignificant, and its magnitude is substantially attenuated. Specifically, if there are other key interpretations for why mapping shaped local agricultural development, the estimated effects in the baseline specification and the TFP effects should be somehow similar. For example, if private investment companies also responded strategically (as suggested in Section

2), NCSS might substantially change local land values, attract private capital, or enhance credit availability, resulting in observed increases in output or revenue that are not fully accounted for by input choices. Similarly, if local governments leveraged these maps to inform tax assessments or revise property-based levies, mapping might alter the fiscal environment that influence land investment, not directly affecting farmers' behaviors. In this sense, the divergence between revenue-based and input-adjusted estimates helps isolate the behavioral margin of farmers as a key pathway of impact.

7.2 Role of Local Agricultural Innovation

Another key mechanism through which NCSS mapping could have increased agricultural productivity is by stimulating the development and diffusion of new agricultural technologies. Two specific scenarios are plausible: First, the NCSS program may have reduced coordination frictions among public agencies and encouraged local institutional innovation (e.g., agricultural experiment station research on soil and crops). In many rural regions, fragmented governance and asymmetric information hindered policy implementation and constrained the quality of technical assistance delivered to farmers. Consequently, by providing a informational foundation, NCSS mapping can provide better coordination among agencies, redirect the targeting of technical advice, and ultimately improve farm-level productivity.[20] Second, access to localized soil quality data could also have stimulated broader innovation activities (e.g., patenting) through more precise identification of agricultural conditions and production constraints. Inventors developed new fertilizers, machinery designs, or crop varieties tailored to local soil characteristics, further facilitating technology diffusion and productivity growth.

To empirically test this channel, I first compile records and bulletins from agricultural experiment stations (1896-1920) to construct a *state/county-by-year* panel of research activities, measured by the number of soil-related experiments (i.e., categorized as research relevant to soil, fertilizers, and crops), conducted at each station. I prefer to focus on the state-level counts because most states have only one primary experiment station, making dis-aggregation to counties infeasible in many cases (Kantor and Whalley, 2019). As a robustness check, I also use the county-year counts, which includes all counties with main agricultural experiment stations, and the primary estimates hold. In Figure 31, I present one such example from the 1907-1908 annual report from Illinois.

These data enable me to examine whether the diffusion of standardized NCSS soil maps

[20] Agricultural experiment stations played a central role in disseminating technological knowledge and extension services to local communities (Kantor and Whalley, 2019). In the absence of standardized soil information, these institutions risked offering misaligned or suboptimal recommendations, thereby diminishing their effectiveness.

coincides with shifts in experimental focus toward region-specific soil types, as well as any corresponding increases in the volume of station-led trials. I further explore whether counties receiving NCSS mapping exhibit a greater alignment between local soil conditions and experiment topics, which would suggest more targeted and responsive research agendas following the dissemination of detailed soil information. The regression specification is listed as follows:

$$log(\#of\ publications)_{(c)st+1} = \alpha + \beta log(\#of\ NCSS)_{(c)st} + \theta_t + \mu_(c)s + \epsilon_{(c)st} \tag{11}$$

where c denotes county and s represents state. Results are presented in Table 15. Panel A presents the estimates based on the state-year sample. I find that states having an increased number of NCSS programs experience significantly higher likelihood of at least one relevant publication (Column 1) and a larger total number of publications (Column 2). In Panel B of this table, I also repeat the exercise at the county level, and the sample is a balanced panel including all counties having the primary experiment station. Compared to the state-year level analysis, focusing on county-level results would better capture the local effects. The results remain robust: having the county NCSS program in place is associated with a 9.8 percentage point increase in the likelihood of observing a publication and a 22.9 log points increase in the number of publications. Together, these findings suggest that the diffusion of standardized soil mapping programs shaped the focus and dissemination of experiment station outputs addressing soil science.

Yet, a key question is then whether these mapping-induced, new technologies got *diffused* and *delivered* to local farmers. I further examine whether the effects of NCSS mapping vary depending on a county's distance to nearby agricultural experiment stations. If NCSS mapping amplifies the benefits of proximity to these stations, it would suggest that the mapping program enhanced the diffusion of *new* localized agricultural research. As shown in Figures 32-37, however, I do not find substantial differences in the outcomes of interest between counties near experiment stations and those located farther away. This finding thus suggests that, in contrast to Kantor and Whalley (2019), the primary mechanism through which NCSS mapping improved productivity was not the geographic diffusion of new research activities, but the reduction of preexisting soil information frictions.

In terms of broader agricultural innovation, I use the number of *patents* related to soil and fertilizer technologies as a proxy. As shown in Table 16, NCSS adoption is in fact associated with a modest decline in both general agricultural and soil-relevant patenting. One possible interpretation is that access to improved soil quality information effectively alters farmers' perceptions of climate-related agricultural risks, thereby reducing the necessity of technolog-

ical adaptation. For example, as discussed in Moscona and Sastry (2022), new agricultural technologies and favorable climatic conditions can act as substitutes in production. This divergence supports the interpretation that the increase in soil-related experiment station research is more likely attributable to the coordination benefits of NCSS mapping, rather than to confounding upward trends in agricultural innovation.

Taken together, these two sets of results suggest that though the NCSS mapping program effectively stimulated institutional research activities on soils and crops, its influence neither translate into broader diffusion of these innovations nor led to a general acceleration of technological development.

7.3 Aggregate Market Demand Shocks

It is important to acknowledge that the documented changes in agricultural outcomes may, in part, reflect aggregate market demand shocks, rather than the direct effects of soil mapping. This concern can arise at both the local and national levels.

First of all, during parts of the sample period, agricultural markets were not fully integrated Kantor and Whalley (2019), so access to better soil quality information could have affected *local* commodity prices in the baseline revenue-based land productivity measure. To mitigate this concern, I construct an alternative, robust measure of crop productivity: the ratio between county-level crop revenue per acre in 1890 and this revenue in subsequent years (1900, 1910, 1920, and 1930). This normalization helps tease out time-varying differences in price levels across counties and focuses more on within-county changes in productivity over time.

$$relative\ crop\ revenue_{ct} = \frac{\sum_1^9 price_{it} \times Q_{ict}}{\sum_1^9 price_{i,1890} \times Q_{ic,1890}} \tag{12}$$

where i refers to crop type (wheat, corn, barley, buckwheat, potato, sweet potato, rye, cotton, and oat), c means county, and t means Census years. $Price_{it}$ is the national price of crop i in year t and Q_{ict} is the yield per acre of crop i in county c and year t. The main advantage of this metric is that it absorbs the influence of persistent local price, since each county's productivity is benchmarked to its own pre-mapping baseline.

Table 17 presents the results. Columns (1)–(2) use 1890 as the benchmark year and include all adjacent counties in the sample, while Columns (3)–(4) use 1900 as the benchmark year to test robustness.[21] Across specifications, the results remain consistent with the baseline: the relative

[21]Sample sizes differ because I exclude extreme values. Specifically, I drop observations above the 99th percentile of the alternative productivity measure.

productivity measure shows very similar effects, suggesting that the findings are unlikely to be driven by local commodity price shocks or incomplete market integration. Figures 38 and 39 further illustrate this with event-study estimates, confirming the robustness of my main results.

Second, aggregate market demand shocks following the rollout of the NCSS mapping program may have occurred at the national level, particularly for staple crops such as corn and wheat, which were widely traded in increasingly integrated markets. If the NCSS program improved the productivity of these nationally dominant crops, the resulting supply expansions could have influenced market-wide prices through national price signals. In such cases, even purely local changes in land productivity may partly reflect broader shifts in aggregate market conditions, rather than genuine improvements attributable to localized soil information. This further highlights the need to interpret revenue-based productivity measures with caution in the presence of national market linkages.

Similarly, Kitchens and Rodgers (2023) documents substantial volatility in agricultural prices during the 1910s and 1920s, especially around World War I, when wartime demand drove sharp commodity price spikes followed by steep declines in the early 1920s. These large, nation-wide price swings generated common shocks across counties, potentially confounding my estimates if not addressed. Put differently, observed increases in county-level crop revenues during this period may have been driven not only by local productivity responses to soil mapping but also by broader national demand booms and busts.

To address this concern, I directly control for the extent to which each county got exposed to national price fluctuations during the sample period. I construct a county-level crop index similar to Kitchens and Rodgers (2023) as follows:

$$CropIndex_{ct} = \sum_{i} price_{it} \times Q_{ic,1890} \tag{13}$$

where i refers to crop type (wheat, corn, barley, buckwheat, potato, sweet potato, rye, cotton, and oat), c means county, and t means Census years. $Price_{it}$ is the national price of crop i in year t and $Q_{ic,1890}$ is the yield per acre of crop i in county c and year 1890. This measure captures the extent to which national price shocks from 1910 to 1930 influenced local crop production of each county based on one's historical crop yield.

The regression model I used in this exercise is specified as:

$$y_{ct} = \alpha + \beta_1 NCSS_{ct} + \beta_2 ln(CropIndex_{ct}) + \theta_c + \omega_t + \epsilon_{ct} \tag{14}$$

In this case, β_1 captures the baseline effect of the NCSS mapping program in counties and β_2 shows the effects from national crop price fluctuations. As shown in Table 18, the main estimates for the NCSS mapping program remain robust after accounting for potential confounding effects of national demand shocks. Panel A presents results using the full county sample, while Panel B restricts the analysis to adjacent counties, which provides a tighter comparison by holding constant regional growing conditions and market access. Across both samples, the coefficients on $NCSS_{ct}$ remain positive and statistically significant.

Importantly, the estimated magnitudes are not only significant but also closely aligned with the baseline findings reported in Tables 1, 2, and 5. This consistency in both direction and size of the effects reinforces the conclusion that the documented changes in local agricultural outcomes are not simply a response to national price swings. Instead, they reflect improvements tied to the dissemination of soil information through the NCSS program. Taken together, these results strengthen the credibility of the baseline estimates: even in periods marked by substantial volatility in agricultural prices, the mapping program's impact persists across different samples and empirical specifications.

7.4 Knowledge Diffusion through (Co-ethnic) Networks

Another alternative explanation for the documented changes in agricultural outcomes is the diffusion of soil information through co-ethnic networks, particularly among immigrant communities. As mentioned in Raz (2025), soil heterogeneity slowed farmers' agricultural learning and prompted those who depended on social networks to migrate elsewhere. In the case of NCSS mapping, social networks in turn could played a key role in facilitating the interpretation and adoption of technical soil data, especially in areas where language barriers or the lack of pre-existing specialization might otherwise hinder access. In other words, the establishment of NCSS mapping program mitigated the effects of soil heterogeneity on social learning. Prior research shows that communities with a higher share of immigrants tend to have greater agricultural productivity (Sequeira, Nunn and Qian, 2019; Boberg-Fazlić and Sharp, 2024).

To test this channel, I examine whether the share of immigrant farmers changed in response to the NCSS mapping program. If co-ethnic diffusion amplified the program's effects, we might expect either selective in-migration of immigrant farmers into mapped counties or differential retention patterns over time. Using Census of Population, I estimate two difference-in-differences models where the outcome is the share of foreign-born individuals reporting agriculture as their primary occupation.

The first DD specification is quite similar to my baseline model

$$\frac{\#immigrant\ farmers_{ct}}{population_{ct}} = \alpha + \beta NCSS_{ct} + \theta_c + \mu_t + \epsilon_{ct} \tag{15}$$

where $\frac{\#immigrant\ farmers_{ct}}{population_{ct}}$ is the share of immigrant farmers per county in year t.[22]

The second specification is shown as follows:

$$\frac{\#immigrant\ farmers_{c,1920}}{population_{c,1920}} - \frac{\#immigrant\ farmers_{c,1880}}{population_{c,1880}} = \alpha + \beta NCSS_{1920} + \theta_c \tag{16}$$

where the outcome is the change in immigrant farmer share in 1920 relative to the baseline year 1880, and the treatment variable captures whether a county had been mapped by 1920. The sample is focused on 1920 and I include county FEs. The key coefficient β would capture the impact of NCSS mapping on the relative change in the share of immigrants.

The main results are reported in Table 19. Columns (1)–(4) present estimates based on Equation (10), while Column (5) reports results from Equation (11). In Columns (1) and (2), the outcome is a binary indicator for the presence of any foreign-born farmers; in Columns (3)–(5), the outcome is the natural logarithm of the number of foreign-born farmers. Because I control for total county population (*log(population)*), these log specifications can be interpreted as effects on the share of immigrants within the farming population. The even-numbered columns include year-region fixed effects to account for region-specific trends and shocks. Across all specifications, I find no statistically significant impact of NCSS mapping on immigrant farming presence or intensity, ruling out this alternative mechanism about co-ethnic networks.

8 Interactions between Data and Other Infrastructures

Data infrastructures, such as soil mapping programs, provide farmers with critical information that improves their ability to exploit market opportunities. Meanwhile, physical infrastructure, such as transportation networks, reduces transaction costs and expands market access, thereby increasing the incentives for farmers to adopt recommended technologies.

These two forms of infrastructure may play complementary roles in promoting local agricultural productivity. Improved market access may not yield substantial benefits unless farmers also have the know-how, inputs, and information necessary to take advantage of such new

[22]Note that birthplace-specific counts are too small, I thus aggregate across origins for the main specifications.

opportunities. Furthermore, integration of villages into the larger network could expose farmers to potential import competitions from producers in connected village. Access to local input information can enable farmers in villages that gain road access to position themselves, through choice of crop varieties or targeting of particular markets, in order to cope with the potentially negative effects of import competition. As shown in Gebresilasse (2023), simultaneous access to both roads and agricultural extension services substantially increases the adoption of new technologies and facilitates specialization in higher-value crops.

In my context, I provide causal estimates on the joint impact of soil mapping and market access by employing the following difference-in-differences (DD) specification:

$$y_{ct} = \alpha + \beta_1 NCSS_{ct} \times ln(MA)_{ct} + \beta_2 NCSS_{ct} + \beta_3 ln(MA)_{ct} + \theta_c + \gamma_t + \epsilon_{ct} \tag{17}$$

where y_{ct} is the outcome variable, and $NCSS_{ct}$ is an indicator on whether the county c has a NCSS mapping program in year t. $ln(MA)_{ct}$ is the natural logarithm of a county's market access. θ_c and γ_t are county and year fixed effects, respectively. I also consider alternative measures of market access by creating an indicator for whether a county's market access lies above the national median. The key coefficient is β_1, which identifies the extent to which the impact of NCSS mapping depends on the level of market access. It thus provides a causal estimate of complementarity between data and physical infrastructure.

Results are presented in Table 20. Based on these coefficient estimates, the interaction term between NCSS mapping and market access is consistently positive and statistically significant across a range of agricultural outcomes. In the full sample (Panel A), counties with both soil mapping and better market access experience higher revenues per acre (0.105 log points), greater investment in machinery (0.156 log points), and more intensive use of fertilizer (0.724 units) and self-own farmland (5.4 percentage points). These effects suggest that mapping programs are substantially more productive when farmers also have access to markets, consistent with the hypothesis that data and physical infrastructure are complementary.

Notably, the main effects of NCSS mapping and market access alone are either negative or statistically insignificant in all specifications, highlighting that productivity gains stem not from either factor individually, but from their joint presence. One plausible explanation is that while mapping can increase farmers' input use and productivity potential, such investments are only economically viable when access to markets allows returns to outweigh costs. For counties with poor market access, these investments may not pay off, helping to explain the negative coefficients associated with $ln(MA)_{ct}$. This pattern reinforces the notion

that information-driven interventions are most effective when embedded within supportive economic infrastructure.

9 Robustness

A key threat to identification is that counties selected for NCSS soil-mapping were not randomly chosen. If selection criteria correlated with unobservables that also influence agricultural development, then simple comparisons using the DD strategy would substantially confound the impacts of NCSS *per se*. Such concerns arise in several ways. For example, the Division of Soil may have prioritized either lagging counties with "catch-up" potential in agricultural development or fast-growing counties where returns to soil quality information were already high. Anticipatory effects could further bias my estimates if local stakeholders lobbied for a NCSS survey to attract emerging opportunities (e.g., new crop varieties, market access) or cope with adverse shocks (e.g., erosion problems, natural disasters).

To address these concerns, I implement several complementary exercises in this section. First, I examine whether pre-treatment county characteristics systematically predict NCSS assignment, thereby documenting/ruling out potential selection patterns. Second, I implement a synthetic control strategy to construct counterfactuals from control counties that closely resemble the treated counties. Third, I leverage alternative research designs, including within-state border comparisons, to exploit plausibly exogenous geographic variation in survey exposure. Further, I apply a double-lasso procedure to the baseline specification as a principled method of covariate selection, ensuring robustness to high-dimensional confounders. Meanwhile, I benchmark my main estimates against a sample restricted to ever-treated counties, which provides a more comparable set of units and mitigates concerns about structural differences between treated and untreated areas. Eventually, I compare the baseline estimates using various

Meanwhile, it is worth noting that the baseline empirical analysis leverages the county-level variation to identify causal effects of NCSS programs on agricultural outcomes. For counties with similar natural conditions and faces the same market prices, the differences in agricultural productivity can be attributed to the differences in accessing to local soil information. The empirical strategy thus relies on the comparison of adjacent counties as described in Section 6. One caveat of this design is that the treatment timing is staggered, and this the treatment effects are likely to vary over time. Recent studies (Goodman-Bacon, 2021; Sun and Abraham, 2021) point out that the traditional difference-in-differences estimation cannot recover the

dynamic treatment effects with staggered adoption and heterogeneous treatment effects.

9.1 Testing Selection Patterns

Following Johnson, Lavetti and Lipsitz (2023), I first test whether county's pre-existing features and characteristics predict the mapping status of a county. To do so, I collect data from multiple resources and construct relevant "predictors" in 1890 as the baseline year. Agricultural outcomes are gathered from Census of Agriculture (i.e., the size of farmland, the use of fertilizers, value of machinery, crop yields); crop-specific suitability index comes from the GAEZ-FAO dataset; the numbers of local residents and farmers are derived from the Census of Population in 1880.

Table 21 presents the estimation results. Column (1) reports the baseline specification without fixed effects, while Columns (2) and (3) sequentially introduce state and year fixed effects. Across all three specifications, I find a consistently positive and statistically significant relationship between county-level population and the likelihood of being mapped. This pattern suggests that more populous counties were systematically more likely to receive early mapping attention, possibly due to their greater perceived economic or administrative importance.

In terms of crop suitability, it is noteworthy that being mapped by the NCSS is positively associated with a county's agro-climatic suitability for barley. This pattern suggests that areas better suited for barley cultivation may have been prioritized in the mapping process, potentially reflecting early federal interest in promoting particular crop types or responding to prevailing regional agricultural practices. To address concerns about potential selection bias, I further examine whether the effects of NCSS mapping differ systematically by a county's actual yield or theoretical suitability for barley, using the regression specification as follows:

$$\begin{aligned} log(crop\ revenue)_{ct} =& \alpha + \beta_1 NCSS_{ct} \times Barley\ Advantage_c + \beta_2 NCSS_{ct} + \mu_t + \phi_c + \\ & Barley\ Advantage_c \times year + \epsilon_{ct} \end{aligned}$$

In this regression, $Barley\ Advantage_c$ is constructed in two ways: (1) whether a county has positive farmland acreage of barley in 1890, and (2) whether a county is considered as "suitable" or "very suitable" for barley based on its GAEZ score.

Table 22 presents the estimates accordingly. In the first three columns, the indicator "$Barley\ Advantage_c$ is for whether a county's actual barley yield is positive in 1890. This proxy allows us to test whether counties already cultivating barley in the baseline period experienced differential impacts of NCSS mapping, shedding light on the extent to which mapping

efforts amplified *existing* agricultural patterns versus enabling new ones. In the last three columns, this indicator is constructed building on a county's GAEZ score for barley. Regression estimates in this table show insignificant changes associated with the interaction term. Instead, we can confirm that most changes in crop revenue still come from the baseline $NCSS_{ct}$ indicator, alleviating the concern of selection.

9.2 Synthetic Control Strategy

I apply an additional synthetic control strategy, which is a good fit for this empirical setting as mentioned in Abadie (2021). First, SC is appropriate for analyzing outcomes from aggregate entities, such as counties. Second, the abrupt introduction of NCSS mapping triggered a rich treatment structure well-suited for the SC methodology. Finally, because the treatment intensity, i.e., the magnitude of soil mapping and agricultural exploration may vary greatly across counties, the best counterfactual counties may differ across treated counties. Thus, a case study-like comparison could be appropriate.

I implemented the SC method for the treated counties in the NCSS mapping programs from 1890 to 1920. The results from this analysis are shown in Appendix Table A1. Notably, theses results closely mirror those reported in Sections 6 and 7 in terms of statistical significance, direction, and magnitude, providing further validation of the main findings.

9.3 Adjacent County Pair Design

To further address the concern of selection, I restrict my attention to all pairs of adjacent counties and implement the following specification:

$$y_{cpt} = \alpha + \beta NCSS_{ct} + \gamma_{cp} + \rho_t + \epsilon_{cpt} \tag{18}$$

where c refers to county, p means pair of adjacent counties, and t represents the year. y_{cpt} is the outcome for county c in pair p in period t. The regression includes county pair fixed effects ($\gamma_c p$) and time fixed effects (ρ_t).

The analysis includes 2111 county pairs, constructed from a total of 726 unique counties.[23] Within this setting, I consider two types of treatment-control comparisons: (1) counties mapped by the NCSS between 1899 and 1930 versus counties that were never mapped (1305 pairs), and (2) counties that were mapped earlier versus those mapped at a later date (806 pairs). This adjacent-county design provides a more credible counterfactual by leveraging

[23]Due to the structure of the data, including pair-by-year fixed effects would absorb all treatment variation introduced by the NCSS mapping program, making such a specification infeasible.

geographic proximity and similar agro-ecological conditions, thereby mitigating the concern of dissimilarities between the treatment and control counties.

Appendix Table A2 presents the corresponding regression estimates. With the exception of the coefficient on agricultural machinery value per acre, the results align closely with the baseline estimates presented in the main text, both in terms of magnitude and statistical significance. Figures A1 through A5 display the corresponding event-study patterns, which further reinforce the robustness of the baseline findings under this geographically constrained specification.

Additionally, I reconstruct the analysis sample to include only adjacent county pairs that form a balanced panel over the treatment window (specifically, 20 years prior to and 20 years following the treatment year). This restriction ensures consistent observation periods across counties within each pair, thereby strengthening the validity of the comparisons and addressing potential concerns related to differential timing or incomplete data coverage. It leaves 2029 adjacent county pairs and the corresponding results are presented in Table A3 and Figures A6-A10. Once again, the findings closely mirror those from the main specifications, reinforcing the robustness of the baseline results under this more restrictive and temporally balanced sample.

9.4 Excluding Never-Treated Counties

Another concern relevant to identification is that including never-treated counties might introduce bias if these counties differ systematically from treated ones in ways that are unobservable or not fully captured by fixed effects. For example, counties that were never selected for NCSS mapping may have had persistently lower agricultural potential, weaker political connections, or limited institutional capacity—factors that could also influence agricultural outcomes independently of the mapping program.

To address this issue, I restrict the comparison sample to counties that were eventually mapped at some point between 1899 and 1930 and exploit variation in the timing of treatment. By focusing on "eventually treated" counties and comparing those mapped earlier to those mapped later, this approach mitigates concerns about baseline structural differences between treated and never-treated areas while still leveraging variation in the rollout of the NCSS programs. Results of this robustness check are in Table A4 and Figures A11-A15, and remain consistent with the baseline findings.

9.5 Double Lasso Estimates

To further strengthen the identification strategy and address potential concerns about omitted variable bias, I implement the Double Lasso procedure proposed by Belloni, Chernozhukov and Hansen (2014). This method is particularly helpful in high-dimensional settings where a large number of covariates may influence both treatment assignment and outcomes. By applying Lasso-based variable selection to both the treatment and outcome equations, the Double Lasso framework helps isolate the causal effect of NCSS mapping while controlling flexibly for a rich set of potentially confounding county characteristics. I include a wide array of baseline covariates (i.e., population, market access, natural endownment/condition, etc.) as candidate controls. The Double Lasso estimates, reported in Appendix Table A5, remain robust and closely aligned with the baseline results in terms of both magnitude and statistical significance, providing further support for the validity of the main findings.

9.6 Alternative TWFE Estimation

Recent advances in the econometrics literature have highlighted potential limitations of the traditional two-way fixed effects (TWFE) estimator in settings with staggered treatment adoption and heterogeneous treatment effects across units and over time. In particular, studies such as Goodman-Bacon (2021) and Sun and Abraham (2021) show that conventional TWFE estimates may be contaminated by negative weighting or inappropriate comparisons between early- and late-treated units, which can bias the estimated average treatment effects.

To address these concerns, I re-estimate the baseline specification using the estimator proposed by Callaway and Sant'Anna (2021), which explicitly accounts for variation in treatment timing and allows for dynamic, group-time average treatment effects (ATTs). This approach constructs proper counterfactuals by comparing each treatment cohort only to units that are not yet treated at a given point in time, thereby avoiding problematic comparisons across already-treated units. The corresponding results, presented in Appendix Table A6, remain consistent with the baseline estimates and confirm the robustness of the treatment effects over time.

9.7 Agricultural Experiment Stations and Extension Service

As pointed out by Kantor and Whalley (2019), proximity to local agricultural experiment stations significantly enhanced land productivity for approximately two decades, leveraging the late nineteenth-century establishment of these stations at preexisting land-grant colleges across the United States. To test whether access to these stations may confound the observed

effects in my main analysis, I conduct several robustness exercises that explicitly account for spatial proximity to land-grant agricultural institutions and the timing of their activities.

First, note that Kantor and Whalley (2019) leveraged the establishment of agricultural experiment stations due to the Hatch Act of 1887. To reduce potential bias from contemporaneous agricultural innovations driven by these stations, I limit my sample period to 1900–1920 after the initial establishment phase, ensuring that the observed effects are less likely to be confounded by early station-induced productivity shocks. County fixed effects then absorb most of the time-invariant geographic advantages associated with proximity to these institutions.

Table A7 presents these results: during 1900–1920, counties with NCSS mapping experienced a 3.9% increase in revenue per acre and a 6.3% rise in the value of agricultural machinery, suggesting gains in land productivity and technology adoption. Fertilizer costs per acre rose by 0.241 units, reflecting increased usage to improved soil information. There were also modest increases in the share of self-owned farms (1.7 percentage points), the SWI, and the Herfindahl Index. All estimates include county and year fixed effects, and results are robust to controlling for market access, which itself is negatively associated with several outcomes. Taken together, these results highlight the productivity-enhancing role of public data infrastructure cannot be fully explained by the establishment of local agricultural experiment stations.

Second, I also examine the heterogeneous effects of NCSS mapping programs based on a county's proximity to its nearest agricultural experiment station. Specifically, I construct a normalized distance measure following Kantor and Whalley (2019), which scales the raw distance by the average distance to all stations in a given year. I then classify counties into "near" or "far" groups based on whether their normalized distance falls below or above the median, respectively. This approach allows me to assess whether the impact of NCSS mapping is moderated by potential access to complementary agricultural research infrastructure.

Table A8 explores the heterogeneous effects of NCSS mapping programs by counties' proximity to agricultural experiment stations, as measured by normalized distance. In both subsamples, the effects of NCSS mapping on revenue per acre and value of machinery remain positive and statistically significant, suggesting widespread productivity and investment gains regardless of proximity. For counties farther from experiment stations (Panel B), NCSS mapping is associated with a larger increase in machinery value (8.5%) and self-ownership (2.4 percentage points), and a significant rise in crop concentration (HHI and SWI), indicating that these regions may benefit more intensively from improved access to standardized soil information. These patterns suggest that standardized soil information provided by NCSS may have partially substituted for the localized research and technical guidance historically delivered by

nearby experiment stations, particularly in more remote regions. Dynamic effects are also shown in Figures **??-??**, showing these estimates between the two groups are quite similar.

9.8 Weather Station and Meteorological Information

Another key confounding factor that can substantially influence local agricultural development is access to meteorological infrastructure, which provides farmers with timely and localized climate information. As documented by Anand and Li (2025), proximity to weather stations is associated with significant improvements in agricultural productivity. Exploiting the staggered historical rollout of over 30,000 weather stations across the United States, the study shows that counties located closer to newly established stations experienced measurable gains in crop output. This suggests that access to climate-related information may operate through similar informational channels as NCSS mapping, raising potential concerns about omitted variable bias if the timing or location of NCSS programs coincided with expansions in meteorological coverage.

To assess the empirical relevance of this potential confounder, I examine the correlation between these weather station access measures and my treatment variable, where an NCSS mapping program was adopted. Between 1890 and 1930, the correlation between NCSS mapping availability and the binary indicator for access to a local weather station is 0.140. The correlation between NCSS status and the continuous distance measure is 0.054. Both correlations are small in magnitude, suggesting that while meteorological infrastructure may play an important role in agricultural development, *a priori* it is unlikely to be a major source of confounding in the identification of NCSS mapping effects.

To directly address this concern, I follow Anand and Li (2025) and collect data from the Master Station History Report (NCEI), which includes detailed station locations, operation dates, and institutional affiliations. In this exercise, I construct two relevant proxies for a county's access to local weather station service.

First, I geo-code the locations of all weather stations that began operating as early as 1799. For each county seat, I compute the distance to its five nearest stations and determine the earliest year in which the county gained access to a nearby station. From this, I construct a binary indicator capturing whether a county had access to local weather station services during the period 1799–1930. Second, I use the annual distance between each county seat and its nearest operational weather station as a continuous proxy for the availability and quality of weather forecasting services. The underlying assumption is that proximity enhances the timeliness and accuracy of climate information available to local farmers. These two measures together

capture both the extensive and intensive margins of meteorological access. As more stations were established over time, the average distance to the nearest station declined, reflecting an expansion in coverage and information availability.

In Table A9, I summarize the results after incorporating these measures of meteorological access. Panel A includes a binary indicator for the availability of local weather stations. The coefficient on this variable (*"access to weather station$_{ct}$"*) suggests that the effects of gaining access to weather forecasting services are somewhat mixed across agricultural outcomes. Specifically, counties with access to a nearby weather station experience a statistically significant increase of 3.05% in total crop revenue per acre. However, the value of agricultural machinery declines by 3.25%, and the SWI decreases by 0.009. These findings indicate that while improved weather information may enhance crop productivity, it could also lead to changes in technology adoption or cropping strategies that influence machinery use and soil concentration patterns.

In Panel B, I replace the binary measure with a continuous alternative: the log of the distance to the nearest operational weather station (*log(distance to station$_{ct}$)*). The results remain broadly consistent. A 10% increase in the distance to a weather station is associated with a 0.12% increase in crop revenue per acre, a 0.12% decline in agricultural machinery value, and significant reductions in both the SWI and HHI, pointing toward lower local crop concentration. While the positive effect on revenue appears counterintuitive, it may reflect nonlinearities or compositional shifts in land use in more remote areas.

Most importantly, across both panels, the coefficient on the $NCSS_{ct}$ indicator remains statistically insignificant and stable in magnitude, closely mirroring the baseline estimates reported in Sections 6 and 7. This consistency reinforces the robustness of the main findings and suggests that the estimated effects of NCSS mapping are not driven by omitted trends in access to weather forecasting services. In other words, even after accounting for variation in meteorological infrastructure, both in terms of presence and proximity, the influence of NCSS programs on agricultural outcomes remains effectively unchanged. This helps rule out a key confounding factor and lends greater credibility to my baseline identification strategy. Results in Figures A16-A21.

9.9 Spatial Spillovers of NCSS Mapping

Another reasonable concern about my identification strategy is spatial spillover: the NCSS mapping in one county may diffuse information to nearby counties via extension agents, market incentives, and social networks, thereby contaminating the control group. For example, if

local farmers in neighboring counties also adopted the agricultural technologies or practices recommended by the mapped county, then control counties would partially benefit from treatment. In this case, positive spatial spillovers would attenuate my estimates, which only capture the relatively increase in agricultural productivity. By contrast, if mapping attracts farmers, capital, or extension effort away from unmapped counties, controls may be adversely affected, introducing upward bias. For example, farmers from controls could relocate to mapped counties and settle down.

To address these concerns, I estimate several *donut* DD specifications that drop at-risk border counties and benchmark these results with the baseline estimates. Specifically, I construct alternative samples by systematically dropping counties whose county seats lie within a certain distance threshold from the treatment boundary. These specifications allow me to benchmark the robustness of the baseline estimates against the possibility that neighboring counties, particularly those in close proximity, may not serve as valid controls due to shared exposure or spatial diffusion of treatment effects.

These samples include (1) non-neighboring counties whose pairwise distance falls within the 25^{th} percentile of all county-to-county distances nationally, (2) non-neighboring counties whose pairwise distance is between the 25^{th} percentile and the 50^{th} percentile the of all county-to-county distances nationally, and (3) non-neighboring counties whose pairwise distance is between the 50^{th} percentile and the 75^{th} percentile the of all county-to-county distances nationally.

As reported in Table A10, I present my main estimates. Panel A removes counties within the bottom 25th percentile of national county-to-county seat distances; Panels B and C progressively exclude counties within the 25–50th and 50–75th percentile ranges, respectively. Across all specifications, the estimated impacts of NCSS programs on agricultural outcomes remain positive, statistically significant, and stable in magnitude. These findings reinforce the robustness of the baseline results and alleviate concerns that estimated treatment effects are driven by spillovers or local diffusion dynamics.

9.10 Local Natural Endowment and Condition

Lastly, I also assess the robustness of my results by incorporating more proxies for local natural endowment and condition. In particular, I assemble information regarding state-level average temperature and total perception, and local historical natural disaster data (i.e., earthquake (Vranes and Pielke Jr, 2009), tornado (Brooks and Doswell III, 2001), and hurricane (Pielke Jr et al., 2008)).

As shown in Table A11, the estimated effects of NCSS mapping remain positive and statistically significant across most key agricultural outcomes—including revenue per acre, machinery use, fertilizer intensity, and the share of self-owned farmland, even after controlling for this extensive set of meteorological and environmental controls. Notably, the inclusion of natural disaster variables does not attenuate the estimated coefficients on the mapping treatment, suggesting that the observed productivity gains are unlikely to be driven by omitted environmental factors. These results thus strengthen the interpretation that NCSS programs contributed directly to improved agricultural decision-making and outcomes, rather than merely reflecting differences in baseline ecological conditions.

10 Concluding Remarks and Future Work

This paper provides novel empirical evidence for the key role of open soil quality information as an input in the agricultural production process. Leveraging the staggered rollout of the National Cooperative Soil Survey (NCSS) mapping program in the early 20^{th} century, I assess whether counties getting mapped by this program experienced an increase in local agricultural productivity. I find that local total crop revenue per acre increased by 3-5% in mapped counties, relative a control group of counties that were not mapped. These productivity gains persisted for approximately 15 years after the mapping program. Meanwhile, getting mapped is also positively associated with a rise in local crop concentration and the likelihood of soil erosion in the 1930s. My results are robust to various sample, measurements, and specifications.

In terms of channels, I show that the productivity gains were largely driven by adjustments in farmers' input choices (i.e., increased adoption of fertilizers, agricultural machinery, self-own farmland, and labor inputs). In the meantime, I also document evidence that such cooperative soil mapping programs facilitated research activities from government agencies, but did not substantially diffuse among farmers. Taken together, the findings highlight the critical role of government-provided technical information in shaping long-run patterns of agricultural development.

Our setting provides a unique opportunity to test the hypothesis that information-providing institutions/programs affect long-term economic development. Unlike many technologies and institutions previously studied in this literature (e.g., the printing press, national science academies), soil data sharing system is a distinctly local institution focused on disseminating key input knowledge access for everyone. My findings suggest that such institutions can generate both short-run benefits and long-run costs. While the NCSS soil mapping program

played a transformative role by standardizing and disseminating localized technical knowledge in the early 20^{th} century, similar large-scale, publicly accessible information infrastructures continue to shape innovation today. For example, modern platforms like Google Maps provide widespread access to specialized geo-spatial data and help reduce informational frictions across users and sectors. These digital systems merit further investigation as drivers of technological progress and economic development.

These findings carry important implications for contemporary policy: investments in open-access information infrastructure, particularly those that lower barriers to localized technical knowledge may continue to generate substantial returns. However, to maximize their welfare-improving potential, such initiatives should be complemented by institutional capacity-building and safeguards designed to mitigate unintended consequences. As policymakers increasingly rely on data-driven tools to promote rural development, climate resilience, and technological innovation, the historical experience of the NCSS highlights the importance of aligning information provision with other supporting systems and coordinated institutional frameworks, such as local soil conservation services.

References

Aaronson, Daniel, Daniel Hartley, and Bhashkar Mazumder. 2021. "The effects of the 1930s HOLC "redlining" maps." *American Economic Journal: Economic Policy*, 13(4): 355–392.

Abadie, Alberto. 2021. "Using synthetic controls: Feasibility, data requirements, and methodological aspects." *Journal of economic literature*, 59(2): 391–425.

Anand, Vaibhav, and Honglin Li. 2025. "Weather Stations and Agricultural Productivity: Evidence from Historical Data in the U.S."

Andrews, Michael J. 2023. "How do institutions of higher education affect local invention? Evidence from the establishment of US colleges." *American Economic Journal: Economic Policy*, 15(2): 1–41.

Belete, Tegegn, and Eshetu Yadete. 2023. "Effect of mono cropping on soil health and fertility management for sustainable agriculture practices: A review." *Plant Sci*, 11: 192–197.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives*, 28(2): 29–50.

Berkes, Enrico, and Peter Nencka. 2024. "Knowledge access: The effects of carnegie libraries on innovation." *Review of Economics and Statistics*, 1–45.

Boberg-Fazlić, Nina, and Paul Sharp. 2024. "Immigrant Communities and Knowledge Spillovers: Danish Americans and the Development of the Dairy Industry in the United States." *American Economic Journal: Macroeconomics*, 16(1): 102–46.

Bockheim, J.G., and A.N. Gennadiyev. 2015. "General state soil maps in the USA." *Geoderma*, 253-254: 78–89.

Brooks, Harold E, and Charles A Doswell III. 2001. "Normalized damage from major tornadoes in the United States: 1890–1999." *Weather and Forecasting*, 16(1): 168–176.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230. Themed Issue: Treatment Effect 1.

Casaburi, Lorenzo, and Ugo Troiano. 2015. "Ghost-House Busters: The Electoral Response to a Large Anti–Tax Evasion Program." *The Quarterly Journal of Economics*, 131(1): 273–314.

Chan, Jeff. 2022. "Farming output, concentration, and market access: Evidence from the 19th-century American railroad expansion." *Journal of Development Economics*, 157: 102878.

Costinot, Arnaud, Dave Donaldson, and Cory Smith. 2016. "Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world." *Journal of Political Economy*, 124(1): 205–248.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives*, 30(4): 171–98.

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

Edwards, Eric C., and Walter N. Thurman. 2025. "Creating American Farmland: Governance Institutions and Investment in Agricultural Drainage." *The Journal of Economic History*, 85(3): 806–843.

Fishback, Price, Jonathan Rose, Kenneth A. Snowden, and Thomas Storrs. 2024. "New Evidence on Redlining by Federal Housing Programs in the 1930s." *Journal of Urban Economics*, 141: 103462.

Fiszbein, Martin. 2022. "Agricultural Diversity, Structural Change, and Long-Run Development: Evidence from the United States." *American Economic Journal: Macroeconomics*, 14(2): 1–43.

Fox, David M. 2018. "The Differential Undercount: When Government Disclosures of Census

Data Undermine Constitutional Guarantees." *Conn. Pub. Int. LJ*, 18: 131.

Furman, Jeffrey L, and Megan J MacGarvie. 2007. "Academic science and the birth of industrial research laboratories in the US pharmaceutical industry." *Journal of Economic Behavior & Organization*, 63(4): 756–776.

Furman, Jeffrey L, Markus Nagler, and Martin Watzinger. 2021. "Disclosure and subsequent innovation: Evidence from the patent depository library program." *American Economic Journal: Economic Policy*, 13(4): 239–270.

Gardner, David Rice. 1998. *The national cooperative soil survey of the United States.* [Washington, DC]: US Department of Agriculture, Natural Resources ....

Gebresilasse, Mesay. 2023. "Rural roads, agricultural extension, and productivity." *Journal of Development Economics*, 162: 103048.

Gilpin, Gregory, Ezra Karger, and Peter Nencka. 2024. "The Returns to Public Library Investment." *American Economic Journal: Economic Policy*, 16(2): 78–109.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277. Themed Issue: Treatment Effect 1.

Haines, Michael, Price Fishback, and Paul Rhode. 2018. "United States Agriculture Data, 1840 - 2012." *Inter-university Consortium for Political and Social Research.*

Haines, Michael R, Inter university Consortium for Political, Social Research, et al. 2010. "Historical, demographic, economic, and social data: The United States, 1790-2002."

Hartemink, Alfred E. 2021. *Soil Science Americana: Chronicles and Progressions 1860-1960.* Springer Nature.

Hausman, Naomi. 2022. "University innovation and local economic growth." *Review of Economics and Statistics*, 104(4): 718–735.

Hornbeck, Richard. 2010. "Barbed Wire: Property Rights and Agricultural Development*." *The Quarterly Journal of Economics*, 125(2): 767–810.

Hornbeck, Richard. 2012. "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe." *American Economic Review*, 102(4): 1477–1507.

Huang, Jingyi. 2024. "Fence Laws: Liability Rules and Agricultural Development."

Hynsjo, Disa, and Luca Perdoni. 2024. "Mapping out institutional discrimination: The economic effects of federal "redlining"."

Johnson, Matthew S, Kurt J Lavetti, and Michael Lipsitz. 2023. "The Labor Market Effects of Legal Restrictions on Worker Mobility." National Bureau of Economic Research Working Paper 31929.

Kang, Sukhun. 2025. "From Outward to Inward: Reframing Search with New Mapping Criteria." *Academy of Management Proceedings*, 2025(1): 10129.

Kantor, Shawn, and Alexander Whalley. 2014. "Knowledge spillovers from research universities: evidence from endowment value shocks." *Review of Economics and Statistics*, 96(1): 171–188.

Kantor, Shawn, and Alexander Whalley. 2019. "Research Proximity and Productivity: Long-Term Evidence from Agriculture." *Journal of Political Economy*, 127(2): 819–854.

Kao, Jennifer. 2024. "Charted territory: Mapping the cancer genome and R&D decisions in the pharmaceutical industry." *UCLA Anderson*.

Karger, Ezra. 2021. "The long-run effect of public libraries on children: Evidence from the early 1900s." *SocArXiv. Doi*, 10.

Katona, Zsolt, Marcus O. Painter, Panos N. Patatoukas, and Jean Zeng. 2025. "On the Capital Market Consequences of Big Data: Evidence from Outer Space." *Journal of Financial and Quantitative Analysis*, 60(2): 551–579.

Kitchens, Carl, and Price Fishback. 2015. "Flip the Switch: The Impact of the Rural Electrification Administration 1935–1940." *The Journal of Economic History*, 75(4): 1161–1195.

Kitchens, Carl T, and Luke P Rodgers. 2023. "The Impact of the WWI Agricultural Boom and Bust on Female Opportunity Cost and Fertility." *The Economic Journal*, 133(656): 2978–3006.

Kline, Patrick, and Enrico Moretti. 2013. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority *." *The Quarterly Journal of Economics*, 129(1): 275–331.

Larcom, Shaun, Ferdinand Rauch, and Tim Willems. 2017. "The benefits of forced experimentation: Striking evidence from the London underground network." *The Quarterly Journal of Economics*, 132(4): 2019–2055.

Luca, Michael, Abhishek Nagaraj, and Gauri Subramani. 2022. "Getting on the Map: The Impact of Online Listings on Business Performance." National Bureau of Economic Research Working Paper 30810.

Mahecha-Vásquez, Germán, Sair Sierra, and Raúl Posada. 2017. "Diversity indices using arbuscular mycorrhizal fungi to evaluate the soil state in banana crops in Colombia." *Applied Soil Ecology*, 109: 32–39.

Mattheis, Ross, and Itzchak Tzachi Raz. 2019. "There's no such thing as free land: The homestead act and economic development." *Unpublished Manuscript*.

Michel-Kerjan, Erwann O. 2010. "Catastrophe Economics: The National Flood Insurance Program." *Journal of Economic Perspectives*, 24(4): 165–86.

Moscona, Jacob, and Karthik A Sastry. 2022. "Does Directed Innovation Mitigate Cli-

mate Damage? Evidence from U.S. Agriculture*." *The Quarterly Journal of Economics*, 138(2): 637–701.

Nagaraj, Abhishek. 2022. "The Private Impact of Public Data: Landsat Satellite Maps Increased Gold Discoveries and Encouraged Entry." *Management Science*, 68(1): 564–582.

Nagaraj, Abhishek, and Matteo Tranchero. 2023. "How Does Data Access Shape Science? The Impact of Federal Statistical Research Data Centers on Economics Research." National Bureau of Economic Research Working Paper 31372.

Nagaraj, Abhishek, and Scott Stern. 2020. "The economics of maps." *Journal of Economic Perspectives*, 34(1): 196–221.

Novaes de Amorim, Arthur. 2023. "Essays on Productivity and Innovation."

Nunn, Nathan, and Nancy Qian. 2011. "The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment*." *The Quarterly Journal of Economics*, 126(2): 593–650.

Olmstead, Alan L., and Paul W. Rhode. 2001. "RESHAPING THE LANDSCAPE: THE IMPACT AND DIFFUSION OF THE TRACTOR IN AMERICAN AGRICULTURE, 1910–1960." *The Journal of Economic History*, 61(3): 663–698.

Petralia, Sergio, Pierre-Alexandre Balland, and David Rigby. 2016. "HistPat Dataset."

Pielke Jr, Roger A, Joel Gratz, Christopher W Landsea, Douglas Collins, Mark A Saunders, and Rade Musulin. 2008. "Normalized hurricane damage in the United States: 1900–2005." *Natural hazards review*, 9(1): 29–42.

Power, JF, and RF Follett. 1987. "Monoculture." *Scientific American*, 256(3): 78–87.

Pradhan, Archana. 2009. *Economic benefits of the National Cooperative Soil Survey Program.* West Virginia University.

Raz, Itzchak Tzachi. 2025. "Soil Heterogeneity, Social Learning, and the Formation of Close-Knit Communities." *Journal of Political Economy*, 133(8): 2643–2691.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian. 2019. "Immigrants and the Making of America." *The Review of Economic Studies*, 87(1): 382–419.

Smith, Cory. 2019. "Land concentration and long-run development: Evidence from the frontier united states." *Unpublished Manuscript*.

Song, Bei Zhou, Hong Ying Wu, Yun Kong, Jie Zhang, Yan Li Du, Jing Hui Hu, and Yun Cong Yao. 2010. "Effects of intercropping with aromatic plants on the diversity and structure of an arthropod community in a pear orchard." *BioControl*, 55(6): 741–751.

Stoll, Steven. 2003. *Larding the lean earth: Soil and society in nineteenth-century America.* Hill and Wang.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199. Themed Issue: Treatment Effect 1.

Suri, Tavneet, and Christopher Udry. 2022. "Agricultural Technology in Africa." *Journal of Economic Perspectives*, 36(1): 33–56.

Vranes, Kevin, and Roger Pielke Jr. 2009. "Normalized earthquake damage and fatalities in the United States: 1900–2005." *Natural Hazards Review*, 10(3): 84–101.

Williams, Heidi L. 2013. "Intellectual Property Rights and Innovation: Evidence from the Human Genome." *Journal of Political Economy*, 121(1): 1–27.

Figures and Tables

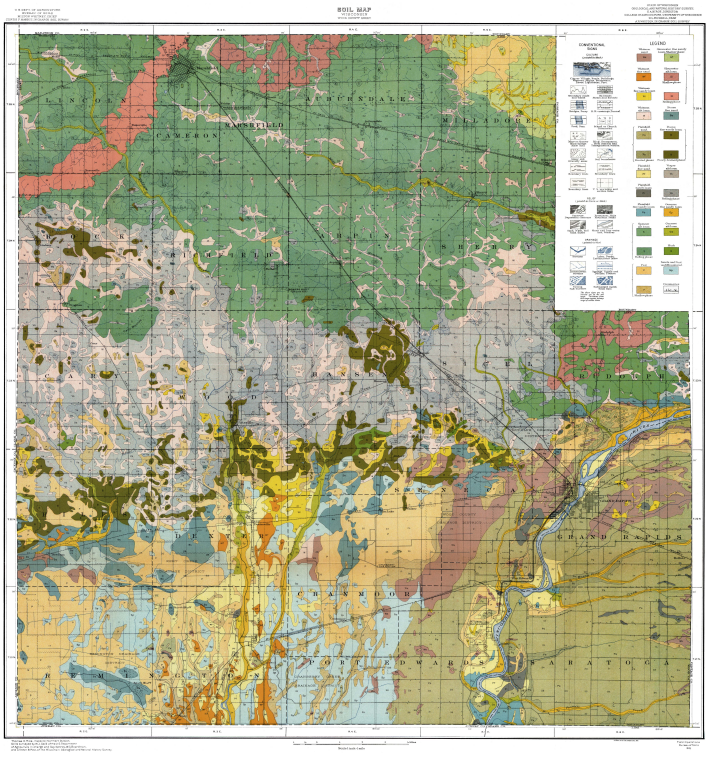


Figure 1: NCSS map in Wood County of Wisconsin, 1917

Areas of different soils.

Soil.	Acres.	Per cent.	Soil.	Acres.	Per cent.
Spencer silt loam	109,440	25.9	Gloucester silt loam, shallow phase	8,000	1.7
Rolling phase	24,384		Gloucester silt loam, rolling phase	512	
Vesper silt loam	60,736	15.0	Whitman fine sandy loam	7,424	1.4
Rolling phase	16,640		Genesee fine sandy loam	7,040	1.4
Peat	49,920	13.1	Gloucester fine sandy loam, shallow phase	6,208	1.2
Shallow phase	17,856		Genesee silt loam	5,312	1.0
Plainfield sand	57,984	11.6	Whitman fine sand	3,840	.7
Eroded phase	2,240		Plainfield sandy loam	2,304	.4
Whitman silt loam	42,112	8.1	Plainfield fine sandy loam	1,216	.2
Sands and Peat (undifferentiated)	28,736	5.6	Muck	960	.2
Boone fine sandy loam	8,896	3.4	Total	517,760	.....
Poorly drained phase	8,896				
Whitman sand	16,640	3.2			
Plainfield fine sand	15,936	3.1			
Boone fine sand	14,528	2.8			

Figure 2: example of NCSS reports for Wood County of Wisconsin, 1917

The soil is especially well adapted to grasses for hay and pasture. With drainage, corn for ensilage would also prove a good crop. Because of the high content of organic matter in most areas of the type, grain would probably lodge badly. Those areas with a very shallow accumulation of organic matter will, upon drainage and cultivation, develop into a soil resembling the more nearly level areas of the Spencer or Vesper silt loams.

Before this soil can be cropped with much success artificial drainage must be supplied. Many of the large areas, because of the level topography, will be rather difficult to drain successfully, and much care will be necessary in laying out a drainage system.

Figure 3: example of NCSS reports for Wood County of Wisconsin, 1917

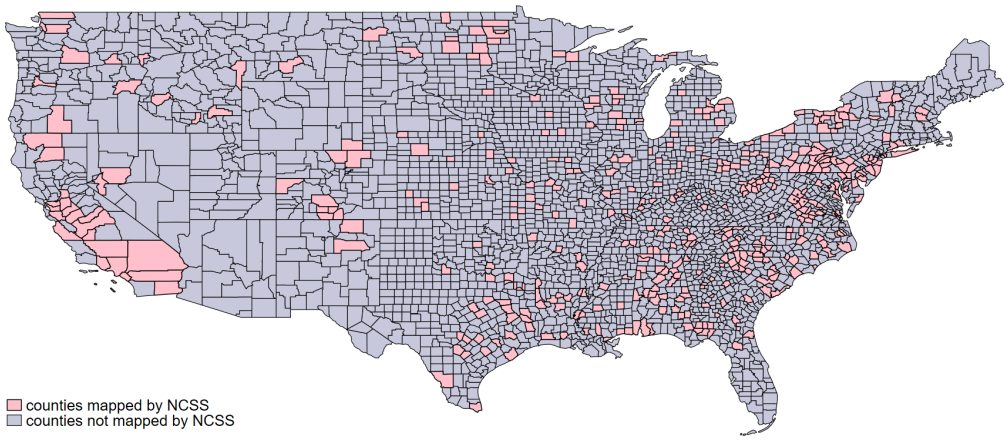


Figure 4: mapping progress as of 1910

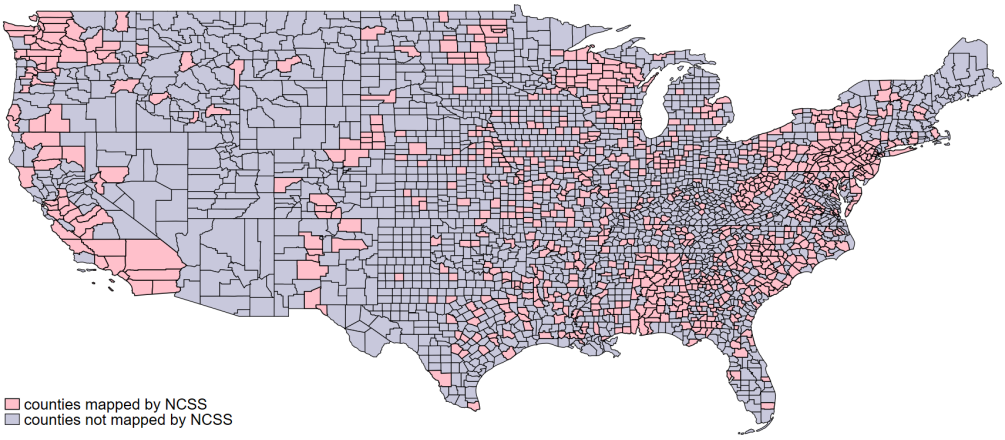


Figure 5: mapping progress as of 1920

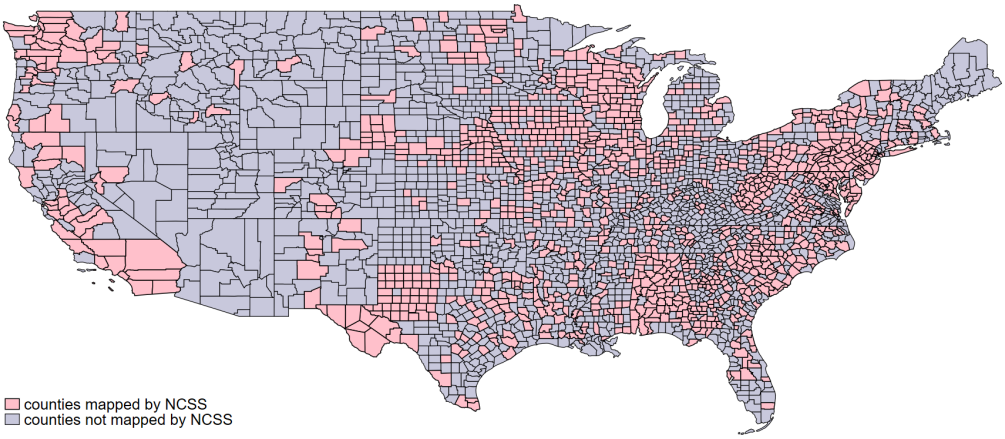


Figure 6: mapping progress as of 1930

Analyses of commercial fertilizers (*Kentucky Sta. Bul. 128, pp. 173–258*).—Analyses and valuations of 651 samples of fertilizers collected during 1906 are reported in tabular form, with brief explanatory notes.

Trade values and fertilizer and soil analyses, C. A. Goessmann and H. D. Haskins (*Massachusetts Sta. Bul. 117, pp. 22*).—This contains a statement of trade values for 1906 and 1907, instructions regarding sampling, and analyses of miscellaneous samples of fertilizers, soils, etc., sent to the station for examination.

Inspection of fertilizers in 1906, F. W. Morse (*New Hampshire Sta. Bul. 130, pp. 8*).—This bulletin reports the results of analysis of 114 samples of fertilizing materials inspected during the year.

Figure 7: a snapshot of soil-relevant station publication (1907)

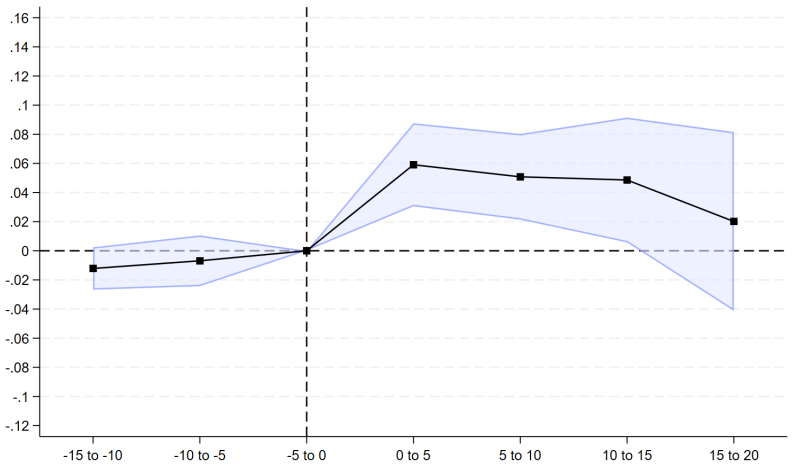


Figure 8: event-study estimates for effects on crop revenue per acre (all counties sample)

Notes: Data are drawn from the Census of Agriculture (1890-1930). The dependent variable is the logarithm of total crop revenue per acre of farmland. Crop revenue is constructed as the total value of production for nine principal crops: wheat, barley, corn, oat, potato, sweet potato, cotton, buckwheat, and rye. For each crop, production is calculated as the quantity harvested multiplied by its corresponding national-level price. The sample includes all counties described in the text (2148 in total). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

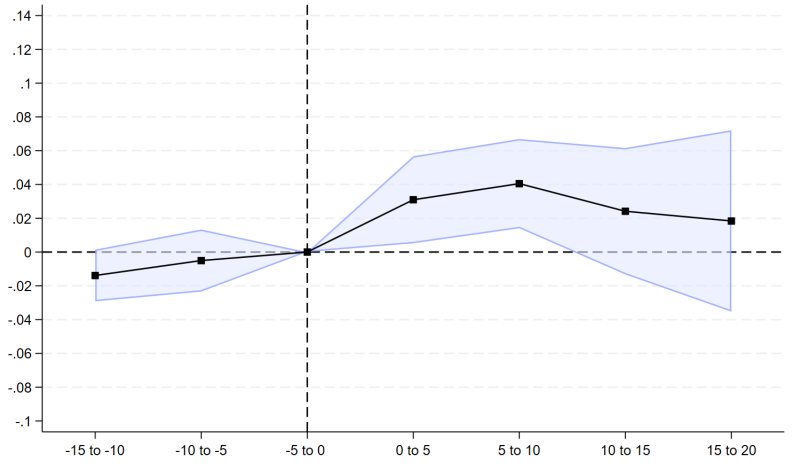


Figure 9: event-study estimates for effects on crop revenue per acre (adjacent counties sample)

Notes: Data are drawn from the Census of Agriculture (1890-1930). The dependent variable is the logarithm of total crop revenue per acre of farmland. Crop revenue is constructed as the total value of production for nine principal crops: wheat, barley, corn, oat, potato, sweet potato, cotton, buckwheat, and rye. For each crop, production is calculated as the quantity harvested multiplied by its corresponding national-level price.The sample includes all adjacent counties (1826 in total). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

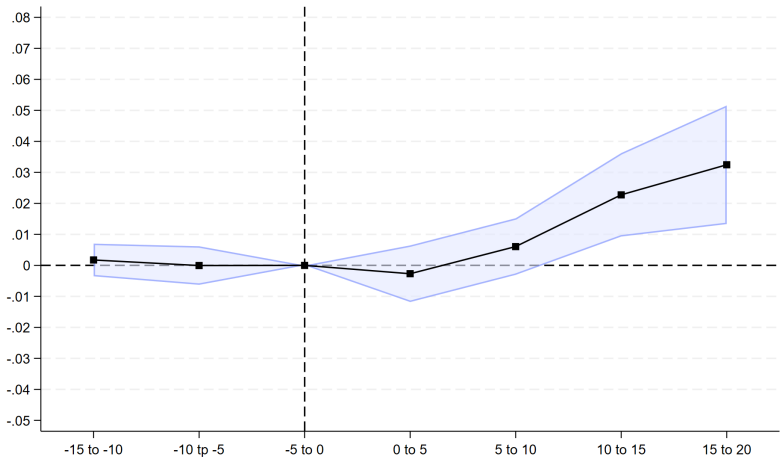


Figure 10: event-study estimates for crop concentration (SWI)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

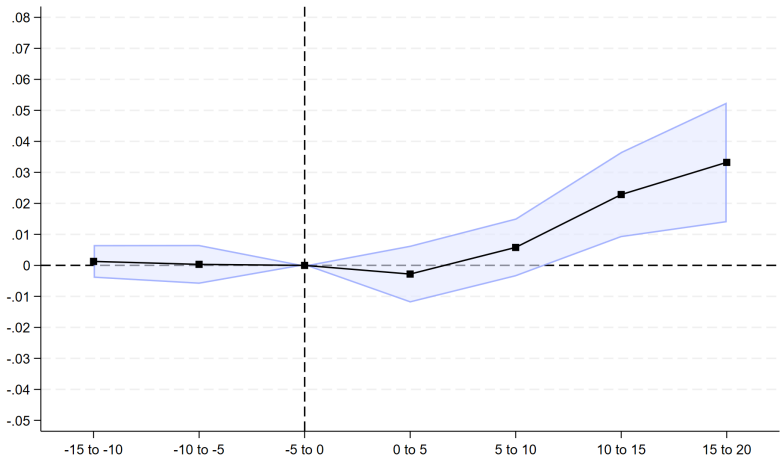


Figure 11: event-study estimates for crop concentration (SWI)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties described in Section 5 (1826 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

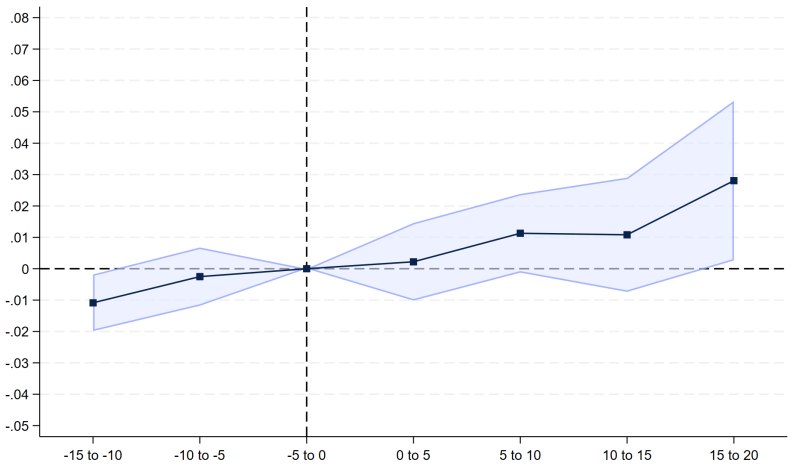


Figure 12: event-study estimates for crop concentration (HHI)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Herfindahl–Hirschman index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

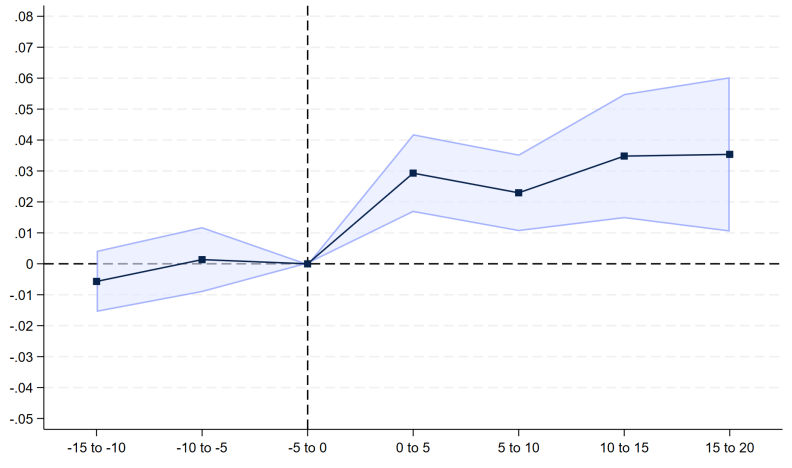


Figure 13: event-study estimates for crop concentration (HHI)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties described in Section 5 (1826 counties). The dependent variable is the Herfindahl–Hirschman index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

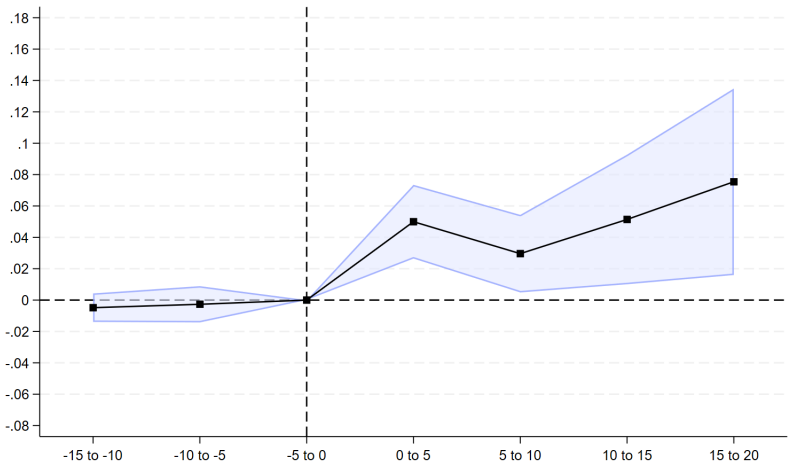


Figure 14: event-study estimates for value of machinery per acre of cropland

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties (1826 in total). The dependent variable is the logarithm of value of agricultural machinery per acre of cropland. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

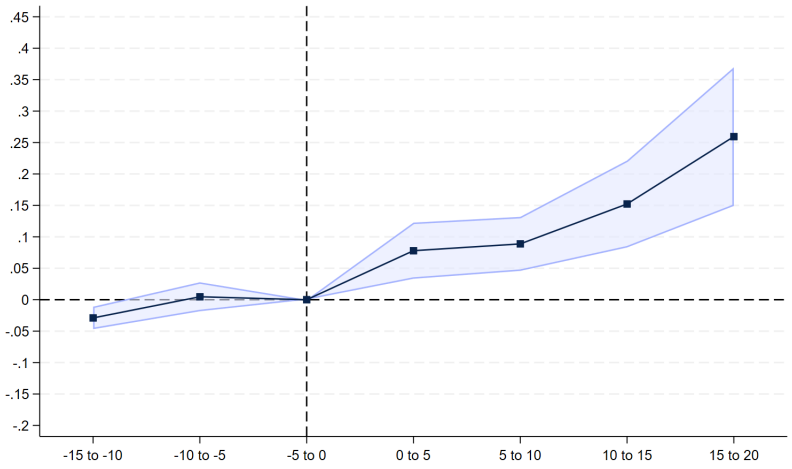


Figure 15: event-study estimates for fertilizer expenditure per acre

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties (1826 in total). The dependent variable is the log of expenses on fertilizer per acre of cropland. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

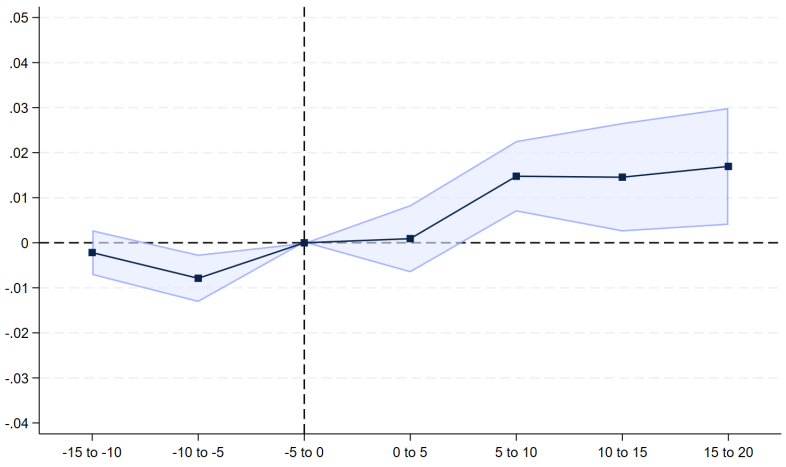


Figure 16: event-study estimates for the share of self-owned farmland

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties (1826 in total). The dependent variable is the share of self-owned farmland per county. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

LIST OF AGRICULTURAL SOCIETIES AND FARMERS' CLUBS

ESTABLISHED TO PROMOTE THE AGRICULTURAL, HORTICULTURAL, AND POMOLOGICAL WELFARE OF THE FARMER, ON THE BOOKS OF THE DEPARTMENT OF AGRICULTURE, JULY 4, 1876, THE CENTENNIAL YEAR OF AMERICAN INDEPENDENCE. GIVING THE YEAR OF ORGANIZATION OF EACH SOCIETY AND OTHER FACTS OF IMPORTANCE AS INDICATED BY THE HEAD-LINES OF THE TABLE, ARRANGED ALPHABETICALLY ACCORDING TO THE COUNTIES IN EACH STATE, TOGETHER WITH THE NAME OF PRESIDENT AND THE NAME AND POST-OFFICE ADDRESS OF THE SECRETARY.

ALABAMA.

	Name.	Year of organization.	Number of members.	Number of meetings held during the year.	Have you a library; if so, the number of volumes?	Do you correspond with the State society?	Do you hold a fair?	Name of president.	Name of secretary.	Post-office address of secretary.	Name of county.	
	State societies.											
	State agricultural society not organized											
	County and district societies.											
1	Autaugaville Agricultural and Hortic'l Society	1871	25	12	No.	No.	No.	T. W. Hall	R. H. King	Autaugaville	Autauga	1
2	Murphric's Valley Agricultural Society	1871	40	12	No.	No.	No.	Stephen Ellis	A. M. Gibson.	Chepultepec	Blount	2
3	Tennessee Valley Agric'l and Mechan'l Ass'n.	1871	125	2	No.	No.	Yes	W. W. Bayless	R. S. McReynolds.	Tuscumbia	Colbert	3
4	Weil's Valley Farmer's and Mechanics' Club	1869	61	4	No.	No.	No.	W. B. Beeson	J. P. Stephens	Greenwood	Etowah	4
5	Ruhama Horti-Agricultural Society	1870	76	12	No.	No.	No.	James Wilson	R. J. Waldrop	Wood's Station	Jefferson	5
6	Madison County Agric'l and Hortic'l Society	1869	31	24	30	No.	No.	George D. Norris	W. F. Whitman	New Market	Madison	6
7	Huntsville Agric'l and Mechanical Assoc'n	1872	125	12	No.	No.	Yes	C. J. Mastin	James H. Bone	Huntsville	...do	7
8	Mobile Cotton Exchange	1872	75	12	No.	No.	No.	A. J. Gillespie	R. H. Bolling	Mobile	Mobile	8
9	Mobile Agric'l, Mech'l, and Hortic'l Assoc'n	1873	300	2	No.	No.	Yes	C. C. Langdon	W. E. Mickle	...do	...do	9
10	Monroe County Agricultural Society	1872	281	6	No.	No.	No.	John W. Leslie	B. L. Hibbard	Monroeville	Monroe	10
11	Scotland Agricultural Society	1876	17	10	No.	No.	No.	W. T. Nettles	W. S. Neall	Kempsville	...do	11
12	High Pine Farmers' Club	1870	32	12	No.	No.	No.	M. J. Mickle	B. F. Weathers	Roanoke	Randolph	12
13	Boiling Spring Agricultural Club	1873	41	12	No.	No.	Yes	A. B. Cooper	J. E. Wilkinson	Prairie Bluff	Wilcox	13

Figure 17: snapshot of the list agricultural societies and farmers' clubs

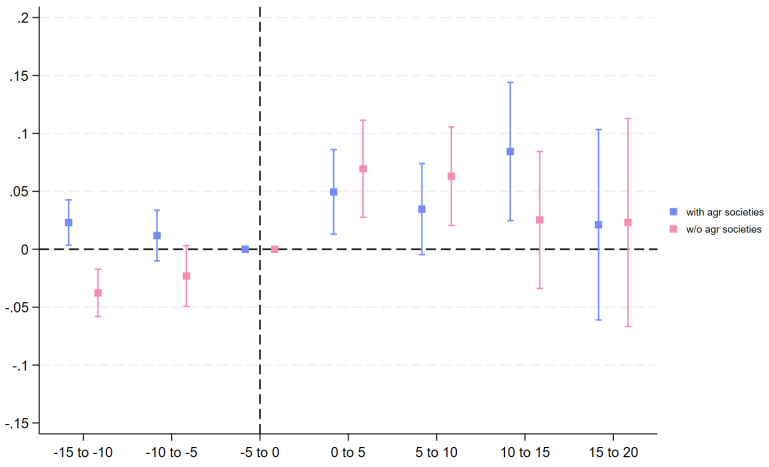


Figure 18: effects on crop revenue, by the existence of county agricultural societies in 1876

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the logarithm of total crop revenue per acre. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

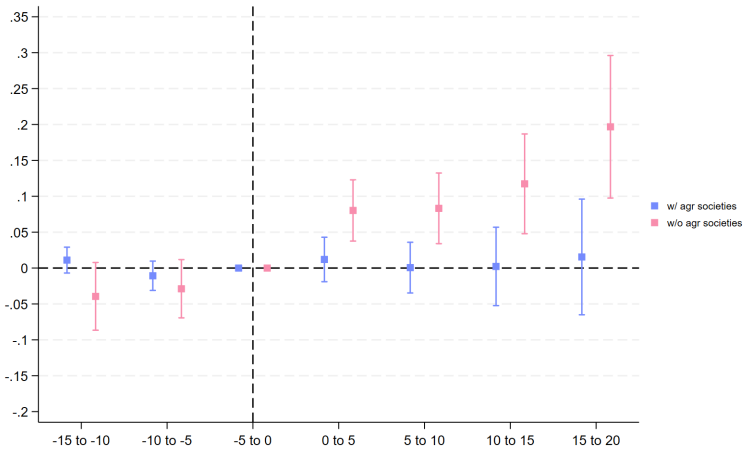


Figure 19: effects on agricultural machinery, by the existence of county agricultural societies in 1876

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the logarithm of value of machinery per acre of farmland. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

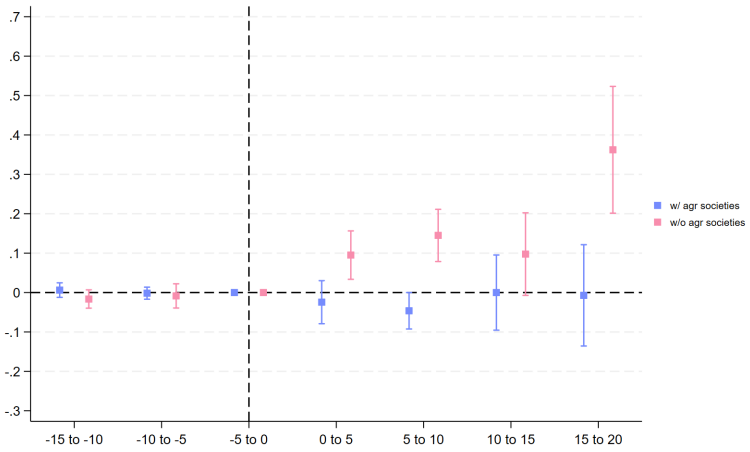


Figure 20: effects on fertilizer expenditures, by the existence of county agricultural societies in 1876

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is fertilizer expenditure per acre of farmland. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

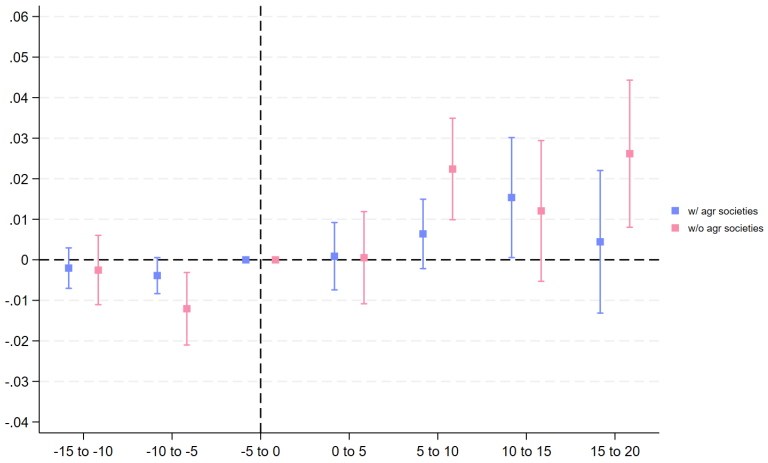


Figure 21: effects on the share of self-own farmland, by the existence of county agricultural societies in 1876

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the share of self-own farmland per county. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

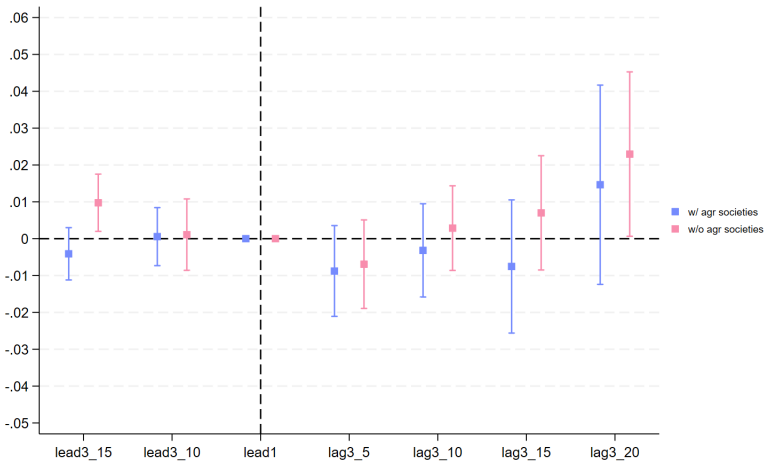


Figure 22: effects on SWI, by the existence of county agricultural societies in 1876

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

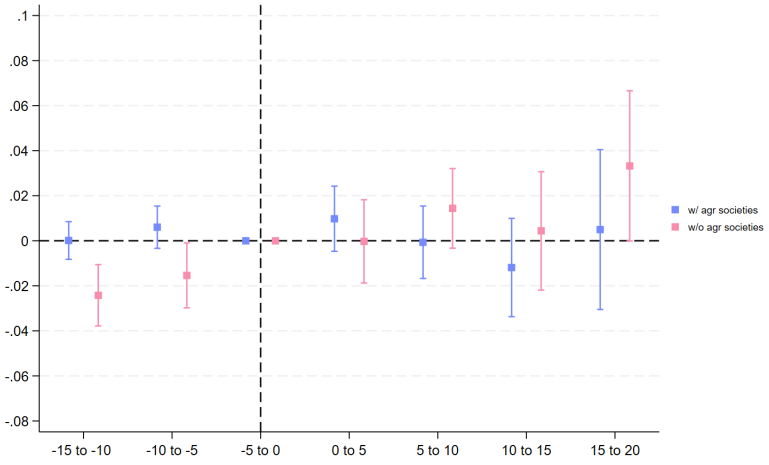


Figure 23: effects on HHI, by the existence of county agricultural societies in 1876

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the HHI in each county. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

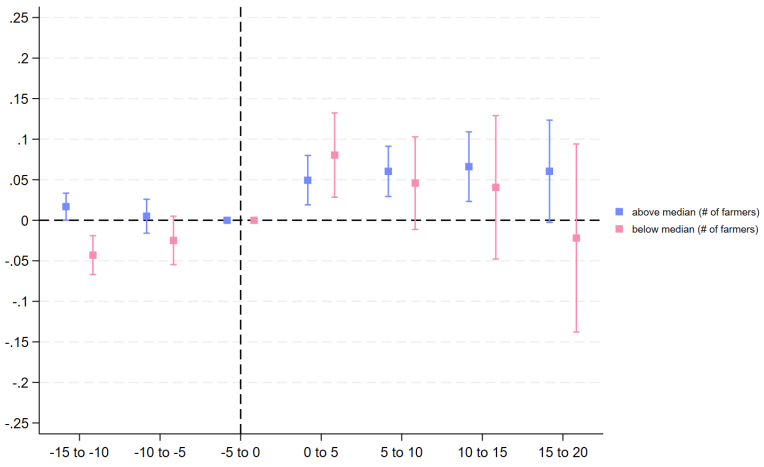


Figure 24: effects on crop revenue, by the number of farmers in 1880

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the logarithm of total crop revenue per acre. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

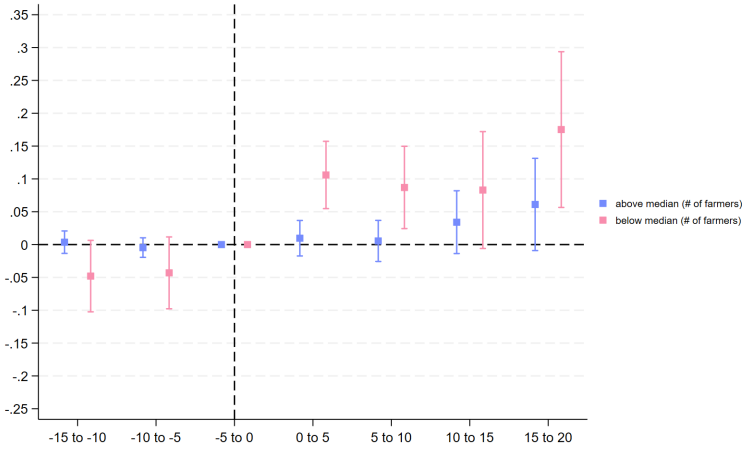


Figure 25: effects on agricultural machinery, by the number of farmers in 1880

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the logarithm of the value of agricultural machinery per acre. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

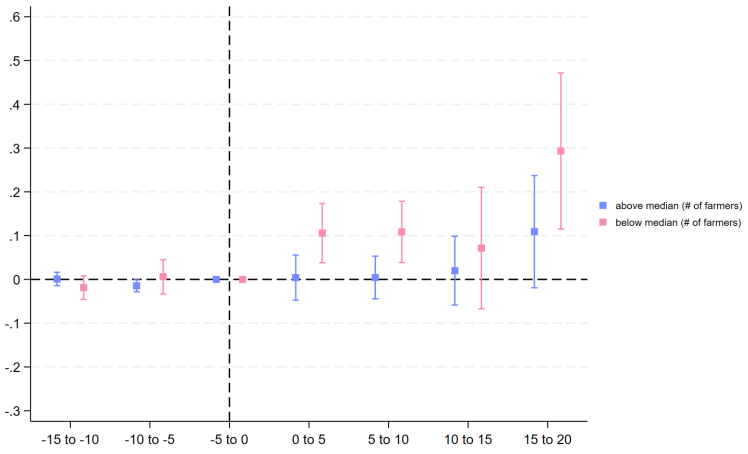


Figure 26: effects on fertilizer expenditures, by the number of farmers in 1880

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the logarithm of fertilizer expenditures per acre. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

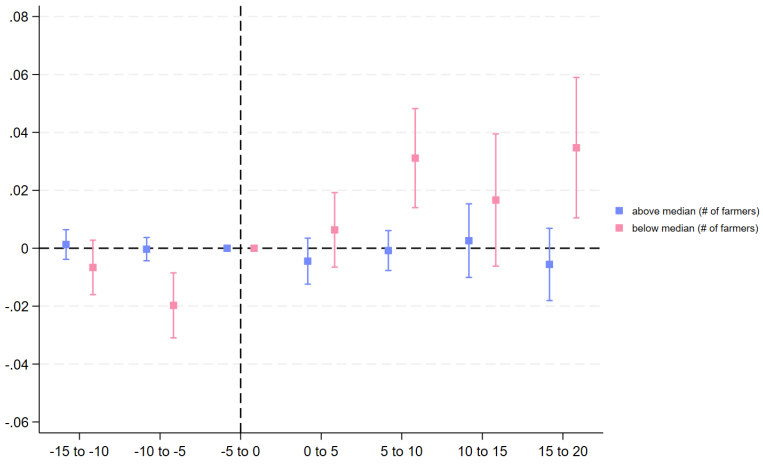


Figure 27: effects on the share of self-own farmland, by the number of farmers in 1880

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the share of self-own farmland per county. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

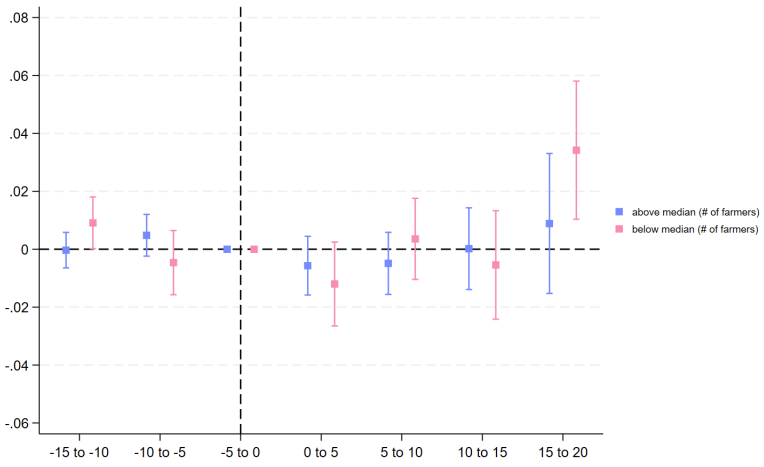


Figure 28: effects on SWI, by the number of farmers in 1880

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

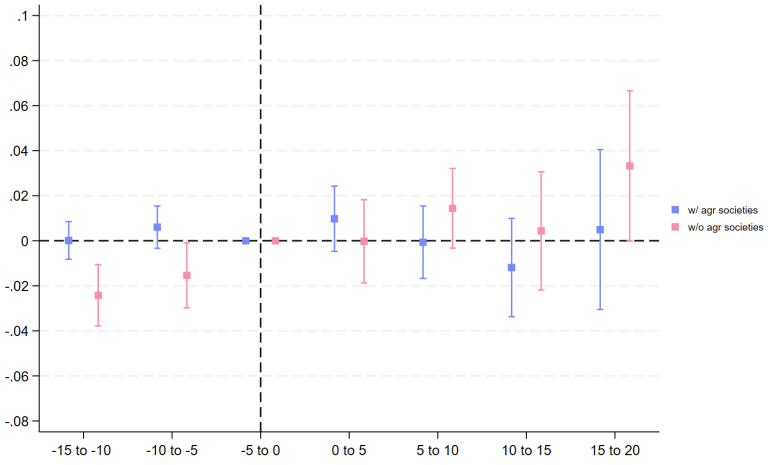


Figure 29: effects on HHI, by the number of farmers in 1880

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the HHI. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

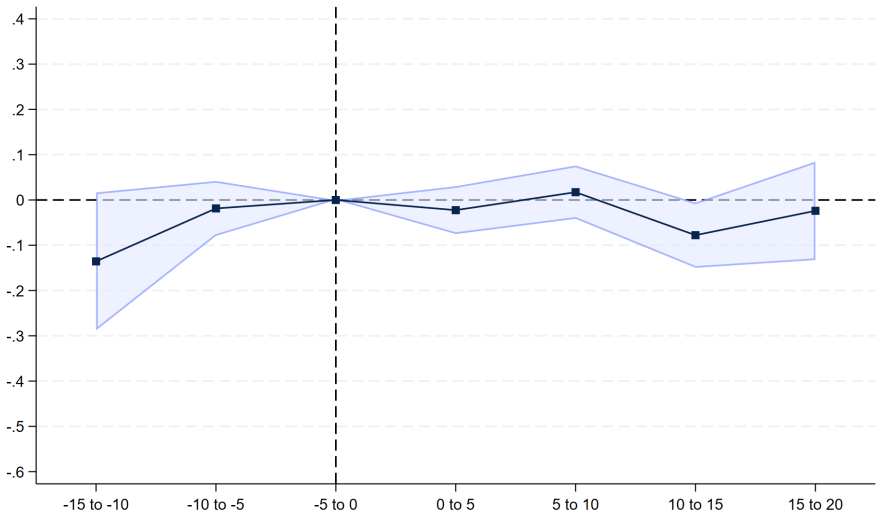


Figure 30: TFP effects of NCSS mapping program

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Soil improvement for the worn hill lands of Illinois, C. G. Hopkins and J. E. Readhimer (*Illinois Sta. Bul. 115, pp. 431–443*).—This bulletin is a report of progress in the land improvement work carried on by the station, chiefly in southern Illinois. A plan of the experiments and earlier results have been previously noted (E. S. R., 17, p. 361).

In summarizing the results with wheat on the Vienna experiment field, taking into account the total yields for the last 3 years, it is shown that untreated land produced 11.8 bu. per acre, while with a leguminous crop the yield increased 11.5 bu., and with lime used in addition to the leguminous crop the increase in yield was 34.3 bu. During this period the use of phosphorus gave a further increase of only 5.6 bu. per acre, and the use of potassium of 10.8 bu. It is pointed out that the direct effect of the lime applied in the form of ground limestone is to increase the growth of the crops for green manuring.

In the corn tests for the last 3 years the untreated land produced 109.2 bu. per acre. The use of a leguminous crop alone increased the yield only 9.8 bu., while with the further addition of lime an increase of 50.7 bu. was secured.

Figure 31: example of soil-relevant records from agricultural experiment station

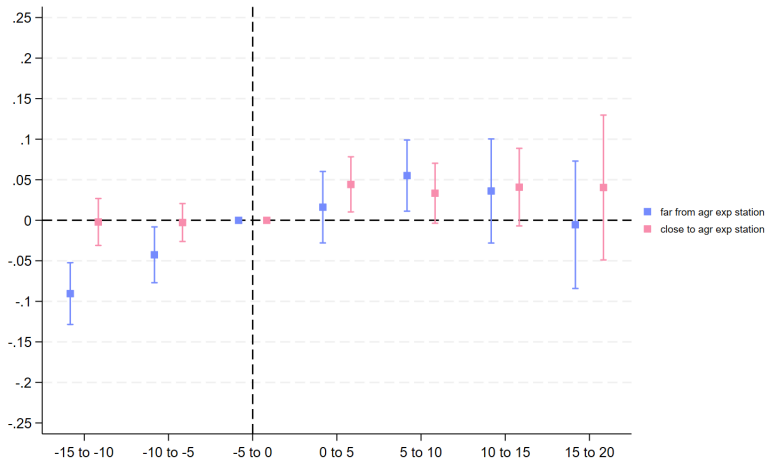


Figure 32: effects on crop revenue, by distance to agricultural experiment station

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the logarithm of total crop revenue per acre. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

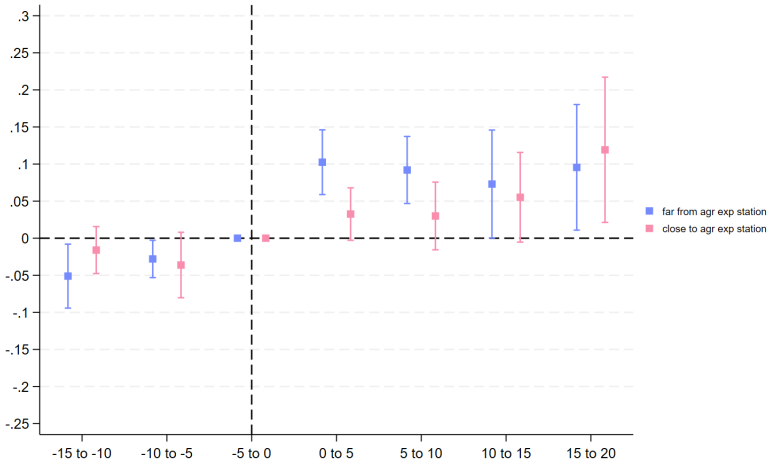


Figure 33: effects on agricultural machinery, by distance to agricultural experiment station

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the logarithm of the value of agricultural machinery per acre. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

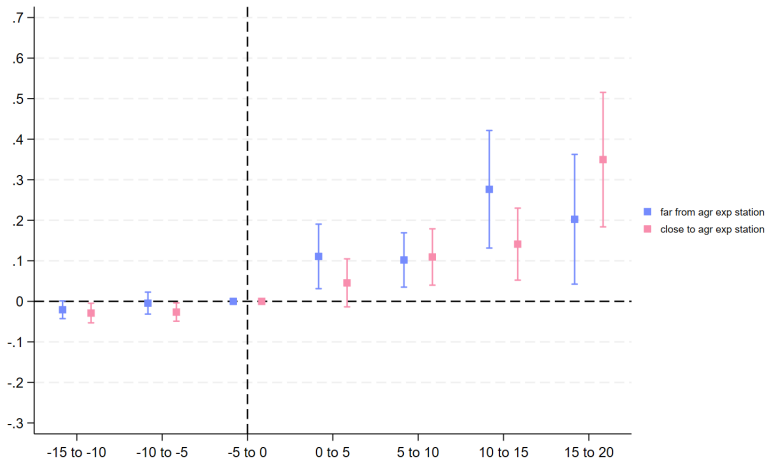


Figure 34: effects on fertilizer expenditures, by distance to agricultural experiment station

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the logarithm of fertilizer expenditures per acre. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

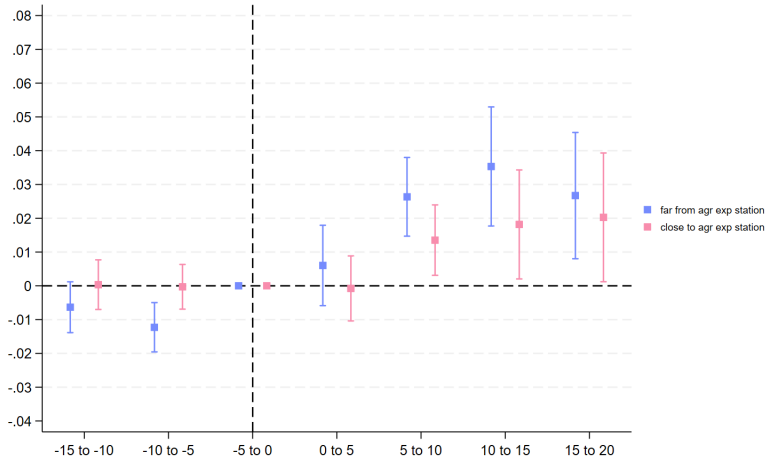


Figure 35: effects on the share of self-own farmland, by distance to agricultural experiment station

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the share of self-own farmland per county. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

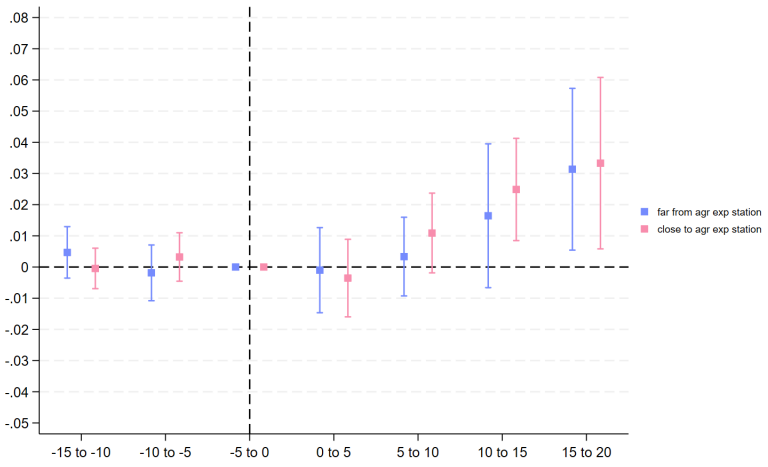


Figure 36: effects on SWI, by distance to agricultural experiment station

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

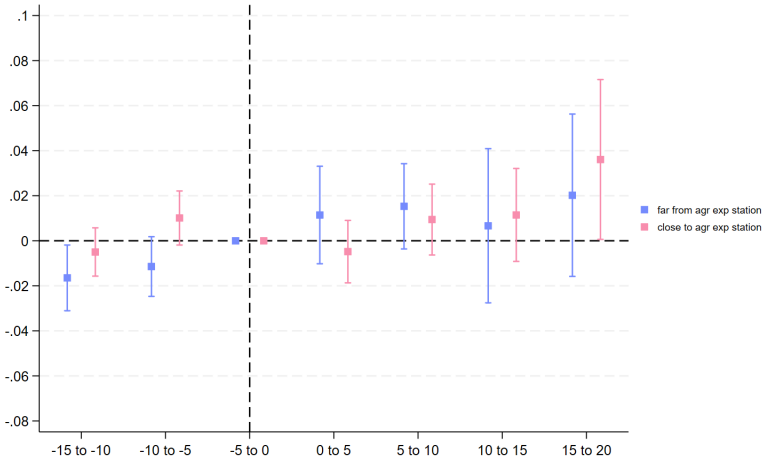


Figure 37: effects on HHI, by distance to agricultural experiment station

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the HHI. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

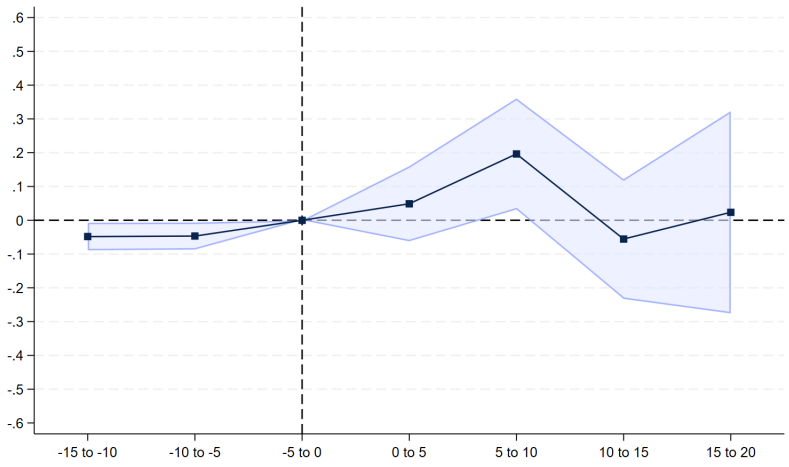


Figure 38: event-study estimates for agricultural productivity (alternative measure)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties (1826 in total). The dependent variable is the normalized total crop revenue per acre of cropland (relative to 1890). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

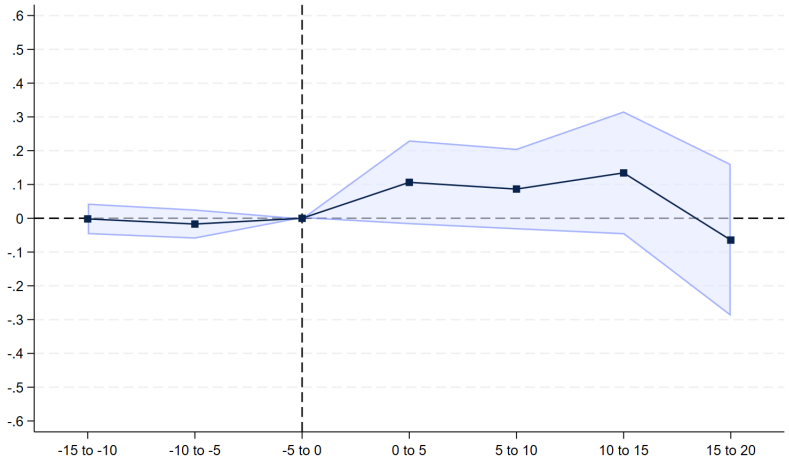


Figure 39: event-study estimates for agricultural productivity (alternative measure)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties (1826 in total). The dependent variable is the normalized total crop revenue per acre of cropland (relative to 1900). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 1. Effects of NCSS Programs on Agricultural Productivity

	dependent variable: log(crop revenue per acre of cropland)									
	all counties sample					*adjacent counties sample*				
$NCSS_{cst}$	0.030**	0.028**	0.052***	0.036***	0.039**	0.032***	0.038***	0.056***	0.051***	0.041***
	(0.014)	(0.013)	(0.011)	(0.011)	(0.015)	(0.011)	(0.010)	(0.009)	(0.009)	(0.013)
county FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year FE	✓		✓			✓		✓		
year×region FE		✓		✓	✓		✓		✓	✓
sample period	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920
mean (1890)	3.422	3.422	3.422	3.422	3.422	4.416	4.416	4.416	4.416	4.416
# of clusters	2148	2148	2148	2148	2148	1826	1826	1826	1826	1826
# of obs.	10740	10740	8592	8592	8592	9130	9130	7304	7304	7304

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). The dependent variable is the logarithm of total crop revenue per acre of farmland. Crop revenue is constructed as the total value of production for nine principal crops: wheat, barley, corn, oat, potato, sweet potato, cotton, buckwheat, and rye. For each crop, production is calculated as the quantity harvested multiplied by its corresponding national-level price. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county (as shown in Columns (5) and (10)).

Table 2. Effects of NCSS Programs on Cropland Concentration

	all counties sample				neighboring counties sample			
	SWI		*HHI*		*SWI*		*HHI*	
$NCSS_{cst}$	0.004	0.007*	0.022***	0.012**	-0.002	0.007*	0.036***	0.029***
	(0.004)	(0.004)	(0.006)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)
log(market access$_{cst}$)	-	0.045***	-	-0.023	-	0.038***	-	-0.053***
	(-)	(0.011)	(-)	(0.017)	(-)	(0.012)	(-)	(0.018)
county FE	✓	✓	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓	✓	✓
sample period	1890-1930	1890-1920	1890-1930	1890-1920	1890-1930	1890-1920	1890-1930	1890-1920
mean (1890)	-0.498	-0.498	0.401	0.401	-0.512	-0.512	0.284	0.284
# of clusters	2148	2148	2148	2148	1826	1826	1826	1826
# of obs.	10740	10740	8592	8592	9130	9130	7304	7304

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). The dependent variables are two indexes for crop concentration: SWI and HHI, taking into eight crop types, wheat, barley, corn, oat, potato, sweet potato, buckwheat, and rye. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county and adding region-year FEs.

Table 3. Correlation between NCSS Programs and Soil Erosion

	presence of soil erosion		share of soil erosion	
$NCSS_{ct}$	0.054***	0.053***	0.098***	0.114***
	(0.013)	(0.017)	(0.017)	(0.020)
weighted by land size in 1930		✓		✓
other state/county controls	✓	✓	✓	✓
# of obs.	2148	2148	2148	2148

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930) and Hornbeck (2012). The dependent variables are the likelihood of having any soil erosion within a county in columns (1) and (2), and the share of eroded soil in the last two columns. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 4. Effects of NCSS Programs on the Choice of Crops

	dependent variable: log(acreage of crop-specific land per farm)									
	all counties sample					*adjacent counties sample*				
$NCSS_{cst} \times Least\ Suitable_i$	-0.084***	-0.064***	-0.080***	-0.079***	-0.108***	-0.087***	-0.061***	-0.088***	-0.083***	-0.113***
	(0.014)	(0.015)	(0.008)	(0.008)	(0.012)	(0.014)	(0.015)	(0.008)	(0.008)	(0.012)
$NCSS_{cst}$	0.008	-0.009	0.007	-0.002	-0.000	0.026***	0.000	0.026***	0.010	0.014
	(0.009)	(0.010)	(0.007)	(0.007)	(0.010)	(0.009)	(0.010)	(0.007)	(0.007)	(0.010)
$Least\ Suitable_i$	-0.618***	-0.606***	-	-	-	-0.692***	-0.686***	-	-	-
	(0.023)	(0.023)	(-)	(-)	(-)	(0.026)	(0.026)	(-)	(-)	(-)
market access		✓		✓	✓		✓		✓	✓
county FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
crop-county FE			✓	✓	✓			✓	✓	✓
crop-year FE			✓	✓	✓			✓	✓	✓
state controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
weighted by acre					✓					✓
# of clusters	2148	2148	2148	2130	2130	1826	1826	1826	1808	1808
# of observations	85920	68736	85920	68160	68160	73040	58432	73040	57856	57856

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). The dependent variable is the logarithm of crop-specific farmland acre per farm. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table 5. Effects of NCSS Programs on Technology and Land Use

	machinery ($)	**fertilizers($)**	**self-owned farm**	**machinery ($)**	**fertilizers($)**	**self-owned farm**
	all counties sample			*neighboring counties sample*		
$NCSS_{cst}$	0.076***	0.134***	0.012***	0.073***	0.117***	0.011***
	(0.013)	(0.020)	(0.003)	(0.014)	(0.018)	(0.003)
county FE	✓	✓	✓	✓	✓	✓
year-region FE	✓	✓	✓	✓	✓	✓
# of clusters	2148	2148	2148	1826	1826	1826
# of obs.	8592	8592	8592	7304	7304	7304

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). "machinery" refers to the logarithm of the value of machinery per acre. "fertilizer" is the logarithm of the amount of fertilizer expenditures per acre. "self-owned farmland" means the share of self-own farmland per county. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 6. Effects of NCSS Programs on Human Capital

	# of farmers per 100,000 people		# of literate farmers per 100,000 people	
Panel A. All Counties Sample				
$NCSS_{cst}$	10.013***	13.785***	10.066***	13.811***
	(2.423)	(3.099)	(2.422)	(3.098)
# of clusters	2148	2148	2148	2148
# of observations	8592	6444	8592	6444
Panel B. Neighboring Counties Sample				
$NCSS_{cst}$	9.578***	13.868***	9.628***	13.884***
	(2.460)	(3.174)	(2.458)	(3.172)
# of clusters	1826	1826	1826	1826
# of observations	7304	5478	7304	5478
county FE	✓	✓	✓	✓
year FE	✓	✓	✓	✓
state controls	✓	✓	✓	✓
market access info		✓		✓
sample period	1880-1930	1880-1920	1880-1930	1880-1920

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Population (1880-1930). "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table 7. Effects of NCSS Mapping by Prior Agricultural Knowledge Accumulation

	log(revenue per acre)	machinery	fertilizer	% of self-own farms	SWI	HHI
Panel A. with agr societies						
$NCSS_{ct}$	0.047***	0.011	0.046*	0.012***	0.005	0.007
	(0.014)	(0.014)	(0.027)	(0.004)	(0.006)	(0.006)
# of clusters	877	877	877	877	877	877
# of obs.	3508	3508	3508	3508	3508	3508
Panel B. w/o agr societies						
$NCSS_{ct}$	0.058***	0.119***	0.192***	0.016***	0.006	0.015**
	(0.016)	(0.019)	(0.028)	(0.005)	(0.005)	(0.007)
# of clusters	1271	1271	1271	1271	1271	1271
# of obs.	5084	5084	5084	5084	5084	5084
Panel C. with agr fairs						
$NCSS_{ct}$	0.054***	0.017	0.067**	0.014***	0.004	0.000
	(0.016)	(0.014)	(0.029)	(0.004)	(0.007)	(0.007)
# of clusters	723	723	723	723	723	723
# of obs.	2892	2892	2892	2892	2892	2892
Panel D. w/o agr fairs						
$NCSS_{ct}$	0.053***	0.108***	0.169***	0.015***	0.007	0.018***
	(0.015)	(0.018)	(0.026)	(0.005)	(0.005)	(0.007)
# of clusters	1425	1425	1425	1425	1425	1425
# of obs.	5700	5700	5700	5700	5700	5700
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table 8. Effects of NCSS Programs on Agricultural Outcomes (Heterogeneity by Agricultural Society)

	log(revenue per acre)	machinery	fertilizer	self-own farms	SWI	HHI
Panel A. agricultural societies						
$NCSS_{ct} \times Agr\ Society_{c,1876}$	-0.007	-0.097***	-0.134***	-0.003	0.002	-0.007
	(0.021)	(0.024)	(0.039)	(0.006)	(0.008)	(0.010)
$NCSS_{ct}$	0.058***	0.120***	0.194***	0.017***	0.006	0.015**
	(0.016)	(0.020)	(0.028)	(0.005)	(0.005)	(0.007)
Agr $\text{Society}_{c,1876}$×year FE	✓	✓	✓	✓	✓	✓
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
Panel B. county fairs						
$NCSS_{ct} \times Agr\ Fair_{c,1876}$	0.003	-0.080***	-0.088**	0.001	0.000	-0.017*
	(0.021)	(0.023)	(0.039)	(0.006)	(0.008)	(0.010)
$NCSS_{ct}$	0.053***	0.109***	0.169***	0.015***	0.007	0.018***
	(0.015)	(0.018)	(0.026)	(0.005)	(0.005)	(0.007)
Agr $\text{Fair}_{c,1876}$×year FE	✓	✓	✓	✓	✓	✓
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
# of clusters	2148	2148	2148	2148	2148	2148
# of obs.	8592	8592	8592	8592	8592	8592

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). Panel A checks the heterogeneity based on the existence of agricultural society/farmers' club. Panel B checks the heterogeneity based on the existence of agricultural fairs. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 9. Effects of NCSS Mapping by Prior Level of Local Farmers

	log(revenue per acre)	machinery	fertilizer	self-own farms	SWI	HHI
above median # of farmers						
$NCSS_{ct}$	0.063***	0.022*	0.085***	0.002	0.006	0.014***
	(0.011)	(0.012)	(0.024)	(0.003)	(0.005)	(0.005)
# of clusters	1107	1107	1107	1107	1107	1107
# of obs.	4428	4428	4428	4428	4428	4428
below median # of farmers						
$NCSS_{ct}$	0.030	0.122***	0.166***	0.025***	-0.000	0.013
	(0.020)	(0.023)	(0.033)	(0.006)	(0.006)	(0.009)
# of clusters	1041	1041	1041	1041	1041	1041
# of obs.	4164	4164	4164	4164	4164	4164
above median # of lit farmers						
$NCSS_{ct}$	0.054***	0.031**	0.108***	0.007*	0.007	0.011**
	(0.012)	(0.012)	(0.040)	(0.003)	(0.005)	(0.005)
# of clusters	1077	1077	1077	1077	1077	1077
# of obs.	4308	4308	4308	4308	4308	4308
below median # of lit farmers						
$NCSS_{ct}$	0.064***	0.088***	0.424***	0.021***	0.003	0.014*
	(0.020)	(0.017)	(0.095)	(0.006)	(0.006)	(0.009)
# of clusters	1071	1071	1071	1071	1071	1071
# of obs.	4284	4284	4284	4284	4284	4284
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table 10. Effects of NCSS Programs on Agricultural Outcomes (Heterogeneity by Local Farmer Density)

	log(revenue per acre)	machinery	fertilizer	self-own farms	SWI	HHI
Panel A. total farmers						
$NCSS_{ct} \times Farmers_{c,1880}$	-0.028**	-0.039***	-0.108**	-0.009**	0.010***	-0.003
	(0.013)	(0.011)	(0.046)	(0.004)	(0.004)	(0.006)
$NCSS_{ct}$	0.127***	0.148***	0.509***	0.032***	-0.020*	0.019
	(0.040)	(0.032)	(0.146)	(0.011)	(0.010)	(0.019)
$Farmers_{c,1880}$×year FE	✓	✓	✓	✓	✓	✓
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
# of clusters	2136	2136	2136	2136	2136	2136
# of obs.	8544	8544	8544	8544	8544	8544
Panel B. literate farmers						
$NCSS_{ct} \times Lit\ Farmers_{c,1876}$	-0.021	-0.038***	-0.132***	-0.005*	0.008**	-0.004
	(0.013)	(0.010)	(0.047)	(0.003)	(0.004)	(0.006)
$NCSS_{ct}$	0.105***	0.137***	0.556***	0.022**	-0.014	0.020
	(0.035)	(0.028)	(0.143)	(0.009)	(0.009)	(0.016)
Literate $Farmers_{c,1880}$×year FE	✓	✓	✓	✓	✓	✓
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
# of clusters	2134	2134	2134	2134	2134	2134
# of obs.	8536	8536	8536	8536	8536	8536

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 11. Effects of NCSS Mapping by Prior Level of Soil Knowledge

	log(revenue per acre)	machinery	fertilizer	self-own farms	SWI	HHI
above median (% of improved farmland)						
$NCSS_{ct}$	0.027**	0.033**	0.010	0.017***	-0.007	-0.014**
	(0.013)	(0.015)	(0.019)	(0.004)	(0.005)	(0.006)
# of clusters	1058	1058	1058	1058	1058	1058
# of obs.	4232	4232	4232	4232	4232	4232
below median (% of improved farmland)						
$NCSS_{ct}$	0.047***	0.117***	0.148***	0.007	0.001	0.033***
	(0.017)	(0.022)	(0.029)	(0.005)	(0.004)	(0.008)
# of clusters	1090	1090	1090	1090	1090	1090
# of obs.	4360	4360	4360	4360	4360	4360
county FE	✓	✓	✓	✓	✓	✓
region-year FE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table 12. Effects of NCSS Programs on Agricultural Outcomes (Heterogeneity by Local Soil Knowledge)

	log(revenue per acre)	machinery	fertilizer	self-own farms	SWI	HHI
$NCSS_{ct} \times Share_{c,1890}$	-0.228***	0.133***	-0.625***	0.000	-0.008	-0.064***
	(0.047)	(0.051)	(0.084)	(0.014)	(0.016)	(0.025)
$NCSS_{ct}$	0.180***	0.008	0.485***	0.015*	0.011	0.048***
	(0.031)	(0.034)	(0.057)	(0.009)	(0.009)	(0.016)
$\text{Share}_{c,1890} \times$year FE	✓	✓	✓	✓	✓	✓
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
# of clusters	2148	2148	2148	2148	2148	2148
# of obs.	8592	8592	8592	8592	8592	8592

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 13. TFP Effects of NCSS Mapping

	dependent variable: log(total crop revenue)									
	all counties					*adjacent counties*				
$NCSS_{cst}$	-0.110***	0.033***	0.036***	-0.026	0.009	-0.029	0.031***	0.024**	-0.005	0.001
	(0.023)	(0.012)	(0.012)	(0.016)	(0.015)	(0.022)	(0.011)	(0.010)	(0.016)	(0.016)
county population	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
market access	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
capital control			✓	✓	✓			✓	✓	✓
total cropland acres		✓	✓				✓	✓		
crop-specific land acres				✓	✓				✓	✓
labor input control					✓					✓
county FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# of clusters	2148	2148	2148	2148	2148	1826	1826	1826	1826	1826
# of obs.	6444	6444	6444	6444	6444	5478	5478	5478	5478	5478

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). The dependent variable is the logarithm of total crop revenue at the county level. Crop revenue is constructed as the total value of production for nine principal crops: wheat, barley, corn, oat, potato, sweet potato, cotton, buckwheat, and rye. For each crop, production is calculated as the quantity harvested multiplied by its corresponding national-level price. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 14. Effects of NCSS Mapping on Monocultural Farming (controlling for inputs)

	SWI			HHI		
$NCSS_{cst}$	0.003	-0.002	-0.002	0.008	0.007	-0.002
	(0.004)	(0.003)	(0.004)	(0.005)	(0.005)	(0.009)
inputs control	✓	✓	✓	✓	✓	✓
county FE	✓	✓	✓	✓	✓	✓
year FE	✓			✓		
region-year FE		✓	✓		✓	✓
weighted by county acre (1900)			✓			✓
# of clusters	2148	2148	2148	2148	2148	2148
# of obs.	8592	8592	8592	8592	8592	8592

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 15. Effects of NCSS Programs on Institutional Research Activities

	presence of pub	log(# of pub)	presence of pub	log(# of pub)
Panel A. state-year panel				
log(# of NCSS)$_{st}$	0.131**	0.190**	-	-
	(0.063)	(0.092)	(-)	(-)
year FE	✓	✓	✓	✓
state FE	✓	✓	✓	✓
Panel B. county-year panel				
NCSS$_{ct}$	-	-	0.098**	0.229**
	(-)	(-)	(0.043)	(0.104)
year FE	✓	✓	✓	✓
county FE	✓	✓	✓	✓
observations	775	775	775	775

Standard errors in parentheses

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Note: Data are drawn from the agricultural experiment station record (1896-1920). The sample in Panel A is a balanced panel at the state-year level; the sample in Panel B is a balanced panel of county and year, which includes all counties having the primary experiment stations. All standard errors are clustered at the state level in Panel A and at the county level in Panel B.

Table 16. Effects of NCSS Programs on Local Agricultural Patenting

	all agricultural patents			soil-relevant patents		
	# of patents	*log(# of patents+1)*	$\mathbb{1}$*(patents>0)*	*# of patents*	*log(# of patents+1)*	$\mathbb{1}$*(patents>0)*
$NCSS_{cst}$	-0.003**	-0.002**	-0.002**	-0.011***	-0.006***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.004)	(0.002)	(0.003)
county-class FE	✓	✓	✓	✓	✓	✓
year-class FE	✓	✓	✓	✓	✓	✓
# of clusters	1920	1920	1920	1920	1920	1920
# of obs.	865920	865920	865920	157440	157440	157440

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from Petralia, Balland and Rigby (2016). County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 17. Effects of NCSS Mapping on Alternative Crop Revenue

	relative crop revenue			
	1890 as baseline		*1900 as baseline*	
$NCSS_{cst}$	0.097**	0.133***	0.072**	0.096**
	(0.038)	(0.051)	(0.035)	(0.048)
log(market access)$_{cst}$	-	-1.046***	-	-0.912***
	(-)	(0.202)	(-)	(0.184)
county FE	✓	✓	✓	✓
year FE	✓	✓	✓	✓
other state/county controls	✓	✓	✓	✓
# of obs.	9050	7240	8685	6948

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties (1826 in total). The dependent variable is the logarithm of total crop revenue per acre of farmland. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 18. Effects of NCSS Mapping (Controlling for Exposure to National Market Demand Shocks)

	log(revenue per acre))	machinery	fertilizer	self-own farms	SWI	HHI
Panel A. all counties						
$NCSS_{ct}$	0.048***	0.080***	0.128***	0.016***	0.007*	0.012**
	(0.011)	(0.013)	(0.019)	(0.003)	(0.004)	(0.005)
$ln(Crop\ Index_{ct})$	0.363***	0.173	0.498***	-0.060***	-0.023*	-0.037*
	(0.070)	(0.112)	(0.071)	(0.012)	(0.012)	(0.022)
# of clusters	2148	2148	2148	2148	2148	2148
# of obs.	8592	8592	8592	8592	8592	8592
Panel B. adjacent counties						
$NCSS_{ct}$	0.053***	0.067***	0.113***	0.014***	0.006*	0.031***
	(0.009)	(0.014)	(0.018)	(0.003)	(0.004)	(0.005)
$ln(Crop\ Index_{ct})$	0.168***	-0.404***	0.321***	-0.048***	-0.015	-0.153***
	(0.049)	(0.087)	(0.063)	(0.012)	(0.013)	(0.016)
# of clusters	1826	1826	1826	1826	1826	1826
# of obs.	7304	7304	7304	7304	7304	7304
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1920). "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table 19. Effects of NCSS Programs on Share of Immigrant Farmers

	$\mathbb{1}$(imm farmers>0)		log(# of imm farmers)		Δimm farmers$_{c,1920-1880}$
$NCSS_{cst}$	0.010	0.011	0.054*	0.030	0.002
	(0.007)	(0.008)	(0.032)	(0.033)	(0.001)
$log(market\ access)_{cst}$	-0.075*	-0.089**	-0.963***	-0.923***	0.009***
	(0.039)	(0.039)	(0.168)	(0.169)	(0.003)
county FE	✓	✓	✓	✓	✓
year-region FE		✓		✓	
# of clusters	2042	2042	2042	2042	2042
# of obs.	6126	6126	6126	6126	2042

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Population (1880-1930). County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table 20. Complemnentarity between Physical Infrastructure and NCSS Mapping

	log(revenue per acre)	machinery	fertilizer	self-own farms	SWI	HHI
Panel A. full sample						
$NCSS_{ct} \times ln(MA)_{ct}$	0.138***	0.140***	0.455***	0.054***	0.036**	0.002
	(0.050)	(0.053)	(0.096)	(0.015)	(0.014)	(0.025)
$NCSS_{ct}$	-1.829**	-2.727***	-12.725***	-0.949***	-0.774***	-0.017
	(0.919)	(0.896)	(4.305)	(0.276)	(0.248)	(0.448)
# of clusters	2148	2148	2148	2148	2148	2148
# of obs.	8592	8592	8592	8592	8592	8592
Panel B. adjacent counties						
$NCSS_{ct} \times ln(MA)_{ct}$	0.179***	0.128***	0.705***	0.050***	0.045***	0.022
	(0.051)	(0.048)	(0.241)	(0.015)	(0.014)	(0.023)
$NCSS_{ct}$	-3.172***	-2.252***	-12.413***	-0.877***	-0.797***	-0.365
	(0.905)	(0.860)	(4.300)	(0.271)	(0.250)	(0.408)
$ln(market\ access_{ct})$	-0.270***	-0.109***	-0.551***	0.028*	0.038***	0.017
	(0.042)	(0.030)	(0.148)	(0.014)	(0.012)	(0.015)
# of clusters	1826	1826	1826	1826	1826	1826
# of obs.	7304	7304	7304	7304	7304	7304
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1920). "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table 21. Do Local Pre-existing Characteristics Predict the Mapping Status?

	being mapped by NCSS?		
log(population in 1890)	0.045***	0.045***	0.045***
	(0.010)	(0.011)	(0.010)
urban status in 1890	0.009	0.015	0.015
	(0.027)	(0.028)	(0.026)
log(# of farms in 1890)	0.001	-0.010	-0.010
	(0.010)	(0.011)	(0.010)
log(value of farmland in 1890)	-0.001	0.007	0.007
	(0.009)	(0.011)	(0.010)
log(value of equipment in 1890)	-0.010	-0.004	-0.004
	(0.010)	(0.011)	(0.010)
log(expenses on fertilizer in 1890)	0.007***	0.003	0.003
	(0.002)	(0.002)	(0.002)
log(value of agricultural output in 1890)	0.009	0.003	0.003
	(0.009)	(0.011)	(0.010)
log(market access value in 1890)	-0.019	-0.009	-0.009
	(0.018)	(0.018)	(0.017)
suitability score of barley	0.033***	0.044***	0.044***
	(0.011)	(0.012)	(0.011)
suitability score of corn	-0.017**	-0.013	-0.013
	(0.008)	(0.009)	(0.008)
suitability score of tobacco	0.010	-0.001	-0.001
	(0.006)	(0.007)	(0.007)
suitability score of cotton	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
suitability score of wheat	-0.018	-0.032**	-0.032***
	(0.013)	(0.013)	(0.012)
# of obs.	6396	6396	6396
state FE		✓	✓
year FE			✓

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the probability of enacting any PA restriction in state s and year t in column (1), the probability of enacting PA restrictions on general SAT services in column (2), and the probability of enacting PA restrictions on medication-assisted treatment (MAT) in column (3). State controls include SAT parity laws, pain clinic regulation, PDMP mandates, state unemployment rates, natural log of the population, natural log of the state GDP, and the effective state-level minimum wage. All standard errors are clustered at the state level. Regressions are weighted by state population.

Table 22. Effects of NCSS Programs on Crop Revenue (by Barley Cultivation)

	dependent variable: log(crop revenue per acre)					
	actual yield in 1890			*barley suitability score*		
$NCSS_{ct} \times Barley\ Advantage_c$	0.036	0.032	0.013	-0.034	-0.009	-0.031
	(0.026)	(0.024)	(0.029)	(0.021)	(0.021)	(0.026)
$NCSS_{ct}$	0.017	0.029**	0.034**	0.054***	0.046***	0.058***
	(0.013)	(0.012)	(0.014)	(0.015)	(0.015)	(0.020)
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
market access		✓	✓		✓	✓
weighted by 1900 acre			✓			✓
# of clusters	1826	1826	1826	1826	1826	1826
# of obs.	9130	7304	7304	9130	7304	7304

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the probability of only offering any low-intensity (i.e., outpatient) services relative to only offering any high-intensity (i.e., intensive outpatient, partial hospital, residential inpatient, and hospital-based inpatient) in columns (1) and (2), and the probability of offering both low- and high-intensity services relative to only providing low-intensity ones in column (3) and column (4). In column (5) and column (6), the dependent variable is the probability of offering both types of services relative to offering only high-intensity ones. Data are obtained from the 2012-2020 N-SSATS. State controls include SAT parity laws, pain clinic regulation, PDMP mandates, state unemployment rates, natural log of the population, natural log of the state GDP, and the effective state-level minimum wage. All standard errors are clustered at the state level.

Appendix

A.1. Tables and Figures from Robustness Tests

Table A1. Results from Synthetic Control Methods

	log(crop revenue per acre)	machinery	fertilizer	self-owned farm	SWI
Panel A. all counties sample					
$NCSS_{cst}$	0.044***	0.034***	0.208***	0.016***	0.012***
	(0.011)	(0.009)	(0.056)	(0.003)	(0.004)
# of clusters	2148	2148	2148	2148	2148
# of obs.	8592	8592	8592	8592	8592
Panel B. neighboring counties sample					
$NCSS_{cst}$	0.039***	0.012	0.155***	0.014***	0.011***
	(0.011)	(0.009)	(0.031)	(0.003)	(0.004)
county FE	✓	✓	✓	✓	✓
year-region FE	✓	✓	✓	✓	✓
# of clusters	1826	1826	1826	1826	1826
# of obs.	7304	7304	7304	7304	7304

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1920). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table A2. Alternative Adjacent County Pair Design

	log(revenue per acre)	value of machinery	cost of fertilizer	% of self-own farms	SWI	HHI
$NCSS_{ct}$	0.045***	-0.351***	0.117***	0.015***	-0.001	0.009**
	(0.008)	(0.077)	(0.028)	(0.003)	(0.003)	(0.004)
county pair FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
mean (1890)	0.818	0.854	0.552	0.608	0.289	
# of clusters	2111	2111	2111	2111	2111	2111
# of obs.	8444	8444	8444	8444	8444	8444

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1920). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

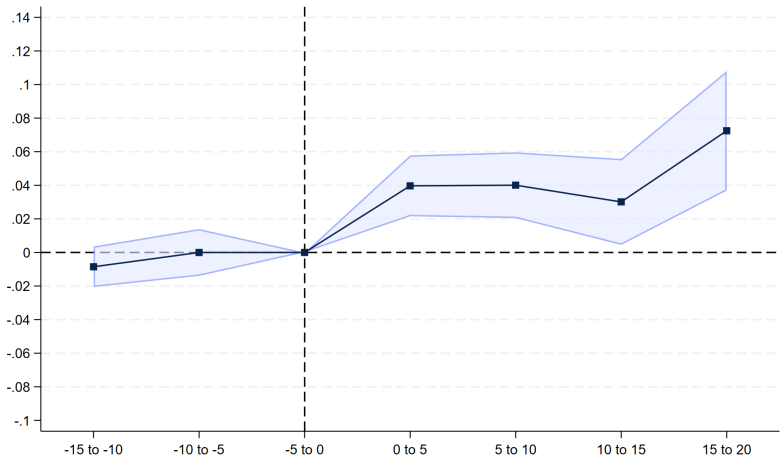


Figure A1: event-study estimates for total crop revenue per acre (adjacent county pair design)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (1826 counties). The dependent variable is the total crop revenue per acre of cropland. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county pair level.

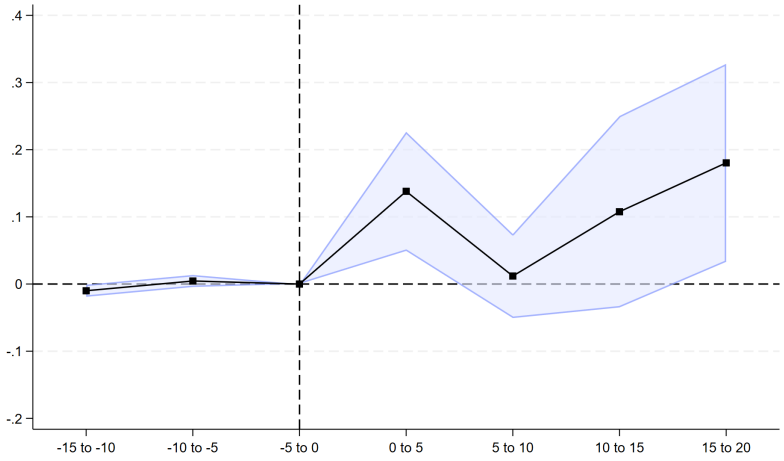


Figure A2: event-study estimates for expenses on fertilizer per acre (adjacent county pair design)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (1826 counties). The dependent variable is the expenditure of fertilizer per acre of cropland. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county pair level.

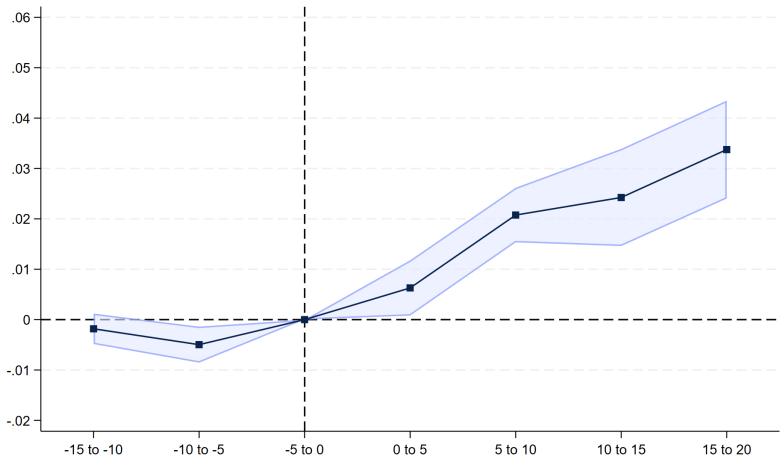


Figure A3: event-study estimates for the share of self-owned farms (adjacent county pair design)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (1826 counties). The dependent variable is the share of self-owned farmland per county. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county pair level.

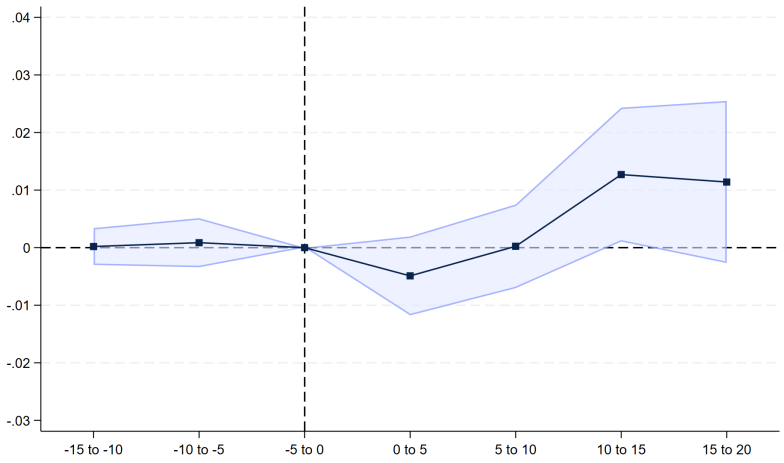


Figure A4: event-study estimates for SWI (balanced sample)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (1826 counties). The dependent variable is SWI. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county pair level.

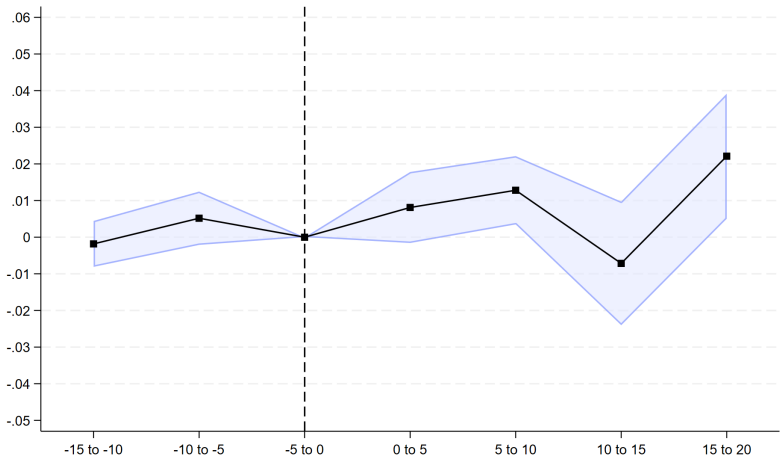


Figure A5: event-study estimates for HHI (balanced sample)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (1826 counties). The dependent variable is HHI. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county pair level.

Table A3. Alternative Balanced Adjacent County Pair Design

	log(revenue per acre))	value of machinery	cost of fertilizer	% of self-own farms	SWI	HHI
$NCSS_{ct}$	0.044***	-0.360***	0.115***	0.015***	-0.003	0.007*
	(0.009)	(0.081)	(0.029)	(0.003)	(0.003)	(0.004)
$ln(market\ access_{ct})$	-0.120***	0.474***	-0.574***	0.044***	0.005	0.005
	(0.034)	(0.137)	(0.097)	(0.011)	(0.009)	(0.018)
county pair FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
# of clusters	2029	2029	2029	2029	2029	2029
# of obs.	8116	8116	8116	8116	8116	8116

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county pair level. Results are robust when weighting the sample using the 1900 acre of land in each county.

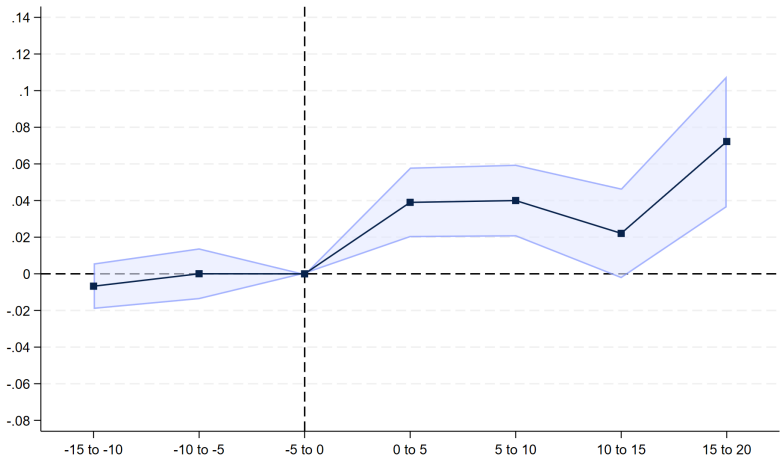


Figure A6: event-study estimates for total crop revenue per acre (balanced sample)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (1826 counties). The dependent variable is the log of total crop revenue per acre. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county pair level.

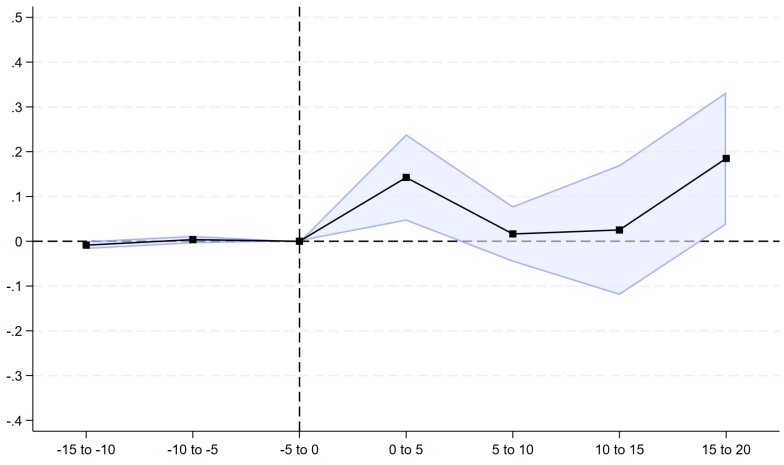


Figure A7: event-study estimates for expenses on fertilizer per acre (balanced sample)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (1826 counties). The dependent variable is fertilizer expenditure per acre. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county pair level.

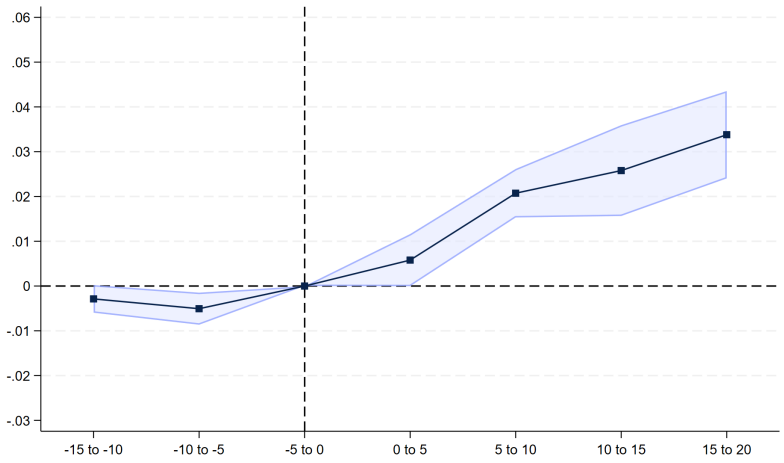


Figure A8: event-study estimates for the share of self-owned farms (balanced sample)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (1826 counties). The dependent variable is self-owned farmland per county. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county pair level.

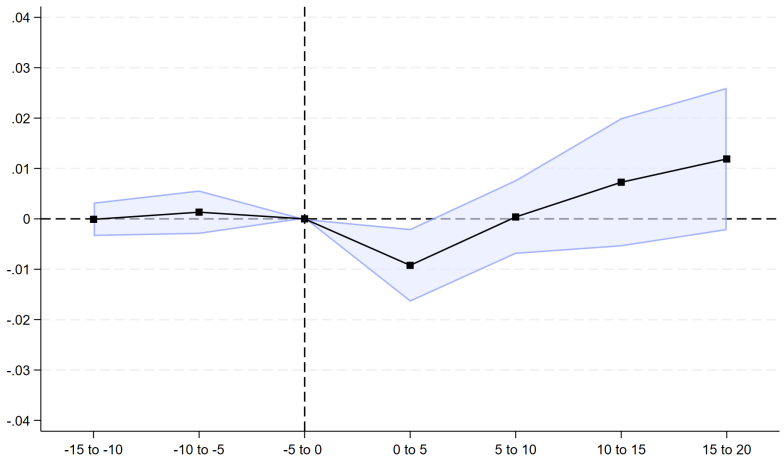


Figure A9: event-study estimates for SWI (balanced sample)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

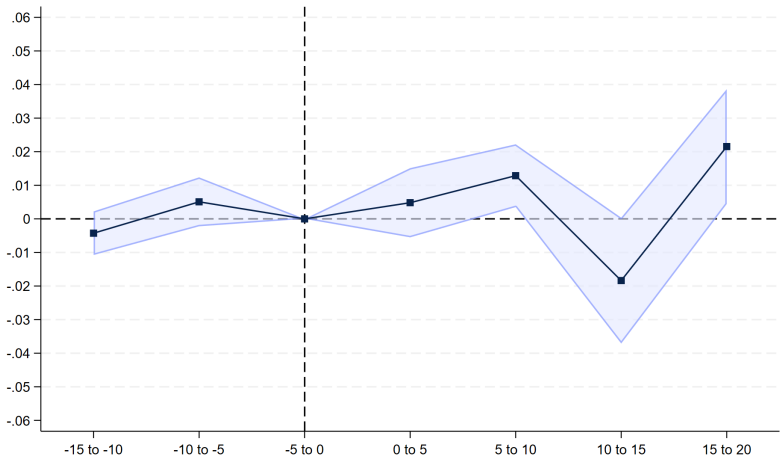


Figure A10: event-study estimates for HHI (balanced sample)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the HHI. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table A4. Effects of NCSS Programs on Agricultural Outcomes (Excluding Never-Treated Counties)

	log(revenue per acre)	**value of machinery**	**cost of fertilizer**	**% of self-own farms**	**SWI**	**HHI**
$NCSS_{ct}$	0.038***	-0.060***	0.184***	0.020***	0.012**	0.012**
	(0.013)	(0.013)	(0.044)	(0.004)	(0.005)	(0.006)
$ln(market\ access_{ct})$	-0.224***	-0.053	-0.404**	0.031*	0.008	0.003
	(0.053)	(0.037)	(0.169)	(0.017)	(0.015)	(0.019)
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
# of clusters	808	808	808	808	808	808
# of obs.	3232	3232	3232	3232	3232	3232

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930) and exclude all counties that have never been mapped by the NCSS. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

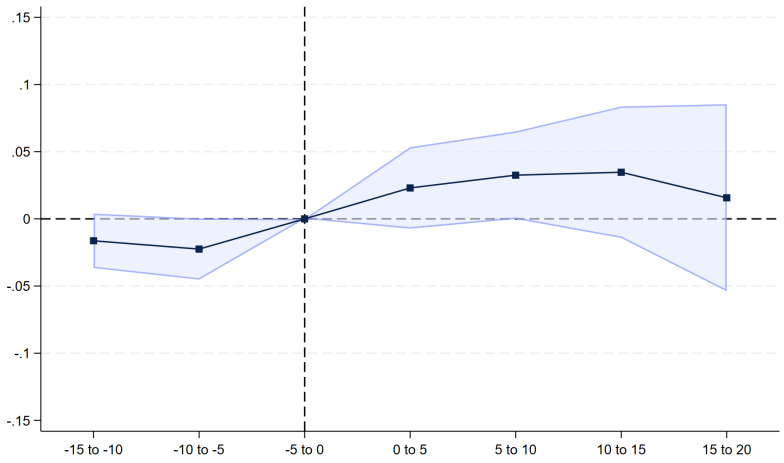


Figure A11: event-study estimates for total crop revenue per acre (excluding never-treated)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties described in Section 5 (1826 counties). The dependent variable is the log of total crop revenue per acre. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

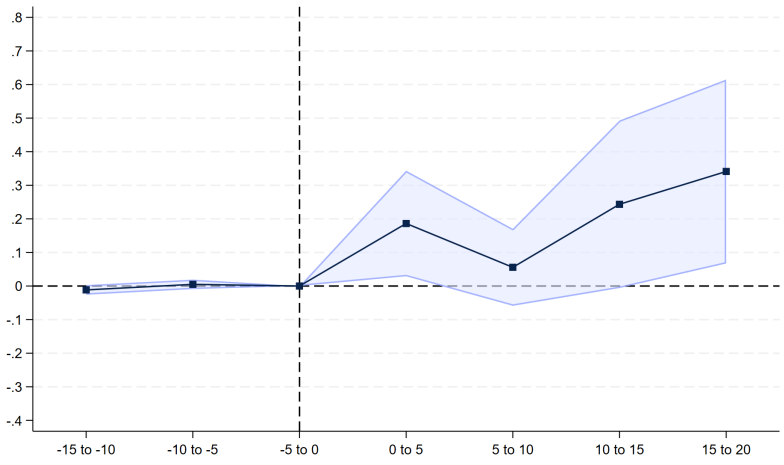


Figure A12: event-study estimates for expenses on fertilizer per acre (excluding never-treated)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties described in Section 5 (1826 counties). The dependent variable is fertilizer expenditure per acre of cropland. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

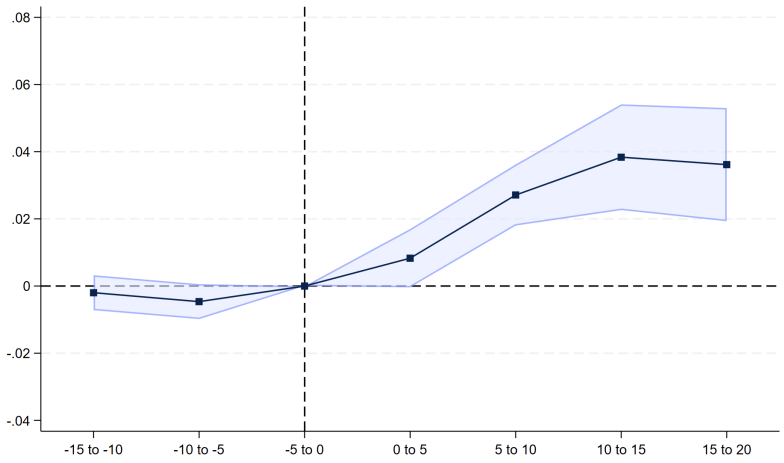


Figure A13: event-study estimates for the share of self-owned farms (excluding never-treated)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties described in Section 5 (1826 counties). The dependent variable is the share of self-own farm per county. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

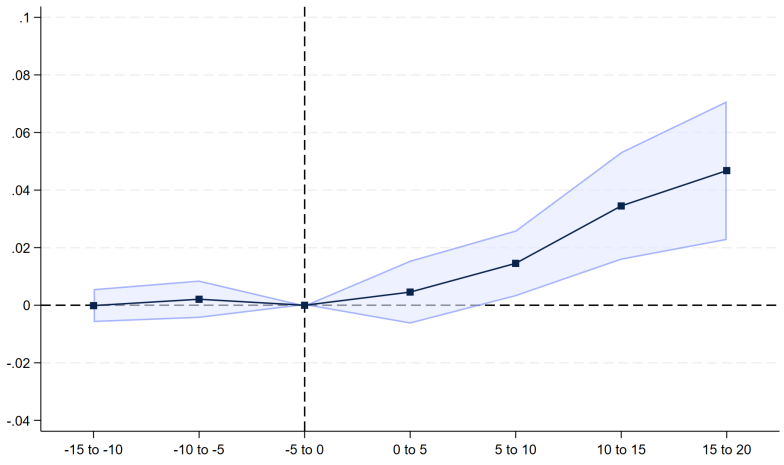


Figure A14: event-study estimates for SWI (excluding never-treated)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties described in Section 5 (1826 counties). The dependent variable is SWI. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

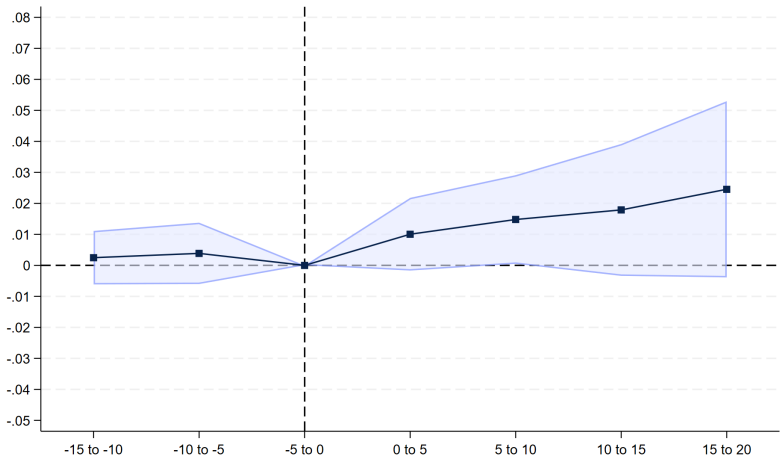


Figure A15: event-study estimates for HHI (excluding never-treated)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all adjacent counties described in Section 5 (1826 counties). The dependent variable is HHI. The event-study specification includes county pair fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table A5. Effects of NCSS Programs on Agricultural Outcomes (Double Lasso)

	log(revenue per acre)	value of machinery	cost of fertilizer	% of self-own farms	SWI	HHI
$NCSS_{ct}$	0.067***	0.329***	0.225***	0.020***	0.008*	0.024***
	(0.012)	(0.091)	(0.043)	(0.004)	(0.004)	(0.005)
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
mean (1890)	0.818	0.854	0.552	0.608	0.289	
# of clusters	1826	1826	1826	1826	808	808
# of obs.	7304	7304	7304	7304	7304	7304

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table A6. Effects of NCSS Programs on Agricultural Outcomes (Alternative TWFE)

	log(revenue per acre)	**machinery**	**fertilizer**	**% of self-own farms**	**SWI**	**HHI**
$NCSS_{ct}$	0.044^{***}	0.033^{***}	0.128^{*}	0.007^{**}	0.009^{**}	0.015^{***}
	(0.012)	(0.012)	(0.077)	(0.003)	(0.004)	(0.006)
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
# of clusters	2148	2148	2184	2184	2184	2184
# of obs.	8592	8592	8592	8592	8592	8592

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table A7. Effects of NCSS Programs on Agricultural Outcomes (1900-1920)

	log(revenue per acre))	log(value of machinery)	cost of fertilizer	% of self-own farms	SWI	HHI
$NCSS_{ct}$	0.039***	0.063***	0.241***	0.017***	0.008**	0.018***
	(0.011)	(0.011)	(0.051)	(0.003)	(0.004)	(0.005)
$log(mkt\ access)_{ct}$	-0.300***	-0.104**	-1.224***	0.034*	0.018	-0.099***
	(0.052)	(0.046)	(0.356)	(0.020)	(0.015)	(0.028)
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
# of clusters	2139	2139	2139	2139	2139	2139
# of obs.	6417	6417	6417	6417	6417	6417

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table A8. Effects of NCSS Programs on Agricultural Outcomes (by (Normalized) Distance to Experiment Stations)

	log(revenue per acre))	log(value of machinery)	cost of fertilizer	% of self-own farms	SWI	HHI
Panel A. below the median						
$NCSS_{ct}$	0.053***	0.054***	0.256***	0.009*	0.008	0.004
	(0.015)	(0.015)	(0.053)	(0.005)	(0.005)	(0.006)
$log(mkt\ access)_{ct}$	-0.292***	-0.093**	-0.675***	0.034*	0.031**	0.003
	(0.048)	(0.043)	(0.211)	(0.019)	(0.014)	(0.022)
# of clusters	1073	1073	1073	1073	1073	1073
# of obs.	4292	4292	4292	4292	4292	4292
Panel B. above the median						
$NCSS_{ct}$	0.055***	0.085***	0.243***	0.024***	0.007	0.020**
	(0.017)	(0.016)	(0.083)	(0.005)	(0.005)	(0.008)
$log(mkt\ access)_{ct}$	-0.309***	-0.198***	-0.647*	-0.002	0.052***	-0.050**
	(0.060)	(0.045)	(0.340)	(0.021)	(0.017)	(0.025)
# of clusters	1075	1075	1075	1075	1075	1075
# of obs.	4300	4300	4300	4300	4300	4300
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table A9. Effects of NCSS Programs on Agricultural Outcomes (Including Meteorological Infrastructure)

	revenue	machinery	fertilizer	% of self-own farms	SWI	HHI
Panel A. access to nearest station						
$NCSS_{ct}$	0.040***	0.037***	0.212***	0.014***	0.007*	0.023***
	(0.011)	(0.011)	(0.043)	(0.003)	(0.004)	(0.005)
access to weather station$_{ct}$	0.030**	-0.033***	-0.002	-0.002	-0.009**	-0.007
	(0.012)	(0.009)	(0.031)	(0.004)	(0.004)	(0.005)
log(mkt access)$_{ct}$	-0.268***	-0.114***	-0.551***	0.027*	0.037***	0.015
	(0.042)	(0.030)	(0.149)	(0.015)	(0.012)	(0.016)
Panel B. distance to nearest station						
$NCSS_{ct}$	0.040***	0.037***	0.213***	0.014***	0.007*	0.023***
	(0.011)	(0.011)	(0.043)	(0.003)	(0.004)	(0.005)
log(distance to station)$_{ct}$	0.012***	-0.012***	-0.003	-0.000	-0.004***	-0.003*
	(0.004)	(0.003)	(0.010)	(0.002)	(0.001)	(0.002)
log(mkt access)$_{ct}$	-0.265***	-0.116***	-0.552***	0.027*	0.036***	0.015
	(0.042)	(0.030)	(0.149)	(0.015)	(0.012)	(0.016)
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓
# of clusters	2148	2148	2184	2184	2184	2184
# of obs.	8592	8592	8592	8592	8592	8592

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

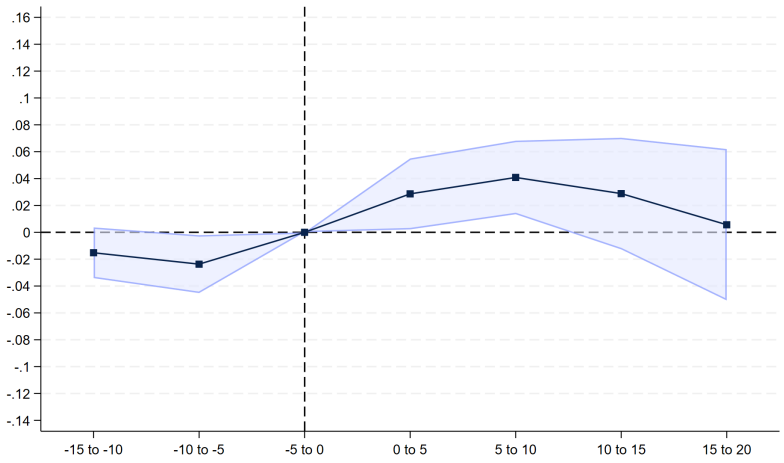


Figure A16: event-study estimates for total crop revenue per acre (including distance to nearest weather station)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

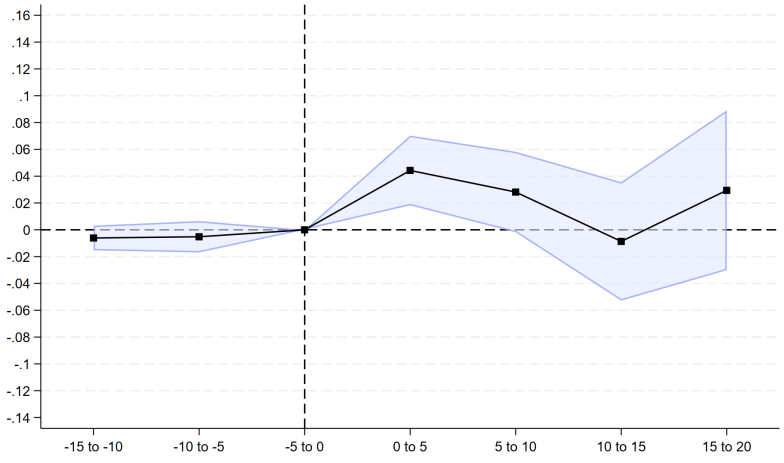


Figure A17: event-study estimates for value of machinery (including distance to nearest weather station)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

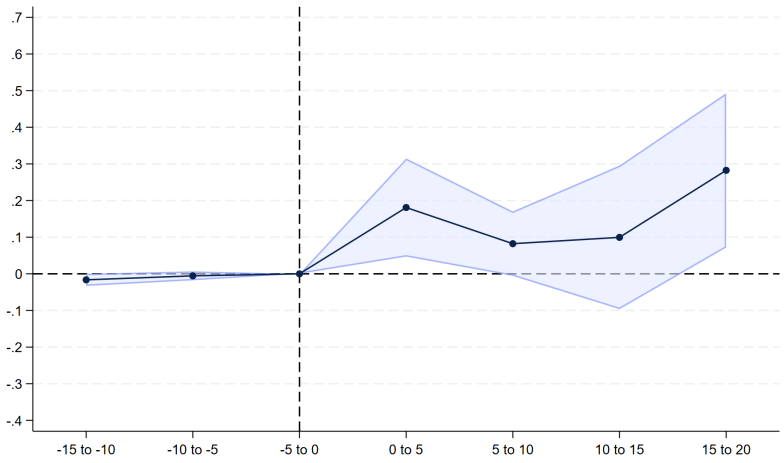


Figure A18: event-study estimates for fertilizer expenditure per acre (including distance to nearest weather station)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

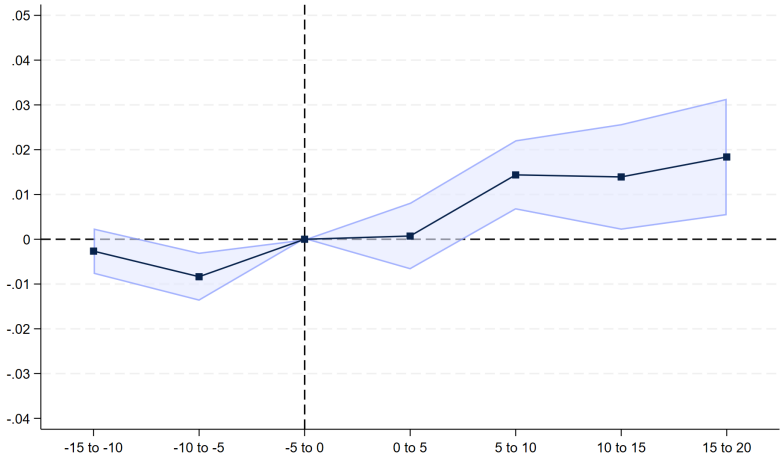


Figure A19: event-study estimates for self-own farmland (including distance to nearest weather station)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

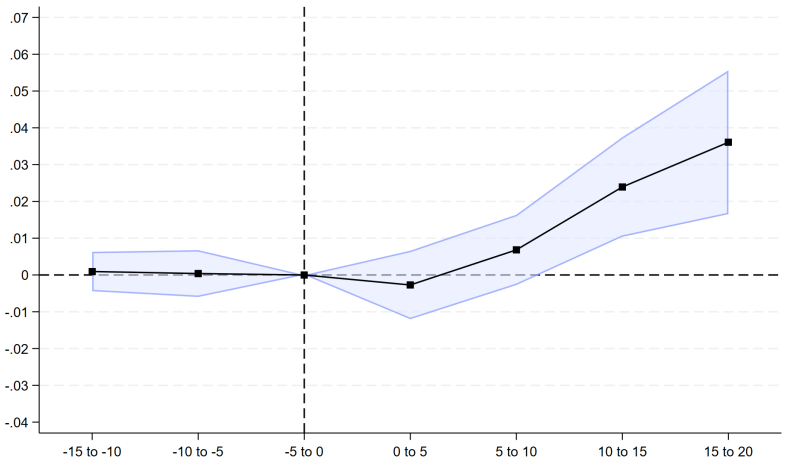


Figure A20: event-study estimates for SWI (including distance to nearest weather station)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

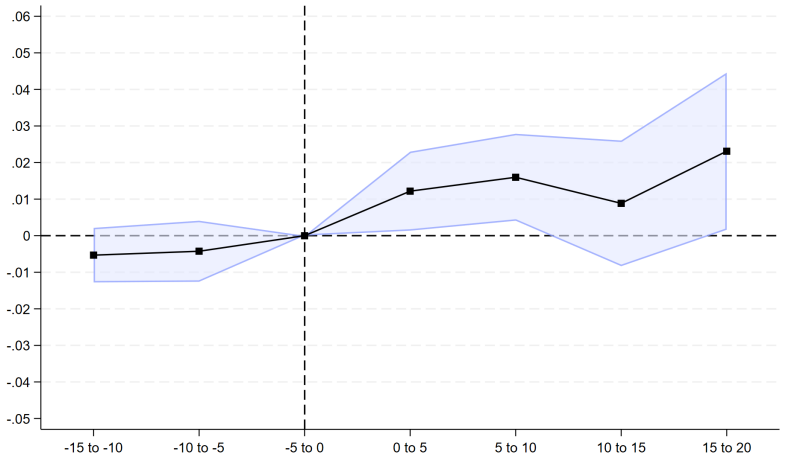


Figure A21: event-study estimates for HHI (including distance to nearest weather station)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is HHI. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table A10. Effects of NCSS Programs on Agricultural Outcomes (by Spatial Distribution)

	revenue	machinery	fertilizer	% of self-own farms	SWI	HHI
Panel A. 25 percentile						
$NCSS_{ct}$	0.054***	0.035***	0.271***	0.015***	0.004	0.009*
	(0.012)	(0.013)	(0.059)	(0.004)	(0.005)	(0.005)
# of clusters	2068	2068	2068	2068	2068	2068
# of obs.	1866480	1866480	1866480	1866480	1866480	1866480
Panel B. 25 to 50 percentile						
$NCSS_{ct}$	0.059***	0.054***	0.260***	0.016***	0.000	0.012**
	(0.013)	(0.013)	(0.057)	(0.004)	(0.004)	(0.006)
# of clusters	2066	2066	2066	2066	2066	2066
# of obs.	1809080	1809080	1809080	1809080	1809080	1809080
Panel C. 50 to 75 percentile						
$NCSS_{ct}$	0.064***	0.056***	0.257***	0.022***	0.007	0.012*
	(0.015)	(0.014)	(0.062)	(0.004)	(0.004)	(0.007)
# of clusters	2065	2065	2065	2065	2065	2065
# of obs.	1619184	1619184	1619184	1619184	1619184	1619184
county pair FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

Table A11. Effects of NCSS Programs on Agricultural Outcomes (Including Other Local Natural Endowment/Conditions)

	revenue	machinery	fertilizer	% of self-own farms	SWI	HHI
Panel A. all counties						
$NCSS_{ct}$	0.036***	0.070***	0.215***	0.015***	0.006*	0.015***
	(0.011)	(0.011)	(0.050)	(0.003)	(0.004)	(0.005)
average temperature$_{st}$	-0.007***	-0.002**	0.026***	0.000	0.001***	0.000
	(0.001)	(0.001)	(0.004)	(0.000)	(0.000)	(0.000)
total precipitation$_{st}$	0.000	-0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$hurricane_{ct}$	0.006	-0.053***	0.191***	-0.002	-0.002	0.019***
	(0.006)	(0.008)	(0.021)	(0.001)	(0.001)	(0.003)
$tornado_{ct}$	-0.081***	-0.103***	-0.095***	0.013***	0.029***	-0.024***
	(0.009)	(0.011)	(0.018)	(0.003)	(0.003)	(0.005)
# of clusters	2148	2148	2148	2148	2148	2148
# of obs.	6444	6444	6444	6444	6444	6444
Panel B. adjacent counties						
$NCSS_{ct}$	0.022**	0.038***	0.169***	0.011***	0.004	0.027***
	(0.011)	(0.011)	(0.044)	(0.003)	(0.004)	(0.005)
average temperature$_{st}$	-0.008***	-0.003**	0.029***	-0.001**	0.002***	0.000
	(0.001)	(0.001)	(0.004)	(0.000)	(0.000)	(0.001)
total precipitation$_{st}$	0.000	-0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$hurricane_{ct}$	-0.050***	-0.054***	0.218***	-0.002	-0.003	0.023***
	(0.007)	(0.008)	(0.022)	(0.001)	(0.002)	(0.003)
$tornado_{ct}$	-0.102***	-0.090***	-0.115***	0.007**	0.028***	-0.023***
	(0.011)	(0.012)	(0.023)	(0.003)	(0.003)	(0.005)
# of clusters	1826	1826	1826	1826	1826	1826
# of obs.	5478	5478	5478	5478	5478	5478
county FE	✓	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county.

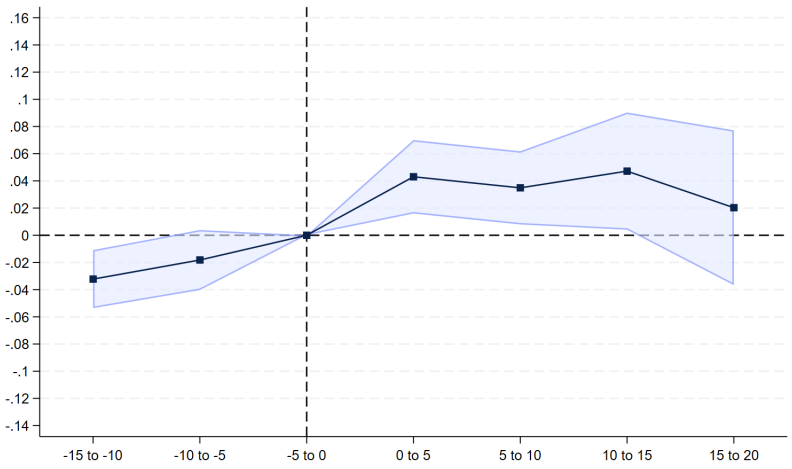


Figure A22: event-study estimates for total crop revenue per acre (including natural endowment/condition)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

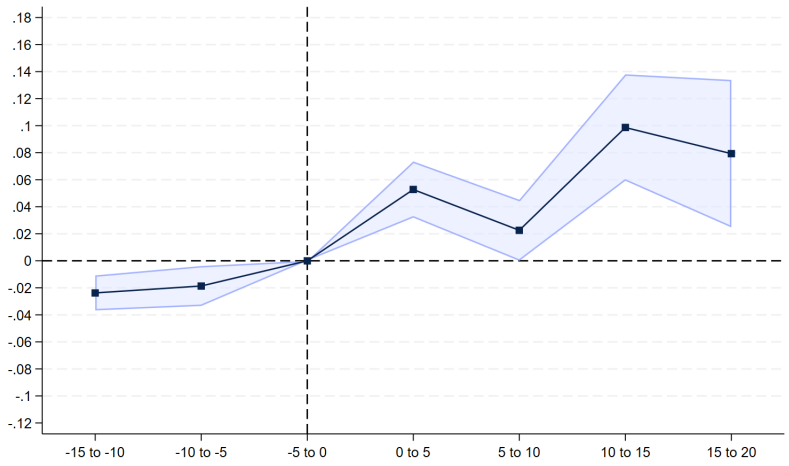


Figure A23: event-study estimates for value of machinery (including natural endowment/condition)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

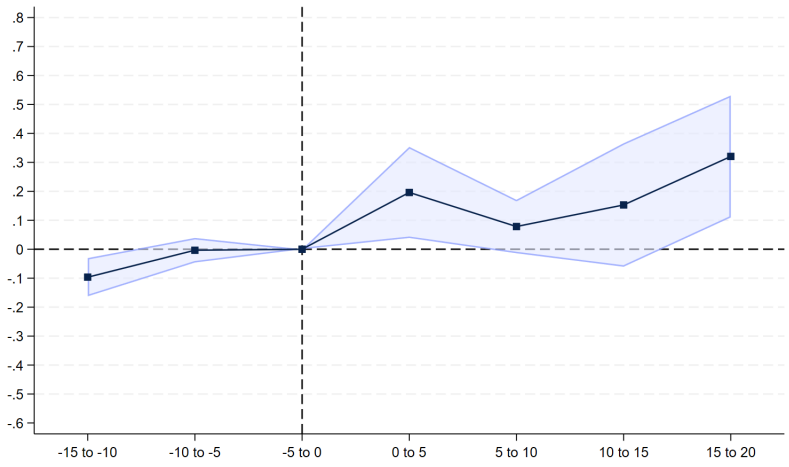


Figure A24: event-study estimates for expenses on fertilizer per acre (including natural endowment/condition)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

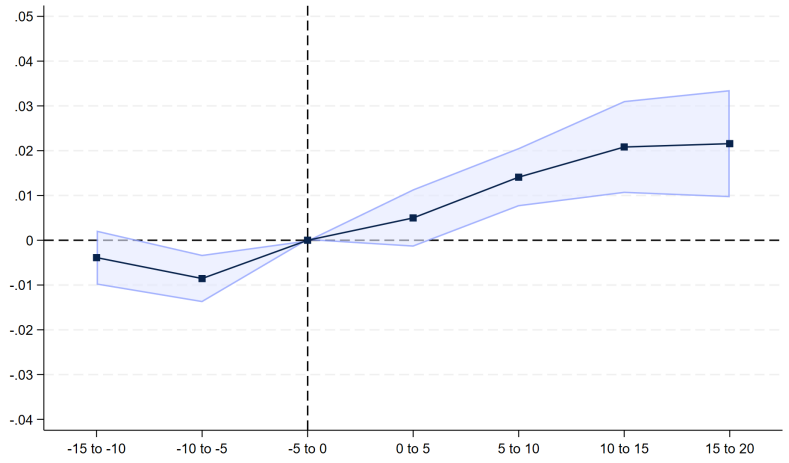


Figure A25: event-study estimates for the share of self-owned farms (including natural endowment/condition)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

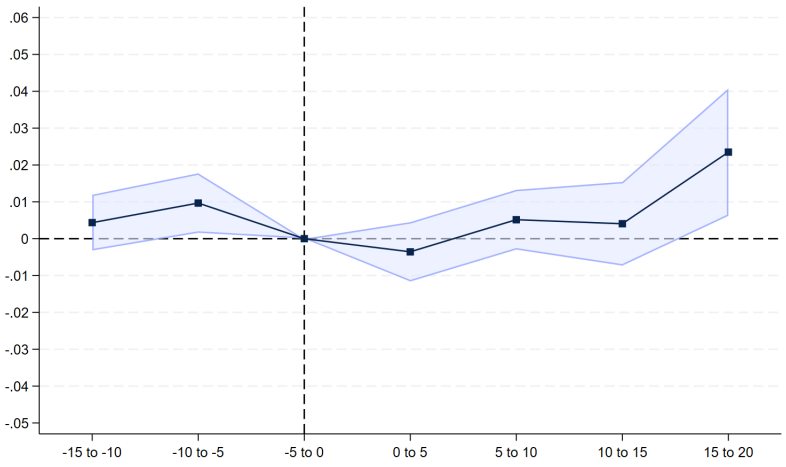


Figure A26: event-study estimates for SWI (including natural endowment/condition)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

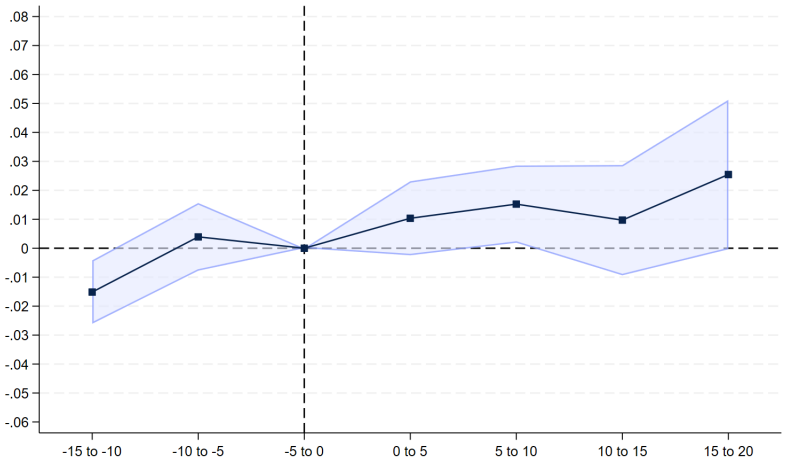


Figure A27: event-study estimates for HHI (including natural endowment/condition)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is HHI. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

A.2. Additional Results

Alternative Measure of Crop Revenue/Productivity. This section presents additional results relevant to alternative measures of agricultural productivity based on crop revenue.

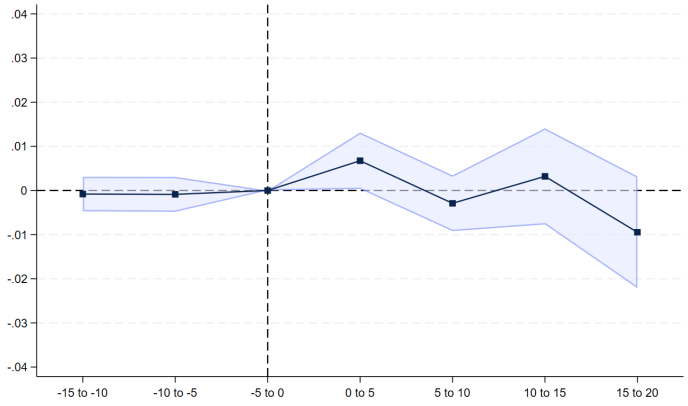


Figure A28: event-study estimates for the ratio between cropland and farmland

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the ratio between cropland acre and farmland acre per county. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

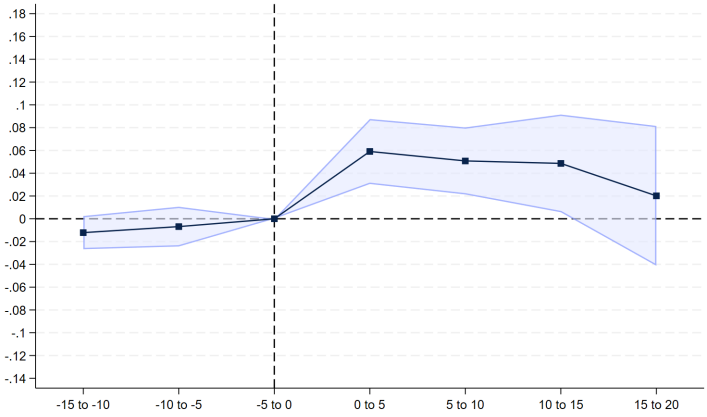


Figure A29: event-study estimates for total crop revenue per farmland acre

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

Table A12. Effects of NCSS Programs on Share of Cropland

	dependent variable: $\frac{cropland}{farmland}$									
	all counties sample					*adjacent counties sample*				
$NCSS_{cst}$	0.002	0.005*	-0.001	0.002	-0.003	0.004	0.007**	-0.002	0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
county FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year FE	✓		✓			✓		✓		
year×region FE		✓		✓	✓		✓		✓	✓
sample period	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920
mean (1890)	3.422	3.422	3.422	3.422	3.422	4.416	4.416	4.416	4.416	4.416
# of clusters	2148	2148	2148	2148	2148	1826	1826	1826	1826	1826
# of obs.	10740	10740	8592	8592	8592	9130	9130	7304	7304	7304

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). The dependent variable is the ratio between cropland acre and farmland acre per county. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county (as shown in Columns (5) and (10)).

Table A13. Effects of NCSS Programs on Agricultural Productivity

	dependent variable: log(crop revenue per acre of farmland)									
	all counties sample					*adjacent counties sample*				
$NCSS_{cst}$	0.014	0.025**	0.055***	0.056***	0.053***	0.021*	0.030***	0.041***	0.038***	0.031**
	(0.013)	(0.013)	(0.011)	(0.011)	(0.013)	(0.011)	(0.011)	(0.011)	(0.010)	(0.012)
county FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year FE	✓		✓			✓		✓		
year×region FE		✓		✓	✓		✓		✓	✓
sample period	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920
mean (1890)	3.422	3.422	3.422	3.422	3.422	4.416	4.416	4.416	4.416	4.416
# of clusters	2148	2148	2148	2148	2148	1826	1826	1826	1826	1826
# of obs.	10740	10740	8592	8592	8592	9130	9130	7304	7304	7304

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). The dependent variable is the logarithm of total crop revenue per acre of farmland. Crop revenue is constructed as the total value of production for nine principal crops: wheat, barley, corn, oat, potato, sweet potato, cotton, buckwheat, and rye. For each crop, production is calculated as the quantity harvested multiplied by its corresponding national-level price. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county (as shown in Columns (5) and (10)).

Table A14. Effects of NCSS Programs on Agricultural Productivity (including area-based NCSS)

	dependent variable: log(crop revenue per acre of cropland)									
	all counties sample					*adjacent counties sample*				
Panel A.										
$NCSS_{cst}$	0.025*	0.036***	0.059***	0.043***	0.050***	0.035***	0.045***	0.059***	0.053***	0.047***
	(0.014)	(0.013)	(0.011)	(0.011)	(0.016)	(0.010)	(0.009)	(0.009)	(0.009)	(0.014)
Panel B.										
$NCSS_{cst}$	-0.003	0.028**	0.050***	0.035***	0.040**	0.004	0.038***	0.044***	0.049***	0.043***
	(0.015)	(0.013)	(0.011)	(0.011)	(0.016)	(0.010)	(0.009)	(0.009)	(0.009)	(0.013)
county FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
year FE	✓		✓			✓		✓		
year×region FE		✓		✓	✓		✓		✓	✓
sample period	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920
mean (1890)	3.422	3.422	3.422	3.422	3.422	4.416	4.416	4.416	4.416	4.416
# of clusters	2361	2361	2361	2361	2361	2088	2088	2088	2088	2088
# of obs.	11805	11805	9444	9444	9444	10440	10440	8352	8352	8352

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). The dependent variable is the logarithm of total crop revenue per acre of farmland. Crop revenue is constructed as the total value of production for nine principal crops: wheat, barley, corn, oat, potato, sweet potato, cotton, buckwheat, and rye. For each crop, production is calculated as the quantity harvested multiplied by its corresponding national-level price. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county (as shown in Columns (5) and (10)). Panel A treats all area-based NCSS mapping as parts of the control group; Panel B define all counties covered by area-based NCSS mapping as the treatment group.

Table A15. Effects of NCSS Programs on Crop Concentration

	all counties sample				adjacent counties sample			
	SWI		*HHI*		*SWI*		*HHI*	
$NCSS_{cst}$	0.008**	0.010**	0.014**	0.007	0.002	0.008**	0.030***	0.023***
	(0.004)	(0.004)	(0.006)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)
county FE	✓	✓	✓	✓	✓	✓	✓	✓
year FE	✓		✓			✓		✓
year×region FE		✓		✓	✓		✓	
sample period	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920	1890-1930	1890-1930	1890-1920
# of clusters	2361	2361	2361	2361	2088	2088	2088	2088
# of obs.	11805	9444	11805	9444	10440	8352	10440	8352

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). The sample defines all counties covered by area-based NCSS mapping as the treatment group. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county (as shown in Columns (5) and (10)).

Table A16. Effects of NCSS Programs on Agricultural Inputs

	all counties sample			adjacent counties sample		
	machinery	*fertilizer*	*self-own*	*machinery*	*fertilizer*	*self-own*
$NCSS_{cst}$	0.114***	0.157***	0.014***	0.085***	0.123***	0.014***
	(0.016)	(0.020)	(0.003)	(0.014)	(0.016)	(0.003)
county FE	✓	✓	✓	✓	✓	✓
year FE	✓		✓			✓
year×region FE		✓		✓	✓	
sample period	1890-1930	1890-1930	1890-1920	1890-1920	1890-1920	1890-1930
# of clusters	2361	2361	2361	2088	2088	2088
# of obs.	9444	9444	9444	8352	8352	8352

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Data are drawn from the Census of Agriculture (1890-1930). The sample defines all counties covered by area-based NCSS mapping as the treatment group. "All-counties sample" includes all 2148 counties of interest in the first five columns, and "adjacent counties sample" only includes the 1826 adjacent counties sharing county borders. County-level controls (Haines et al., 2010) include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level. Results are robust when weighting the sample using the 1900 acre of land in each county (as shown in Columns (5) and (10)).

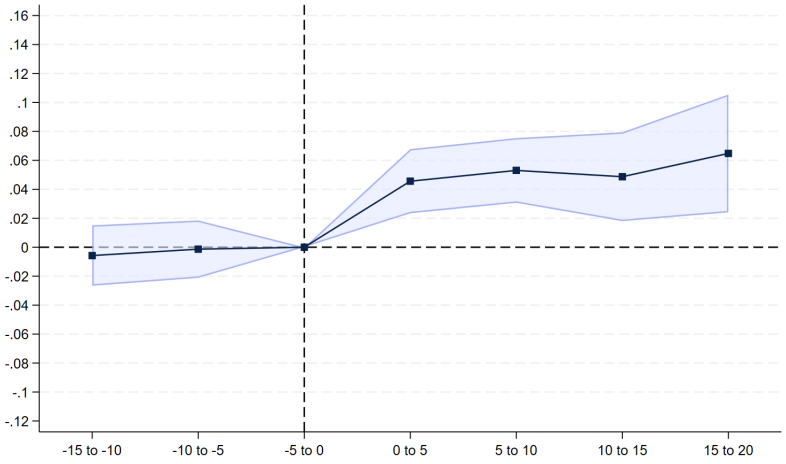


Figure A30: event-study estimates for crop revenue (including area-based surveys)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

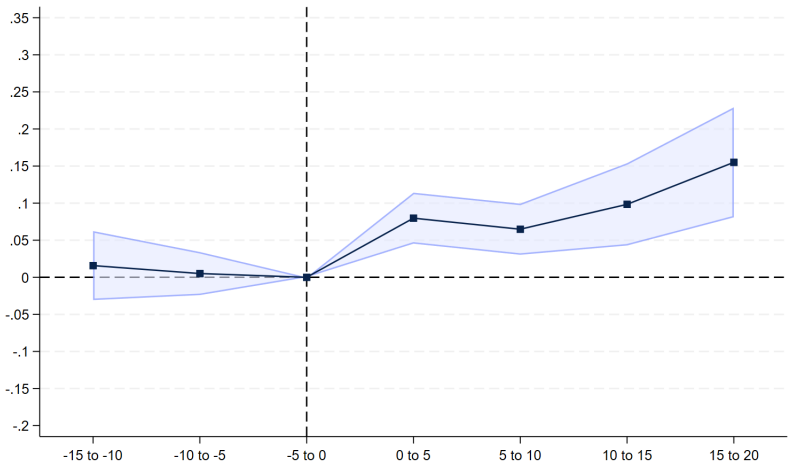


Figure A31: event-study estimates for the value of machinery (including area-based surveys)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

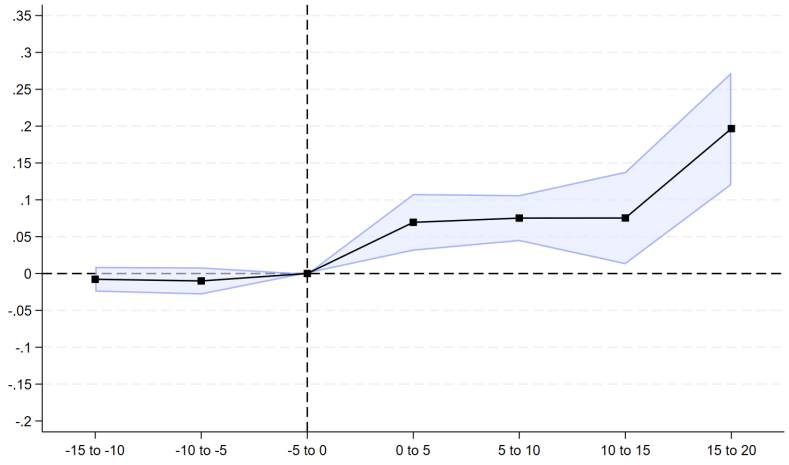


Figure A32: event-study estimates for fertilizer costs (including area-based surveys)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

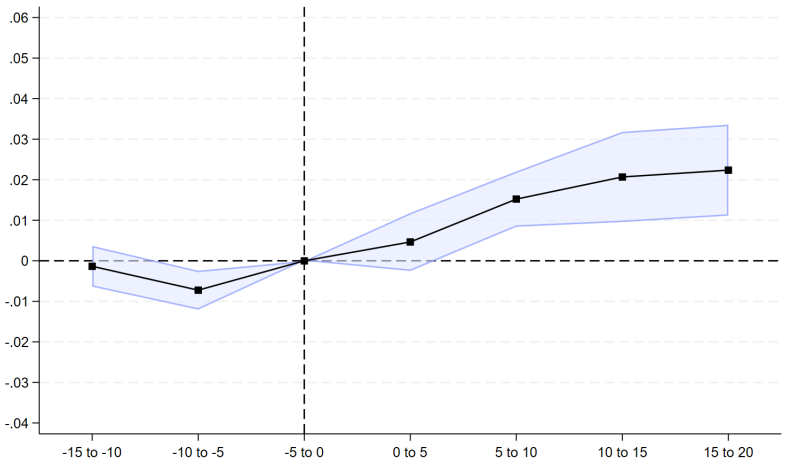


Figure A33: event-study estimates for the share of self-owned farmland (including area-based surveys)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

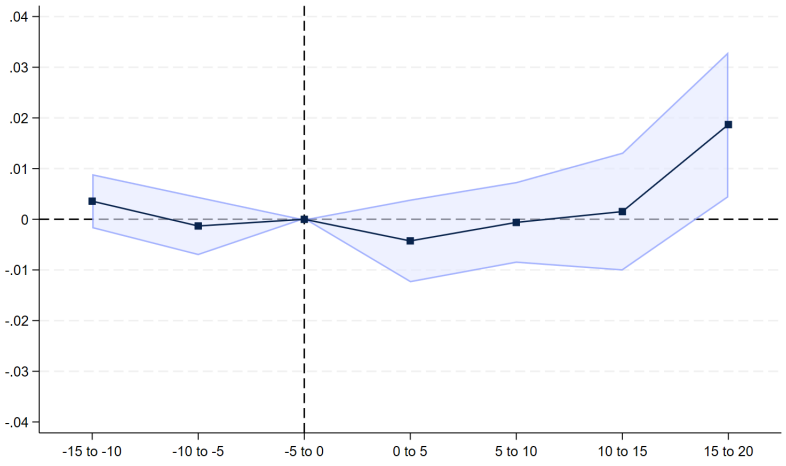


Figure A34: event-study estimates for SWI (including area-based surveys)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is the Shannon-Weaver index. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

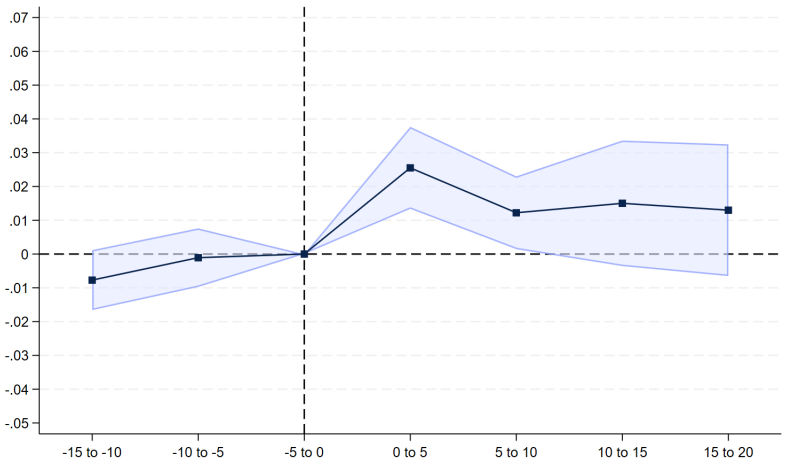


Figure A35: event-study estimates for SWI (including area-based surveys)

Notes: Data are drawn from the Census of Agriculture (1890-1920). The sample includes all counties described in Section 5 (2148 counties). The dependent variable is HHI. The event-study specification includes county fixed effects and year fixed effects. County controls include natural log of the population, natural log of market access information, and a county's urban status. All standard errors are clustered at the county level.

A.3. Additional Figures

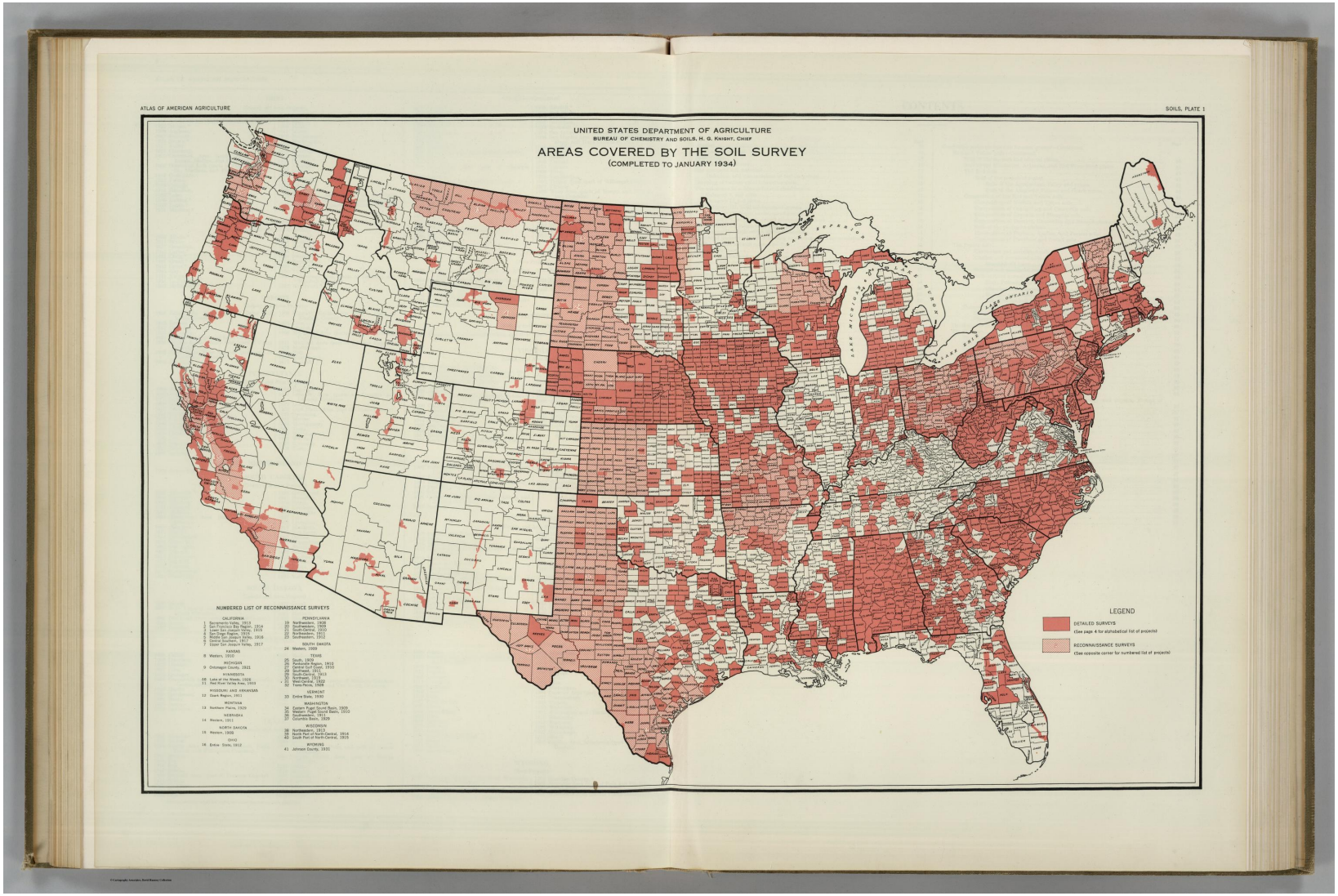
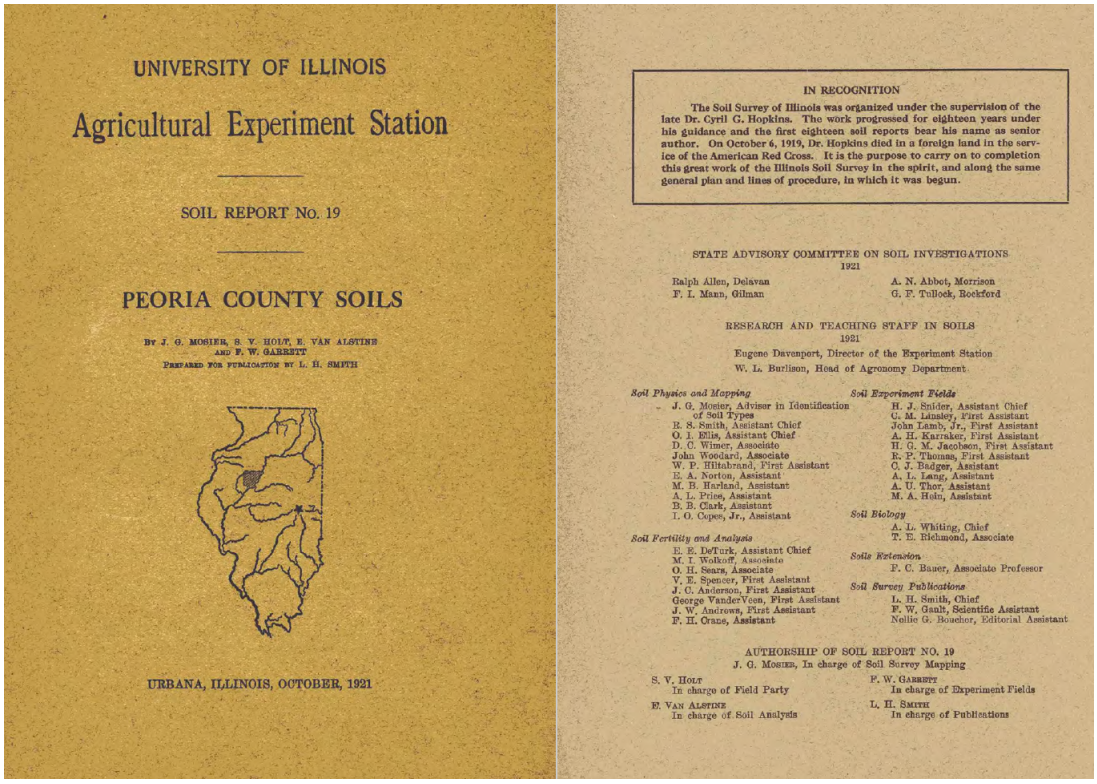


Figure A36: snapshot of the progress of NCSS



UNIVERSITY OF ILLINOIS

Agricultural Experiment Station

SOIL REPORT No. 19

PEORIA COUNTY SOILS

BY J. G. MOSIER, S. V. HOLT, E. VAN ALSTINE AND F. W. GARRETT

PREPARED FOR PUBLICATION BY L. H. SMITH

URBANA, ILLINOIS, OCTOBER, 1921

IN RECOGNITION

The Soil Survey of Illinois was organized under the supervision of the late Dr. Cyril G. Hopkins. The work progressed for eighteen years under his guidance and the first eighteen soil reports bear his name as senior author. On October 6, 1919, Dr. Hopkins died in a foreign land in the service of the American Red Cross. It is the purpose to carry on to completion this great work of the Illinois Soil Survey in the spirit, and along the same general plan and lines of procedure, in which it was begun.

STATE ADVISORY COMMITTEE ON SOIL INVESTIGATIONS
1921

Ralph Allen, Delavan — A. N. Abbot, Morrison
F. I. Mann, Gilman — G. F. Tullock, Rockford

RESEARCH AND TEACHING STAFF IN SOILS
1921

Eugene Davenport, Director of the Experiment Station
W. L. Burlison, Head of Agronomy Department

Soil Physics and Mapping
J. G. Mosier, Advisor in Identification of Soil Types
B. S. Smith, Assistant Chief
O. I. Ellis, Assistant Chief
D. C. Wimer, Associate
John Woodard, Associate
W. P. Hittabrand, First Assistant
E. A. Norton, Assistant
M. B. Harland, Assistant
A. L. Price, Assistant
B. B. Clark, Assistant
I. O. Copes, Jr., Assistant

Soil Fertility and Analysis
E. E. DeTurk, Assistant Chief
M. I. Wolkoff, Associate
O. H. Sears, Associate
V. E. Spencer, First Assistant
J. C. Anderson, First Assistant
George VanderVeen, First Assistant
J. W. Andrews, First Assistant
F. H. Crane, Assistant

Soil Experiment Fields
H. J. Snider, Assistant Chief
C. M. Linsley, First Assistant
John Lamb, Jr., First Assistant
A. H. Karraker, First Assistant
H. G. M. Jacobson, First Assistant
E. P. Thomas, First Assistant
C. J. Badger, Assistant
A. L. Lang, Assistant
A. U. Thor, Assistant
M. A. Hohn, Assistant

Soil Biology
A. L. Whiting, Chief
T. E. Richmond, Associate

Soils Extension
F. C. Bauer, Associate Professor

Soil Survey Publications
L. H. Smith, Chief
F. W. Gault, Scientific Assistant
Nellie G. Bouchor, Editorial Assistant

AUTHORSHIP OF SOIL REPORT NO. 19
J. G. Mosier, In charge of Soil Survey Mapping

S. V. Holt — In charge of Field Party
E. Van Alstine — In charge of Soil Analysis
F. W. Garrett — In charge of Experiment Fields
L. H. Smith — In charge of Publications

Figure A37: snapshot of NCSS reports

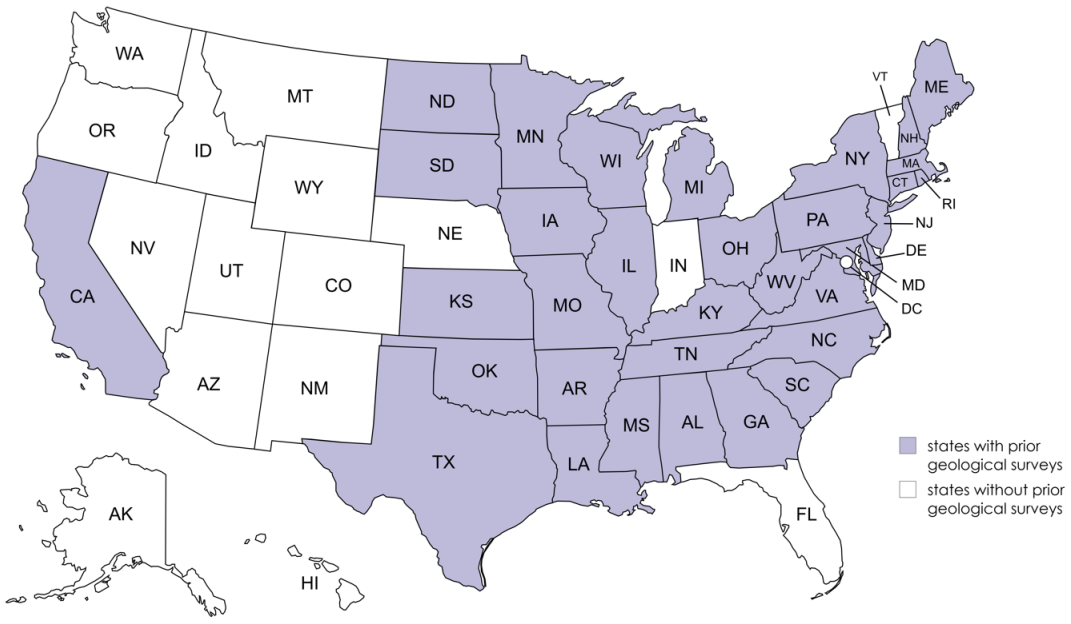
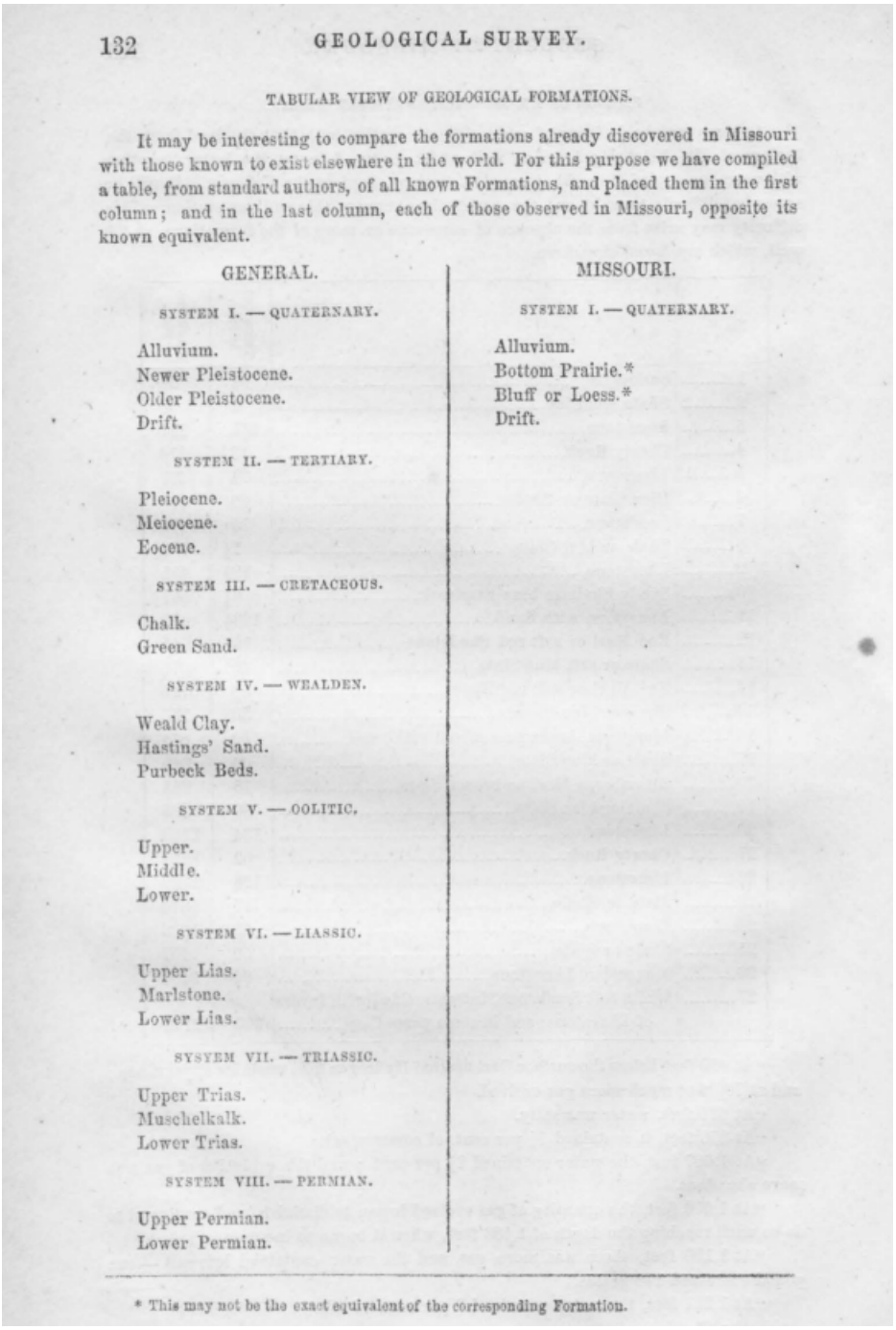


Figure A38: distribution of state geological surveys before 1899



132 GEOLOGICAL SURVEY.

TABULAR VIEW OF GEOLOGICAL FORMATIONS.

It may be interesting to compare the formations already discovered in Missouri with those known to exist elsewhere in the world. For this purpose we have compiled a table, from standard authors, of all known Formations, and placed them in the first column; and in the last column, each of those observed in Missouri, opposite its known equivalent.

GENERAL.	MISSOURI.
SYSTEM I. — QUATERNARY.	SYSTEM I. — QUATERNARY.
Alluvium.	Alluvium.
Newer Pleistocene.	Bottom Prairie.*
Older Pleistocene.	Bluff or Loess.*
Drift.	Drift.
SYSTEM II. — TERTIARY.	
Pleiocene.	
Meiocene.	
Eocene.	
SYSTEM III. — CRETACEOUS.	
Chalk.	
Green Sand.	
SYSTEM IV. — WEALDEN.	
Weald Clay.	
Hastings' Sand.	
Purbeck Beds.	
SYSTEM V. — OOLITIC.	
Upper.	
Middle.	
Lower.	
SYSTEM VI. — LIASSIC.	
Upper Lias.	
Marlstone.	
Lower Lias.	
SYSYEM VII. — TRIASSIC.	
Upper Trias.	
Muschelkalk.	
Lower Trias.	
SYSTEM VIII. — PERMIAN.	
Upper Permian.	
Lower Permian.	

* This may not be the exact equivalent of the corresponding Formation.

Figure A39: example of state geological surveys before 1899

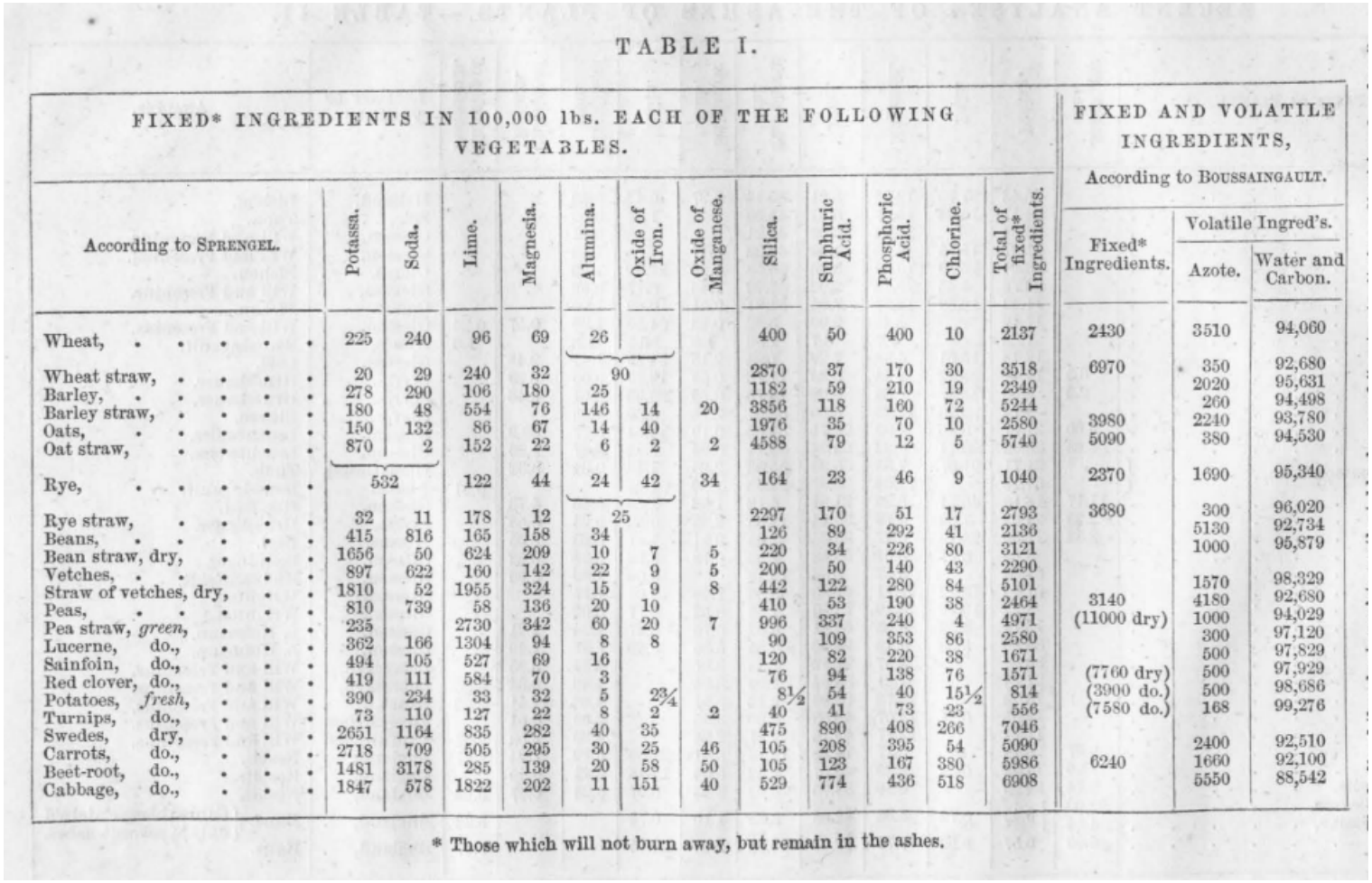


TABLE I.

According to SPRENGEL.	Potassa.	Soda.	Lime.	Magnesia.	Alumina.	Oxide of Iron.	Oxide of Manganese.	Silica.	Sulphuric Acid.	Phosphoric Acid.	Chlorine.	Total of fixed* Ingredients.	FIXED AND VOLATILE INGREDIENTS, According to BOUSSAINGAULT. Fixed* Ingredients.	Volatile Ingred's. Azote.	Volatile Ingred's. Water and Carbon.
Wheat,	225	240	96	69	26			400	50	400	10	2137	2430	3510	94,060
Wheat straw,	20	29	240	32	90			2870	37	170	30	3518	6970	350	92,680
Barley,	278	290	106	180	25			1182	59	210	19	2349		2020	95,631
Barley straw,	180	48	554	76	146	14	20	3856	118	160	72	5244		260	94,498
Oats,	150	132	86	67	14	40		1976	35	70	10	2580	3980	2240	93,780
Oat straw,	870	2	152	22	6	2	2	4588	79	12	5	5740	5090	380	94,530
Rye,	532		122	44	24	42	34	164	23	46	9	1040	2370	1690	95,340
Rye straw,	32	11	178	12	25			2297	170	51	17	2793	3680	300	96,020
Beans,	415	816	165	158	34			126	89	292	41	2136		5130	92,734
Bean straw, dry,	1656	50	624	209	10	7	5	220	34	226	80	3121		1000	95,879
Vetches,	897	622	160	142	22	9	5	200	50	140	43	2290			
Straw of vetches, dry,	1810	52	1955	324	15	9	8	442	122	280	84	5101		1570	98,329
Peas,	810	739	58	136	20	10		410	53	190	38	2464	3140	4180	92,680
Pea straw, *green*,	235		2730	342	60	20	7	996	337	240	4	4971	(11000 dry)	1000	94,029
Lucerne, do.,	362	166	1304	94	8	8		90	109	353	86	2580		300	97,120
Sainfoin, do.,	494	105	527	69	16			120	82	220	38	1671		500	97,829
Red clover, do.,	419	111	584	70	3			76	94	138	76	1571	(7760 dry)	500	97,929
Potatoes, *fresh*,	390	234	33	32	5	2¾		8½	54	40	15½	814	(3900 do.)	500	98,686
Turnips, do.,	73	110	127	22	8	2	2	40	41	73	23	556	(7580 do.)	168	99,276
Swedes, dry,	2651	1164	835	282	40	35		475	890	408	266	7046			
Carrots, do.,	2718	709	505	295	30	25	46	105	208	395	54	5090		2400	92,510
Beet-root, do.,	1481	3178	285	139	20	58	50	105	123	167	380	5986	6240	1660	92,100
Cabbage, do.,	1847	578	1822	202	11	151	40	529	774	436	518	6908		5550	88,542

* Those which will not burn away, but remain in the ashes.

Figure A40: example of state geological surveys before 1899

SOIL SURVEY OF THE WAYCROSS AREA, GEORGIA.

By M. EARL CARR and W. E. THARP.

DESCRIPTION OF THE AREA.

The Waycross area is located in the heart of the "wire grass" section of southeast Georgia, about 60 miles inland from the coast, and covers the main part of Ware County. The thirty-first parallel of north latitude forms the southern boundary of the area and separates it from Charlton County and a southern extension of Ware County. Clinch County lies on the southwest, and Coffee County on the west and also north of that portion extending westward on the south side of the Satilla River and Red Bluff Creek, which separate this westward extension from it. Appling County forms the extreme northern boundary and Pierce County the entire eastern boundary. The eastern boundary north of the river has been made a straight north and

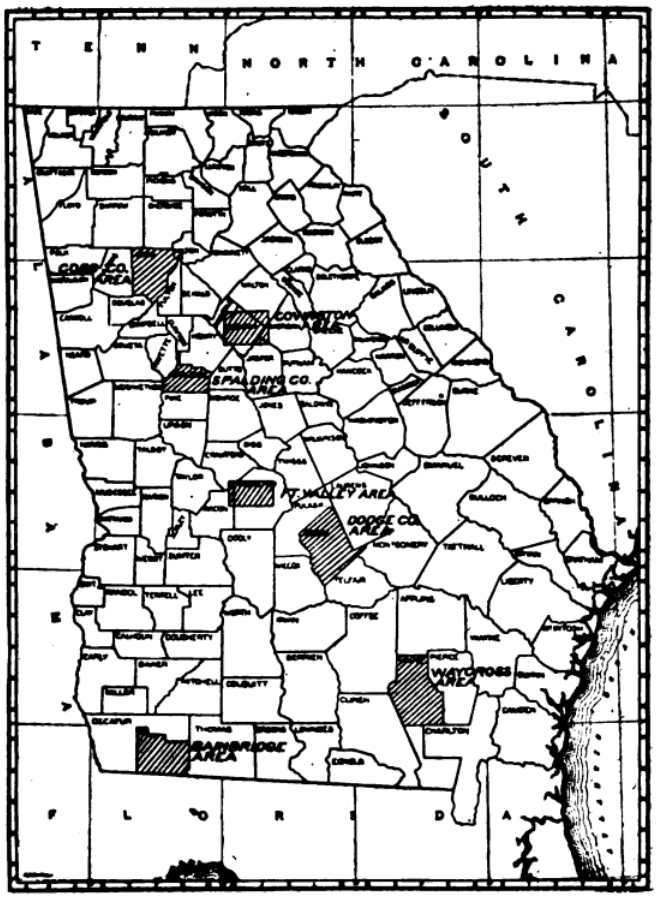


Figure A41: an example of small-area-based NCSS (Waycross Area, Georgia, 1906)

SOIL SURVEY OF THE MODESTO-TURLOCK AREA, CALIFORNIA, WITH A BRIEF REPORT ON A RECONNAISSANCE SOIL SURVEY OF THE REGION EAST OF THE AREA.

By A. T. SWEET, J. F. WARNER, and L. C. HOLMES.

DESCRIPTION OF THE AREA.

The Modesto-Turlock area lies approximately in the center of the State of California from east to west and slightly north of the center from north to south. It is about 100 miles east of San Francisco, 300 miles north of Los Angeles, and near the center of the great valley which extends from northwest to southeast between the Sierra Nevada Mountains on the east and the Coast Range on the west. The upper part of this valley is drained to the south by the Sacramento River and the southern part is drained to the north by the San Joaquin River, the two uniting and emptying into Suisun Bay. Rising in the Sierra Nevada Mountains along the east side of this valley, flowing to the west and emptying into the San Joaquin, are three rivers of considerable importance, the Merced, the Tuolumne, and the Stanislaus. The Modesto-Turlock area lies on the east side of the San Joaquin and is drained



FIG. 34.—Sketch map showing location of the Modesto-Turlock area, California.

Figure A42: an example of Reconnaissance NCSS (1908)