Research Proximity and Productivity: Long-Term Evidence from Agriculture[*]

Shawn Kantor[†] Alexander Whalley[‡]

May 2014

Abstract

The spatial concentration of ideas is central to economic geography. Yet, how proximity to research affects productivity is not well studied. We use the late 19th century establishment of agricultural experiment stations in the United States to estimate the importance of proximity to research for productivity growth. Our analysis of county-level agricultural census data from 1870 to 2000 reveals three results. First, research proximity effects from permanent station opening grew for about 20 years and then subsequently declined until becoming largely absent today. Second, proximity to station-based innovations affected local farmers' productivity for 20 to 40 years after the discovery. Third, research proximity effects remain today where stations historically focused on basic research and where nearby farmers were producing with frontier technology. Persistence in research proximity effects depend not just on research infrastructure, but also persistence in idea production and the cumulative effects of learning.

Keywords: Ideas, Innovation, Geography, Productivity
JEL Classifications: R11, O33, N50

---

[*] We thank Peter Berck, Mark Bils, Greg Caetano, David Card, Selim Gulesci, Alain de Janvry, Richard Hornbeck, Caroline Hoxby, Bill Kerr, Josh Kinsler, Megan MacGarvie, Enrico Moretti, Petra Moser, Alan Olmstead, Paul Rhode, Jeff Smith, Mitchell Stevens, Bill Sundstrom, Andy Toole, Sarah Turner, Romain Wacziarg, Fabian Waldinger, Bruce Weinberg, Heidi Williams, Diego Ubfal, John Van Reenen, and David Zilberman for helpful comments. We are equally grateful to seminar participants at Berkeley, Binghamton, Bocconi, Brown, BU, Harvard, LSE, Maryland, Rochester, Santa Clara, Stanford, Syracuse, UCL, USDA-ERS, and Warwick and to participants at the All-UC Economic History, Cliometrics, NBER-DAE Summer Institute, RCEF, and Urban Economics Association conferences. We also thank Jeremy Atack, Price Fishback, and Paul Rhode for sharing data. We gratefully acknowledge financial support form the National Science Foundation under grants SES-0851788 and SBE-1158794. Chris Abrescy, Aly Englert, Tony Hua, Nicole Nguyen, Michael Wall, and Marko Zivanovic provided excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the NSF.

[†] Rensselaer Polytechnic Institute & NBER.

[‡] University of California, Merced & NBER.

1 Introduction

Economists have long stressed the local nature of idea flows (Marshall 1890), and local economic outcomes increasingly depend on local idea generation (Moretti 2012). While access to goods markets can lead to long-term regional development (Krugman 1991), what about access to ideas? Despite the high spatial correlation of income and research, the role ideas play in regional economic growth is not well studied.[1]

Two central challenges have limited progress. Distinguishing the effects of local ideas from fundamentals is a tall order. Measurement problems arise because those who produce ideas locally also receive them, and persistent unobserved factors can lead to the co-location of idea producers and highly productive firms. In addition, as regional adjustments take many decades to manifest (Davis and Weinstein 2002; Collins and Margo 2007; Hornbeck 2012), separating long- from short-term effects requires data typically unavailable.

By examining the productivity effects of proximity to research in agriculture, we are able to address both challenges. We propose that the establishment of federal agricultural experiment stations in the late 19th century serves as a source of exogenous variation in the location of knowledge production. Because the stations were opened at pre-determined land grant university locations in response to nationwide concerns about agriculture, they created a positive shock to research virtually independent of local economic conditions. Furthermore, detailed historical productivity data allow research proximity effects to be estimated over a time horizon possible in few other contexts. While estimates of the aggregate return to agricultural research spending are very large – typically over 40% (Alston et. al. 2000) – no research to date has explored how research proximity has affected regional productivity in agriculture over the very long-term.[2]

---

[1] To get a sense of the connection between income and knowledge production, consider that OECD countries account for 80% of global expenditures on R&D overall (National Science Board 2010), and that high-income countries account for about 90% of expenditures on R&D in agriculture (Alston and Pardey 2014). Similarly, the five MSAs that produce 30% of U.S. patents have per capita income 35% above the U.S. average (USPTO 2010).

[2] There is a large literature in agricultural economics that has estimated the spillover effects from agricultural research. See Alston et. al. (2010) for a review. We differ by focusing on variation in research spending unlikely to be driven by an unobserved local agricultural shock, and by studying county-level research proximity over a very long time period.

Experiment station researchers discovered biological innovations (i.e., new and improved crop varieties) that powered productivity growth in agriculture.[3] Farmers themselves required knowledge of appropriate growing conditions and complementary inputs to profitably adopt these technologies. That knowledge was often diffused through social interaction, which remains the case today in developing counties (Foster and Rosenzweig 2010), suggests a strong role for proximity-based knowledge transfer. Proximity effects, however, were unlikely to be static because knowledge accumulates and the pace of discovery can change.

How would proximity effects change over the short- to long-term? To see, we consider the dynamic implications of non-rival ideas in a spatial endogenous growth framework. In equilibrium farmers near research facilities acquire sufficient technological know-how to adopt new discoveries quickly and become more productive. Proximity to research is beneficial in the short term. Whether proximity effects grow or decline over time is less clear. On the one hand, discovery may speed up as scientists learn from increasingly knowledgeable local farmers, making long-term proximity effects stronger.[4] On the other hand, diminishing returns to research could reduce the pace of innovation, leading to weaker proximity effects over the long-term.[5] As the dynamics of discovery and learning are unclear, how proximity effects differ over the short- and long-term is an empirical question.

Our empirical analysis begins by examining how proximity to experiment station research affected land productivity. We find that research proximity effects started out small and

---

[3] Olmstead and Rhode (2008, 61) conclude that about half of the increase in wheat productivity from 1839 to 1909 can be attributed to biological innovation. Similarly, Johnson and Gustafson (1962) show that new varieties accounted for 60 percent of the increase in yields for U.S. grain production from 1928 to 1954 in the West.

[4] Many modern public agricultural research systems seek to make use of farmer knowledge. There are some examples of this phenomenon before the extension programs of the early 20th century were in full effect when station agents travelled to outlying areas to learn from and to teach farmers. In the late 19th century Japan incorporated farmers' expertise into the discovery process directly (Frederico 2009, 106). By having leading farmers develop technologies that were then tested at experiment stations the government sought to develop appropriate innovations after the importation of inappropriate western technologies had failed.

[5] That biological innovation before the discovery of genetics may have been bounded by the stock of genetic material suggests returns to applied biological research fell over time. As Frederico (2009, 86) puts it, "Biological innovation via worldwide transfer is bound to reach a limit, when all known plants and animals have been tested in all reasonably suitable environments. This limit was reached towards the end of the nineteenth century. Since then, almost 'all' varieties of existing plants have been produced with systematic hybridization, and later, genetic engineering (new plants need genetic engineering and, so far, they are the realm of laboratory science)."

grew, peaking between 20-30 years after the experiment stations opened. While research expenditures continued to grow, research proximity effects declined, however, dying out 30-50 years after station opening.[6] These results are robust to the inclusion of controls for distance to minor research stations in some states, urban markets, or railways, controlling for fixed state characteristics over time, and the use of alternative land productivity measures.

We explore three opportunities to evaluate the credibility of our approach. We first examine whether research proximity effects appeared before the opening of the experiment stations. Then we test whether distance to other research facilities, opened by the United States Department of Agriculture much later, show a similar pattern of proximity estimates around the land grant experiment stations' opening date. Last, we examine whether manufacturing output per worker also reveals an agricultural experiment station proximity effect, where no direct effect would be expected. We find little evidence from any of these analyses that our identification strategy is undermined.

Next, we seek to understand the limited persistence of research proximity effects. Limited persistence may arise if farmers' knowledge mattered little for discovery or station research was subject to diminishing returns. Our results reveal larger and more persistent proximity effects in counties that produced a different main crop from the experiment station county. Crop-specific local knowledge does not seem to be central to advancing research. However, we do find stronger and more persistent proximity effects where farmers were initially more productive, suggesting that farmers' general knowledge may have accelerated discovery.

The stations publication of experiment results allow us to document where and when applied innovations occurred. Consistent with diminishing returns we find that even though stations spent more on research over time the frequency of applied innovations generally fell. The effects of proximity to these innovations did persist, however, for 20-40 years after the innovation. Taken together, declining frequencies of applied innovations accounted for 40% of the decline in research proximity effects from their peak. We also find proximity effects

---

[6] The real annualized growth rate of public research spending in the United States from 1890 to 1990 was 4.2% (Huffman and Evenson 1993, 93), so the opening of the experiment stations represented a permanent shock to the location of agricultural research.

persist to the present day where stations historically focused on basic research, which may be less subject to diminishing returns (i.e., Furman and Stern 2011). Diminishing returns to applied research, therefore, appear to have reduced proximity effects over time.

While much literature has emphasized that property rights regimes or various market imperfections can generate cross-area agricultural productivity gaps, our estimates indicate that the spatial distribution of research also plays an economically significant role. To estimate how significant, we compare productivity between the observed spatial distribution of experiment stations and a counterfactual geographic distribution that maximizes farmers' proximity to research. This back of the envelope calculation implies that reallocating research to maximize geographic proximity would have increased national land productivity by 7.7% and national total factor productivity in agriculture by 3.0% in 1910. Unavoidably, our calculation abstracts from many features that may have impacted agricultural productivity, so should be viewed as a first step towards a full welfare analysis of the social returns to reallocating public research closer to potential users of such knowledge.

Our research relates to three different literatures associated with knowledge spillovers or regional development. Closest is the literature on external economies of local ideas. One branch follows Henderson (1974) in considering local knowledge spillovers as exogenous externalities (e.g., Lucas 2001 and Desmet and Rossi-Hansberg 2014). Another has focused on linking density to learning (Glasear 1999; Duranton and Puga 2001; Davis and Dingel 2012). Empirical evidence of the productivity effects of access to ideas is mixed. Within the United States cross-location learning-by-doing productivity effects appear to be small (Thornton and Thompson 2001), yet access to foreign knowledge has important effects on own-country productivity (Keller 2002; Griffith, Harrison, and Van Reenen 2006; Keller and Yeaple 2013). By identifying the causal effects of research proximity on productivity directly we seek to tackle the "black box" nature of local knowledge spillovers head on.

Our work is also related to the emerging literature on the persistence of regional shocks. Recent evidence indicates that the early concentrations of population attracted to portage locations lead to higher populations today (Bleakley and Lin 2012). Similarly, the temporary shock that the U.S. Civil War had on Britain's cotton textile industry led to persistent effects on the location of producers in other industries (Hanlon 2013). Kline and Moretti (2014) show effects of the TVA in manufacturing, where agglomeration economies could generate increasing returns, but not in agriculture, where agglomeration economies are less strong. Our findings highlight that whether permanent research infrastructure has persistent regional effects depends on where highly productive firms are, and what types of ideas (basic science versus applied) are produced.

Finally, our paper relates to recent studies of the time-varying effects of proximity to knowledge producers. The dismissal of German scientists during the Nazi era had effects on innovation that persisted at least 30 years in Germany (Waldinger 2013) and in the United States (Moser, Voena, and Waldinger 2014). Firms located near universities receive spillovers in the short- to medium-term (Aghion, Boustan, Hoxby, and Vandenbussche 2009; Hausman 2012; Kantor and Whalley 2014), as well as the long-term (Liu 2013). We contribute by showing that where knowledge is produced has implications for national productivity, and research proximity effects can be highly dynamic.

2 Historical Background

Station Establishment. Advocates for federal experiment stations had been planning since at least the early 1860s.[7] The mid-1880s contraction amplified the concerns of farmers who were already adjusting to long-term increases in global competition and the rapid expansion of Midwestern agriculture.[8] The "farm problem" of the day created an opening for research station advocates to argue for productivity-enhancing public investments rather than protectionist barriers. Opposition emerged from those concerned about states rights and federal oversight of station finances.[9] Ultimately, the potentially broader benefits to

---

[7] Excellent surveys of the role of public research in U.S. agriculture spanning both the 19th and 20th centuries include Huffman and Evenson (2006) and Alston et. al. (2010). American experiment stations were based on the model of the Moeckern plan in Saxony (Knoblauch et. al. 1962).

[8] As Davis, Hanes and Rhode (2009) show, cotton, wheat, and corn harvests all significantly fell in the mid-1880s.

[9] Rosenberg (1964, 3) argues that "President Cleveland had grave doubts as to the constitutionality of the Hatch Act, the first direct cash grant-in-aid to individual states." Others objected to the idea of granting permanent funding, preferring instead that research be funded on a case-by-case basis as was done

agriculture carried the day and the Hatch Act of 1887 passed with widespread support.[10]

The Hatch Act provided federal funding for the establishment of an experiment station in each state, specifically at the land grant college established under the Morrill Act of 1862. The Hatch Act's "combination plan" touted the complementarities that could be achieved by co-locating teaching and research activities, a hallmark of the land-grant mission.

Research and Innovation. The federal experiment stations provided each land grant university with the capability to generate biological innovations to drive agricultural productivity growth (Olmstead and Rhode 2008). Early research focused on finding new crop varieties, plant breeding, and animal husbandry. Notable biological innovations were achieved quite soon after the stations were established.[11] These innovations are readily apparent in the much higher crop yields achieved in station experiments relative to local farmers, as shown in Figure 1. In addition, as crop varieties depreciated fairly rapidly, new varieties were required to simply keep productivity from falling (Olmstead and Rhode 2002).[12] Other innovations in agriculture, such as the use of appropriate crop rotations and commercial fertilizer, were built from applying new knowledge of chemistry to agricultural practice.

Competing visions quickly emerged regarding the ideal objective of station scientists (Knoblauch et. al. 1962, 82). Many farmers believed the Hatch Act's "practical information" mandate meant that scientists should have primarily investigated farmers' specific concerns (i.e., recommending customized fertilizer given a specific soil type or responding to specific threats related to pests). Experiment station scientists, often trained in chemistry or botany, by contrast, saw the Hatch Act's call for "original researches" as a mandate for basic science. After much debate the Adams Act was passed in 1906 which significantly strengthened the

---

in Michigan at the time (Knoblauch et. al. 1962, 47).

[10] Petitions favoring passage came from 34 states (Harding 1947, 175) and the Act passed both chambers of Congress on a voice vote, based on Poole and Rosenthal's collection of all roll-call votes in Congress. See http://www.voteview.com.

[11] By the end of 1888, 46 states had established stations employing 369 staff nationwide (True 1888). Prior to the Hatch Act, 12 states had already established their own experiment stations and a number of others reported such work without having a formal structure governing the research (Carstensen 1960, 13). We refer to the states with pre-Hatch stations as the "non-Hatch states" in our empirical analysis.

[12] Crop varieties depreciate because pests and fungi adapt to planted varieties. As a result many varieties are only in use for a short period of time. For example, in 1919 roughly 80 percent of wheat acreage consisted of varieties new to North America since 1873 (Olmstead and Rhode 2008, 32).

stations' basic research mission and enhanced the role of the Office of Experiment Stations as overseer of a national research program (Ferleger 1990, 20).

Entrepreneurs and Farmers. While seed breeders were actively producing new crop varieties, such as Fultz wheat in 1862 or Robert Reid's Yellow Dent corn in 1847 (Olmstead and Rhode 2008, 29), well before the stations opened, entrepreneurs soon drew on the discoveries of the stations.[13] Kingsbury (2009, 159) summarizes, "Research into the laws of inheritance fed directly into plant breeding practice." The increased prominence of basic research eventually led scientists to acquire knowledge of genetics, thus enabling breeders to produce important breakthroughs like hybrid corn (Griliches 1957).

More productive farmers were quick to make use of station discoveries. These farmers were active in agricultural societies, agricultural fairs, and read agricultural periodicals that highlighted new ideas from home and abroad. Other less productive farmers who initially derided scientific agriculture as "book farming" lacking real world applicability eventually saw the value of experiment station work. "Intelligent farmers long since ceased to scoff at the work of the agricultural experiment stations, and the unintelligent ones may revise their opinions, " the *Christian Advocate* declared in 1901. "Years ago doubts of the value of experiment station work were commonly expressed among farmers . . . No one now doubts that such discoveries amply justified the trouble and expense involved" (Wiancko 1922).

Some farmers were reluctant to adopt technologies they had not experienced first hand. According to Cornell University's Isaac P. Roberts in 1889, "Farmers, like other people, hesitate to believe and act on theories, or even facts, until they see with their own eye the proof of them in material form" (Scott 1970, 3). To enhance direct farmer experience with advanced technology, a national agricultural extension program was developed about two decades after the widespread opening of research stations. The Smith-Lever Act of 1915 was

---

[13] One scientist described the Funk Brothers Seed Company as "a seed company with the direct purpose of following up on a large commercial scale with the work I have been doing on the breeding of seed corn..." Fitzgerald (1990, 21). Similarly, "While in 1918 and 1920 Donald Jones [of the Connecticut experiment station] theorized about this possibility [double-cross hybrids] in academic publications, Funk demonstrated it on his farm. Indeed, Funk provided Jones with the large scale evidence he lacked" (Fitzgerald 1990, 149). Biological innovations came from other sources as well. Immigrants brought seeds from abroad that led to varieties that were well adapted to United States conditions.

rolled out quickly, in support of the war effort, with over 2,600 counties having extension agents by 1918 (Smith and Wilson 1930).[14]

3 Research Proximity and Productivity

3.1 Conceptual Framework

We adapt an Aghion and Howitt (1992) growth model to capture the effects of proximity to station research on agricultural productivity. A region has firms (farmers) that produce final output using an intermediate good (seeds) as an input. Entrepreneurs invest in research and development to discover the next vintage of the intermediate good. Station research increases entrepreneurs' research productivity, speeding up the pace of innovation ("discovery"). Station research also increases the knowledge stock of local farmers ("learning"), enabling them to utilize frontier technology, thus increasing their productivity.[15]

Production and Innovation. There is a sequence of time periods $t = 1, 2, ...$ indexing the vintage of the technology. People consume a single composite agricultural commodity with price normalized to one, produced by farmers endowed with a single acre of land.[16] Farmers use intermediate goods to produce agricultural output Y according to $Y = Ax^{\alpha}$, where A is the productivity of the intermediate input used, x is the quantity of the intermediate product used, and $0 < \alpha < 1$. New innovations increase productivity in agricultural production, so that $A_f = \gamma A_o$, where $\gamma > 1$ is the size of the innovation, f is the frontier vintage and o is the previous frontier vintage.[17]

---

[14] A number of states had already begun experimenting with small-scale extension programs, such as offering short courses for farmers off campus (Smith and Wilson 1930, 31) before the federal policy was enacted. Seamen Knapp is generally credited with developing the first large scale co-operative demonstration program at the Porter Farm in Texas in 1903 in efforts to combat the boll weevil.

[15] The insight that R&D has two "faces," learning and discovery, was articulated in Cohen and Levinthal (1989). Evidence on the two faces of R&D in OECD data has been found by Griffith, Redding, and Van Reenen (2004). In a similar vein recent work by Akcigit and Kerr (2010) incorporates both exploration and exploitation innovations into an endogenous growth framework.

[16] That farmers are endowed with a single unit of land makes farm output and productivity identical here.

[17] To ease exposition we drop the vintage subscripts as we consider steady state equilibria. We use subscripts f to refer to frontier technology and o to refer to the previous frontier vintage, as necessary.

Entrepreneurs have the opportunity to attempt an innovation that would create a more productive version of the intermediate good each period.[18] The probability that an innovation will occur in a period, $I(\cdot)$, depends positively on the amount of the final good spent on research by entrepreneurs (P), by stations (S), and on the stock of knowledge of local final goods producers (E_l).[19] Assuming a Cobb-Douglas innovation function and denoting $p = \frac{P}{A^\star}$ and $s = \frac{S}{A^\star}$, with $A^\star_f = \gamma A_o$ being the productivity of the intermediate product if the innovation succeeds, gives

$$I(s) = E_l{}^{\eta} p^{\sigma} s^{\lambda}, \tag{1}$$

where $0 < \eta < 1$, $0 < \sigma < 1$, and $0 < \lambda < 1$. The probability that an innovation occurs increases with the knowledge stock of local farmers and the research spending of entrepreneurs and stations.

Learning. A farmer's knowledge stock depends on cumulative experience with frontier technology. We define farmers with a high stock of knowledge as having at least K units of cumulative experience, while those with a low knowledge stock have less than K units. The technology farmers use in their production process depends on their type. High knowledge stock farmers use technology vintage f, so that $A^H = A_f$. Low knowledge stock farmers use vintage o so that $A^L = A_o$.

Farmers can learn about frontier technologies as a result of proximity to station research. Social learning requires personal interaction with frontier technology that declines with distance to station research.[20] The diffusion of knowledge decays linearly with distance, so that a farmer d miles from a station receives $\frac{s}{d}$ units of experience from station research. A farmer's cumulative knowledge stock is $E(s) = [\sum_0^t (\frac{1}{1-\rho})^t \frac{s}{d}]^{\phi}$, where ρ is the rate of depreciation of the knowledge stock and ϕ is the cumulative learning effect from station research. Innovation depends on the stock of knowledge of local farmers, $E_l(s)$, who we normalize to

---

[18] For ease of exposition we relegate the derivation of the relative input expression that comes from solving the entrepreneur's problem to the Appendix, section A.1 (see equation A2).

[19] We treat station expenditure as exogenously determined. This assumption fits our empirical setting where station research in agriculture was federally funded.

[20] Comin, Dmitriev, and Rossi-Hansberg (2013) study the spatial implications of social learning for technology adoption in a cross-country setting. That social learning is central to agricultural technology adoption has been well established (see Foster and Rosenzweig 2010 for a review).

be $d = 1$ miles from the station, so that: $E_l(s) = [\sum_0^t (\frac{1}{1-\rho})^t s]^\phi$.

Productivity. Solving for the equilibrium distribution of high and low knowledge stock farm types is straightforward. A unique threshold distance determines farm type because experience accumulation declines monotonically with distance and farm types depend on a cumulative experience threshold. We denote the threshold distance $d^\star(s)$.

Farmers located at $d > d^\star(s)$ are low knowledge stock types and those located at $d < d^\star(s)$ are high knowledge stock types. The threshold distance is

$$d^\star(s) = \frac{\sum_0^t (\frac{1}{1-\rho})^t s}{K^{\frac{1}{\phi}}}, \tag{2}$$

with $d^{\star\prime}(s) = \frac{\sum_0^t (\frac{1}{1-\rho})^t}{K^{\frac{1}{\phi}}} > 0$.[21] The effect of station research on the threshold distance increases with the strength of the cumulative learning effect from station research (ϕ). Farmer distances to station research are uniformly distributed between 0 and D, i.e. $d \sim U[0, D]$. The fraction of farms with a high knowledge stock is $\frac{d^\star(s)}{D}$ and a low knowledge stock is $1 - \frac{d^\star(s)}{D}$.

Expected land productivity is

$$Y = I(s)\left[\frac{d^\star(s)}{D} A_f x_f^\alpha + \left[1 - \frac{d^\star(s)}{D}\right] A_o x_o^\alpha\right] + [1 - I(s)] A_o x_o^\alpha, \tag{3}$$

where $I(s)$ is the probability that an innovation occurs in period t.[22] We use equation (3) to obtain the relationship between distance to a station and expected land productivity ΔY_D, the spatial productivity differential. Using the relative input expression (see Appendix

---

[21] Empirical evidence for such a threshold effect is clear. We estimated a regression of 1880 crop revenue per acre on 25 measures of climatic, geographic, and soil quality for each county. The residuals, or the non-geography explained portion of productivity, are plotted in Appendix Figure A1 according to distance from the experiment station. At roughly 119 miles from the station, farmers become less productive than their county geography would suggest.

[22] Realized land productivity depends on whether or not an innovation occurs. Without an innovation $Y = A_o x_o^\alpha$, as both the high and low knowledge stock farmers use the vintage o technology. With an innovation $Y = \frac{d^\star(s)}{D} A x^\alpha + \left[1 - \frac{d^\star(s)}{D}\right] A_o x_o^\alpha$. Here high knowledge stock farmers use the newly developed vintage, but low knowledge stock farmers do not.

equation A2) to substitute for x_f we obtain

$$\Delta Y_D = \frac{\partial Y}{\partial D} = I(s)d^{\star}(s)\Psi < 0, \tag{4}$$

where $\Psi = \frac{A_o x_o^{\alpha}(1-\gamma^{\frac{1}{1-\alpha}})}{D^2} < 0$ since $\gamma > 1$ and $0 < \alpha < 1$. Equation (4) is the spatial productivity differential.

Research Proximity. The comparative static of the spatial productivity differential with respect to station research is

$$\frac{\partial \Delta Y_D}{\partial s} = [I'(s)d^{\star}(s) + I(s)d^{\star\prime}(s)]\Psi. \tag{5}$$

Expression (5) is strictly negative when $I'(s)d^{\star}(s) + I(s)d^{\star\prime}(s) > 0$.

Our model delivers a clear implication that proximity to research enhances productivity. The intuition is straightforward. Station research speeds up the pace of innovation, increasing the productivity of high knowledge stock farmers who are located near the station. This *discovery effect* of station research is represented by the first term in brackets in expression (5). Station research also provides nearby farmers with the technological experience necessary to adopt the current frontier technology. This *learning effect* of station research is represented by the second term in brackets in equation (5).

Short- versus Long-Term. How will research proximity effects change over time? To see we assume that (i) farmer knowledge is fixed in the short-term; and (ii) marginal returns to station research change over the short- to long-term. Both discovery and learning effects, then, will change over time. Discovery effect dynamics are captured by the probability of innovation, equation (1), differing over time: $I(s)^{ST} = \bar{E_l}^{\eta} p^{\sigma} s^{\lambda_{ST}}$; and $I(s)^{LT} = E_l(s)^{\eta} p^{\sigma} s^{\lambda_{LT}}$. Learning effect dynamics manifest in an absence of learning in the short term, since $E_l(s) = \bar{E_l}$ implies $d^{\star\prime}(s) = 0$. We can now express the short- and long-term comparative statics in (5) as

$$\frac{\partial \Delta Y_D^{ST}}{\partial s} = [\underbrace{\lambda_{ST} p^{\sigma} s^{\lambda_{ST}-1} \bar{E_l}^{\eta} d^{\star}(s)}_{Discovery}]\Psi; \tag{6}$$

$$\frac{\partial \Delta Y_D^{LT}}{\partial s} = [\underbrace{\lambda_{LT}\eta\phi p^{\sigma} s^{\lambda_{LT}\eta\sigma-1}[\textstyle\sum_0^t(\frac{1}{1-\rho})^t]^{\phi\eta} d^{\star}(s)}_{Discovery} + \underbrace{I(s)^{LT} d^{\star\prime}(s)}_{Learning}]\Psi. \tag{7}$$

Expressions (6) and (7) reveal that whether proximity effects are larger in the short- or long-term depends on the *discovery* and *learning* effects. Positive discovery effects may decline over time with strong diminishing returns to cumulative research (i.e., $\lambda_{ST} > \lambda_{LT}$), leading proximity effects to weaken over time. Positive discovery effects may grow over time with cumulative knowledge significantly enhancing discovery (i.e., $\eta\phi > 1$), leading proximity effects to strengthen over time. Similarly, the positive learning effect, only present in the long term, leads long-term proximity effects to be stronger than short-term proximity effects.

Our model implies that research proximity effects will be more persistent when discovery effects persist and farmers have much to learn. The microfoundations of the research proximity dynamics we emphasize stem from local producers both learning from and contributing to the discovery process only over the long-term.[23]

3.2 Measurement Framework

To examine how research proximity affected farmers' productivity over the course of 130 years, we use the opening of the agricultural experiment stations as a source of exogenous variation. We estimate the equation

$$Y_{it} = \sum_{t=1870:1890}^{2000} \beta_t Distance_i \times \gamma_t + \theta_i + \gamma_t + \gamma_{Cit} + \epsilon_{it}. \tag{8}$$

Y_{it} is the agricultural outcome in county i in years t=1870,1880,...,2000. γ_t is a set of year fixed effects (1880 serves as the reference year) that flexibly control for national time series trends in agricultural outcomes and θ_i is a set of county fixed effects which absorb time-

---

[23] We make a number of assumptions that depart from the prior literature to focus on the effects of interest. Our single sector model does not allow local innovation to affect regional development through the cross-industry churning effects as modeled by Duranton (2007). In our model entrepreneurs sell embodied technology at profit-maximizing prices, but we do not incorporate the time costs of acquiring tacit knowledge (as in Davis and Dingel 2012). We also abstract from the effects of producer allocation across a continuum of locations, as studied in Desmet and Rossi-Hansberg (2014). Our emphasis on feedback effects of producer learning on discovery do echo the microfoundations of communication externalities in Charlot and Duranton (2004), however.

invariant characteristics across counties. $\gamma_{C_i t}$ is a set of county characteristics in 1880 interacted with year fixed effects that may be included depending on the specification. $Distance_i$ is the linear distance in miles between the experiment station and county i based on the geographic coordinates of each county seat. We normalize all the distance variables to have a mean of zero and a standard deviation of 1. ϵ_{it} is random error.

The event-study specification we use describes the dynamics of the research proximity effects flexibly. Our parameters of interest are $\beta_{1890}, ..., \beta_{2000}$ that measure how the relationship between distance to the station and productivity differed from the reference year in 1880. If stations generate proximity effects that persist indefinitely, then $\beta_t < 0$ for all years $t > 1890$. If, on the other hand, proximity effects disappear then $\beta_t = 0$ for t beyond some critical time period.

Our empirical approach utilizes variation in research that is plausibly exogenous to local agriculture. Because the stations were opened at pre-determined land grant university locations in response to nationwide concerns about agricultural productivity, they create a positive shock to research virtually independent of local economic conditions. Our central identifying assumption is that changes in station research at land grant colleges are unrelated to changes in unobserved determinants of local agricultural development. While it is not possible to test our identification assumption directly, estimates of β_{1870} and examining the non-agricultural sector allow us to test for "effects" where none is expected.

A few other estimation details are worth noting. Our analysis is based on the balanced panel of counties that report agricultural output every year. Each county is weighted by its total land area in 1880 so that estimates capture the effects for a typical acre of land rather than county. In all specifications, to address the possibility of persistent autocorrelation in outcomes within a county, we cluster the standard errors at the county level.

4 Data and Descriptive Statistics

Agricultural, Demographic, and Geographic Data. The primary data we use to estimate the impact of agricultural-related research on local agricultural development from

1870 to 2000 come from the Agricultural Census (Haines 2005).[24] We use county-level data on crop revenue, farm value, farm acreage, improved acres, and equipment value for the full sample period. Our main outcome of interest is total crop revenue per acre of farmland, a measure of the revenue productivity of land.[25] The advantages or disadvantages of a revenue-based productivity measure depend on the determinants of price variation across producers (Syverson 2011). When price variation reflects differences in product quality, a revenue-based measure may be preferred. To the extent that station research led to higher quality crops, this is a strength of our measure. Alternatively, if price variation reflects market power, then changes in market integration could be confused with changes in productivity (Costinot and Donaldson 2014). To address this concern we also consider another crop revenue measure that is the summation of the quantity of each of four crops times their respective national prices. We use corn, wheat, oats and barley to create this alternative measure of crop revenue as national prices for these crops are available in each year of our sample and account for a large share of planted acreage in our sample states.[26] Finally, starting in 1880 we also have data on acreage planted in wheat and corn, allowing us to calculate crop-specific yields, but ruling out the possibility of testing for prior trends before the stations opened in 1887.

We merge into our dataset variables reported at the county level from the decennial Population Census and the Fishback, Horrace and Kantor (2005, 2006) (FHK) county geography database. We create a time-invariant "land suitability" index that measures each county's expected land productivity given its geographic endowment and 1880 production technology. To construct our agricultural suitability measure we take the prediction from a regression of 1880 crop revenue per farm acre on all 12 FHK county-level soil quality measures and all 13

---

[24] The sources and details of construction for all variables are in the Data Appendix.

[25] For the years 1870, 1880, 1890 and 1900 total farm revenue, but not crop revenue, is reported. Our crop revenue measure for these years is interpolated using the reported level of total farm revenue times the average fraction of revenue from crops from the county in the years where both total and crop revenue are reported. The years where both are reported are from 1950 to 2000. The average fraction of revenue from crops in our sample is 37.6%.

[26] Our alternative measure also addresses two related concerns. First, reported crop revenue does not typically include the value of crops fed to livestock as it is based on revenue received in market transactions. Our alternative measure is based on the quantities of crops grown rather than transacted. Second, how crop revenue was enumerated by the census changed over time. In 1920 the census switched from reporting crop revenue directly enumerated from farmer reports to, instead, constructing revenue from the quantities enumerated from farmers times regional prices from another source. We thank Paul Rhode for making us aware of this change in enumeration.

FHK county-level climate, water access, elevation, and geographic coordinate measures.

Locational Data. The locations of the federal experiments stations are obtained from the United States Office of Experiment Stations (1910, 300). While nearly all of the states in our sample have a single federal experiment station located at their land grant college, there are a couple of exceptions. Connecticut, Missouri, and New York had two stations in different locations in operation in 1910.[27] For these states we denote one station as primary and the other as secondary based on the number of publications produced by the respective stations.[28] We use these locations to compute each county's distance from the primary and secondary (if applicable) experiment stations in its state.[29] We specify the distance to the experiment station in our estimation equation (8) as the distance to the primary station and add year times distance to a secondary station as additional controls as a sensitivity analysis. We also take into consideration the locations of USDA research stations, railway lines, and large urban markets and calculate distances to these places for each county.

Experiment Station Data. We use experiment station reports to create a measure of when applied innovations occurred at our sample stations. Our approach builds on the development literature that considers productivity gaps between farmers and scientists as reflecting differences in technology (i.e., Udry 2010, Figure 2.13). When yields are much higher at the experiment station relative to farmers facing the same agricultural conditions, then this event would reflect a discovery. To measure the presence of an applied innovation at a station, we compare the experimental crop yields for corn and wheat reported by station scientists to those reported by farmers in the same county in the same year.[30] Using these

---

[27] New Jersey also had two stations. However, as they were both in the land grant college county the existence of multiple stations does not affect how we measure station location.

[28] The primary and secondary stations for these states are as follows: Connecticut – Storrs (primary) and New Haven (secondary); Missouri – Columbia (primary) and Mountain Grove (secondary); New York – Geneva (primary) and Ithaca (secondary). Most states that had multiple stations were in the South (i.e., Louisiana) or far west (i.e., California and Montana) and are not in our sample.

[29] The location of the primary and secondary stations in our sample states is given in appendix Table A1. We are aware of only one change in station location during the time period under consideration. The Ohio experiment station was established at the Ohio State University campus in Columbus in 1870, but was moved to Wooster in 1892, and then back to Columbus in 1948. In our estimations we used the location of the Ohio station in 1910 (Wooster, OH). Later, we examine the robustness of using distance to the closest experiment station, regardless of which state it happened to be located.

[30] We report the full set of experiment station results we use in Appendix Table A2. These data are for all corn and wheat experiments reported in the annual *Experiment Station Record* for our sample states. We

gaps we are able to study if the timing of discoveries affected the timing of localized productivity growth. One limitation of our approach is that we can only measure innovations in years for which we have county-level agricultural census data. The full distribution of farmer-experiment station gaps is displayed visually in Appendix Figure A2. We code an "innovation" as occurring when the yield gap between scientists and farmers was in the top 25% of this distribution.

Sample Selection and Descriptive Statistics. To create our balanced panel of counties we impose the following sample restrictions. We first keep states east of the Continental Divide so that we are working with counties that were relatively settled in 1870 and did not experience large boundary changes over the long time period we consider. We drop southern states (as defined in Olmstead and Rhode (2002) who use the regional definitions from Parker and Klein (1966)) because attributing Hatch Act funding to specific locations is harder to document where Historically Black Colleges received some funding. Moreover, the locations of the colleges were established in 1890 and may have been endogenous. This sample of 20 states in the Northeast and Midwest has 1,277 counties based on their boundaries in 1880 (Horan and Hargis 1995).[31] We drop the 54 counties with large boundary changes during our sample period based on Horan and Hargis (1995). Finally, we drop 160 counties that did not report crop revenue in every year of our sample. Our final sample contains a balanced panel of 1,063 counties based on consistent county boundaries defined as of 1880.

Table 1 provides a first look at summary statistics of relevant measures. We report the 1880 summary statistics for the full sample and then stratify the counties based on whether their geographic distance to their state experiment station was below- or above-median. As shown in Panel A, agricultural outcomes in below-median distance counties were quite different from their counterparts in counties that were farther away. The comparisons indicate many differences existed early on with the closer counties having higher levels of

---

drop those experiments where we are unable to convert the units to bushels per acre (as reported in the agricultural census) or that have missing yields. We code those results reported in ranges of yields based on their mid point. As the Office of Experiment Stations ceased publication of this compilation in 1946 we are unable to obtain a census of crop trials to match to agricultural census data after the Agricultural Census published in 1940.

[31] The 20 states contained in our sample are: CT, IA, IL, IN, KS, MA, ME, MI, MN, MO, NE, NH, NJ, NY, OH, PA, RI, TX, VT, and WI.

crop revenue and land value per acre. These differences can also be seen visually in the map of crop revenue per acre in 1880 displayed in Figure 2. Similarly, farms farther from the station had lower land quality, input intensities, fraction of county acreage in farming, and smaller rural populations. We also see that counties close to the experiment station had higher levels of literacy and were closer ethnically to the experiment station county population, potentially reducing the cost of knowledge transfer.

5 Results

Main Results. We present the main results in Table 2. Column (1) of the table reports the year-by-year distance interaction estimates under the baseline model. To capture underlying trends the model in column (1) includes year fixed effects interacted with three measures of initial conditions in 1880: woodland above 4% in 1880 to capture the diffusion of barbed wire at this time (Hornbeck 2010); and the fraction of land in farms in 1880 and total population in 1880 to capture secular trends in settlement patterns. In column (2) we exclude these year × initial county characteristic fixed effects. In column (3) we estimate the baseline model for the 10 states that had no permanent state operated experiment stations before the passage of the Hatch Act in 1887. With the opening of experiment stations in these states dictated by federal policy alone, this sample may be especially likely to satisfy our exclusion restriction.

The results in Table 2 indicate that proximity effects manifested quite quickly, first emerging in some specifications in 1890. Proximity effects grew until 1910 – more than 20 years after the experiment stations first opened. They then declined and became statistically insignificant by 1920 or 1950 depending on the specification.[32] Figure 3 visually depicts the year-by-year interactions for all three models along with the 95th percentile confidence

---

[32] In Appendix Table A3 we report estimates using different inference procedures: (1) clustered standard errors at the state level; (2) clustered at the year × distance to experiment station category level; and (3) allowing for cross-sectional spatial dependence of the standard errors. The conclusions about statistical significance are very similar across these approaches. The one exception is that the standard errors are smaller when we cluster on state (Table A3, Column (1)) than on county (Table 2, Column (1)). As there are relatively few states in our sample (20), we prefer to use the more conservative county clustered standard errors to draw inferences.

interval for the first model's estimates.

Table 2 also demonstrates that none of the coefficients of the 1870-distance interactions is statistically significant, all have small magnitudes and their signs are specification dependent. These first results indicate that the opening of federal experiment stations did not follow changes in local agricultural production – an important falsification test for our specification.

The Economic Significance of the Effect. To put the size of the estimated proximity effect into some perspective, we first consider the counterfactual experiment of hypothetically reallocating research stations across space to maximize their average proximity to farmers. We find that had stations been located to maximize proximity, instead of simply placed at land grant colleges, national crop productivity would have been 7.7% higher in 1910.[33] Another way to interpret the magnitude of the effect is to consider that if a county were one standard deviation closer to a research station, its land productivity would increase by 36 percent. In other words, this one standard deviation increase in research proximity would account for about 10 percent of the productivity dispersion between the 90th and 10th percentiles of land productivity.[34] While each calculation is subject to many caveats (i.e., spatial reallocation might have many effects we do not capture), they both indicate that the research proximity effects we estimate are economically significant.

Robustness. We present a range of estimates to examine the robustness of our main results using a more parsimonious specification in Table 3. Specifically, we collapse the 1940 to 2000 year dummies into a single dummy variable which allows us to study the dynamic effects of the stations' opening during the 50 years following the Hatch Act in a flexible way, while summarizing the magnitudes and joint statistical significance of any persistence

---

[33] To implement this experiment we calculated how far the average county would be from an experiment station if their locations were chosen to maximize proximity. To obtain this distance we used a bootstrap procedure that chose 20 counties at random (as we have 20 stations) and find the minimum average distance out of 1,000 possible 20-county draws. Doing so yields an average proximity of 87.84 miles. Our 1910 estimate in Table 2 column (1) and the distance to true experiment station of 99.97 miles with a standard deviation of 56.58 can be used to calculate the national productivity effects of reallocating research. The calculation is: $0.36 \times \frac{(99.97-87.84)}{56.58} = 0.077$.

[34] In our sample the 90th - 10th percentile ratios of land productivity in 1880 is 4.5:1. This productivity dispersion is closer to modern developing economies than present day U.S. firms. Syverson (2011) notes that within-industry 90th - 10th percentile ratios of TFP dispersion are close to 2 in the U.S. Hsieh and Klenow (2009) find within-industry TFP 90-10 ratios of 5:1 in China and India.

beyond that period in a single estimate. In Table 3, column (1), we report estimates from our baseline parsimonious model. We then present models where we further include state × year dummy interactions (column (2)), 1880 cattle per acre × year fixed effects (column (3)), and percent voting Democrat in 1880 × year fixed effects (column (4)).[35] Our results are robust to these specification changes. Similarly, the estimates in column (5) indicate that measuring distance to the nearest station, regardless of whether it was in the same state or not, does little to alter the main results.

Agricultural markets may not have been fully integrated at this time, so experiment station opening could have affected local commodity prices in our revenue-based land productivity measure.[36] To address this issue we construct an alternative measure of crop productivity based on county-level quantities of the four crops reported consistently throughout our sample (corn, wheat, oats and barley) times their national prices, which were unaffected by local shocks. When we estimate our model using this measure of revenue productivity (in column (6)) we obtain similar but slightly larger estimates (in absolute terms), perhaps indicating that positive productivity effects pushed down local prices in markets that were not fully integrated. Similarly, the results in columns (7), (8) and (9) of Table 3 show that adding controls for year fixed effects × distance to minor stations, railways, or urban markets does little to change the results.

In the last two columns of Table 3 we examine the effects of research proximity on productivity by crop. There are two challenges here. First, while information on output of corn and wheat is reported in 1870, the acreage planted for each crop is not. Thus, we cannot calculate the crop specific yield in 1870, which prohibits the examination of prior trends. A second challenge is that agricultural crops are geographically specialized in certain areas based on the suitability of the soil and climatic conditions. We thus examine the responses of each crop within its region of production using Parker and Klein's (1966) definition of

---

[35] One important issue in the 1880 Presidential election was railroad regulation, often pushed by agricultural interests at the time to lower transportation prices. We thus use vote share in 1880 to measure voters' agriculture-related policy preferences before the Hatch Act passed to address potential policy endogeneity.

[36] Experiment stations were small relative to the state or even local economy. For example, in our sample states the average experiment stations employed 24 workers in 1910 (Office of Experiment Stations 1910, Table 2). This compares to an average rural population in the experiment station counties of 21,791 people in 1910 so little direct fiscal impact on local farmers would be expected.

"corn states" for our corn productivity analysis and "grain states" for the wheat analysis.[37] The effects of research proximity on crop productivity in columns (10) and (11) are relatively modest, likely reflecting farmer responses to research proximity on crop quality or land use margins. That the effects are stronger and earlier for corn, while appearing later for wheat, is consistent with historians' accounts.[38]

Falsification. We conduct two additional analyses to evaluate the possible concern that producers near a station may have been poised to experience productivity gains even if experiment stations had not opened. We first examine whether each county's distance to other research stations, opened much later by the USDA, display similar dynamics around the experiment stations' opening date in 1887.[39] The results for this analysis are reported in the first three columns of Table 4 where the distance included in the model is now the distance to the nearest USDA station open in 1929. We find little evidence of similar effects, where little would be expected.

We next examine whether manufacturing productivity responded to experiment station opening, as agriculturally-specific research would have had little direct effect.[40] We report

---

[37] We do this so that we are focusing on counties where the crop is likely to be planted each year and the experiment stations are likely to specialize in research related to that crop for time-invariant reasons. We follow Olmstead and Rhode (2008) in using the Parker and Klein (1966) crop region classifications. The corn states in our sample are: IA, IL, IN, MO, and OH. The grain states in our sample are: KS, NE and TX.

[38] May (1980) summarizes agricultural experiment station contributions to wheat productivity as: "James Malin's *Winter Wheat in the Golden Belt of Kansas* discusses the slow, dismal picture of experiment stations' efforts. Malin believed that as late as World War II agricultural scientists engaged in a selective breeding process that demonstrated that their only real creative work was the hybrid variety Tanmarq." (May 1980, 183). In contrast, experiment stations' contributions to corn "is an example of the use of experiment station research by commercial companies to achieve practical results unobtainable by either the experiment stations or the scientists themselves . . . The eventual results of corn hybridizing were clearly beneficial to American agriculture." (May 1980, 188). Similarly, a recent meta analysis finds mean rates of return for corn research are more than double those for wheat research (Alston et. al. 2000, Table 15).

[39] We measure USDA station location as of 1929, the earliest year we have been able to find a comprehensive source. While the dates that the facilities opened is not recorded, the legislative history of USDA research suggests most facilities would have opened after the Bureau of Plant Industry was established in 1901. Most openings probably occurred 10 to 20 years after this as the USDA began to play a larger role in inspection and conservation. As the opening dates of these facilities are typically not reported we are not able to date them with more precision.

[40] Of course, linkages between the agriculture and manufacturing sectors could lead to indirect effects. Because both sectors have tradable output, indirect effects may not be spatially concentrated. While Davis, Hanes and Rhode (2009) demonstrate that large national or multi-state agricultural shocks did affect industrial production at this time, market integration would lead sub-state changes in agricultural productivity to have little sub-state implications.

the result of fitting equation (8) with manufacturing output per worker as the outcome variable. As reported in columns (4)-(6) of Table 4, we find statistically significant negative point estimates early in our sample period. However, the manufacturing results indicate very stable proximity effects over time and, hence, show little effect from the stations' actual opening. The estimates in Table 4 are displayed visually in Appendix Figure A3.

Total Factor Productivity. How important were producers' input adjustments in explaining the effects of research proximity on land productivity? While our use of county-level production data over a very long time renders the analysis necessarily more crude than a study of individual firm productivity with modern data, we follow Greenstone, Hornbeck, and Moretti (2010) and examine the total factor productivity (TFP) effects.

We estimate the effects of proximity to research on total factor productivity in Table 5. To do so we estimate a modified version of equation (8) that does not include variables in per acre terms, but rather county totals, and controls for improved acres, rural population, farm equipment, fertilizer expenditure, and farmland acres as inputs. The results in column (1) differ from the land productivity results above in two key ways. First, research proximity TFP effects took more time to manifest than the land productivity effects, and display less persistence. A second feature of the TFP effects is that the magnitude of the 1910 point estimate is less than half of the overall land productivity results above. These results imply that the counterfactual experiment of reallocating experiment station research across space to maximize farmer proximity to research would have increased national total factor productivity in agriculture in 1910 by 3.0%.[41]

We address the endogeneity of input choices by using the adjustment of investment (Olley and Pakes 1996) and material inputs (Levinsohn and Petrin 2003) as controls. To do so we add fourth degree polynomials of capital (log farm equipment and log improved acres) and investment (changes in log farm equipment and changes in log improved acres) and their interactions in column (2) and add fourth degree polynomials of log equipment, log cropland, and log fertilizer, as well as their interactions to the model in column (3). In

---

[41] Using the baseline estimate in column (1) of Table 5 in the calculation gives: $0.14 \times \frac{(99.97-87.84)}{56.58} = 0.030$. See footnote 33 for further details of this counterfactual analysis.

columns (4) and (5) we add controls for log farm size and log education to control for changes in farmer composition. The results are very similar to the baseline estimates, indicating that producers' adjustments were highly responsive to research proximity.

Geography and Farmer Knowledge. Producers in counties close and far from experiment stations differed along a number of dimensions that can be related to regional productivity growth. Might these other differences account for the research proximity effect? In Table 6 we present the results of an analysis that interacts year fixed effects with both geographic proximity and other types of proximity from the experiment station for each county. We calculate each county's distance from the experiment station county based on initial productivity, literacy, ethnicity, and land quality. We find that adding year fixed effects × other non-geographic proximity to the model does little to alter our main results, indicating a central role for geographic proximity.

Does farmer knowledge stock amplify proximity effects? In columns (1) and (2) of Table 7 we stratify counties based on whether their similar agro-climatic conditions would lead them to grow similar crops to the station county.[42] If research drew on farmers' knowledge, then we might expect to find proximity effects only in counties with similar crop growing conditions. The results in columns (1) and (2) both reveal proximity effects, suggesting that local farmers' crop-specific knowledge was not central to advancing research.[43]

In columns (3) and (4) of Table 7 we stratify counties based on whether their agricultural output per farm acre was more or less in 1880 than we would expect given their climatic, geographic, and soil qualities. We find that in areas with high concentrations of frontier producers (column (3)), strong proximity effects persist to the present day. In contrast, in areas with low concentrations of frontier producers (column (4)), persistent proximity effects fail to emerge. These results indicate a learning process in which knowledge is most benefi-

---

[42] We calculate the expected productivity in each county in 1880 based on its geographic and climatic endowment for each crop. We then compute crop revenue based on 1880 national prices for each crop. We treat a county as "similar" to the experiment station county if its revenue maximizing crop was the same for both counties. See the Data Appendix for more detail.

[43] Marcus (1988) argues that the research directions of the experiment stations were the outcome of a political process with many vested interests across the state having a stake in the station's efforts. If scientists devoted significant resources to investigate how to expand productivity on marginal land different from the station counties, then the results in Table 7 would not be surprising.

cially absorbed by producers closest to the productivity frontier. Further, they suggest that highly productive producers nearby mutually benefit scientists' research efforts to advance the technological frontier, which then feeds back to further benefit farmers.

Station Innovation and Research Focus. To measure experiment station innovations we compute yield gaps between the experimental yields achieved in trials at the station relative to farmers' yields in the same county and same year. We define an innovation as occurring when the yield gap is in the top 25% of the distribution.[44] Our first finding is that farmer-station yield gaps decline over time, suggesting diminishing returns to applied research. Figure 4 shows that yield gaps peaked in 1890, the first year of data, and declined until 1940, when the data are no longer available.

We estimate how proximity to these innovations affected agricultural productivity in Table 8.[45] In columns (1) and (2) of the table we see that proximity to a station innovation increased land and total factor productivity when the yield gap occurs – but not before – and persisted for about 40 years.[46] In column (3) we find similar results for corn productivity, though the effects occur with a longer lag. We find little effect for wheat productivity, as pre-innovation proximity estimates are similar to post-innovation proximity estimates. These estimates are depicted visually in Appendix Figure A4.

---

[44] We define large yield gaps as being in the top 25% of the yield gap distribution for corn and wheat trials reported in 1889, 1899, 1909, 1919, 1929, and 1939. Experiment station data are reported in Appendix Table A2. The yield gap distribution and top 25% cutoff is displayed visually in Appendix Figure A2.

[45] To do so we estimate the specification,

$$Y_{it} = \sum_{j=d-20}^{d+50} \pi_j YG_j \times Distance_i + \sum_{j=d-20}^{d+50} \delta_j YG_j + \theta_i + \gamma_{Cit} + \epsilon_{it}, \tag{9}$$

where YG_{d-20}, YG_{d-10} are dummy variables that take a value one and zero otherwise in the 20 and 10 years before a large yield gap exists, $YG_d, ..., YG_{d+50}$ are dummy variables that take a value of one in the large yield gap year,d, and each decade there after, and a value of zero otherwise.

[46] The negative proximity estimate for contemporaneous yield gaps in columns (1) and (2) of Table 8 could reflect a limitation of our approach. We cannot precisely measure stations innovations that occur in non-census years between $d - 10$ and d. We can only compute yield gaps in years when the agricultural census is available. Because farmer crop yields are highly sensitive to soil and climatic conditions higher frequency data at the state level may not capture highly localized soil conditions near the stations. Similarly, interpolation between county level census observations over time will not capture year to year fluctuations in weather conditions. In Appendix Table A4 we conduct a falsification analysis of station yield gaps with: (i) USDA station distance and (ii) manufacturing productivity. That we find little evidence of similar patterns reinforces the validity of the station-innovation approach.

Our results in Table 8 imply that the full diffusion of typical experiment station innovations takes substantial time. Blockbuster crop varieties appear to have diffused more quickly. New rust resistant wheat varieties introduced in 1900 diffused to cover 87.4 percent of wheat production by 1921 (Olmstead and Rhode 2008, 30). Similarly, hybrid corn diffused in just 4 to 10 years according to Manuelli and Seshardi's (2013) calculations using Grilliches (1957) estimates. Our estimates are quite close to those in Comin and Hobijn (2010) who estimate the typical diffusion lag of a technology is 45 years.[47]

Our station-innovation proximity estimates can also be compared to the spatial technology diffusion literature. Keller (2002) finds that technology diffusion falls by 50% for every 1,200 Kilometres (745 Miles) in his cross-country analysis. Our contemporaneous estimate in column (1) implies that agricultural technology diffusion falls by 50% for every 205 Miles in the United States.[48] While more rapid technology diffusion within countries has been documented (Branstetter 2001), that technology diffusion within a country is at least three times faster than across countries provides a new comparative perspective.

The results in Table 8 shed light on the limited persistence of the research proximity estimates in Table 2. Using the applied innovation frequency time-series in concert with the estimates in Table 8 implies that decreasing innovation frequencies can account for almost 40 percent of the decline in proximity effects from 1910 to 1940 in Table 2 column (1).[49] Decreasing returns to applied research are economically significant and appear to be key to

---

[47] Our results line up closely with the six technologies they study that were invented between 1880 and 1940. These technologies fully diffused between 30 to 73 years. The six technologies are: (i) *Cars* developed in 1885 and fully diffused in 73 years; (ii) *Trucks* developed in 1885 fully diffused in 58 years; (iii) *Aviation passengers* developed in 1903 and fully diffused in 50 years; (iv) *Aviation freight* developed in 1903 fully diffused in 30 years; (v) *Telegraph* developed in 1835 fully diffused in 44 years; and (vi) *Telephone* developed in 1876 and fully diffused in 64 years (Comin and Hobijn 2010, Table 2).

[48] To make this comparison we first note that the modal size of the technology shock in the top 25% of our yield gap distribution is a 116% increase in land productivity. The distance that experiences a 58% increase in productivity receives 50% of the technology. Our contemporaneous yield gap estimate in Table 8 column (1) indicates that a one standard deviation increase in distance (56.58 miles) reduces land productivity by 16%. Thus, a distance of $\frac{0.58}{0.16} \times 56.58 = 205$ miles is required for 50% of the technology to diffuse.

[49] The fraction of counties with a station applied innovation in the top 25% in each year is: 0.28 in 1890, 0.33 in 1900, 0.09 in 1910, 0.19 in 1920, 0 in 1930 and 0 in 1940. Multiplying these fractions by the year d to $d + 40$ proximity coefficient estimates in Table 8 column (1) we obtain proximity effects in each year as: -0.05 in 1890, -0.12 in 1900, -0.14 in 1910, -0.13 in 1920, -0.13 in 1930, -0.08 in 1940, and thereafter is it not defined due to data limitation. These estimates indicate decreasing frequency of applied innovations accounts for almost 40 percent of the 1910 to 1940 decline in proximity effects in Table 2 column (1).

the limited persistence we find.

Is proximity to basic research, where diminishing returns may be less strong (Furman and Stern 2011; Waldinger 2013), more valuable than applied research? To see, we stratify stations (and their corresponding states) based on their respective basic versus applied research orientation. We measure research orientation by the fraction of station research output published in academic journal articles relative to farmers bulletins from 1900 to 1940. In states where stations focused on basic research (column (1) of Table 9), proximity effects emerge slowly, but persist to the present day. In contrast, in states where stations focused on applied research (Table 9 column (2)), proximity effects emerged relatively soon after station opening, but ultimately declined and did not persist indefinitely.[50] The persistence of the research proximity effect depends on the research focus, with a discovery effect that persists only for basic research. Our back of the envelope calculation suggests that national productivity would be 4.1% higher today if stations focusing primarily on basic research were allocated to maximize proximity.[51]

6 Conclusion

We show that proximity to where new ideas are produced can enhance final goods producers' productivity. The concentration of research activity in highly productive areas today constrains convergence in living standards, as less productive areas do not receive the know how to utilize frontier technology. While the idea that research infrastructure is central to creating a highly productive "knowledge hub" has been widely touted, our work indicates it is not so simple. For permanent research infrastructure, such as experiment stations, to be a source of persistent comparative advantage, the research process must produce the new ideas needed for frontier producers to become more productive, and facilitate the mutual learning of scientists from frontier producers. When these effects are present the spatial distribution of research can have implications for nationwide productivity growth.

---

[50] In Appendix Table A5 we conduct a similar analysis with the manufacturing productivity outcome and find none of the specifications reveal a clear change in the proximity relationship when the stations opened.

[51] Using the baseline estimate in column (1) of Table 9 in the calculation described in footnote 33 gives: $0.19 \times \frac{(99.97-87.84)}{56.58} = 0.041$.

Two limitations of our study suggest exciting directions for future work. Our back of the envelope calculation on how spatial reallocation of research affects national productivity is a very tentative first step. Future work could broaden the range of effects considered for a more complete accounting of how spatial research allocation affects social welfare. That we study a single sector opens up a range of questions about how research proximity effects may differ in other sectors, or generate sectoral spatial reallocation. To highlight just a couple: the effects found in widely dispersed agricultural production may be quite different from those in spatially concentrated sectors, where perhaps even stronger learning effects may be present; or industries that co-locate with knowledge producers may experience quite different outcomes. Finally, long-term persistence in research proximity effects may depend on whether the spatial dynamics of industrial mobility amplify the benefits of proximity to the knowledge frontier.

7 References

Aghion, Philippe, Leah Boustan, Caroline Hoxby, and J. Vandenbussche (2009) "The Causal Impact of Education on Economic Growth: Evidence from U.S.," *Harvard University*, Working Paper.

Aghion, Philippe and Paul Howitt (1992) "A Model of Growth Through Creative Destruction," *Econometrica*, 60(2): 323-351.

Akcigit, Ufuk and William R Kerr (2010) "Growth through Heterogeneous Innovations," *National Bureau of Economic Research*, Working Paper no. 16443.

Alston, Julian M., M.A. Andersen, J.S. James and P.G. Pardey (2010) *Persistence Pays: U.S. Agricultural Productivity Growth and the Benefits from Public R&D Spending*, Springer: New York, NY.

Alston, Julian M., C. Chan-Kang, M.C. Marra, P.G. Pardey, and T.J. Wyatt (2000), "A Meta Analysis of Rates of Return to Agricultural R&D: Ex Pede Herculem?" Washington D.C., IFPRI Research Report No. 113.

Alston, Julian M., and Philip G. Pardey (2014) “Agriculture in the Global Economy,” *Journal of Economic Perspectives*, 28(1): 121-146.

Bleakley, Hoyt and Jeffrey Lin (2012) “Portage and Path Dependence,” *Quarterly Journal of Economics*, 127(2): 587-644.

Branstetter, Lee (2001) “Are Knowledge Spillovers International or Intranational in Scope? Microeconometric Evidence from the U.S. and Japan,” *Journal of International Economics*, 53: 53-79.

Carstensen, Vernon (1960) “The Genesis of an Agricultural Experiment Station, *Agricultural History*, 34(Jan): 13-20.

Charlot, Sylvie and Gilles Duranton (2004) “Communication Externalities in Cities,” *Journal of Urban Economics*, 56(3): 581-613.

Christian Advocate (1901) “Experiments with Corn”, June 27, 1901: 76(26): 1031.

Cohen, Wesley M. and Daniel A. Levinthal (1989) “Innovation and Learning: The Two Faces of R & D,” *Economic Journal*, 99(397): 569-596.

Collins, William J. and Margo, Robert A. (2007) “The Economic Aftermath of the 1960s Riots in American Cities: Evidence from Property Values,” *Journal of Economic History*, 67(4): 849-883.

Comin, Diego, Mikhail Dmitriev and Esteban Rossi-Hansberg (2013) “The Spatial Diffusion of Technology,” *Princeton University,* Working Paper.

Comin, Diego and Bart Hobijn (2010) “An Exploration of Technology Diffusion,” *American Economic Review*, 100(5): 2031-2059.

Costinot, Arnaud and David Donaldson (2014)“How Large are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-1997.”*Massachusetts Institute of Technology*, Working Paper.

Davis, Donald R. and Jonathan I. Dingel (2012) “A Spatial Knowledge Economy,” *National Bureau of Economic Research*, Working Paper no. 18188.

Davis, Donald R. and David E. Weinstein (2002) “Bones, Bombs, and Break Points: The Geography of Economic Activity,” *American Economic Review*, 92(5): 1269-1289.

Davis, Joseph H., Christopher Hanes and Paul W. Rhode (2009) “Harvests and Business Cycles in Nineteenth-Century America,” *Quarterly Journal of Economics*, 124(4): 1675-1727.

Desmet, Klaus and Rossi-Hansberg, Esteban (2014) “Spatial Development,” *American Economic Review*, forthcoming.

Duranton, Gilles (2007) “Urban Evolutions: The Fast, the Slow, and the Still,” *American Economic Review*, 97(1): 197-221.

Duranton, Gilles and Diego Puga (2001) “Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products,” *American Economic Review*, 91(5): 1454-1477.

Ferleger, Louis (1990) “Uplifting American Agriculture: Experiment Station Scientists and the OES in the Early Years After the Hatch Act,” *Agricultural History*, 64(2): 5-23.

Fishback, Price V., Horrace, William C. and Shawn Kantor (2005) “Did New Deal Grant Programs Stimulate Local Economies? A Study of Federal Grants and Retail Sales During the Great Depression,” *Journal of Economic History*, 65(1): 36-71.

Fishback, Price V., Horrace, William C. and Shawn Kantor (2006) “The Impact of New Deal Expenditures on Mobility During the Great Depression,” *Explorations in Economic History*, 43(2): 179-222.

Fitzgerald, Deborah (1990) *The Business of Breeding: Hybrid Corn in Illinois, 1890-1940*, Cornell University Press: Ithaca, NY.

Foster, Andrew D. and Mark Rosenzweig (2010) “Microeconomics of Technology Adoption,” *Annual Reviews of Economics*, 2(2): 395-424.

Frederico, Giovanni (2009) *Feeding the World: An Economic History of Agriculture, 1800-2000*, Princeton University Press: Princeton, NJ.

Furman, Jeffrey L. and Scott Stern (2011) “Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research”, *American Economic Review*, 101(5): 1933-1963.

Glasear, Edward L. (1999) “Learning in Cities,” *Journal of Urban Economics*, 46(2), pp. 254-277.

Greenstone, Michael, Richard Hornbeck and Enrico Moretti (2010) “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” *Journal of Political Economy*, 118(3): 536-598.

Griffith, Rachel, Rupert Harrison and John Van Reenen (2006) “How Special is the Special Relationship? Using the Impact of U.S. R&D Spillovers on U.K. Firms as a Test of Technology Sourcing,” *American Economic Review*, 96(5): 1859-1875.

Griffith, Rachel, Stephen Redding and John Van Reenen (2004) “Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries,” *Review of Economics and Statistics*, 86(4): 883-895.

Griliches, Zvi (1957) “Hybrid Corn: An Exploration in the Economics of Technological Change,” *Econometrica*, 25(4): 501-522.

Haines, Michael R. (2005) “Historical, Demographic, Economic, and Social Data: The United States, 1790-2002,” ICPSR02896-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21.

Hanlon, W. Walker (2013) “Industry Connections and the Geographic Location of Economic Activity,” *Working Paper*, UCLA.

Harding, T. Swann (1947) *Two Blades of Grass: A History of Scientific Development in the U.S. Department of Agriculture*, University of Oklahoma Press: Norman, OK.

Hausman, Naomi (2012) “University Innovation, Local Economic Growth, and Entrepreneurship,” US Census Bureau Center for Economic Studies Paper No. CES-WP- 12-10.

Henderson, J. Vernon (1974) “The Sizes and Types of Cities,” *American Economic Review*, 64(4): 640-56.

Horan, Patrick M., and Peggy G. Hargis (1995) “County Longitudinal Template, 1840-1990,” ICPSR06576-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

Hornbeck, Richard (2010) “Barbed Wire: Property Rights and Agricultural Development,” *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.

Hsieh, Chang-Tai and Peter J. Klenow (2009) “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, 124(4): 1403-1448.

Huffman, W.E. and R.E. Evenson (2006) *Science for Agriculture: A Long Term Perspective*, Blackwell Publishing: Ames, IA.

Johnson, D. Gale and Robert L. Gustafson (1962) *Grain Yields and American Food Supply: An Analysis of Yield Changes and Possibilities*, University of Chicago Press: Chicago, IL.

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.

Keller, Wolfgang (2002) “Geographic Localization of International Technology Diffusion .” *American Economic Review*, 92(1): 120-142.

Keller, Wolfgang and Stephen Yeaple (2013) “Gravity in the Knowledge Economy,” *American Economic Review*, 103(4): 1414-1444.

Kingsbury, Noel (2009) *Hybrid: The History and Science of Plant Breeding*, University of Chicago Press: Chicago, IL.

Kline, Patrick and Enrico Moretti (2014) "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," *Quarterly Journal of Economics*, 129(1): 275-331.

Knoblach, H.C, E.M. Law, and W.P. Meyer (1962) *State Experiment Stations: A History of Research Policy and Procedure*, United States Department of Agriculture, Miscellaneous Publication 904; Washington, DC.

Krugman, Paul (1991) "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99(3): 483-499.

Levinsohn, James and Amil Petrin (2003) "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(April): 317-342.

Liu, Shimeng (2013) "Spillovers from Universities: Evidence from the Land-Grant Program," *Working Paper*, Syracuse University.

Lucas, Robert E. (2001) "Externalities in Cities," *Review of Economics Dynamics,* 4(2): 245-274.

Manuelli, Rodolfo E. and Seshadri, Ananth (2013) "Frictionless Technology Diffusion: The Case of Tractors," *American Economic Review*, forthcoming.

Marcus, Alan I. (1988) "The Wisdom of the Body Politic: The Changing Nature of Publicly Sponsored American Agricultural Research since the 1830s," *Agricultural History*, 62(2), pp. 4-26.

Marshall, A. (1890) *Principles of Economics*, MacMillan and Co.; New York, NY.

May, Irwin M. (1980) "Research in Land-Grant Universities: The Agricultural Experiment Station." in Peterson, Turdy Huskamp (1980) *Farmers, Bureaucrats, and Middlemen: Historical Perspectives on American Agriculture,* Howard University Press; Washington, DC.

Moretti, Enrico (2012) *The New Geography of Jobs.* Mariner Books; New York, NY.

Moser, Petra, Alessandra Voena, Fabian Waldinger (2014) “German-Jewish Emigres and U.S. Invention,” *American Economic Review*, forthcoming.

National Science Board (2010) *Science and Engineering Indicators 2010*, Arlington, VA: National Science Foundation (NSB 10-01).

Olley, Steven G., and Ariel Pakes (1996) “The Dynamics of Productivity in Telecommunications Equipment Industry,” *Econometrica*, 64 (November): 1263-98.

Olmstead, Alan L. and Paul W. Rhode (2002) “The Red Queen and the Hard Reds: Productivity Growth in American Wheat, 1800 1940,” *Journal of Economic History*, 62(4): 929-966.

Olmstead, Alan L. and Paul W. Rhode (2008) *Creating Abundance: Biological Innovation and American Agricultural Development*, Cambridge University Press: New York, NY.

Parker, William N. and Judith L.V. Klein (1966) “Productivity Growth In Grain Production in the United States, 1840-60 and 1900-10,” in *Output, Employment and Productivity in the United States After 1800*, 20: 532-80. Columbia University Press; New York, NY.

Rosenberg, Charles E. (1964) “The Adams Act: Politics and the Cause of Scientific Research,”*Agricultural History*, 38 (Jan): 3-12.

Scott, Roy V. (1970) *The Reluctant Farmer: The Rise of Agricultural Extension to 1914,* University of Illinois Press: Urbana, IL.

Smith, Clarence Beaman, and Meredith Chester Wilson (1930) *The Agricultural Extension System of the United States*, John Wily and Sons: New York, NY.

Syverson, Chad (2011) “What Determines Productivity?,” *Journal of Economic Literature*, 49(2): 326-65.

Thornton, Rebecca Achee and Peter Thompson (2001) “Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding,” *American Economic Review*, 91(5): 1350-1368.

True, Alfred C. (1888) “Origin and Development of the Agricultural Experiment Stations in the United States,”*Annual Report*, U.S. Department of Agriculture, Washington, DC.

Udry, Christopher (2010) “The Economics of Agriculture in Africa: Notes Toward a Research Program,” *Working Paper*, Yale University.

United States Office of Experiment Stations (1910) *Annual Report of the Office of Experiment Stations, June 30, 1910. Washington*, U.S. Department of Agriculture. UNT Digital Library. http://digital.library.unt.edu/ark:/67531/metadc5000/. Accessed October 17, 2012.

United States Patent and Trademark Office (2011) “Patenting In Technology Classes Breakout by Origin, U.S. Metropolitan and Micropolitan Areas,” accessed at $http://www.uspto.gov/web/offices/ac/ido/oeip/taf/cls_cbsa/allcbsa_gd.htm$ Accessed April 1, 2014.

Waldinger, Fabian (2013) “Bombs, Brains, and Science - The Role of Human and Physical Capital for the Creation of Scientific Knowledge in Universities,” *Warwick University*, Working Paper.

Wiancko, A.T. (1922) “Farm Practices Based on Results of Experimental Work,” *Indiana Farmer’s Guide*, Feb 18, 1922; 34(7): 157.

TABLE 1: Descriptive Statistics, 1880

	Full Sample	Distance to Experiment Station: Low	Distance to Experiment Station: High	*t*-stat (2)-(3) [p-value]
	(1)	(2)	(3)	(4)
Panel A: Crop Revenue, Productivity and Farm Values				
Crop Revenue Per Farm Acre	4.15	4.94	3.51	-5.69
	(2.68)	(2.36)	(2.76)	[0.000]
Corn Yield	28.24	32.47	24.78	-6.82
	(11.17)	(8.02)	(12.14)	[0.000]
Wheat Yield	12.22	13.62	10.98	-6.04
	(4.68)	(4.26)	(4.68)	[0.000]
Farm Value Per Farm Acre	20.68	25.87	16.51	-6.55
	(18.26)	(18.28)	(17.14)	[0.000]
Panel B: Agricultural Inputs				
Land Suitability	1.12	1.44	0.87	-4.64
	(0.84)	(0.46)	(0.98)	[0.000]
Farm Acre Per County Acre	0.60	0.71	0.51	-5.68
	(0.31)	(0.25)	(0.33)	[0.000]
Improved Acre Per Farm Acre	0.56	0.64	0.49	-6.30
	(0.23)	(0.18)	(0.25)	[0.000]
Farm Equipment Value per Farm Acre	0.86	1.05	0.71	-6.33
	(0.60)	(0.54)	(0.60)	[0.000]
Fertilizer Expenditure Per Farm Acre	0.03	0.04	0.02	-3.43
	(0.08)	(0.10)	(0.06)	[0.001]
Rural Population Per Farm Acre	0.07	0.07	0.08	0.71
	(0.08)	(0.04)	(0.10)	[0.475]
Woodland Acre Per Farm Acre	0.15	0.16	0.14	-1.95
	(0.11)	(0.10)	(0.11)	[0.051]
Panel C: Standardized Geographic Distances				
Distance to Main Station	0.33	-0.76	1.21	10.03
	(1.47)	(0.44)	(1.41)	[0.000]
Distance to Minor Station	0.28	-0.10	0.58	4.46
	(1.22)	(0.91)	(1.34)	[0.000]
Distance to Railway	0.06	-0.21	0.28	5.36
	(1.17)	(0.28)	(1.52)	[0.000]
Distance to Urban Center	0.31	-0.12	0.68	4.67
	(1.29)	(0.68)	(1.54)	[0.000]
Panel D: Other Characteristics				
Literacy Rate	0.84	0.90	0.80	-4.33
	(0.18)	(0.12)	(0.20)	[0.000]
Ethnic Distance to Station County Population	0.55	-0.03	1.03	2.06
	(2.40)	(0.75)	(3.07)	[0.039]
Manufacturing Output per Worker	2307	2272	2343	0.49
	(1060)	(972)	(1143)	[0.627]
County Observations	1063	532	531	

Notes and sources: Authors' calculations with 1880 county data unless indicated otherwise, as described in the text. Variable definitions and sources are described in the data appendix. All distance variables are standardized to have a sample mean of zero and variance of one. The unit of observation is a county. All statistics are computed with 1880 county land area weights. The main entries in column (1) present the mean of the selected variables for all counties. The main entries in column (2) present the mean of the selected variables for counties below the median distance from the experiment station. The main entries in column (3) present the mean of the selected variables for counties above the median distance from the experiment station. The standard deviations of the selected variable are presented in parenthesis in columns (1)-(3). The main entries in column (4) present the test statistics for a test of differences in means between column (2) and (3), with the p-value of the test presented in square brackets. All monetary values are expressed in 1880 $.

TABLE 2: Research Proximity and Land Productivity

Dependent Variable=	Log(Crop Revenue Per Farm Acre)		
Specification:	Baseline	No Initial Conditions	Baseline: Hatch
	(1)	(2)	(3)
1870 × Station Distance	0.01	0.06	-0.02
	(0.05)	(0.06)	(0.06)
1890 × Station Distance	-0.07*	-0.04	-0.07*
	(0.04)	(0.04)	(0.04)
1900 × Station Distance	-0.24***	-0.21**	-0.28***
	(0.09)	(0.09)	(0.10)
1910 × Station Distance	-0.36***	-0.27**	-0.40***
	(0.11)	(0.12)	(0.10)
1920 × Station Distance	-0.26**	-0.14	-0.29**
	(0.12)	(0.13)	(0.13)
1930 × Station Distance	-0.20*	-0.07	-0.20
	(0.11)	(0.12)	(0.12)
1940 × Station Distance	-0.20*	-0.12	-0.22*
	(0.11)	(0.12)	(0.11)
1950 × Station Distance	-0.05	0.01	-0.03
	(0.12)	(0.11)	(0.14)
1960 × Station Distance	-0.06	0.00	-0.03
	(0.10)	(0.09)	(0.12)
1970 × Station Distance	-0.13	-0.06	-0.10
	(0.10)	(0.10)	(0.12)
1980 × Station Distance	-0.13	-0.08	-0.10
	(0.10)	(0.11)	(0.14)
1990 × Station Distance	-0.13	-0.08	-0.09
	(0.10)	(0.10)	(0.12)
2000 × Station Distance	-0.17*	-0.12	-0.14
	(0.10)	(0.10)	(0.12)
1880 IC × Year FEs	yes	yes	yes
Sample States	all	all	hatch
R^2	0.92	0.91	0.92
County Observations	1063	1063	723

Notes and sources: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources are described in the data appendix. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area. The main entries in the first row of columns (1)-(3) report estimates of β_t from equation (8) in the text. Standard errors clustered at the county level are reported in parentheses. The excluded year interaction is 1880. The models in all columns include additional controls for year and county fixed effects. The model in column (1) and (3) also includes controls for: 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 farm acre per county acre × year fixed effects. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

TABLE 3: Research Proximity and Land Productivity: Alternative Specifications

Dependent Variable=	Log(Crop Revenue Per Farm Acre)									Log (Corn Yield)	Log(Wheat Yield)
							Year-Distance To x Interaction:				
Specification:	Baseline	State × Year	1880 Cattle × Year	1880 PE × Year	Closest Main Station	National Price Revenue	Minor Station	Railway	Urban Center	Baseline	Baseline
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1870 × Station Distance	0.01	-0.06	0.01	-0.04	0.03	0.00	-0.02	0.00	0.06		
	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)		
1890 × Station Distance	-0.07*	-0.07**	-0.07*	-0.07**	-0.05	-0.07	-0.08***	-0.07*	-0.09**	0.01	-0.11*
	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.07)	(0.03)	(0.04)	(0.04)	(0.01)	(0.06)
1900 × Station Distance	-0.24***	-0.26***	-0.25***	-0.26***	-0.22**	-0.47***	-0.27***	-0.24***	-0.22***	-0.02*	-0.09
	(0.09)	(0.10)	(0.09)	(0.09)	(0.10)	(0.15)	(0.09)	(0.09)	(0.08)	(0.01)	(0.07)
1910 × Station Distance	-0.36***	-0.35***	-0.35***	-0.36***	-0.36***	-0.45***	-0.33***	-0.36***	-0.35***	-0.04***	0.03
	(0.11)	(0.10)	(0.11)	(0.10)	(0.10)	(0.14)	(0.10)	(0.10)	(0.09)	(0.01)	(0.04)
1920 × Station Distance	-0.26**	-0.28**	-0.25**	-0.27**	-0.25**	-0.20	-0.27**	-0.26**	-0.25**	-0.05***	-0.06
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.17)	(0.11)	(0.11)	(0.11)	(0.02)	(0.05)
1930 × Station Distance	-0.20*	-0.20**	-0.20*	-0.21**	-0.20*	-0.45*	-0.24**	-0.21*	-0.19*	-0.02	-0.13***
	(0.11)	(0.10)	(0.11)	(0.10)	(0.12)	(0.24)	(0.10)	(0.11)	(0.11)	(0.02)	(0.04)
[1940-2000] × Station Distance	-0.12	-0.10	-0.13	-0.14	-0.11	-0.31*	-0.18**	-0.13	-0.06	-0.01	-0.04*
	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.19)	(0.08)	(0.10)	(0.11)	(0.01)	(0.02)
1880 IC × Year FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Sample States	all	all	all	all	all	all	all	all	all	corn	grain
R^2	0.92	0.93	0.92	0.92	0.92	0.88	0.92	0.92	0.92	0.93	0.76
County Observations	1063	1063	1063	1063	1063	1063	1063	1063	1063	493	214

Notes and sources: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources are described in the data appendix. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area. The main entries in the first row of columns (1)-(11) report estimates of β_t from a modified version of equation (8) in the text. Standard errors clustered at the county level are reported in parentheses. The excluded year interaction is 1880. The models in all columns include additional controls for year and county fixed effects, 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 fraction of land in farms × year fixed effects. The model in column (2) also includes controls state × year fixed effects. The model in column (3) also includes 1880 cattle per farm acre × year fixed effects. The model in column (4) also includes 1880 democratic vote share × year fixed effects. The model in column (5) defines station distance based on the nearest station regardless of state boundaries. The model in column (6) defines crop revenue based on the summation of crop quantity × national prices. The model in column (7) includes additional controls for distance to secondary station × year fixed effects. The model in column (8) includes additional controls for distance to an 1880 railway × year fixed effects. The model in column (9) includes additional controls for distance to an 1880 urban center × year fixed effects. The sample in column (10) is the Parker-Klein (1966) Corn States: IA, IL, IN, MO, and OH. The sample in column (11) is the Parker-Klein (1966) Grain States: KS, NE and TX. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

TABLE 4: Research Proximity and Productivity: Future Research Locations and Manufacturing Productivity Specifications

Dependent Variable=	Log(Crop Revenue Per Farm Acre)			Log(Manufacturing Revenue Per Worker)		
Distance to Station Type:	USDA (1929) Station			Agricultural Experiment Station		
Specification:	Baseline	No Initial Conditions	Baseline: Hatch	Baseline	No Initial Conditions	Baseline: Hatch
	(1)	(2)	(3)	(4)	(5)	(6)
1870 × Station Distance	-0.08*	-0.01	-0.13**	-0.07***	-0.10***	-0.09***
	(0.05)	(0.04)	(0.06)	(0.02)	(0.02)	(0.02)
1890 × Station Distance	-0.07*	-0.05	-0.12**	-0.06**	-0.07**	-0.09**
	(0.04)	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)
1900 × Station Distance	-0.03	-0.05	-0.07	-0.03	-0.01	-0.06**
	(0.09)	(0.08)	(0.12)	(0.02)	(0.03)	(0.02)
1910 × Station Distance	-0.05	-0.03	-0.12	-0.05*	-0.04	-0.07**
	(0.13)	(0.15)	(0.18)	(0.03)	(0.03)	(0.03)
1920 × Station Distance	-0.09	-0.01	-0.17	-0.06*	-0.05	-0.07**
	(0.12)	(0.11)	(0.16)	(0.03)	(0.03)	(0.03)
1930 × Station Distance	-0.14	-0.03	-0.22	-0.06*	-0.03	-0.05
	(0.11)	(0.11)	(0.15)	(0.04)	(0.03)	(0.04)
1940 × Station Distance	-0.11	-0.06	-0.20	0.01	0.03	0.01
	(0.11)	(0.11)	(0.15)	(0.02)	(0.02)	(0.03)
1950 × Station Distance	-0.16*	-0.10	-0.24*	-0.01	-0.01	-0.04
	(0.09)	(0.06)	(0.13)	(0.02)	(0.02)	(0.03)
1960 × Station Distance	-0.17**	-0.10*	-0.22*	-0.03	-0.03	-0.05*
	(0.09)	(0.06)	(0.12)	(0.03)	(0.03)	(0.03)
1970 × Station Distance	-0.15	-0.09	-0.19	-0.03	-0.04	-0.06*
	(0.09)	(0.06)	(0.12)	(0.03)	(0.03)	(0.03)
1980 × Station Distance	-0.13	-0.09	-0.18	-0.03	-0.03	-0.07**
	(0.10)	(0.07)	(0.14)	(0.03)	(0.03)	(0.03)
1990 × Station Distance	-0.15	-0.10	-0.19	-0.05*	-0.06**	-0.08***
	(0.09)	(0.07)	(0.12)	(0.03)	(0.03)	(0.03)
2000 × Station Distance	-0.14	-0.11	-0.20	-0.04	-0.05**	-0.08***
	(0.10)	(0.08)	(0.13)	(0.03)	(0.03)	(0.02)
1880 IC × Year FEs	yes	no	yes	yes	no	yes
Sample States	all	all	hatch	all	all	hatch
R^2	0.92	0.91	0.91	0.96	0.96	0.96
County Observations	1063	1063	723	670	670	413

Notes and sources: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources are described in the data appendix. The sample in columns (4)-(6) is the balanced panel of counties with manufacturing revenue reported in every year of the panel. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area. The main entries in the first row of columns (1)-(6) report estimates of β_t from equation (8) in the text. Standard errors clustered at the county level are reported in parentheses. The excluded year interaction is 1880. The models in all columns include additional controls for year and county fixed effects. The model in column (1), (3),(4) and (6) also includes controls for: 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 farm acre per county acre × year fixed effects. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

TABLE 5: Research Proximity and Total Factor Productivity

Dependent Variable=	Log (Crop Revenue)				
Specification:	Baseline	Investment-Capital Interactions	Fertilizer-Capital Interactions	Baseline + Farm size	Baseline + Farm Size + Education
	(1)	(2)	(3)	(4)	(5)
1870 × Station Distance	-0.03		-0.02	-0.04	-0.04
	(0.09)		(0.06)	(0.09)	(0.09)
1890 × Station Distance	0.01	-0.01	0.01	0.01	0.01
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
1900 × Station Distance	0.02	-0.01	0.02	0.02	0.02
	(0.07)	(0.03)	(0.05)	(0.07)	(0.07)
1910 × Station Distance	-0.14***	-0.16***	-0.14***	-0.15***	-0.15***
	(0.04)	(0.03)	(0.04)	(0.05)	(0.05)
1920 × Station Distance	-0.06	-0.09*	-0.05	-0.07	-0.07
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)
1930 × Station Distance	-0.04	-0.01	-0.01	-0.04	-0.04
	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)
[1940-2000] × Station Distance	0.02	0.02	0.03	0.02	0.01
	(0.05)	(0.03)	(0.03)	(0.05)	(0.05)
Input Controls	yes	yes	yes	yes	yes
Farm Size Controls	no	no	no	yes	yes
Education Controls	no	no	no	no	yes
1880 IC × Year FEs	yes	yes	yes	yes	yes
R^2	0.94	0.97	0.95	0.94	0.94
County Observations	1063	1063	1063	1063	1063

Notes and sources: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources are described in the data appendix. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area. The main entries in the first row of columns (1)-(5) report estimates of β_t from a modified form of equation (8) in the text. Standard errors clustered at the county level are reported in parentheses. The excluded year is 1880. The models in all columns include additional controls for year and county fixed effects, 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 farm acre per county acre × year fixed effects. The models in columns (1)-(5) control for log(improved acres), log(rural population), log(farm equipment), log(fertilizer expenditure), and log(farmland acres). The model in column (2) also control for forth order polynomials in log(farm equipment), log(improved acres), Δ log(farm equipment), and Δ log(improved acres) fully interacted. The model in column (3) also controls for forth order polynomials in log(farm equipment), log(improved acres), and log(fertilizer expenditure) fully interacted. The model in column (4) also controls for log(farm size). The model in column (5) also controls for log(farm size) and log(years of education). * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

TABLE 6: Research Proximity and Land Productivity: Geographic Versus Other Distances

Dependent Variable=	Log(Crop Revenue Per Farm Acre)			
Station Geographic Distance Measure:	Miles	Miles	Miles	Miles
Other Distance Measure Included:	1880 Productivity	1880 Literacy	1880 Ethnic	Land Suitability
	(1)	(2)	(3)	(4)
1870 × Station Geographic Distance	0.01	0.01	-0.03	-0.09*
	(0.04)	(0.05)	(0.06)	(0.06)
1890 × Station Geographic Distance	-0.06*	-0.07*	-0.05**	-0.06***
	(0.04)	(0.04)	(0.02)	(0.03)
1900 × Station Geographic Distance	-0.24***	-0.24***	-0.22**	-0.22***
	(0.09)	(0.09)	(0.10)	(0.08)
1910 × Station Geographic Distance	-0.35***	-0.36***	-0.27***	-0.33***
	(0.10)	(0.11)	(0.10)	(0.08)
1920 × Station Geographic Distance	-0.25**	-0.26**	-0.23**	-0.27***
	(0.11)	(0.12)	(0.11)	(0.10)
1930 × Station Geographic Distance	-0.20*	-0.20*	-0.23***	-0.21**
	(0.11)	(0.11)	(0.07)	(0.09)
[1940-2000] × Station Geographic Distance	-0.12	-0.13	-0.22***	-0.14
	(0.10)	(0.10)	(0.07)	(0.09)
1880 IC × Year FEs	yes	yes	yes	yes
Other Distance × Year FEs	yes	yes	yes	yes
R^2	0.92	0.92	0.93	0.92
County Observations	1063	1063	1063	1063

Notes and sources: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources are described in the data appendix. All distance variables are standardized to have a sample mean of zero and variance of one. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area. The main entries in the first row of columns (1)-(4) report estimates of β_t from a modified version of equation (8) in the text, with the two sets of distance-year interactions indicated. Standard errors clustered at the county level are reported in parentheses. The excluded year interaction is 1880. The models in all columns include additional controls for year and county fixed effects, 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 farm acre per county acre × year fixed effects, and 1880 other distance measure × year fixed effects. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

TABLE 7: Research Proximity and Land Productivity, by Land and Producer Type

Dependent Variable=	Log(Crop Revenue Per Farm Acre)			
	Revenue Maximizing Crop:		1880 Crop Productivity - Agricultural Suitability Gap	
Sample:	Same As Station County	Different From Station County	Positive	Negative
	(1)	(2)	(3)	(4)
1870 × Station Distance	0.00	0.01	-0.02	0.19***
	(0.08)	(0.02)	(0.04)	(0.04)
1890 × Station Distance	-0.03	-0.14**	-0.08*	0.05
	(0.03)	(0.06)	(0.04)	(0.04)
1900 × Station Distance	-0.25*	-0.20***	-0.33***	0.06
	(0.13)	(0.09)	(0.06)	(0.06)
1910 × Station Distance	-0.30**	-0.51***	-0.39***	-0.07
	(0.14)	(0.19)	(0.07)	(0.05)
1920 × Station Distance	-0.19	-0.42**	-0.32***	0.12*
	(0.18)	(0.15)	(0.08)	(0.06)
1930 × Station Distance	-0.10	-0.46**	-0.28***	0.22*
	(0.16)	(0.14)	(0.03)	(0.11)
[1940-2000] × Station Distance	-0.07	-0.23***	-0.23***	0.28*
	(0.16)	(0.06)	(0.02)	(0.15)
1880 IC × Year FEs	yes	yes	yes	yes
R^2	0.91	0.94	0.94	0.92
County Observations	553	510	570	493

Notes and sources: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources are described in the data appendix. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area. The main entries in the first row of columns (1)-(4) report estimates of β_t from a modified version of equation (8) in the text. Standard errors clustered at the county level are reported in parentheses. The excluded year interaction is 1880. The models in all columns include additional controls for year and county fixed effects, 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 farm acre per county acre × year fixed effects. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

TABLE 8: Applied Innovation Proximity and Productivity

Dependent Variable=	Log(Crop Revenue Per Farm Acre)	Log(Crop Revenue)	Log (Corn Yield)	Log (Wheat Yield)
Productivity:	Land	TFP	Land	Land
	(1)	(2)	(3)	(4)
Yield Gap$_{d-20}$ × Station Distance	0.08	-0.04	0.01	-0.11***
	(0.10)	(0.07)	(0.02)	(0.04)
Yield Gap$_{d-10}$ × Station Distance	0.01	-0.02	0.01	-0.19***
	(0.09)	(0.04)	(0.01)	(0.05)
Yield Gap$_{d}$ × Station Distance	-0.16***	0.00	-0.01	-0.05
	(0.04)	(0.03)	(0.01)	(0.04)
Yield Gap$_{d+10}$ × Station Distance	-0.24***	-0.15***	-0.01	-0.26***
	(0.06)	(0.05)	(0.01)	(0.03)
Yield Gap$_{d+20}$ × Station Distance	-0.16***	-0.08**	-0.04**	-0.04
	(0.05)	(0.04)	(0.01)	(0.03)
Yield Gap$_{d+30}$ × Station Distance	-0.10**	-0.05	-0.04**	-0.07*
	(0.05)	(0.03)	(0.02)	(0.04)
Yield Gap$_{d+40}$ × Station Distance	-0.11***	-0.08**	-0.02	-0.01
	(0.04)	(0.04)	(0.01)	(0.03)
Yield Gap$_{d+50}$ × Station Distance	0.03	0.10***	-0.02	0.04
	(0.02)	(0.02)	(0.01)	(0.03)
Yield Gap$_{d-20}$	0.10**	0.05*	-0.10***	-0.26***
	(0.03)	(0.03)	(0.02)	(0.05)
Yield Gap$_{d-10}$	0.15***	0.16**	-0.14***	-0.21***
	(0.03)	(0.03)	(0.02)	(0.09)
Yield Gap$_{d}$	0.13***	0.18**	-0.09***	0.00
	(0.03)	(0.02)	(0.02)	(0.09)
Yield Gap$_{d+10}$	-0.05	0.00	-0.12***	-0.52***
	(0.04)	(0.04)	(0.02)	(0.08)
Yield Gap$_{d+20}$	-0.01	0.07**	-0.14**	-0.02
	(0.03)	(0.03)	(0.02)	(0.08)
Yield Gap$_{d+30}$	-0.09***	-0.02	-0.10***	-0.18***
	(0.04)	(0.04)	(0.02)	(0.07)
Yield Gap$_{d+40}$	-0.12**	-0.05	-0.02	-0.05
	(0.03)	(0.03)	(0.01)	(0.07)
Yield Gap$_{d+50}$	0.02*	0.15***	-0.12	-0.14***
	(0.02)	(0.03)	(0.01)	(0.05)
1880 IC × Year FEs	no	no	no	no
Sample States	all	all	corn	grain
R^2	0.92	0.95	0.93	0.76
County Observations	1063	1063	493	214

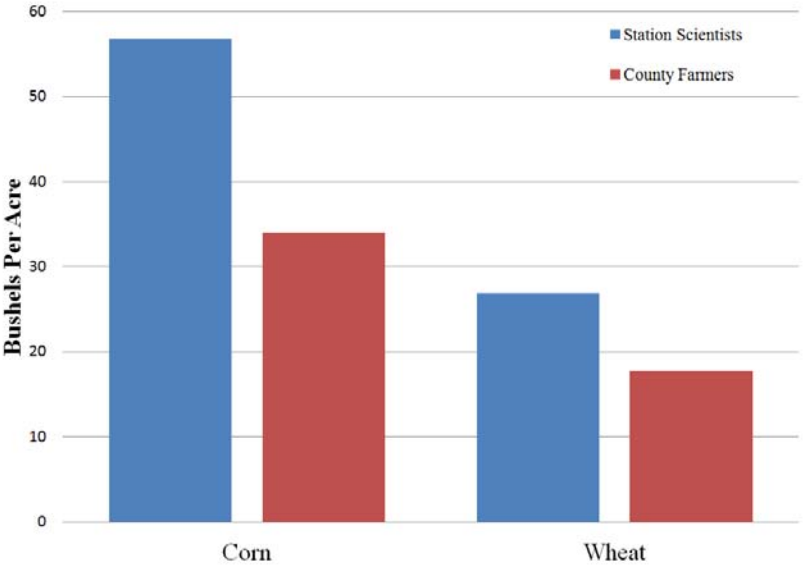
Notes and sources: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources are described in the data appendix. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area. The main entries in the first row of columns (1)-(4) report estimates of β_{d-t} from equation (9) in the text. Standard errors clustered at the county level are reported in parentheses. The excluded year interaction is *d+60*. The models in all columns include additional controls for year and county fixed effects. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

TABLE 9: Research Proximity and Land Productivity, By Research Type

Dependent Variable=	Log(Crop Revenue Per Farm Acre)	
Stratification Variable=	Journal Article to Bulletin Ratio	
Sample:	Above median	Below median
	(1)	(2)
1870 × Station Distance	-0.06	0.01
	(0.08)	(0.07)
1890 × Station Distance	-0.05	-0.07*
	(0.03)	(0.04)
1900 × Station Distance	-0.11*	-0.28***
	(0.06)	(0.11)
1910 × Station Distance	-0.13***	-0.41***
	(0.04)	(0.11)
1920 × Station Distance	-0.17***	-0.27*
	(0.04)	(0.14)
1930 × Station Distance	-0.17***	-0.20
	(0.05)	(0.13)
[1940-2000] × Station Distance	-0.19***	-0.11
	(0.06)	(0.14)
1880 IC × Year FEs	yes	yes
Sample States	all	all
R^2	0.94	0.91
County Observations	577	486

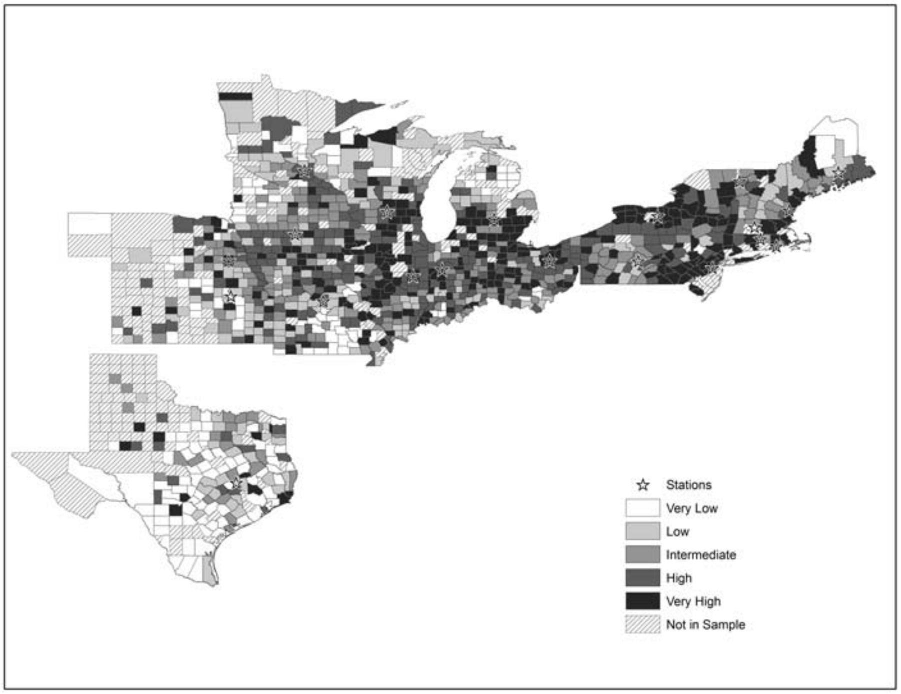
Notes and sources: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources are described in the data appendix. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area for the sample indicated. The main entries in the first row of columns (1)-(4) report estimates of β_t from a modified form of equation (8) in the text. Standard errors clustered at the county level are reported in parentheses. The excluded year interaction is 1880. The models in all columns include additional controls for year and county fixed effects, 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 farm acre per county acre × year fixed effects. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

FIGURE 1: Corn and Wheat Crop Yields: Station Trials and Local Farmers Compared



Notes: Authors' tabulation of experiment station reported crop yields from crop trials and agricultural census data for the primary experiment station county from 1889, 1899, 1909, 1919, 1929, and 1939.

FIGURE 2: Station Locations and Crop Revenue Per Farm Acre, 1880

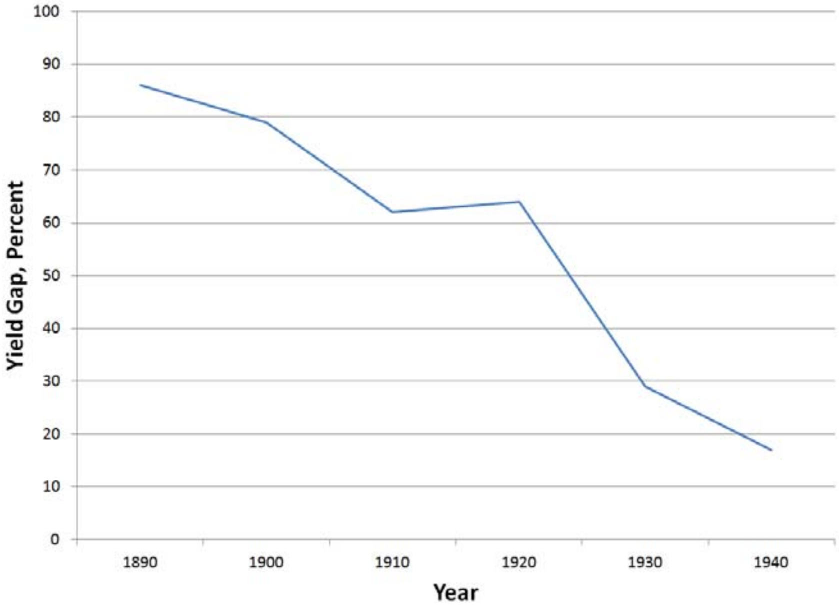


Notes: The map reflects station locations and crop revenue per acre in 1880 for the 1880 county boundaries for the balanced panel of counties that do not have large border changes during the sample period and report crop revenue in every year of the sample. Station locations denote the primary federal experiment station for a state.

FIGURE 3: Research Proximity and Land Productivity

Notes: Authors' calculation from fitting equation (7). Each solid line plots the year-by-year distance interaction estimates from a different specification with crop revenue per acre as the outcome variable. Model 1 plots estimates from Table 2 column (1). Model 2 plots estimates from Table 2 column (2). Model 3 plots estimates from Table 2 column (3). The dotted lines depict 95% confidence intervals from fitting Model 1.

FIGURE 4: Experiment Station – Farmer Yield Gap for Corn and Wheat, 1880-1940



Notes: Authors' calculation with data from the agricultural census and experiment station reports for the years indicated.

Online Appendix

A.1 Research by Entrepreneurs

Entrepreneurs have the opportunity to attempt an innovation to create a new more productive version of the intermediate good each period. If they succeed the productivity of the leading variety is $A_f = \gamma A_o$. If they fail, the leading intermediate good remains the same so that $A_f = A_o$.

A successful innovator becomes a monopolist who maximizes its profit, Π_f, measured in units of the final good:

$$\Pi_f = p_f x_f - x_f \tag{A1}$$

where p_f is the relative price of the intermediate good. Their revenue is $p_f x_f$, and cost their input of the final good which must equal their output x_f. Monopolist entrepreneurs charge a price equal to marginal product of their input in final goods production, so that $p_f = \alpha A_f x_f^{\alpha-1}$. Solving the entrepreneurs problem in (A1) using this price yields: $x_f = (A_f \alpha^2)^{\frac{1}{1-\alpha}}$, and relative input expression

$$x_f = x_o \gamma^{\frac{1}{1-\alpha}}. \tag{A2}$$

A.2 Data Appendix

I. Agriculture

Crop Revenue Per Farm Acre: Total crop revenue per acre of farmland. Crop revenue is not reported in 1870, 1880, 1890 and 1900 at the county level. County level data for these years is based on reported total farm revenue scaled by the fraction of total farm revenue from crops for a county averaged over years 1950-2000. *Sources: Agricultural Census, Haines (2010).*

Corn Yield: Total bushels of corn per acre planted in corn. *Sources: Agricultural Census, Haines (2010).*

Wheat Yield: Total bushels of wheat per acre planted in wheat *Sources: Agricultural Census, Haines (2010).*

Farm Value Per Farm Acre: Total value of the farm per acre of farmland. *Sources: Agricultural Census, Haines (2010).*

Farm Acre Per County Acre: Total acres of farmland per total county acreage. *Sources: Agricultural Census, Haines (2010).*

Improved Acre Per Farm Acre: Total acres of improved land per total acreage of farmland. *Sources: Agricultural Census, Haines (2010).*

Farm Equipment Value Per Farm Acre: Total value of farm equipment per acre of farmland. The total value of farm equipment is not reported in 1950 and 1960. For these years the total value of farm equipment is linearly interpolated using data from 1940 and 1970 for a county. *Sources: Agricultural Census, Haines (2010).*

Fertilizer Expenditure Per Farm Acre: Total expenditure on fertilizer per acre of farmland. Total expenditure on fertilizer is not reported in 1880 and 1960. For these years total expenditure on fertilizer is linearly interpolated. *Sources: Agricultural Census, Haines (2010).*

Rural Population Per Farm Acre: Total rural population per acre of farmland. Total rural population is defined as total population minus urban population. Percent urban population is not reported in 1960 and 1970. We linearly interpolate urban population in these years. *Sources: Agricultural Census, Haines (2010).*

1880 Woodland Acre Per County Acre: Total woodland acres per total county acreage in 1880. *Sources: Agricultural Census, Haines (2010).*

1880 Cattle Per County Acre: Number of milch cows per farm acre in 1880. *Sources: Agricultural Census, Haines (2010).*

1880 Crop Productivity - Agricultural Suitability Gap: The difference between reported Crop Revenue Per Farm Acre and Agricultural Suitability in 1880. Agricultural Suitability is calculated by taking the predictions of a regression of Crop Revenue Per Farm Acre on 25 measures of county geography using Census of Agriculture data from the 1880 county cross-section, as described in more detail below.*Sources: Agricultural Census, Haines (2010), authors calculations.*

Crop Revenue Per Farm Acre, National Price Revenue Based: We calculated the total revenue from selected crops based on national prices, per acre of farmland. Crop revenue for this measure is equal to $P_{Corn} \times Q_{Corn} + P_{Wheat} \times Q_{Wheat} + P_{Oats} \times Q_{Oats} + P_{Barley} \times Q_{Barley}$, where prices are national prices (dollars per bushel) and quantities are county specific (measured in bushels). *Sources: Agricultural Census, Haines (2010) and National Historical Statistics.*

II. Geographic Distances

Distance to Main Station: Each countys distance to its states primary agricultural research station, standardized to have a sample mean of zero and variance of one. Most states have only a primary station. *Source: Authors calculations using locations from the United States Office of Experiment Stations (1910), p.300.*

Distance to Minor Station: For states with more than one experiment station receiving federal aid, we calculated each countys distance to its states minor agricultural research station, standardized to have a sample mean of zero and variance of one. The primary and minor stations for these states are as follows: Connecticut - Storrs (Primary) and New Haven (Minor); Missouri - Columbia (Primary) and Mountain Grove (Minor); New York - Geneva (Primary) and Ithaca (Minor). *Source: Authors calculations using locations from the United States Office of Experiment Stations (1910), p.300.*

Distance to 1880 Railway: Each countys distance to a county with a railway connection in 1880, standardized to have a sample mean of zero and variance of one. The presence of a railroad in 1880 is referenced against the NHGIS boundary file for 1880 based upon the digitized image from the Library of Congress of "Colton's railroad map of the United States, published in 1882 in the manner described in Atack, Bateman, Haines and Margo (2010). We thank Jeremy Atack for making these data available to us. *Sources: Atack, Bateman, Haines and Margo (2010).*

Distance to Closest Station: Each countys distance to the nearest primary agricultural research station from a county regardless of state borders, standardized to have a sample mean of zero and variance of one. *Source: Authors calculations using locations from the United States Office of Experiment Stations (1910), p.300.*

Distance to 1880 Urban Center: Each countys distance to nearest urban center, standardized to have a sample mean of zero and variance of one. Urban center counties are defined as being in the top 5% of the urban population distribution in 1880. *Sources: Authors calculations with data from the 1880 Population Census.*

Distance to USDA (1929) Station: Each countys distance to the nearest United State Department of Agriculture research center open in 1929, standardized to have a sample mean of zero and variance of one. *Source: United States Department of Agriculture (1929).*

III. Non-Geographic Distances

Ethnic Distance to Station County Population: We utilize a similar approach to Spolaore and Wacziarg (2009) based on the place of birth of the population of each county in 1880. We use the 100% sample of the 1880 census to construct a measure of the ethnic distance of the population of the county county in question from the population of the experiment station county in the same state. To do so we first create country of origin shares in each county as well as the experiment station county. We then use country of origin shares from the census data and the genetic distance between each country of origin to measure the weighted genetic distance of county i from experiment station county e as $F_{ie}^{ST} = \sum_{j=1}^{n} \sum_{k=1}^{n} (s_{j,e} \times s_{k,i} \times d_{jk})$ for birthplaces $j, k = 1, ...n$ with genetic distances between them of d_jk. This measure is standardized to have a sample mean of zero and variance of one. *Sources: Spolaore and Wacziarg (2009) and 1880 Population Census, Ruggles et al. (2010).*

Ethnic Distance to US Population: We follow the same procedure as for the Ethnic Distance to Station County Population variable with one alteration. Here we compute the genetic distance to the US population rather than the experiment station county. This measure is standardized to have a sample mean of zero and variance of one. *Sources: Spolaore and Wacziarg (2009) and 1880 Population Census, Ruggles et al. (2010).*

Agricultural Suitability Distance: Each countys distance from its states primary agricultural research station county in terms of land suitability for agriculture, standardized to have a sample mean of zero and variance of one. “Agricultural suitability captures geographic suitability for agriculture, described in section VI below. *Sources: Sources: Authors calculations with Fishback, Horrace, and Kantor (2005, 2006) and 1880 Agricultural Census, Haines (2010).*

1880 Productivity Distance: Each countys distance from its states primary agricultural research station county in terms of crop productivity in 1880, standardized to have a sample mean of zero and variance of one. Crop productivity is measured as crop revenue per farm acre as defined in section I above. *Source: 1880 Agricultural Census, Haines (2010)*

1880 Literacy Distance: *Sources:* Each countys distance to its states primary agricultural research station county in terms of literacy rate in 1880, standardized to have a sample mean of zero and variance of one. Literacy is measured by the population literacy rate defined in section IV below. *Source: 1880 Population Census, Haines (2010)*

IV. Demographics and Other Sectors

Manufacturing Output per Worker: Total manufacturing revenue per worker employed in manufacturing. Manufacturing revenue is interpolated in 1970 using the reported value added in manufacturing in 1970 times the ratio of value added to total revenue in 1980 in the county. *Source: Census, Haines (2010)*

1880 Democratic Vote Share: Fraction of votes cast for Winfield Scott Hancock the Democratic candidate in the Presidential election. *Source: Clubb et al. (1987)*

1880 Literacy Rate: Fraction of the population aged 10 and older that can either read or write *Source: Census, Haines (2010)*

Years of Education: Constructed from literacy rates and years of education at the county-level. Use the 1940 census reported national means of literacy and years of education to construct years of education before 1940. Nation illiteracy rate in 1940 is 3% and years of school is 8.6 years. We use the ratio of literacy to years of schooling to convert pre-1940 census data that only reports literacy into years of schooling. *Sources: Census, Haines (2010), Ruggles et al. (2010), and National Assessment of Adult Literacy (2013)*

Population: Total population of the county *Sources: Haines (2010)*

V. Experiment Station Innovations and Research

Yield Gap: Gap between reported experiment station crop yields and experiment station county farmers crop yields in a year that are in the top 25% of the yield gap distribution. Computed for both wheat and corn yields for the years 1889, 1899, 1909, 1919, 1929, and 1939. *Sources: Agricultural Census, Haines (2010) and United States Department of Agriculture, Office of Experiment Stations (Various).*

Journal Article to Bulletin Ratio: Average ratio of scholarly articles attributable to land grant campus to farmers bulletins from experiment station, 1910-1940, state level *Sources: Web of Science and United States Department of Agriculture, Office of Experiment Stations (Various).*

Yield Gap Frequency: Frequency of top 25% station-farmer yield gap, state-level *Sources: Agricultural Census, Haines (2010) and United States Department of Agriculture, Office of Experiment Stations (Various).*

VI. Agricultural Land Suitability

Agricultural Suitability: This measure is calculated by taking the predictions of a regression of Crop Revenue Per Farm Acre on 25 measures of county geography using Census of Agriculture data from the 1880 county cross-section. The 25 geography measures are: latitude of the county seat, longitude of the county seat, the latitude - longitude of the county seat interaction, average precipitation in the 1920s, average temperature in the 1920s, average precipitation - average temperature in the 1920s interaction, months severe drought (1895-1929, 10 year Average), months severe wet (1895-1929, 10 year Average), average number of small rivers (that pass through 11-20 counties) in a county, average number of medium rivers (that pass through 20-50 counties) in a county, average number of large rivers (that pass through at least 51 counties) in a county, minimum elevation in the county, maximum elevation in the county, soil: available water capacity, soil: percentage of clay in the soil, soil: annual k factor, soil: organic material in the soil, soil: permeability of the soil, soil: soil thickness, soil: hydrological grouping, soil: quality of the drainage, soil slope of the component, spoil: liquid limit, soil: hydric soil rating, and soil: annual flood frequency. *Sources: Authors calculations with Fishback, Horrace, and Kantor (2005, 2006) and 1880 Agricultural Census Data.*

Geography Determined Yield, by Crop: This measure is calculated separately for each of five crops by taking the predictions of a regression of crop level productivity on 5 measures of county geography using Census of Agriculture data from the 1880 county cross-section. The five crops are: corn, wheat, oats, rye and barley. The 25 geography measures are listed above in Agricultural Suitability. *Sources:Authors calculations with Fishback, Horrace, and Kantor (2005, 2006) and 1880 Agricultural Census Data.*

Revenue Maximizing Crop: This value takes a value of one for the crop that maximizes expected revenue per acre based on the Geography Determined Yield and national prices. Revenue per acre = Geography Determined Yield × average national price of the crop from 1870-1880. *Sources: Authors calculations with Fishback, Horrace, and Kantor (2005, 2006), 1880 Agricultural Census, and National Historical Statistics Data.*

Revenue Maximizing Crop, Same As Station County: Takes a value of one if the Revenue Maximizing Crop in a county is the same as that in the experiment station county in the same state; zero otherwise. *Sources: Authors calculations with Fishback, Horrace, and Kantor (2005, 2006) and 1880 Agricultural Census Data.*

A.3 Data References

American Association of Nurserymen, Florists and Seedsmen (1880) “Proceedings at the Annual Meeting of the American Association of Nurserymen, Florists and Seedsmen,” Chicago, IL; p.67 Membership list.

Clubb, Jerome M., William H. Flanigan, and Nancy H. Zingale, (1987) “Electoral Data for Counties in the United States: Presidential and Congressional Races, 1840-1972,” [Computer file]. Compiled by Jerome M. Clubb, University of Michigan, William H. Flanigan, University of Minnesota, and Nancy H. Zingale, College of St. Thomas. ICPSR08611-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 2006-11-13.

Fishback, Price V., Horrace, William C. and Kantor, Shawn (2006) “The impact of New Deal expenditures on mobility during the Great Depression,” *Explorations in Economic History*, 43(2): 179-222.

Fishback, Price V., Horrace, William C. and Kantor, Shawn (2005) “Did New Deal Grant Programs Stimulate Local Economies? A Study of Federal Grants and Retail Sales During the Great Depression,” *Journal of Economic History*, 65(1): 6-71.

Haines, Michael R.(2010) “Historical, Demographic, Economic, and Social Data: The

United States, 1790-2002.” ICPSR02896-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21.

Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek (2010) *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota.

United States Department of Agriculture (1929) “List of Technical Workers in the Department of Agriculture, and Outline of Department Functions 1929,” Miscellaneous Publication 63, United States Government Printing Office, Washington D.C.

United States Department of Agriculture, Office of Experiment Stations (Various) *Experiment Station Record*, United States Government Printing Office, Washington D.C.

TABLE A1: Experiment Station Locations

State	City	County	University	Pre-Hatch Station Open Date	Station Type
CT	New Haven	New Haven	Yale University	1875	Minor
CT	Storrs	Tolland	University of Connecticut	None	Main
IL	Urbana	Champaign	University of Illinois - Urbana-Champaign	None	Main
IN	Lafayette	Tippecanoe	Purdue University	None	Main
IA	Ames	Story	Iowa State University	None	Main
KS	Manhattan	Riley	Kansas State University	None	Main
ME	Orono	Penobscot	University of Maine	1885	Main
MA	Amherst	Hampshire	University of Massachusetts Amherst	1882	Main
MI	East Lansing	Ingham	Michigan State University	None	Main
MN	Saint Paul	Ramsey	University of Minnesota	1885	Main
MO	Columbia	Boone	University of Missouri–Columbia	None	Main
MO	Mountain Grove	Texas	Missouri State University	1906	Minor
NE	Lincoln	Lancaster	University of Nebraska - Lincoln	1884	Main
NH	Durham	Strafford	University of New Hampshire	None	Main
NJ	New Brunswick	Middlesex	Rutgers University	1880	Main
NY	Geneva	Ontario	Cornell University - College of Agriculture and Life Sciences	1882	Main
NY	Ithaca	Tompkins	Cornell University - Ithaca	1879	Minor
OH	Wooster	Wayne	Ohio State University*	1882	Main
PA	State College	Centre	Pennsylvania State University	None	Main
RI	Kingston	Washington	University of Rhode Island	None	Main
TX	College Station	Brazos	Texas A and M University	None	Main
VT	Burlington	Chittenden	University of Vermont	1886	Main
WI	Madison	Dane	University of Wisconsin-Madison	1883	Main

Source: *Annual Report*, Office of Experiment Stations (1910), Government Printing Office, Washington: DC. *Established at Ohio State Campus in Columbus in 1870, moved to Wooster in 1892 and returned to Columbus in 1948.

TABLE A2: Experiment Station Corn and Wheat Trials Used to Construct Yield Gaps, 1889-1939

Crop	Volume	Page	State	Census Year	Units	Yield in Trial Number:										
						One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten	Eleven
Corn	1	30	IL	1890	Bushels per acre	92.8	93.4	87.4	87.8	60.1	66	57.6	60.2			
Corn	1	96	MN	1890	Tons per acre	6.2	4.4									
Corn	1	96	MN	1890	Tons per acre	26.8	13									
Corn	1	96	MN	1890	Bushels per acre	40										
Corn	1	265	NY	1890	Pounds per acre	215	237									
Corn	1	265	NY	1890	Pounds per acre	525	512									
Corn	1	265	NY	1890	Pounds per acre	1443	1650									
Corn	1	265	NY	1890	Pounds per acre	580	751									
Corn	1	265	NY	1890	Pounds per acre	390	634									
Corn	1	265	NY	1890	Pounds per acre	909	1077									
Corn	1	265	NY	1890	Pounds per acre	817	611									
Corn	1	265	NY	1890	Pounds per acre	2696	3073									
Corn	1	265	NY	1890	Pounds per acre	125	188									
Corn	1	265	NY	1890	Pounds per acre	5004	5660									
Corn	1	265	NY	1890	Pounds per acre	20302	19351									
Corn	1	265	NY	1890	Pounds per acre	25306	25011									
Corn	1	140	OH	1890	Tons per acre	19.5										
Corn	10	540	IL	1900	Bushels per acre	54										
Corn	10	142	KS	1900	Bushels per acre	19.57										
Corn	10	136	MI	1900	Pounds per acre	19134	17210									
Corn	10	428	NE	1900	Bushels per acre	30	18	38								
Corn	11	529	MA	1900	Bushels per acre	20.6	18.5	19.8	30	10.9	41.2	55.9	67.7	116.9	16.3	
Corn	11	233	TX	1900	Bushels per acre	40.7	39.5	37.2								
Corn	21	328	IA	1910	Bushels per acre	59.8	57.8									
Corn	21	35	IN	1910	Bushels per acre	43.8	65.1	56.8	57.4	48.4	52.8	49.1	43.7	44.2	44.8	40.4
Corn	21	732	KS	1910	Bushels per acre	43.98	47.62	50.75	41.77							
Corn	21	729	NH	1910	Bushels per acre	19										
Corn	21	39	TX	1910	Bushels per acre	33.15	31.85	33.8								
Corn	40	319	KS	1920	Bushels per acre	23	42.25									
Corn	40	329	KS	1920	Pounds per acre	2100	1900	1800								
Corn	40	329	KS	1920	Bushels per acre	58.6	41.5	16.3	7.5							
Corn	40	731	MI	1920	Tons per acre	2.21										
Corn	40	334	OH	1920	Bushels per acre	75.06	71.02	72.12	74.21							
Corn	40	736	TX	1920	Bushels per acre	21.98	20.1									
Corn	41	32	KS	1920	Bushels per acre	22.6	44.9	32.8	22.7							
Corn	41	32	KS	1920	Tons per acre	19.2	12.6	12.1								
Corn	41	636	MO	1920	Tons per acre	6.25	4.58	4.35								
Corn	41	636	MO	1920	Bushels per acre	81.3	63	43.7	69.3	58.3	47.3					
Corn	41	636	MO	1920	Bushels per acre	38	33	32	30	59	70	73	75	84.2		
Corn	41	433	NE	1920	Tons per acre	4.5	9.1									
Corn	60	534	CT	1930	Percentage Multiplier	104.8	103.9									
Corn	60	222	IL	1930	Percentage Multiplier	190										
Corn	60	717	TX	1930	Percentage Multiplier	147.5										
Corn	61	129	NE	1930	Percentage Multiplier	112	113	107								
Corn	61	129	NE	1930	Bushels per acre	33.9	36.8	45.4	48.7	46	42.9	51.4	52.3	51.6	50.3	
Corn	81	38	IA	1940	Percentage Multiplier	116.9										
Wheat	1	206	IN	1890	Bushels per acre	24.5										
Wheat	1	206	IN	1890	Bushels per acre	24.12	16.65	7.47								
Wheat	1	206	IN	1890	Bushels per acre	10.7	15.5	4.8								
Wheat	1	214	KS	1890	Bushels per acre	18.25										
Wheat	1	289	OH	1890	Bushels per acre	40.5	37.4	37	38							
Wheat	10	843	IN	1900	Bushels per acre	29.08	28.94									
Wheat	10	41	PA	1900	Bushels per acre	42.93	29.49	30								
Wheat	11	638	MN	1900	Bushels per acre	24.2	20.6	20.8	20.8	23.1	22	23.3	23.6			
Wheat	11	638	MN	1900	Bushels per acre	26.6	23.6	28.4	27.7	21.5	21.1					
Wheat	11	731	PA	1900	Bushels per acre	41.42	37.97	37.57	37.15	26.87	26.8	26.43				
Wheat	11	731	PA	1900	Bushels per acre	31.82	31.77	31								

Wheat	21	35	IN	1910	Bushels per acre	21	19.3	18.4	13.6	13.6						
Wheat	21	732	KS	1910	Bushels per acre	12.34	16.61	16.49	14.91							
Wheat	21	132	MN	1910	Bushels per acre	26.7	26.4									
Wheat	21	129	NE	1910	Bushels per acre	42	38.35	41	59	24.4						
Wheat	21	129	NE	1910	Bushels per acre	57	66.9	53	59	20.86	41					
Wheat	21	129	NE	1910	Bushels per acre	54.84	55.48	56.24	31.4	24.25	28.15	31.6				
Wheat	21	729	NH	1910	Bushels per acre	30.75	18.2									
Wheat	21	634	PA	1910	Bushels per acre	36.4	32.6									
Wheat	21	634	PA	1910	Pounds per acre	4191	4129	3153								
Wheat	40	443	IL	1920	Bushels per acre	29.3	34.8									
Wheat	40	735	IN	1920	Bushels per acre	13.7	38.7									
Wheat	40	329	KS	1920	Bushels per acre	8.25	31.1	26.5	25.9	34.3	24.2					
Wheat	40	329	KS	1920	Bushels per acre	14	42.55									
Wheat	40	731	MN	1920	Bushels per acre	25.6										
Wheat	40	732	MN	1920	Bushels per acre	27.7	12.5	30.7								
Wheat	40	732	MN	1920	Bushels per acre	15.9	16.6	25.8	23.7	25.5	20.9					
Wheat	40	734	MN	1920	Bushels per acre	11.9										
Wheat	40	738	OH	1920	Bushels per acre	17.9										
Wheat	41	32	KS	1920	Bushels per acre	17.3	15.73	13.38	14.88	11.85	20.5					
Wheat	41	636	MO	1920	Bushels per acre	23.8	30.6	23.5	35	27.8	42	34.3	33.6	37.1		
Wheat	41	636	MO	1920	Bushels per acre	41.1	42.2	65.3	69	20.5						
Wheat	41	644	MO	1920	Bushels per acre	0.2	10.9	20.7	27.4	13.5	31	30	39.4			
Wheat	41	644	MO	1920	Bushels per acre	9.5	17.4	17.9	20.7	25.9	29.1					
Wheat	41	433	NE	1920	Bushels per acre	8.9	7	17.5	3.7	5.6	19	8.3	26.2			
Wheat	41	19	NJ	1920	Bushels per acre	7.33	12.33									
Wheat	41	19	NJ	1920	Pounds per acre	780	1420									
Wheat	60	24	MI	1930	Bushels per acre	18.73	30.3									

Source: *Experiment Station Record*, Government Printing Office; Washington DC. Authors' compilation from volume and page number indicated.

TABLE A3: Research Proximity and Land Productivity: Alternative Inference Procedures

Dependent Variable=	Log(Crop Revenue Per Farm Acre)		
Inference Procedure:	State Clusters	Year and Distance to Station Clusters	Driscoll-Kraay Spatial Standard Errors
	(1)	(2)	(3)
1870 × Station Distance	0.01	0.01	0.01
	(0.03)	(0.03)	(0.06)
1890 × Station Distance	-0.07**	-0.07*	-0.07
	(0.01)	(0.03)	(0.04)
1900 × Station Distance	-0.24***	-0.24***	-0.24**
	(0.07)	(0.09)	(0.10)
1910 × Station Distance	-0.36***	-0.36***	-0.36***
	(0.09)	(0.11)	(0.11)
1920 × Station Distance	-0.26**	-0.26**	-0.26**
	(0.06)	(0.12)	(0.13)
1930 × Station Distance	-0.20***	-0.20*	-0.20
	(0.03)	(0.11)	(0.12)
1940 × Station Distance	-0.20***	-0.20*	-0.20*
	(0.03)	(0.11)	(0.12)
1950 × Station Distance	-0.05**	-0.05	-0.05
	(0.02)	(0.11)	(0.13)
1960 × Station Distance	-0.06*	-0.06	-0.06
	(0.03)	(0.10)	(0.12)
1970 × Station Distance	-0.13***	-0.13	-0.13
	(0.02)	(0.10)	(0.11)
1980 × Station Distance	-0.13***	-0.13	-0.13
	(0.03)	(0.11)	(0.11)
1990 × Station Distance	-0.13***	-0.13	-0.13
	(0.04)	(0.10)	(0.11)
2000 × Station Distance	-0.17***	-0.17*	-0.17
	(0.04)	(0.10)	(0.12)
1880 IC × Year FEs	yes	yes	yes
Sample States	all	all	all
R^2	0.92	0.07	0.92
County Observations	1063	1063	1063

Notes: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources as described in the data appendix. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area. The main entries in the first row of columns (1)-(3) report estimates of β_t from equation (8) in the text. Standard errors clustered at the state level are reported in parentheses in column (1), clustered at the year and distance to station categories in column (2) and Driscoll-Kraay spatial standard errors in column (3). The excluded year interaction is 1880. The models in all columns include additional controls for year and county fixed effects, 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 farm acre per county acre × year fixed effects. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

TABLE A4: Applied Innovation Proximity and Productivity: Future Research Locations and Manufacturing Productivity Specifications

Dependent Variable=	Log(Crop Revenue Per Farm Acre)	Log(Manufacturing Revenue Per Worker)
Location Type:	USDA (1929) Station	Agricultural Experiment Station
	(1)	(2)
Yield Gap$_{d-20}$ × Location Distance	0.09	0.06**
	(0.07)	(0.03)
Yield Gap$_{d-10}$ × Location Distance	0.01	-0.02
	(0.05)	(0.03)
Yield Gap$_{d}$ × Location Distance	0.04	0.04***
	(0.06)	(0.01)
Yield Gap$_{d+10}$ × Location Distance	0.01	0.02*
	(0.12)	(0.01)
Yield Gap$_{d+20}$ × Location Distance	0.02	0.01
	(0.08)	(0.01)
Yield Gap$_{d+30}$ × Location Distance	0.01	0.03*
	(0.07)	(0.01)
Yield Gap$_{d+40}$ × Location Distance	0.00	0.08**
	(0.08)	(0.03)
Yield Gap$_{d+50}$ × Location Distance	-0.03	0.04***
	(0.03)	(0.01)
Yield Gap$_{d-20}$	0.08**	0.05***
	(0.04)	(0.03)
Yield Gap$_{d-10}$	0.19***	0.03
	(0.03)	(0.03)
Yield Gap$_{d}$	0.10***	0.01
	(0.03)	(0.03)
Yield Gap$_{d+10}$	-0.01	-0.07**
	(0.05)	(0.03)
Yield Gap$_{d+20}$	-0.01	-0.09**
	(0.03)	(0.03)
Yield Gap$_{d+30}$	-0.06	-0.11**
	(0.04)	(0.03)
Yield Gap$_{d+40}$	-0.15***	-0.10***
	(0.04)	(0.02)
Yield Gap$_{d+50}$	-0.02*	-0.07***
	(0.03)	(0.02)
1880 IC × Year FEs	no	no
Sample States	all	all
R^2	0.91	0.96
County Observations	1063	670

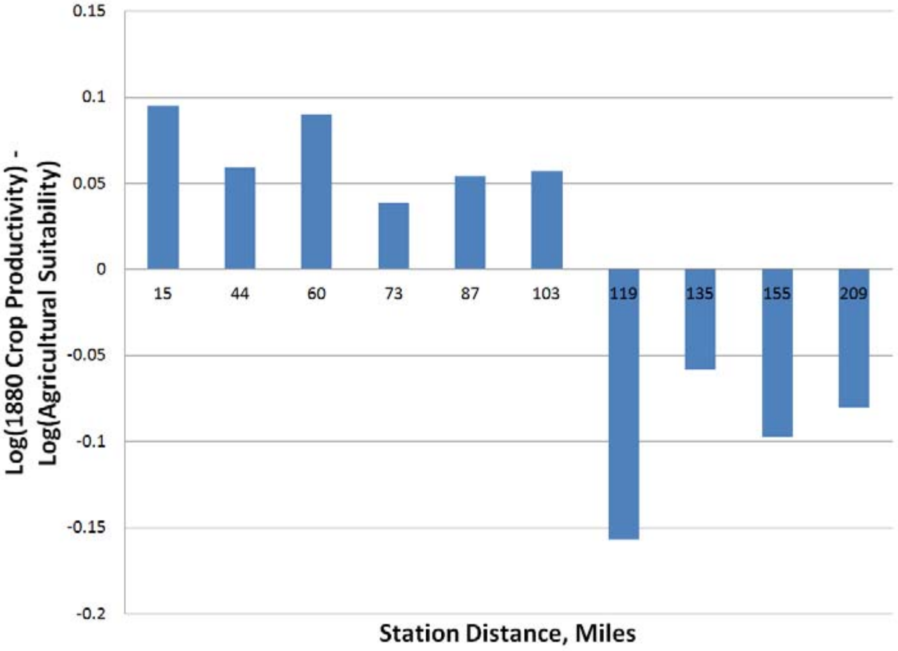
Notes: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources as described in the data appendix. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area. The main entries in the first row of columns (1)-(2) report estimates of β_{d-t} from equation (9) in the text. Standard errors clustered at the county level are reported in parentheses. The excluded year interaction is *d+60*. The models in all columns include additional controls for year and county fixed effects, 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 farm acre per county acre × year fixed effects. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

TABLE A5: Research Proximity and Productivity, Discovery and Learning Stratifications: Manufacturing Productivity Specifications

Dependent Variable=	Log(Manufacturing Revenue Per Worker)			
Location Type:	Agricultural Experiment Station			
Stratification Variable=	Journal Article to Bulletin Ratio		1880 Crop Productivity - Agricultural Suitability Gap	
Sample:	High	Low	Positive	Negative
	(1)	(2)	(3)	(4)
1870 × Location Distance	-0.04	-0.07***	-0.08***	-0.07**
	(0.05)	(0.02)	(0.03)	(0.03)
1890 × Location Distance	0.00	-0.07***	-0.05*	-0.08
	(0.02)	(0.02)	(0.03)	(0.07)
1900 × Location Distance	-0.01	-0.04	-0.06***	0.01
	(0.03)	(0.02)	(0.03)	(0.02)
1910 × Location Distance	-0.02	-0.07***	-0.08***	0.00
	(0.04)	(0.03)	(0.03)	(0.02)
1920 × Location Distance	-0.02	-0.08**	-0.09***	0.00
	(0.05)	(0.04)	(0.03)	(0.03)
1930 × Location Distance	-0.02	-0.07*	-0.09***	0.02
	(0.05)	(0.04)	(0.03)	(0.03)
[1940-2000] × Location Distance	-0.05	-0.02	-0.06***	0.05**
	(0.03)	(0.03)	(0.02)	(0.02)
1880 IC × Year FEs	yes	yes	yes	yes
Sample States	all	all	all	all
R^2	0.96	0.97	0.96	0.96
County Observations	367	303	412	258

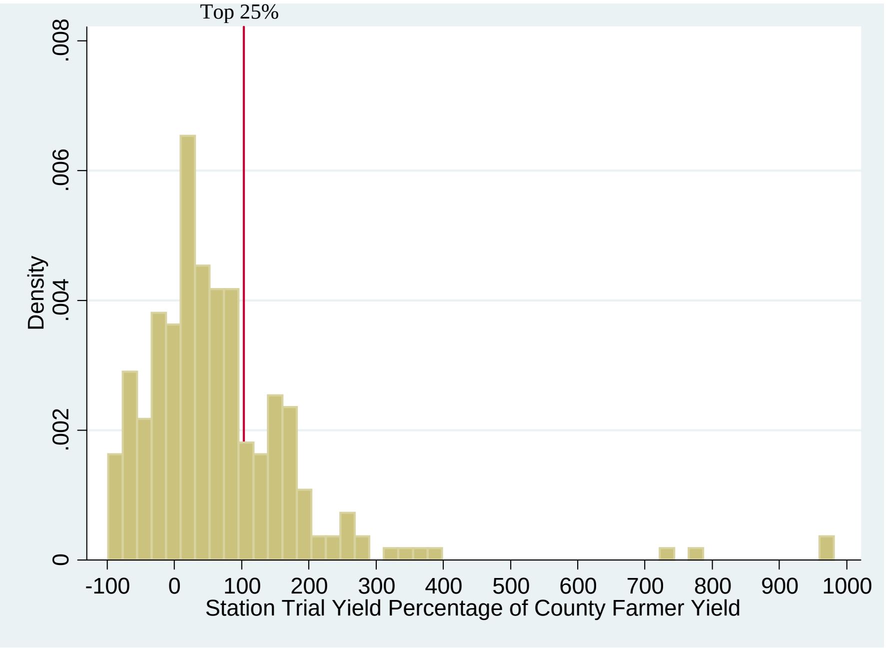
Notes: Authors' calculations with county data from 1870 to 2000 as described in the text. Variable definitions and sources as described in the data appendix. The unit of observation is a county-year. Each column reports the results from one regression, weighted by 1880 county land area for the sample indicated. The main entries in the first row of columns (1)-(4) report estimates of β_t from equation (8) in the text. Standard errors clustered at the county level are reported in parentheses. The excluded year interaction is 1880. The models in all columns include additional controls for year and county fixed effects, 1880 woodland fraction × year fixed effects, 1880 population × year fixed effects, and 1880 farm acre per county acre × year fixed effects. * indicates significance at the 10 percent level, ** significance at the 5 percent level and *** significance at the 1 percent level.

FIGURE A1: Station Distance and 1880 Crop Productivity - Agricultural Suitability Gap



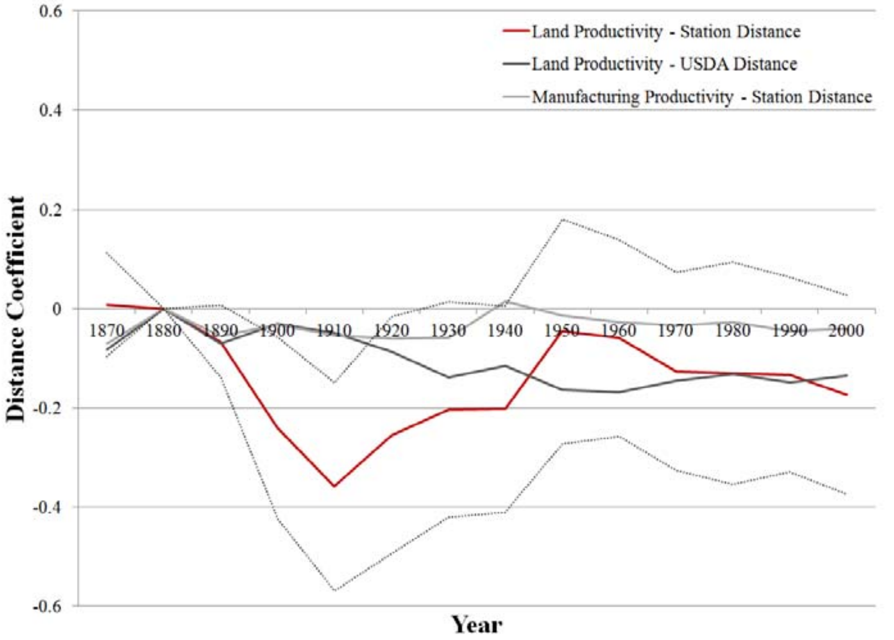
Source: Authors' calculations with 1880 agricultural census data. The height of each bar is the mean of the unexplained productivity in 1880 for the distance to experiment station decile . The value on the x- axis is the mean distance to the experiment station for the decile.

FIGURE A2: Applied Innovation Definition: Top 25th Percentile Station Trial-Farmer Percentage Yield Gap



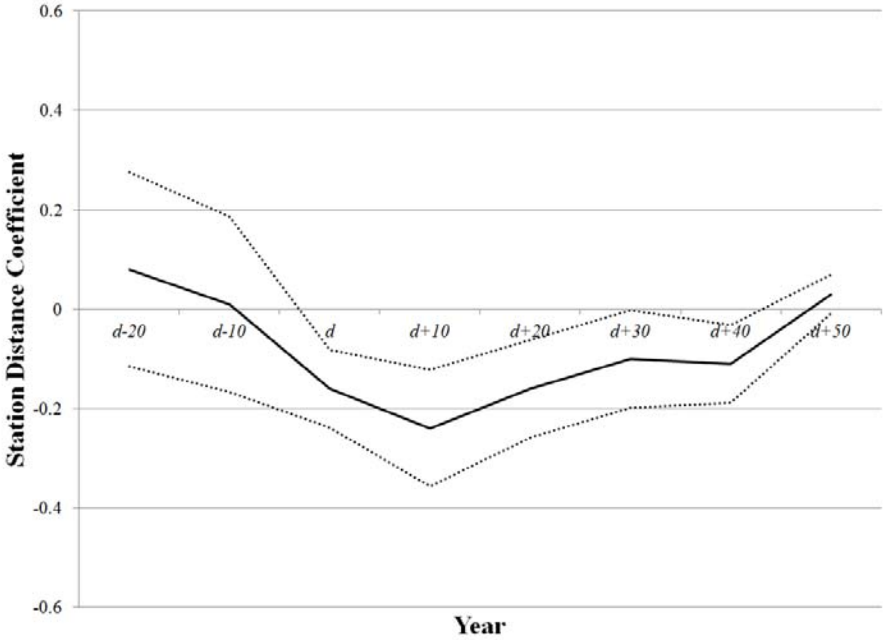
Notes: Source: Experiment station data reported in Table A2, and agricultural census yield data. Authors' calculations from experiment station reported crop yields from corn and wheat trials minus agricultural census data for the primary experiment station county in the same year from 1889, 1899, 1909, 1919, 1929, and 1939. Figure A2 drops the 4 observations with a station-farmer yield gap percentage greater than 2000.

FIGURE A3: Research Proximity and Productivity: Future Research Locations and Manufacturing



Notes: Authors' calculation from fitting equation (8).Each solid line plots the year-by-year distance interaction estimates from a different specification with crop revenue per acre as the outcome variable. Land Productivity - Station Distance plots estimates from Table 2 column (1). Land Productivity – USDA Distance plots estimates from Table 4 column (1). Manufacturing Productivity – Station Distance plots estimates from Table 4 column (4). The dotted lines depict 95% confidence intervals from the Land Productivity - Station Distance model in Table 2 column (1).

FIGURE A4: Applied Innovation Proximity and Land Productivity



Notes: Authors' calculation from fitting equation (7). Each solid line plots the year-by-year distance interaction estimates with crop revenue per acre as the outcome variable from Table 8 column (1). The dotted lines depict 95% confidence intervals from fitting the model.