AGRICULTURAL DIVERSITY, STRUCTURAL CHANGE AND LONG-RUN DEVELOPMENT:
EVIDENCE FROM THE U.S.

Martin Fiszbein

Agricultural Diversity, Structural Change and Long-run Development: Evidence from the U.S.
Martin Fiszbein
NBER Working Paper No. 23183
February 2017, Revised June 2017
JEL No. N11,N12,N51,O13,O14

## **ABSTRACT**

This paper examines the role of agricultural diversity in the process of development. Using data from U.S. counties and exploiting climate-induced variation in agricultural production patterns, I show that mid-19th century agricultural diversity had positive long-run effects on population density and income per capita. Examining the effects on development outcomes over time, I find that early agricultural diversity fostered structural change during the Second Industrial Revolution. Besides stimulating industrialization, agricultural diversity boosted manufacturing diversification, patent activity, and new labor skills, as well as knowledge- and skill-intensive industries. These results are consistent with the hypothesis that diversity spurs the acquisition of new ideas and new skills because of the presence of cross-sector spillovers and complementarities.

Martin Fiszbein
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and NBER
fiszbein@bu.edu

# 1   Introduction

At stages of development when agriculture employs a large share of the population, the characteristics of agricultural production can strongly influence the evolution of the economy. Economists have extensively studied the impact of agricultural productivity on development (Matsuyama, 1992; Restuccia et al., 2008; Gollin, 2010; Hornbeck and Keskin, 2015; Bustos et al., 2016). The effects of certain types of agricultural products have also attracted considerable interest (Goldin and Sokoloff, 1984; Engerman and Sokoloff, 2002; Nunn and Qian, 2011; Vollrath, 2011; Bruhn and Gallego, 2012). In contrast, the role of diversity (the variety and balance of the production mix) in agriculture has remained largely unexplored. This paper shows that agricultural diversity can be an important driver of structural change and economic growth.

To examine how agricultural diversity affected the process of development, I use data from U.S. counties over 140 years and propose a novel instrumental variable (IV) strategy. I find significant positive effects of mid-19th century agricultural diversity on development following the onset of the Second Industrial Revolution (usually dated 1870-1920). Early agricultural diversity—measured in 1860, when the American economy was still predominantly agricultural—did not have a significant impact on contemporaneous development across U.S. counties. But in the decades that followed, while the nation was becoming the world's industrial leader, county-level industrialization and population density were positively affected by early agricultural diversity. The effects were persistent and sizable: according to my IV estimates, a one standard deviation increase in 1860 agricultural diversity led to an increase of 73% in 2000 population density and a gain of 6% in 2000 per capita income.

In theory, production diversity may affect development in either direction, through various channels. Diversity can hinder development, as it may imply foregoing gains from specialization based on comparative advantage or scale economies. On the other hand, if there are complementarities among inputs or skills, diversity may yield productivity gains. Moreover, diverse economic environments may favor technological change and skill formation. Diversity could also foster development through other channels, e.g., by reducing risk and volatility. Empirically identifying the causal effects of diversity is challenging: while diversity may affect development in various ways, the latter may also affect the former, and there are third factors that may affect both. In this paper, I address the challenge exploiting climate-induced variation in agricultural production patterns, and assess the plausibility of different mechanisms drawing on a wealth of historical data.

I start with a preliminary exploration of the relationship between early agricultural diversity and development outcomes through ordinary least squares (OLS) regressions, and

find significant positive correlations that are robust to controlling for state fixed effects, land productivity measures and several other geo-climatic features, the dominance of cotton and other specific agricultural products, and a number of socio-economic initial conditions. But despite their robustness, the correlations could be driven by heterogeneity in unobservables (e.g., preferences or technology). For instance, if diversity in early agriculture partly reflected openness to new ideas and this trait accelerated subsequent growth, OLS estimates would be biased upwards. A downward bias generated by other omitted variables is also possible.

To identify the causal effects of agricultural diversity, I propose a strategy that exploits variation in the composition of agricultural production generated by climatic features. I construct an IV for agricultural diversity using high resolution spatial data on climate-based potential yields for many different crops (from FAO, the Food and Agriculture Organization of the United Nations). The IV is based on the estimation of a fractional multinomial logit (FML) model of crop choice, in which the county-level shares of agricultural products in agricultural output are functions of the crop-specific potential yields. With the predicted shares for each crop, I compute an index of potential agricultural diversity, which I then use as an IV for actual diversity. Controlling for geographic and climatic features that might have direct effects on development mitigates concerns about the exclusion restriction in the IV estimation. The findings are robust to adding higher-order terms of geo-climatic variables and to controlling for the importance of specific crops in a variety of ways.

After establishing the positive effects of agricultural diversity on development, I investigate the underlying mechanisms. My proposed explanation emphasizes that agricultural diversity can foster industrial diversification and the acquisition of new ideas and new skills. To motivate the hypothesis in the context of the Second Industrial Revolution, I offer a collection of examples that illustrate the ramifications of early agricultural diversity in U.S. history. Agricultural products required different skills and technologies in production, storage, packaging, marketing, and transportation. In addition, they had linkages with various manufacturing activities, which in turn required different technologies and skills, and had a variety of other linkages. Thus, directly and indirectly, agricultural diversity may have increased the local diversity of products, ideas and skills. In turn, this diversity may have favored technological change and skill formation because of the presence of complementarities and cross-sector spillovers.[1]

To assess this hypothesis, I estimate the impacts of agricultural diversity on industrial diversity, patent activity, and the share of manufacturing employment in new occupations. The results show that these key intermediate variables, measured in the decades following the

---

[1] In an appendix, I present a formal model of these mechanisms that emphasizes the interconnected roles of entry into new activities and the acquisition of ideas and skills in the process of economic growth.

onset of the Second Industrial Revolution, were all positively affected by early agricultural diversity. The proposed mechanisms have an additional testable implication: the effects of diversity ought to be larger in activities where skill complementarities and cross-sector knowledge spillovers are more relevant. Consistent with this prediction, I find that beyond its effect on overall industrialization, agricultural diversity positively affected the shares of manufacturing workers employed on knowledge- and skill-intensive industries.

There are other channels that could account for the positive effects of agricultural diversity on development: changes in agricultural productivity (which could in turn boost industrialization), reduced exposure to product-specific shocks (which would decrease volatility and possibly foster economic performance), or lower land concentration (which could affect local institutions, e.g., schools and banks). I assess the relevance of these channels and do not find evidence supporting any of them.

My findings contribute to a growing body of knowledge on the deep roots of comparative development (see Galor, 2011; Nunn, 2014; Spolaore and Wacziarg, 2013), offering novel insights on how the production structure at early stages of development can affect long-run performance. Like the recent contributions of Bleakley and Lin (2012), Hornbeck and Naidu (2014), and Glaeser et al. (2015), I find evidence from U.S. history indicating significant long-run effects of local conditions that are no longer directly relevant. Among the many contributions that study how geo-climatic factors have affected contemporary outcomes by shaping historical development paths, my research closely echoes those highlighting the role of early agricultural production patterns (e.g., Engerman and Sokoloff, 2002). But the focus on agricultural diversity, instead of agricultural productivity or specific crops, provides a new perspective on the role of agriculture in the development process.[2]

This paper also adds to the macro-development literature. My estimation of diversity's causal effects and the emphasis on skills formation and technological change as the underlying channels (along the lines of Hausmann and Hidalgo, 2011) complement the influential contributions of Acemoglu and Zilibotti (1997), Imbs and Wacziarg (2003), and Koren and Tenreyro (2013), which link diversity and development in different ways. Moreover, my results and interpretation have implications about the mechanics of structural change: they suggest that analyzing complementarities and cross-sector spillovers (e.g., along the lines of Jones, 2011 or Hanlon and Miscio, 2017) in models with finer levels of aggregation than standard two- or three-sector frameworks may produce novel insights.

---

[2]While many important contributions study diversity along other dimensions (e.g., ethnic, cultural, or genetic diversity), the closest precedents to my focus on agricultural diversity are Michalopoulos (2012) and Fenske (2014), who examine the effects of different measures of diversity in natural endowments on ethnolinguistic fragmentation and state centralization, respectively (both in the context of Africa).

Finally, my results are relevant for the urban and regional economics literature. The mechanisms that I emphasize echo the ideas of Jacobs (1969), which were quantitatively studied in the seminal contribution of Glaeser et al. (1992) and in numerous follow-up studies (see Beaudry and Schiffauerova, 2009, for a review). While many of these studies present suggestive evidence about the relationship between diversity and development, a causal interpretation is generally subject to endogeneity concerns. In this paper, the focus on agriculture offers a useful lever for identification, as it allows me to isolate climate-induced variation in diversity.

The paper is organized as follows. The next section discusses the theories that link diversity and development. The third section describes the data and historical background, and the fourth section presents the estimating equation and exploratory results from OLS estimation. Section five outlines the instrumental variable strategy and discusses the IV estimates. Section six investigates the mechanisms underlying the long-term effects of agricultural diversity. Section seven concludes.

## 2    Theories Linking Economic Diversity and Development

The relationship between economic diversity and development has been addressed by a number of theories with contrasting implications about the sign and direction of causality. This section briefly describes this multiplicity of theories and elaborates on those that generate positive effects of diversity because of the presence of complementarities and knowledge spillovers. This subset of theories informs my interpretation of how diversity in early agricultural production affected the process of development in U.S. history, as explained in detail throughout the paper.

Well-known pieces of economic wisdom suggest that specialization, as opposed to diversification, is good for growth. First, any form of increasing returns to scale operating only at the sector- or product-level implies that specialization yields productivity gains; within urban economics, for instance, the idea of localization externalities points to the benefits that firms derive from being co-located with other firms in the same sector. In addition, the principle of comparative advantage is often used to make a case for specialization. Finally, an elementary consideration of modern portfolio theory may suggest that growth is diminished by diversification insofar as (with the purpose of attenuating risk) it reduces expected returns.

In contrast, several theories imply positive effects of diversity. A number of different macro-development models feature complementarities (i.e., imperfect substitutability) among inputs in the production function—intermediate goods, skills, tasks, or technologies,

depending on the case—with the implication that diversity increases productivity (in the sense of lowering unit costs). Some models (e.g., endogenous growth models with expanding variety *à la* Romer (1990)) highlight the productivity gains from adding new production inputs. Others focus on the role of balance among required inputs—e.g., Jones (2011), who sheds light on the importance of "weak links" in the presence of strong complementarities.

There are also theories that generate *dynamic* advantages of diversity. Jacobs (1969) eloquently argued that diverse economic environments promote the acquisition of new ideas and skills. These effects may be explained by cross-product knowledge spillovers, particularly by recombinant technological progress—the production of new ideas by combining multiple old ideas. These forces were emphasized early on by Usher (1929), Schumpeter (1934), and Schmookler (1966). Recombinant innovation was first incorporated in a growth model by Weitzman (1998). Later contributions include Olsson (2000, 2005), Berliant and Fujita (2008, 2011), and Akcigit et al. (2013). Van den Bergh (2008) and Zeppini and Van den Bergh (2013) show that even if there are increasing returns that favor specialization, diversity's stimulus to recombinant innovations can make it efficient in the long-run.

Other models that generate dynamic advantages emphasize diversity in inputs and skills. Duranton and Puga (2001) build a model in which diversified cities, characterized by a broad range of intermediate inputs and labor skills, allow an entrepreneur with a new project to find the ideal productive process. In a model developed by Helsley and Strange (2002), the diversity of input suppliers reduces the cost of bringing new ideas to fruition. Hausmann and Hidalgo (2011) emphasize diversity in skills, or more broadly, capabilities: if each product requires a set of complementary capabilities, and capability sets for different products partially overlap, then diversity increases the return to acquiring new capabilities, insofar as they complement previously existing ones.

Various elements of the theories described above are closely related and can be jointly incorporated in a stylized model, as I do in Appendix C. The model features skill complementarities and cross-sectoral spillovers, and generates positive effects of diversity on both the acquisition of new ideas and the formation of new skills. Each sector requires a number of complementary skills. The expected level of efficiency for each skill positively depends—through spillovers—on the variety of established skills in other sectors of the local economy. Higher diversity reflects a wider range of skills, which increases expected returns and thus entry into new sectors. In turn, expanding the set of active industrial sectors goes along with the adoption of new technologies and the formation of new skills. This subsequently fosters entry into additional activities, boosting the process of growth and structural change.

Positive effects of diversity could also arise from reduced risk and volatility. Diversification limits the direct impact of negative product-specific shocks and also facilitates sub-

stitution away from negatively affected products. Moreover, if agents are risk averse, being able to enter a wide array of projects may be a precondition for carrying out risky projects with high returns. These insights are suggested by Acemoglu and Zilibotti (1997) and Koren and Tenreyro (2013), two important contributions characterizing the endogenous evolution of diversification and volatility in the process of development.

A positive relationship between diversity and development might also be driven, under some special conditions, by trade integration. While the standard theories predict that trade induces specialization and also increases income, Imbs et al. (2012) show that at low income levels and under indivisibilities in some sectors, enlarging market access can make minimum scales feasible and thus induce entry into new sectors, which would increase diversity and also income.[3]

Finally, diversity and development might be connected by an effect of the latter on the former: Li (2013) shows that higher income can lead to increased variety in consumption (which may in turn induce higher diversity in production).

In sum, there are multiple ways in which diversity can affect development, as well as ways in which development might affect diversity or a third variable might affect both. The theories reviewed in this section do not have a focus on *agricultural* diversity, but their insights remain relevant in the context studied in this paper. In 1860 agriculture was still the main sector of the U.S. economy. In addition, the mechanisms that I emphasize may have been activated by agricultural diversity directly, or indirectly, through industrial diversity induced by early agricultural diversity. The empirical analysis in section 5 and the model in Appendix C provide a precise account of my interpretation of agricultural diversity's effects on development in U.S. history.

## 3   Data and Historical Background

The empirical analysis is based on a sample of 1,821 U.S. counties that represented over 97% of national population in 1860. The sample excludes all Western states (most of which were still territories by 1860 and had only started to be partitioned into counties), as well as Dakota Territory, Kansas Territory, Nebraska Territory and Indian Territory (later Oklahoma).[4] To

---

[3]Imbs et al. (2012) put together this prediction and the usual one to show that trade integration occurring first between regions (at relatively low income levels) and subsequently between countries (at higher income levels) can generate a country-level inverted U shape of diversification in relation to income (Imbs et al., 2012, and the earlier contribution by Imbs and Wacziarg, 2003, empirically document this pattern with data from 1960 onward).

[4] See Figure A1 in Appendix A. Within the remaining states, 85 counties are dropped because of missing data, many of them in Michigan and Texas. The sample includes El Paso, Texas, which is about 400 miles

have consistent units of observation over time in spite of changes in county boundaries, I adjust all data to conform to 1860 boundaries; the procedure is explained in Appendix A.

To measure early agricultural diversity, I use data on production of 36 items comprising agricultural output from the 1860 Census of Agriculture, in combination with data on state-level prices from Atack and Bateman (1987) and Craig (1993) (both available from Minnesota Population Center, 2016). Table 1 reports the sample average and maximum percentages of county-level agricultural production for each of these products, as well as the percentages of counties in which each product was dominant and in which it represented over 50% of agricultural output (average shares are weighted by counties' agricultural output, so they coincide with shares in agricultural production for the sample as a whole).

TABLE 1. AGRICULTURAL PRODUCTION DATA

| Product | % of Agri.Output | | % of Counties | | Product | % of Agri.Output | | % of Counties | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Max | Dominant | >50% | | Mean | Max | Dominant | >50% |
| Corn | 23.80 | 98.89 | 43.73 | 11.28 | Barley | 0.44 | 10.54 | 0.00 | 0.00 |
| Ginned cotton | 16.03 | 94.11 | 20.08 | 13.86 | Clover seed | 0.30 | 12.93 | 0.00 | 0.00 |
| Animals slaughtered | 13.08 | 95.16 | 4.35 | 5.50 | Rice | 0.27 | 86.38 | 0.51 | 0.28 |
| Hay | 11.86 | 73.97 | 17.22 | 1.65 | Dew-rotted hemp | 0.26 | 64.86 | 0.55 | 0.55 |
| Wheat | 10.39 | 92.86 | 8.42 | 0.44 | Cane molasses | 0.25 | 19.58 | 0.00 | 0.00 |
| Butter | 4.22 | 25.36 | 0.66 | 0.00 | Sorghum molasses | 0.25 | 7.19 | 0.00 | 0.00 |
| Oats | 3.92 | 98.89 | 0.66 | 0.00 | Honey | 0.23 | 8.41 | 0.11 | 0.00 |
| Irish potatoes | 3.10 | 59.87 | 0.77 | 0.11 | Maple sugar | 0.22 | 36.26 | 0.00 | 0.00 |
| Tobacco | 2.34 | 57.27 | 2.59 | 1.65 | Hops | 0.17 | 19.51 | 0.00 | 0.00 |
| Sweet potatoes | 1.26 | 30.15 | 0.22 | 0.00 | Grass seed | 0.16 | 9.39 | 0.00 | 0.00 |
| Orchards | 1.18 | 24.18 | 0.00 | 0.00 | Hemp (other) | 0.09 | 22.04 | 0.00 | 0.00 |
| Cane Sugar | 1.15 | 88.43 | 0.77 | 0.66 | Maple molasses | 0.06 | 7.43 | 0.00 | 0.00 |
| Wool | 1.01 | 100.00 | 0.22 | 0.22 | Flax | 0.05 | 4.83 | 0.00 | 0.00 |
| Rye | 0.92 | 13.86 | 0.00 | 0.00 | Flaxseed | 0.04 | 3.35 | 0.00 | 0.00 |
| Market gardens | 0.88 | 94.36 | 0.66 | 0.11 | Beeswax | 0.02 | 1.03 | 0.00 | 0.00 |
| Peas and beans | 0.72 | 22.95 | 0.00 | 0.00 | Wine | 0.02 | 4.72 | 0.00 | 0.00 |
| Cheese | 0.70 | 35.70 | 0.22 | 0.00 | Water-rotted hemp | 0.02 | 7.17 | 0.00 | 0.00 |
| Buckwheat | 0.56 | 15.31 | 0.00 | 0.00 | Silk cocoon | 0.01 | 2.86 | 0.00 | 0.00 |

*Notes:* Based on county-level production data from the 1860 Census of Agriculture and state-level price data from Atack and Bateman (1987) and Craig (1993). The statistics for the 36 products included in agricultural production are for the sample of 1,821 counties considered in this paper. The first and second columns indicate the average and the maximum share of each product in agricultural output (average shares are weighted by county-level total agricultural production value, so that they coincide with shares in total agricultural production in the sample); the third and fourth column indicate, for each product, the percentages of counties in which it had the largest share in agricultural output, and in which it represented over 50% of agricultural output, respectively.

There were five products (corn, cotton, animals slaughtered, hay, wheat) representing shares of agricultural output above 10% each, but almost all 36 products represented sizable percentages of agricultural output in at least one county (for instance, rice represented less

---

away from the closest county in the sample; excluding it does not qualitatively change the results. For convenience, El Paso is omitted from the maps (Figures 1, 3, and 6).

than 0.3% in the sample but over 86% in Georgetown, South Carolina). This indicates that there was considerable variation in agricultural production patterns across counties, which allows me to identify the effects of diversity beyond the importance of specific crops.

Using these data, I calculate a measure of agricultural diversity as 1 minus a Hirschman-Herfindahl index of the shares of each product in gross agricultural output: $\text{Ag.Diversity}_c = 1 - \sum_i \theta_{ic}^2$, where $\theta_{ic}$ (with $i = 1, 2, ..., 36$) is the share of product $i$ in county $c$'s agricultural production (in value terms).[5] Figure 1A shows the spatial distribution of agricultural diversity in 1860: it was high in the Northeast and Midwest, but also in the Southeastern seaboard, along the Appalachian Mountains, and in Northern Texas. Beyond regional disparities, there were sizable differences across counties in each state, so I can identify the effects of agricultural diversity relying solely on within-state variation.

In 1860, the agricultural sector employed over 55% of the labor force and produced around 45% of total output in the U.S. economy (Weiss, 1992). Within my sample of 1,821 counties, over 80% of the population lived in rural areas. Three out of four of these counties did not have any urban population. Rural areas had some industrial production, but the bulk of manufacturing employment was located in major urban centers—in my sample, over 50% of total manufacturing employment in 1860 was concentrated in 25 counties.

Between 1860 and 1920, the share of the U.S. labor force employed in agriculture halved. The economy grew faster than ever before, both in absolute and per capita terms, making the U.S. the biggest and most productive economy in the world. The U.S. overtook the U.K. in terms of labor productivity in the 1890s, and subsequently widened the gap, as a result of a larger shift of labor away from agriculture (the sector with the lowest productivity) and faster productivity growth in non-agricultural activities (Broadberry and Irwin, 2006).

The average county-level share of population in manufacturing in my sample more than doubled its 1860 value by 1900, but there was wide variation across counties in the advance of the industrialization process. Small local economies had an important role in the takeoff of this period (Meyer, 1989; Page and Walker, 1991). While the importance of large industrial urban centers in manufacturing peaked by 1870 (their subsequent growth was led by services), some counties with little or no manufacturing production in 1860 became thriving mid-sized industrial localities by the early 20th century. Among the counties with zero manufacturing production in 1860 in the sample (about 11% of the 1,821 counties), one out of three had

---

[5] Appendix B.1 considers a number of alternative measures of agricultural diversity. The results of the paper are not driven by the specific functional form of the Hirschman-Herfindahl index. Moreover, the results are robust to using measures that address potential concerns about mismeasurement in the agricultural production data (e.g., in the value of animal products and cotton, for reasons specific to the 1860 data) or about the use of gross output data rather than value added data.

a share of population in manufacturing above the median value in 1900. Figure 1B shows the spatial distribution of this measure of industrialization; here again, besides the sharp disparity between the North (particularly the Northeast) and the South, there was significant variation within states.

The period circa 1870-1920, sometimes called the Second Industrial Revolution, was characterized by a series of massive transformations in technologies, spanning electricity, chemicals, steel, transportation, communications, management, and marketing (see Rosenberg, 1972; Chandler, 1977). Electric power and continuous-process methods, two salient innovations of the time, were associated with the origins of the complementarity between technology and skill that would mark the 20th century (Goldin and Katz, 1998). The share of high-skill workers in the labor force increased, and there was a rapid expansion of primary and secondary education (Goldin and Katz, 2009; Katz and Margo, 2014).

The technological advances of the Second Industrial Revolution would continue to fuel growth in the following decades (David, 1990; Gordon, 2016). Innovations diffused progressively across different activities and types of firms as the growing supply of capital goods embodying the new technologies increased their profitability, and as the acquisition of new skills and knowledge facilitated adoption. The effects on productivity were enhanced by subsequent incremental innovations based on learning-by-doing and by the introduction of new complementary technologies.

As measures of long-run development, I consider population density and income per capita in 2000 (in logs). Figures 1C and 1D show their spatial distribution. Again, the highest levels (both for population density and income per capita) are concentrated in the Northeastern coast, but there is also considerable within-state variation. Note that while income per capita would be the natural outcome of interest in a context without labor flows across localities, population density may be the more meaningful outcome in a context of relatively high labor mobility, where income differentials (unless compensated by other factors) induce population flows.

Historical Census data on manufacturing employment and population for 1860-1940 are drawn from Minnesota Population Center (2016) and for later periods from County Data Books (Haines and ICPSR, 2010). I use the share of the total population employed in the manufacturing sector as a measure of industrialization because labor force data is not consistently available before 1940.[6] Data on income per capita for 2000 come from the local

---

[6] For the years in which labor force data can be constructed using employment micro-data from Ruggles et al. (2010), the correlation between the share of the *labor force* in manufacturing and the share of *population* in manufacturing is above 0.9; the empirical analysis in the following sections yields similar results using either outcome for the years in which both are available.

area statistics of the Bureau of Economic Analysis (2015). Full details on variable definitions and sources are provided in Appendix A.

FIGURE 1. AG.DIVERSITY, INDUSTRIALIZATION, AND LONG-RUN DEVELOPMENT

A. AGRICULTURAL DIVERSITY, 1860          B. INDUSTRIALIZATION, 1900



C. LN POPULATION DENSITY, 2000          D. LN INCOME PER CAPITA, 2000
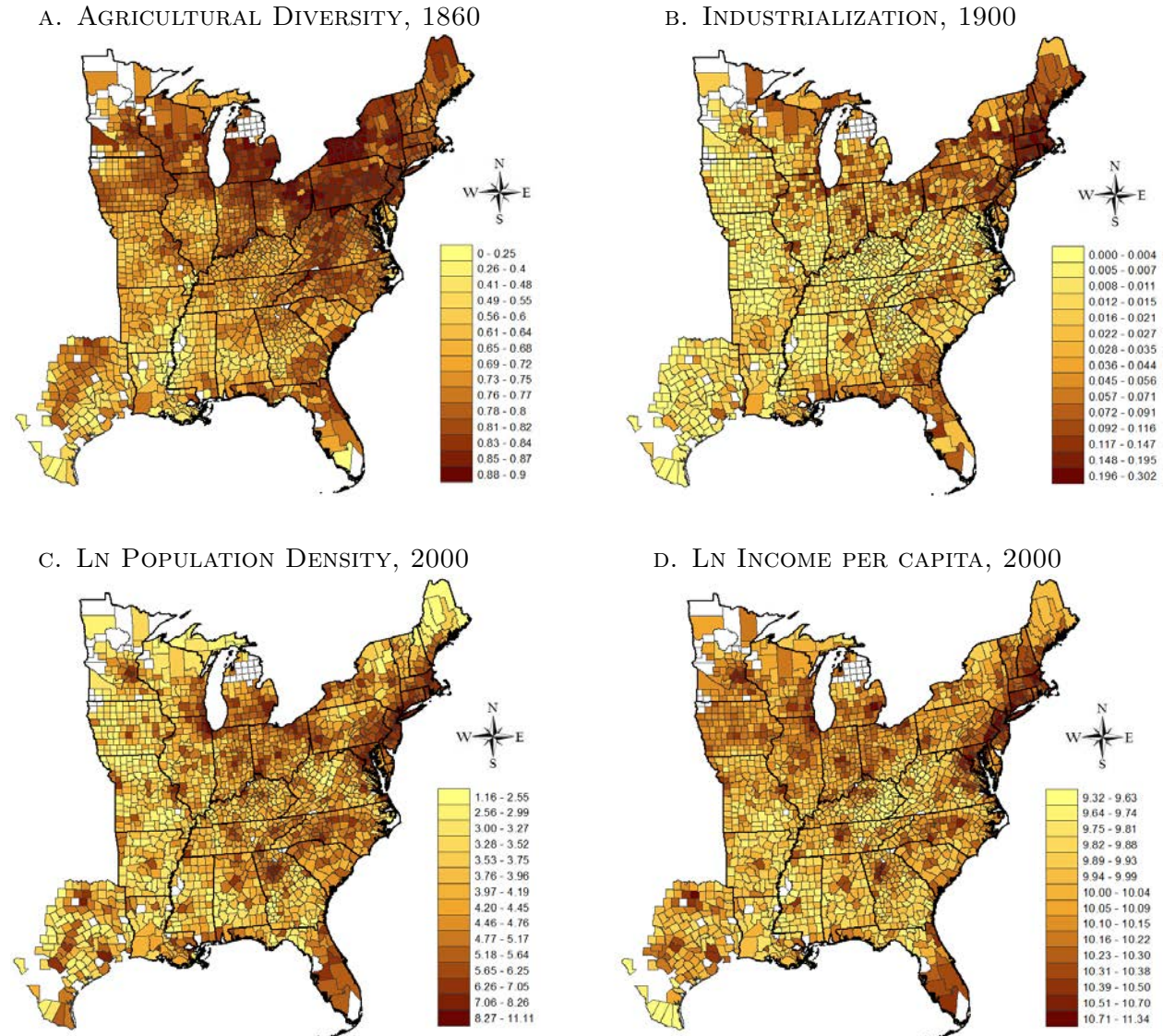


*Notes:* The maps display the spatial distribution of (A) agricultural diversity in 1860, (B) industrialization (the average number of workers employed in manufacturing throughout the year over total county population) in 1900, (C) the log of population density in 2000, and (D) the log of income per capita in 2000. See Appendix A.1 for variable definitions and sources.

## 4  Estimating Equation and OLS Results

This section presents the estimating equation and a preliminary exploration of the relationship between early agricultural diversity and development through OLS estimation. An

instrumental variable strategy to identify the causal effects of agricultural diversity is presented in the next section.

The estimating equation is

$$y_c = \alpha + \beta \text{Ag.Diversity}_{c,1860} + \delta_s + \gamma^{'} \mathbf{X}_c + \varepsilon_c \ ,$$

where $y_c$ is a development outcome for county $c$ (in the main regressions, income per capita, population density, or the share of population in manufacturing, at different points in time), $\delta_s$ is a state fixed effect, $\mathbf{X}_c$ is a vector of control variables, and $\varepsilon_c$ is an error term.

Throughout the paper, I consider different specifications that sequentially expand the set of controls to include state fixed effects, geo-climatic controls, crop-specific controls, and socio-economic initial conditions. I discuss these subsets of control variables in detail below.

Table 2 reports OLS estimates of the coefficient on 1860 agricultural diversity in regressions for three different outcome variables: the log of population density in 2000 (Panel A), the log of income per capita in 2000 (Panel B), and the share of the population employed in manufacturing in 1900 (Panel C). The first column displays the coefficient on agricultural diversity when this is the only regressor, while the second column reports results from regressions that also include state fixed effects. State fixed effects ensure that the estimated coefficient on agricultural diversity does not pick up the effects of other factors that display sharp regional differences like those displayed by agricultural diversity, e.g. between the North and the South. The estimated coefficients are considerably lower when only within-state variation is considered, but they are still significant, positive and sizable.

In columns (3)-(5), as I sequentially expand the set of controls, the coefficients on agricultural diversity remain significant and stable in magnitude; the estimates in these columns imply that an increase of one standard deviation in agricultural diversity (0.1234) is associated (without a causal interpretation) with increases of about 26-29 log points (30-34%) in 2000 population density, 3-5% in 2000 income per capita, and 4-5 percentage points in the share of population in manufacturing in 1900. Appendix B.1 shows that the results are robust to considering indexes of agricultural diversity with alternative functional forms and adjustments that address potential concerns about measurement error in the agricultural production data.

Geo-climatic controls include county area, various measures of potential agricultural productivity, mean annual temperature, rainfall, latitude and longitude, and distance to the ocean or the Great Lakes.[7] Including geo-climatic controls is important as they may be

---

[7]As measures of agricultural productivity I use the max and the mean of 22 product-specific climate-based (normalized) measures of potential yield provided by FAO-GAEZ. The data are explained in detail in section 5.1 and Appendix A. I normalize each product-specific potential yield by the maximum attained in the sample to make yields for different products comparable.

correlated with agricultural diversity and may also have direct effects on development. Agricultural productivity's impacts on development are the subject of a large literature (see Gollin, 2010), while proximity to waterways is a key determinant of access to markets, which in standard trade models increases both specialization and income levels. Appendix B.2 shows that the results are robust to controlling for a wider array of geo-climatic conditions (including additional land productivity measures, distance to the fall line, etc) as well as higher order terms.

Crop-specific controls capture the dominance of specific agricultural products in 1860. Since the diversity index is based on the squared values of individual shares, and some agricultural products (wheat, corn, hay, cotton, animals slaughtered) represented large shares of agricultural output, these shares may be strongly correlated with agricultural diversity. Thus, the estimated coefficient for diversity could pick up the positive or negative effects of being specialized in one of those particular products. To address this, I include dummies for each of the five major agricultural products that take a value of 1 when the product has the largest share in a county's agricultural production. I also include a dummy that takes a value of 1 when the combined share of plantation crops other than cotton (tobacco, sugarcane, and rice) is larger than the share of any other individual crop.[8] Appendix B.3 shows that the results are robust to controlling for specialization in particular crops in different ways.

Socio-economic controls (in all cases measured in 1860) include the urbanization rate; farm output per improved acre (in logs); the shares of the population corresponding to slaves, foreigners, people below 15 years of age and people above 65; distance to the nearest railroad; distance to steam-boat navigated rivers; and a measure of "market potential." Market potential in county $c$ is calculated, following the classic definition from Harris (1954), as $\sum_{k \neq c} d_{c,k}^{-1} N_k$, where $k$ is the index spanning other counties, $d_{c,k}$ is the distance between county $c$ and county $k$, and $N_k$ is the population of county $k$ (here, in 1860).[9]

Crop-specific controls and socio-economic controls (all of them measured in 1860) capture initial conditions that are predetermined with respect to the outcome variables, but they could be endogenous. They are not predetermined with respect to the regressor of interest, and thus may be "bad controls," as explained by Angrist and Pischke (2008). If these variables capture mechanisms through which diversity affects outcomes, including them as controls would mask the true effects. Therefore, I prefer the specification that includes state

---

[8] Specialization in plantation crops is considered to be a crucial determinant of the U.S. North-South divide (Engerman and Sokoloff, 1997, 2002), and it could also affect within-state variation in development.

[9]Donaldson and Hornbeck (2016) show that Harris' *ad hoc* measure is similar to a first-order approximation to market access derived from an Eaton-Kortum trade model; in the latter, neighboring populations are weighted by the inverse of trade costs elevated to the trade elasticity, for which they use a baseline value of 3.8. Measuring market potential as $\text{Market}_c = \sum_{k \neq c} d_{c,k}^{-3.8} N_k$ does not qualitatively affect my findings.

fixed effects and geo-climatic controls but not these initial conditions (Column (4) in Table 2). However, including them may be appropriate insofar as there can be correlations between them and agricultural diversity that do not reflect causality from the latter to the former. Thus, the robustness of the results to controlling for those variables may be considered reassuring.

In line with the approach suggested by Bester et al. (2011) to address spatial serial correlation, standard errors in all specifications are clustered on 60-square-mile grid squares that completely cover the counties in the sample. These standard errors are larger than Huber-White standard errors in all specifications, and the results are qualitatively similar to those obtained with clustering at the level of 1970 county groups or 1940 state economic areas or with Conley (1999) standard errors for various distance cutoffs.

TABLE 2. AGRICULTURAL DIVERSITY AND LONG-RUN DEVELOPMENT: OLS RESULTS

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| **Panel A.** *Dependent variable: Ln Population Density 2000* | | | | | |
| Ag.Diversity$_{1860}$ | 2.994*** | 2.020** | 2.370*** | 2.206*** | 2.137*** |
|  | (0.312) | (0.584) | (0.539) | (0.521) | (0.257) |
| $R^2$ | 0.095 | 0.277 | 0.354 | 0.372 | 0.612 |
| | | | | | |
| **Panel B.** *Dependent variable: Ln Income per capita 2000* | | | | | |
| Ag.Diversity$_{1860}$ | 0.547*** | 0.241** | 0.249*** | 0.221*** | 0.269*** |
|  | (0.0559) | (0.0899) | (0.0828) | (0.0816) | (0.0652) |
| $R^2$ | 0.102 | 0.297 | 0.361 | 0.373 | 0.478 |
| | | | | | |
| **Panel C.** *Dependent variable: Share of Population in Manufacturing 1900* | | | | | |
| Ag.Diversity$_{1860}$ | 0.0985*** | 0.0358** | 0.0404*** | 0.0334*** | 0.0333*** |
|  | (0.0108) | (0.00949) | (0.00919) | (0.00887) | (0.00825) |
| $R^2$ | 0.087 | 0.439 | 0.461 | 0.468 | 0.618 |
| | | | | | |
| State FE | N | Y | Y | Y | Y |
| Geo-climatic controls | N | N | Y | Y | Y |
| Crop-specific controls | N | N | N | Y | Y |
| Socio-economic conditions in 1860 | N | N | N | N | Y |
| Observations | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 |

*Notes:* See Appendix A.1 for variable definitions and sources. The means of the dependent variables in panels A, B, and C are 4.37, 10.06, and 0.034, respectively. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

To conclude this preliminary exploration, I examine the statistical association of 1860 agricultural diversity with population density and the share of population in manufacturing

at different points in time (I focus on these development outcomes as data are available all the way back to 1860). Figure 2 shows estimates of the coefficient on Ag.Diversity$_{1860}$, with the corresponding 95% confidence intervals, for my preferred specification. For both development outcomes, the estimated coefficients on 1860 agricultural diversity increase over the Second Industrial Revolution; for the share of population in manufacturing, the association becomes significant in the late 19th century, and for population density, in the early 20th century. Thereafter, the statistical associations remain positive and significant.[10]

FIGURE 2. EARLY AGRICULTURAL DIVERSITY AND DEVELOPMENT
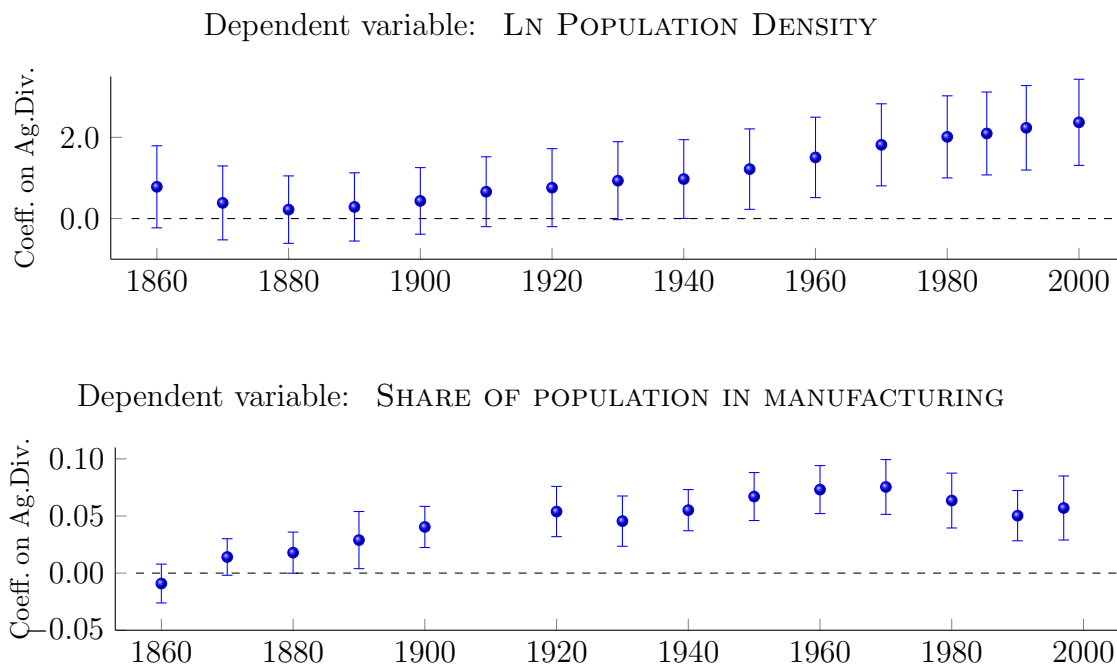
Dependent variable: LN POPULATION DENSITY



Dependent variable: SHARE OF POPULATION IN MANUFACTURING



*Notes:* The figure displays the estimated coefficients on 1860 agricultural diversity from OLS regressions with the log of population density and the share of population in manufacturing at different times as outcomes variables, controlling for state fixed effects and geo-climatic conditions. Also shown are the 95% confidence intervals based on robust standard errors clustered on 60-square-mile grid squares. See Appendix A.1 for variable definitions and sources.

## 5   Instrumental Variable Strategy and Results

The estimations presented in the previous section show suggestive correlations, but they do not establish a causal relationship. The positive association between agricultural diversity and development could be driven by omitted variables that induced higher diversity in 1860 and also improved subsequent economic performance. For example, diversity might reflect a high propensity to adopt new ideas, or higher levels of human capital, which would also

---

[10] I run a separate regression for each outcome in each year; the estimates are the same as those obtained from pooling observations from all time periods and letting all regressors have time-varying effects.

boost industrial development. Diversity in agricultural production may also reflect a more diversified local demand driven by higher income in 1860, which could in turn be positively correlated with long-run development.

On the other hand, if there are omitted variables that are negatively correlated with diversity and positively associated with subsequent economic growth, the OLS estimates would be negatively biased. For instance, diversity might partly reflect risk aversion, which could in turn discourage investments in manufacturing if this sector was perceived as highly risky. Diversity might also reflect the prevalence of traditional agriculture (whereby farmers grow most of what they need for their own subsistence), which may in turn be associated with poor economic outcomes. Market access (if not adequately captured by the controls) could also generate a negative bias, since larger potential gains from trade would induce both specialization and growth. Finally, measurement error in agricultural production may introduce attenuation bias in the OLS estimates.

With the aim of identifying the causal effects of agricultural diversity on economic development, this section introduces an instrumental variable strategy that exploits exogenous variation in agricultural production patterns generated by climatic features. I start by explaining the construction of an instrumental variable for agricultural diversity, and then present the IV estimation results.

## 5.1  Potential Agricultural Diversity

The proposed identification strategy relies on the basic insight that agricultural production patterns are partly induced by climatic features. In particular, agricultural diversity is influenced by the dispersion of product-specific potential yields determined by climate. Intuitively, a county that has similar levels of productivity for many different crops is likely to be more diversified than a county with productivity for one crop much higher than for all others.

The FAO's Global Agro-Ecological Zones (GAEZ) project provides measures of maximum attainable yields (in tons per hectare per year) for different crops based on high spatial resolution climatic data and crop-specific characteristics. These measures of crop-specific potential productivity are based on expert knowledge of climatic features affecting agricultural production processes; they do not rely on a statistical analysis of production patterns observed across the world. In addition, though based on climatic records for 1961-1990, they provide good proxies for historical conditions (see Nunn and Qian, 2011). I use attainable yields for rain-fed conditions and intermediate levels of inputs/technology, as these correspond most closely to the context under consideration. Figure 3 displays the county-level average values of potential yields for selected crops (wheat, corn, potato, and tobacco).

## Figure 3. Potential Yields for Selected Crops

Potential Wheat Yields

Potential Corn Yields

Potential Potato Yields

Potential Tobacco Yields
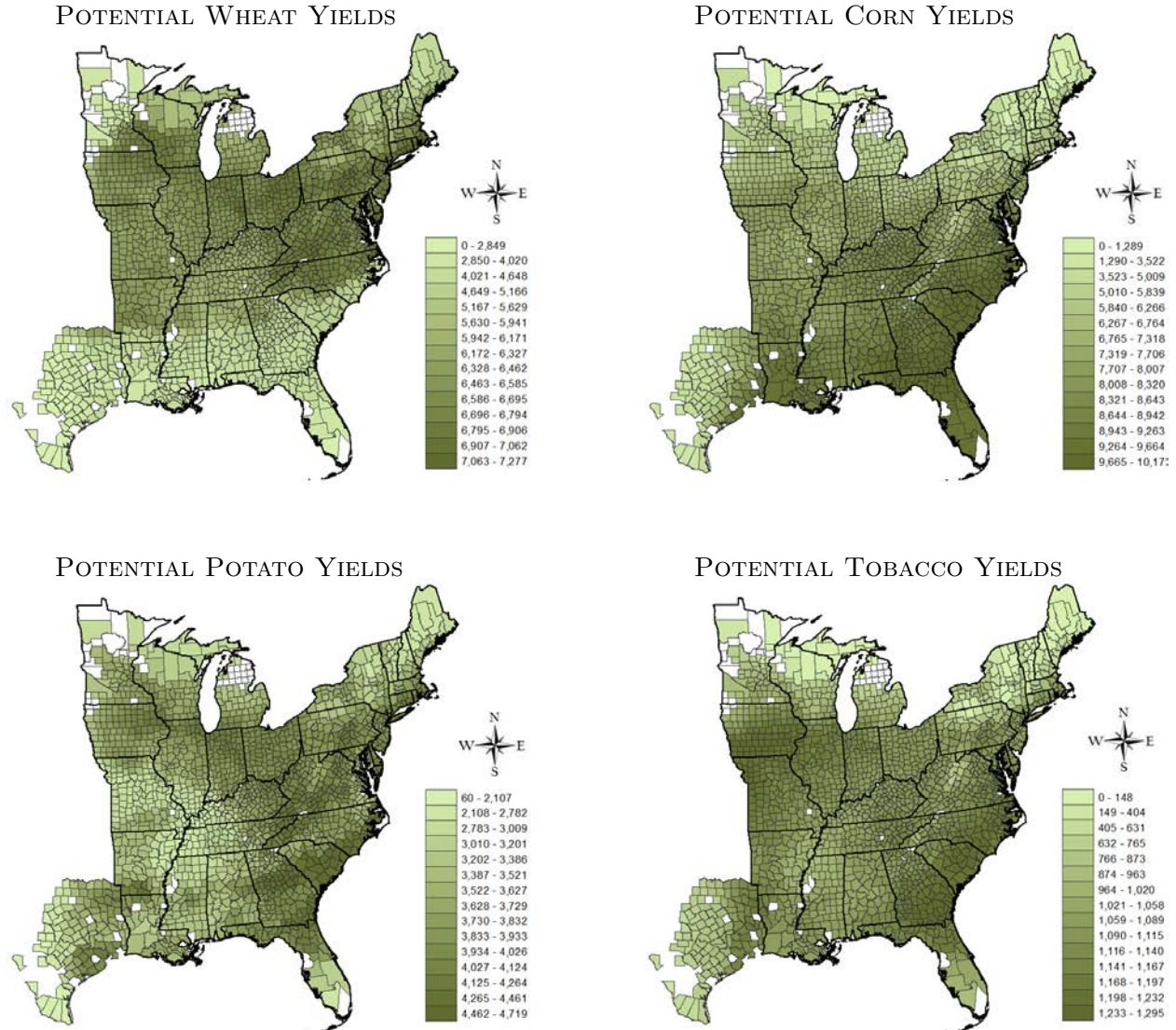


*Notes:* The maps displays county-level means of agro-climatic attainable yields from IIASA/FAO (2012) for wheat, corn, potato, and tobacco, in tons per hectare per year, for rain-fed conditions and intermediate levels of inputs/technology. See Appendix A.1 for variable definitions and sources.

To construct an instrumental variable based on crop-specific attainable yields from FAO-GAEZ, I use a fractional multinomial logit (FML) framework (Sivakumar and Bhat, 2002), which generalizes the fractional logit model (Papke and Wooldridge, 1996) to an arbitrary number of choices (for recent discussions, see Ramalho et al., 2011; Mullahy, 2015).

In the context under consideration, the FML model is specified as a system of equations in which the outcome variables are the shares of each agricultural product $i$ in total agricultural output in county $c$ (i.e., $\theta_{ic}$, for $i = 1, 2, ..., 36$) and the regressors are the crop-specific

16

potential yields $\mathbf{A}_c$.[11]

The functional form of the FML model is

$$\hat{\theta}_{ic} = E[\theta_{ic}|\mathbf{A}_c] = \frac{e^{\beta'_i \mathbf{A}_c}}{1 + \sum_{j=1}^{I-1} e^{\beta'_j \mathbf{A}_c}} \quad .$$

By construction, $\sum_{i=1}^{I} \hat{\theta}_{ic} = 1$ , i.e. the predicted shares for each county add up to 1. The parameters are estimated by quasi-maximum-likelihood.

This econometric framework can be motivated by a simple model of optimal crop choice. Recasting the conditional logit framework of choice behavior by McFadden (1974) in terms of profit maximization rather than utility maximization, assume that profits obtained when choosing crop $i$ for a unit of farm resources $k$ are $\pi_{ik} = \beta'_i \mathbf{A}_k + \mu_{ik}$. Farmers are assumed to be price-takers. The estimated parameters reflect the price and cost differentials among agricultural products, as well as any other factors that affect profits for different crops. If the error term $\mu_{ik}$ is assumed to be *iid* with type I extreme value distribution, then choice $i$ is optimal (i.e. $\pi_{ik} \geq \pi_{i'k}$ for all $i'$) with probability $\frac{e^{\beta'_i \mathbf{A}_k}}{1 + \sum_{j=1}^{I-1} e^{\beta'_j \mathbf{A}_k}}$.
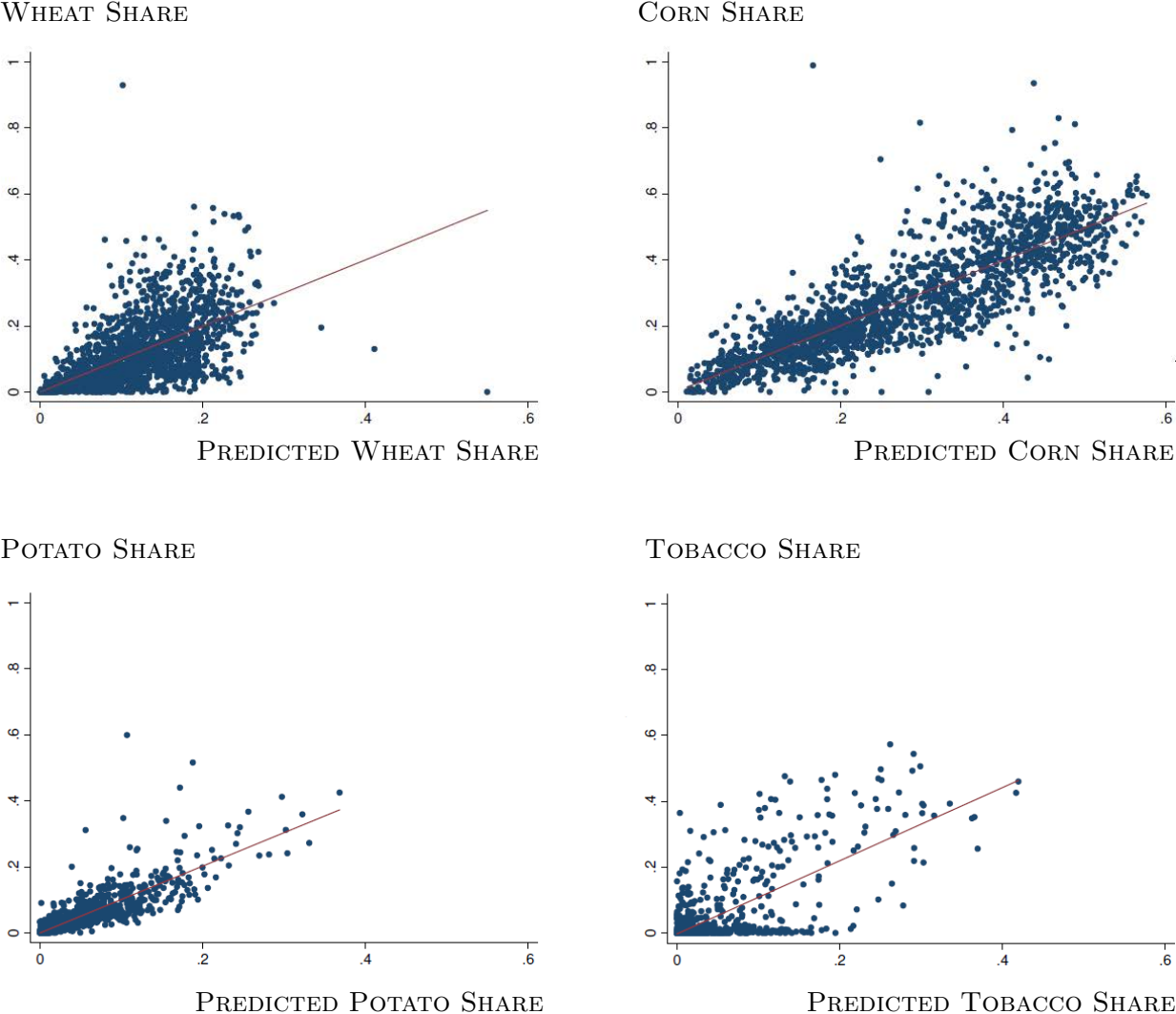
Once the coefficients of the FML have been estimated, they are used (in combination with county-level product-specific productivities) to calculate predicted shares for each agricultural product. In all cases, the predicted shares account for a large fraction of the variation in actual shares. Regressing actual on predicted shares for each crop, the coefficients on the latter are always positive and highly significant, even when controlling for state fixed effects and geo-climatic conditions, and the partial $R^2$'s are large (in most cases, higher than 0.3). Figure 4 plots actual against predicted shares for selected agricultural products.

Using the predicted shares from the FML model for all agricultural products, I calculate an index of potential agricultural diversity. This index is defined by the same formula used for actual diversity but substituting predicted shares for actual shares, that is, Potential Ag.Diversity$_c = 1 - \sum_i \hat{\theta}_{ic}^2$. As the predicted shares are good predictors of the actual shares, the measure of potential diversity is a good predictor of actual diversity. Figure 5 shows the scatter plot of the latter against the former. Figure 6 shows the spatial distributions

---

[11] In the estimation of the model, the vector of potential yields includes all relevant crop-specific productivities available from FAO: alfalfa, barley, buckwheat, cane sugar, carrot, cabbage, cotton, flax, maize, oats, onion, pasture grasses, pasture legumes, potato, pulses, rice, rye, sorghum, sweet potato, tobacco, tomato, and wheat. Note that for some of the crops comprised in the agricultural production of U.S. counties, the FAO does not provide measures of attainable yields. The results are very similar if I include measures of overall land productivity in $\mathbf{A}_c$ to serve as proxies for these crops' potential yields. Even without overall productivity measures, the model can provide good predictions for the shares of these crops insofar as their potential yields are correlated with some of the available ones and/or they display production complementarities with specific crops.

of potential agricultural diversity and its values after partialling out state fixed effects and geo-climatic controls, to give a sense of the variation that is used to identify the causal effects of agricultural diversity in the IV estimation with my preferred specification.

FIGURE 4. ACTUAL AND PREDICTED SHARES OF SELECTED CROPS

WHEAT SHARE

CORN SHARE



PREDICTED WHEAT SHARE

PREDICTED CORN SHARE

POTATO SHARE

TOBACCO SHARE

PREDICTED POTATO SHARE

PREDICTED TOBACCO SHARE

*Notes:* The figure displays scatter plots of the actual and predicted shares of wheat, corn, potato, and tobacco in agricultural production (the predicted shares are obtained from the FML model).

## 5.2 The Effects of Agricultural Diversity on Development

I estimate the causal effect of agricultural diversity using the index of potential agricultural diversity as an IV. Table 3 reports the results of the IV estimation using the same specifications considered in the OLS estimation, for the same development outcomes: the log of

income per capita in 2000 (Panel A), the log of population density in 2000 (Panel B), and the share of population in manufacturing in 1900 (Panel C). Panel D displays the results from the first stage regressions (with Kleibergen-Paap F-statistics to assess the possibility of weak instruments).

FIGURE 5. ACTUAL AND POTENTIAL AGRICULTURAL DIVERSITY

AG.DIVERSITY



POTENTIAL AG.DIVERSITY
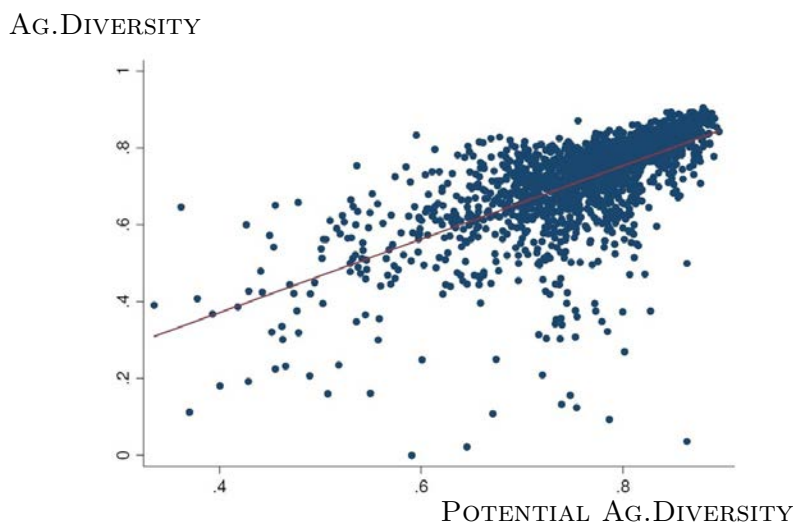
*Notes:* The figure displays a scatter plot of actual and potential agricultural diversity (the latter is calculated with the predicted shares from the FML model).

The identifying assumption is that the measure of potential agricultural diversity, based on the estimation of the FML model, only affects development outcomes through actual agricultural diversity. To alleviate possible concerns about the validity of the exclusion restriction, recall that the FAO measures of potential yields for different crops do not rely on a statistical analysis of observed production patterns. Moreover, note that any determinants of crop choice other than the climate-based productivity measures have their effects loaded onto the residuals of the FML model, and thus do not affect the IV estimates of the effects of agricultural diversity. Still, potential agricultural diversity might be correlated with geo-climatic features or socio-economic conditions with direct effects on development outcomes, which would bias the IV estimates. Mitigating this concern, the results are robust to controlling for a wide array of geo-climatic features (further expanded in Appendix B2 to include flexible polynomials of land productivity measures, latitude and longitude, etc) as well as a number of socio-economic conditions.

For all specifications, the IV estimates indicate significant positive effects of 1860 agricultural diversity on all outcomes, and the IV has strong predictive power. According to the results from my preferred specification (including state fixed effects and geo-climatic

controls), reported in column (3), increasing agricultural diversity by one standard deviation (0.1234) led to an increase of 0.46 standard deviations in the log of 2000 population density, which amounts to 55 log points, or 73% in density levels. By comparison, Bleakley and Lin (2012) estimate that being close to a historical portage site increased county-level population in 2000 by about 77-94 log points (depending on the specification), and Michaels (2011) estimates that oil abundance in the U.S. South led to about 42-102 log points higher 1990 population. Regarding the other outcomes, an increase of one standard deviation in agricultural diversity led to increases of 0.29 standard deviations in the log of 2000 income per capita (a 6% increase in its level), and 0.25 standard deviations (1 pp.) in the share of population in manufacturing in 1900.

FIGURE 6. POTENTIAL AGRICULTURAL DIVERSITY

POTENTIAL AG.DIVERSITY

POTENTIAL AG.DIVERSITY RESIDUALS
(PARTIALLING OUT STATE F.E. AND GEOCLIMATIC CONTROLS)
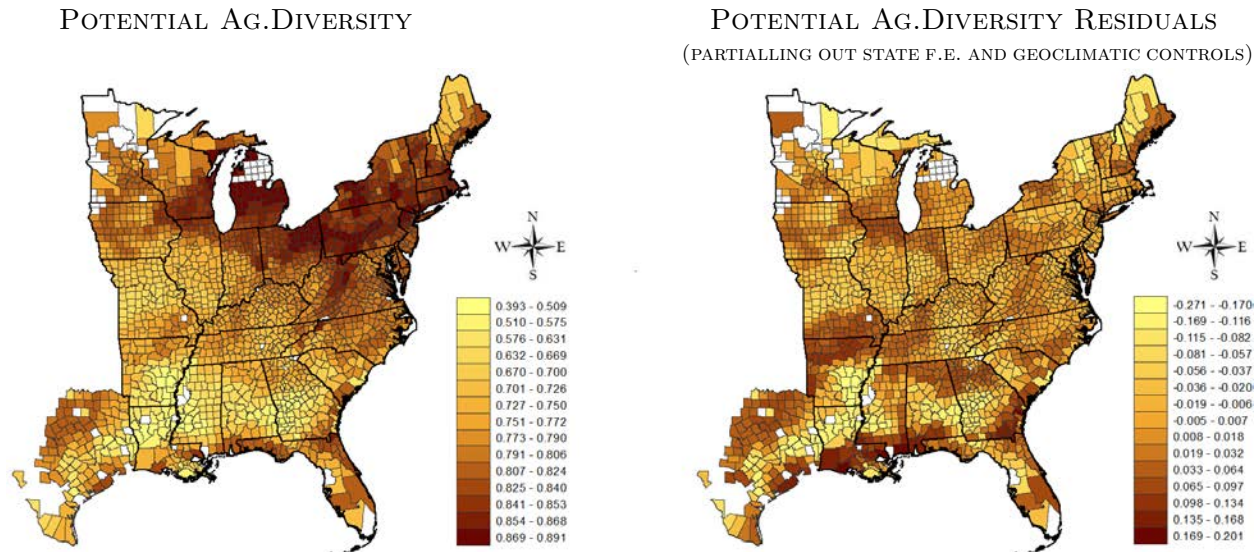


*Notes:* The maps display the spatial distribution of potential agricultural diversity (calculated with the predicted shares from the FML model) and its values after partialling out state fixed effects and geo-climatic controls.

The observed effects of agricultural diversity on long-run income per capita may be smaller, or larger, than what they would be if labor was immobile across counties. Without labor flows, productivity gains translate into differences in income per capita, with no effects on population density (except through fertility). In contrast, with relatively high labor mobility, like in the context of U.S. counties, income differentials induce labor flows, and thus productivity gains arising from early agricultural diversity would induce differences in population density. Differentials in income can only remain in equilibrium insofar as they are compensated by differentials in living costs and amenities. As people move to

places with higher productivity and initially higher income per capita, congestion pushes up living costs, and may also erode some of the initial productivity differential (if there are decreasing returns of some sort) or augment it (through agglomeration forces). In any case, the expected effects of productivity gains on income per capita and population density go in the same direction, but it is important to note the the magnitude of the observed effects on income and population do not reflect only the direct impacts but also the ensuing movements toward spatial equilibrium.

TABLE 3. EFFECTS OF AGRICULTURAL DIVERSITY ON DEVELOPMENT: IV ESTIMATES

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| **Panel A.** *Dependent variable: Ln Population Density 2000* | | | | | |
| Ag.Diversity$_{1860}$ | 4.539*** | 3.641*** | 4.477*** | 4.276*** | 4.463*** |
|  | (0.619) | (0.973) | (1.045) | (1.190) | (1.142) |
| $R^2$ | 0.069 | 0.261 | 0.329 | 0.349 | 0.589 |
| **Panel B.** *Dependent variable: Ln Income per capita 2000* | | | | | |
| Ag.Diversity$_{1860}$ | 0.915*** | 0.449*** | 0.491*** | 0.415** | 0.492** |
|  | (0.0982) | (0.160) | (0.160) | (0.187) | (0.204) |
| $R^2$ | 0.056 | 0.289 | 0.350 | 0.367 | 0.474 |
| **Panel C.** *Dependent variable: Share of Population in Manufacturing 1900* | | | | | |
| Ag.Diversity$_{1860}$ | 0.169*** | 0.0719*** | 0.0830*** | 0.0639** | 0.0884*** |
|  | (0.0220) | (0.0219) | (0.0213) | (0.0251) | (0.0309) |
| $R^2$ | 0.042 | 0.433 | 0.452 | 0.464 | 0.607 |
| **Panel D.** First Stage *Dependent variable: Ag.Diversity$_{1860}$* | | | | | |
| Potential Ag.Diversity | 0.937*** | 0.624*** | 0.561*** | 0.459*** | 0.346*** |
|  | (0.0517) | (0.0590) | (0.0632 ) | (0.0676) | (0.0604) |
| Kleibergen-Paap Wald rk F-stat | 384.232 | 103.428 | 86.529 | 51.029 | 43.057 |
| State FE | N | Y | Y | Y | Y |
| Geo-climatic controls | N | N | Y | Y | Y |
| Crop-specific controls | N | N | N | Y | Y |
| Socio-economic conditions in 1860 | N | N | N | N | Y |
| Observations | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 |

*Notes:* See Appendix A.1 for variable definitions and sources. The means of the dependent variables in panels A, B, and C are 4.37, 10.06, and 0.034, respectively. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

I continue to follow the approach suggested by Bester et al. (2011) to address spatial serial correlation, clustering standard errors on 60-square-mile grid squares that completely cover the counties in the sample. A sufficient condition for standard errors to be correct when using a generated instrumental variable (here, potential agricultural diversity) is that the expectation of the error term in the estimating equation conditional on the variables used in the IV construction (here, the crop-specific potential yields) is equal to zero (see Wooldridge, 2010). This sufficient condition is satisfied insofar as the estimating equation adequately controls for measures of overall land productivity and climatic variables that may have direct effects on development outcomes (Appendix B.2 shows that the results are robust to flexibly controlling for an expanded set of land productivity measures and geo-climatic conditions in a number of different specifications).

The IV estimates are larger than the OLS estimates. This difference may be due to a negative bias in the OLS estimation generated by omitted variables that display correlations of opposite signs with 1860 agricultural diversity and with subsequent economic performance (e.g., risk aversion, the prevalence of traditional agriculture, or market access, as discussed above). It could also be due to attenuation bias generated by measurement error.

I also examine the effects of early agricultural diversity on population density and the share of population in manufacturing at different points in time. Figure 7 shows estimates of these effects (with the corresponding 95% confidence intervals) for my preferred specification.

FIGURE 7. EFFECTS OF AGRICULTURAL DIVERSITY ON DEVELOPMENT OUTCOMES
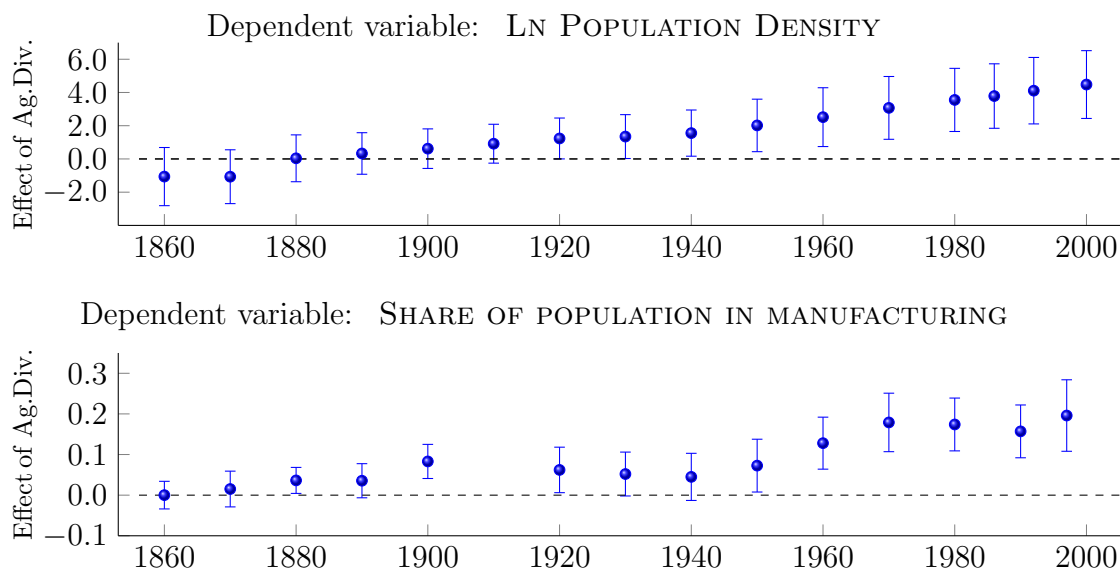


*Notes:* The figure displays the estimated coefficients on 1860 agricultural diversity from IV regressions with the log of population density and the share of population in manufacturing at different times as outcomes variables, controlling for state fixed effects and geo-climatic conditions. Also shown are the 95% confidence intervals based on robust standard errors clustered on 60-square-mile grid squares. See Appendix A.1 for variable definitions and sources.

The results indicate that early agricultural diversity had positive effects on development outcomes that emerged in the early 20th century, during the Second Industrial Revolution.[12] The next section examines the mechanisms underlying these effects.

# 6 Channels: Diversity, Ideas, and Skills in the Second Industrial Revolution

Having established the significant positive effects of agricultural diversity on long-run development, I proceed to investigate the underlying mechanisms. This section puts forth the hypothesis that these effects are explained by the presence of complementarities and cross-sector spillovers, and empirically shows that early agricultural diversity fostered industrial diversification, technological change, and formation of new skills during the Second Industrial Revolution.

A number of theoretical contributions, summarized in section 2, provide relevant insights for my proposed interpretation. If there are complementarities among production inputs, diversity can reduce unit costs. From a dynamic perspective, variety in skills may spur the formation of new skills, insofar as they complement previously existing ones. Moreover, in the presence of cross-sector knowledge spillovers and combinatorial learning (the production of new ideas through combinations of old ideas), diversity can foster technological change. In the context of comparative development across U.S. counties, the relevance of these forces requires significantly high costs of moving goods, skills or ideas even at a small spatial scale.

The proposed mechanisms can be captured in a formal model, as I do in Appendix C. Featuring a combination of the ideas mentioned above, the model highlights that the outcomes considered in this section's empirical analysis—industrial diversification, technological change, and skills formation—are interconnected dimensions of the process of structural transformation. With a stylized representation of complementarities and cross-sector spillovers, I show that diversity in skills (initially given by agricultural diversity) favors entry into new activities; this brings new ideas and new skills, which fosters subsequent diversification, boosting the process of growth and structural change.

Highlighting an important aspect of my interpretation, the model portrays growth as driven by structural change, not only from agriculture to manufacturing, but also within manufacturing—it is through entry into new industrial activities that the local economy acquires new skills. In addition, the model starkly states a key assumption underlying my explanation of the effects of agricultural diversity across U.S. counties and their long-run

---

[12] Income per capita data are not available before the mid-20th century. Estimated effects on (log) income per capita in 1960, 1970, 1980, and 1990 are positive, significant, and similar in magnitude to those for 2000.

persistence: skills and spillovers are highly localized. Furthermore, the model generates an additional testable implication: diversity, beyond increasing industrialization and industrial diversification, favors sectors that require more skills. I examine this prediction in section 6.3.

My study of channels focuses on the 1860-1940 period, encompassing the Second Industrial Revolution, over which the effects of early agricultural diversity emerged. The technological transformations of these decades were accompanied by rapid changes in the industrial composition and the occupational structure. There was widespread entry into new industrial activities and rapid growth of skill-intensive sectors. Meanwhile, there was a shift away from traditional occupations and a diversification of labor skills that included the emergence of many new occupations.[13]

A collection of historical examples throughout this section illustrates the variety of skills and technologies required by different agricultural products, their multiple input-output linkages with industrial activities, and the cross-sector spillovers and complementarities underlying the positive effects of diversity on technological change and skill formation. Agricultural diversity may have increased the local diversity of products, ideas and skills directly, or indirectly, by favoring entry into various agro-processing industries. A neat example of recombinant innovation comes from the history of Ford Motors' assembly line, which integrated insights about continuous flow production from several agro-processing industries.

To empirically examine my explanation of agricultural diversity's long-run effects, I consider a set of key intermediate variables capturing the channels mentioned above. Section 6.1 shows that early agricultural diversity favored subsequent diversification across manufacturing activities (Appendix B.4 explores the relevance of input-output linkages). In section 6.2, I present evidence indicating positive effects of agricultural diversity on patent activity and the share of industrial workers with new skills. Finally, Section 6.3 shows that the shares of industrial workers employed in skill- and knowledge-intensive activities were also positively affected by early agricultural diversity, lending further support to the proposed interpretation.

There are other channels that may explain the effects of agricultural diversity on development. First, agricultural diversity could be associated with agricultural productivity, which could in turn affect the industrialization process. Second, in the presence of product-specific

---

[13]Between 1880 and 1940, the county-level average number of active manufacturing sectors within my sample increased from 22.3 to 37.9, while the county-level average number of existing occupations in manufacturing increased from 22.8 to 43.3. Between 1860 and 1940, the share of carpenters, blacksmiths and shoemakers—the thee most prominent traditional occupations—in national employment across all craftsmen occupations declined from over 44% to under 15%.

shocks, higher diversity in agriculture would reduce volatility in the value of farm production, which may improve economic performance. Third, agricultural diversity may be inversely related to land concentration, which could in turn affect local institutions (e.g., schools and banks) through political economy mechanisms. My empirical assessment of these channels (briefly summarized in section 6.4, and presented in detail in Appendix D), does not support the relevance any of them.

## 6.1  Diversity, from Agriculture to Manufacturing

Given the presence of multiple linkages between different agricultural products and non-agricultural products, diversity in agriculture may induce diversity in manufacturing. As Hirschman (1981) remarked, "development is essentially the record of how one thing leads to another, and the linkages are that record." Loosely speaking, a wider range of (agricultural) things at early stages of development could lead to a more diverse set of (industrial) things later on.

Input-output linkages between agricultural products and industrial sectors in the late 19th century U.S. are documented by Kuhlmann (1944) and Page and Walker (1991). Textile production used cotton and wool. The manufacturing of tobacco products was a sizable industrial activity. Flour milling, the nation's largest industrial sector in production value between 1850 and 1880, was the main source of demand for wheat. Distilling industries used rye, barley, and corn to make spirits. Breweries made beer using hops and barley, sometimes other grains too. The growing meatpacking industry and leather industries were supplied by cattle raisers.

In turn, agro-processing industries had multiple forward and backward linkages. Grain processing supplied bakeries, confectionery establishments, and other food industries. Meat-packing establishments produced not only meat products but also leather, lard, candle wax, glue, and fertilizer, which were used in other activities. In addition, different agricultural products and agro-processing had various linkages with producers of tools and machines, packaging supplies, storage equipment and transportation equipment.

The presence of various agricultural products may favor entry into various manufacturing sectors, which could in turn favor entry into other manufacturing sectors. Insofar as linkages—including not only input-output relationships but also those generated by common labor skills or knowledge spillovers—affect the local structure of production, early agricultural diversity may be conducive to subsequent manufacturing diversity.[14] Appendix B.4 shows that a number of industrial activities with known input-output connections with spe-

---

[14] Using recent U.S. data, Ellison et al. (2010) show that input-output relationships, common labor skills, and knowledge spillovers are all important sources of industry co-location.

cific agricultural products were more likely to be present the larger the importance of those agricultural products at the local level.

To assess whether early agricultural diversity boosted diversification across industrial activities, I use Census employment microdata to construct a measure of diversity across 59 manufacturing sectors (defined in the 1950 classification system of the Census Bureau). For each county $c$ and time period $t$, I compute the share of each sector $s$ in total industrial employment, $\vartheta_{s,c,t}$, and then use the standard index to measure manufacturing diversity, $1 - \sum_s \vartheta_{s,c,t}^2$.[15]

Figure 8 displays IV estimates of the effects of early agricultural diversity on manufacturing diversity at different points in time between 1860 and 1940, using my preferred regression specification (controlling for state fixed effects and geo-climatic conditions).[16] The time pattern shows how the effects of early agricultural diversity on manufacturing diversity emerged over the course of the Second Industrial Revolution. An increase of one standard deviation in 1860 agricultural diversity led to an increase of 0.48 standard deviations in manufacturing diversity in 1920, when the magnitude of the effects reached a peak.

FIGURE 8. EFFECTS OF AG.DIVERSITY ON MANUFACTURING DIVERSITY
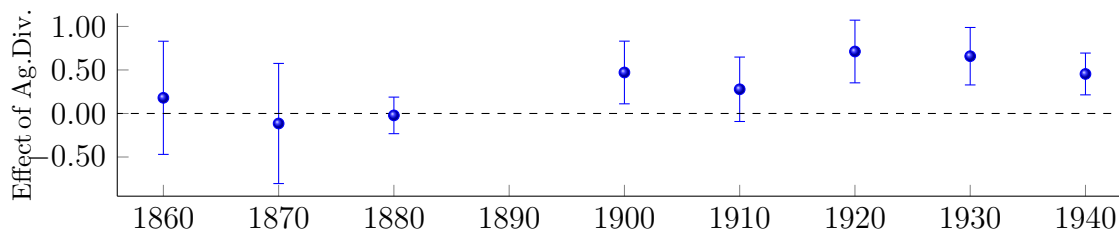


*Notes:* The figure displays the estimated coefficients on 1860 agricultural diversity from IV regressions with manufacturing diversity at different times as the outcome variable, controlling for state fixed effects and geo-climatic conditions. Also shown are the 95% confidence intervals based on robust standard errors clustered on 60-square-mile grid squares. See Appendix A.1 for variable definitions and sources. The manufacturing diversity index is constructed from complete count data for 1880, 1920, 1930, and 1940, from 1-in-100 samples for 1860, 1870, 1900, and 1910, and from a 5-in-100 sample in 1900 (in all cases, the largest sample available from Ruggles et al., 2010).

A local economy with higher diversity has a broader availability of local production inputs. Thus, particularly if transport costs are high, imperfect substitutability between inputs

---

[15] Complete count data is available for 1880, 1920, 1930, and 1940; for the other years the measure of industrial diversity is calculated from census samples, which reduces sample sizes (there are 1,010 counties with non-missing data in 1860, 1,159 in 1870, 1,789 in 1900, and 1,668 in 1910).

[16] The results are similar for other specifications. OLS estimates also produce positive and significant coefficients for 1900, 1920, 1930 and 1940, with lower magnitudes. The results for pre-1920 outcomes have to be interpreted with caution. For 1860, 1870, 1900, and 1910, the lack of complete count Census data (see previous footnote) reduces sample size and creates measurement error. Moreover, as emphasized by Katz and Margo (2014), the IPUMS data on sector of employment before 1910 was imputed from data on occupation, which implies an additional source of measurement error.

may imply positive effects of agricultural and/or manufacturing diversity on productivity (in the sense of reducing unit costs). Perhaps more relevant, though, are the dynamic implications of economic diversity for the acquisition of new skills and new ideas, to which I now turn.

## 6.2 New Ideas and New Skills

Diversity can foster entry into new sectors, acquisition of new ideas and formation of new skills. These are interconnected dimensions of the process of development in a local economy as described by Jacobs (1969): "[t]he greater the sheer number and varieties of divisions of labor already achieved in an economy, the greater the economy's inherent capacity for adding still more kinds of goods and services." Along the same lines, Jacobs describes development as the process of "adding new work to old work," where the notion of "new work" includes new skills as well as new ideas (encompassing both innovation and imitation).

The positive link between diversity and technological change can be explained by the presence of cross-sector spillovers, recombination and complementarities. A well-known historical example of cross-product spillovers is the origination of the bow in the bow drill. Classic examples of recombination include windmills (which combine the principles of watermills and sails), the cotton spinning mule (which relied on the moving carriage from Hargreaves' spinning jenny and the rollers of Arkwright's water frame), and incandescent light bulbs (which reinvented candles on the basis of electricity) (Desrochers, 2001; Akcigit et al., 2013). An important case of technological complementarities from the Second Industrial Revolution was the interdependence of improvements in power generation and transmission networks (Rosenberg, 1979).

Historical research shows that in 19th century America, agricultural production and its linkages were technologically dynamic and diverse. Different agricultural products were associated with different technologies and skills in production as well as in storage, packaging, transportation, contractual arrangements, and so on. Olmstead and Rhode (2008) offer an impressive account of the rich paths of biological innovation characterizing wheat, corn, cotton, tobacco, livestock, draft animals, and dairy. Two cases of storage and transportation innovations directed to specific agricultural products—the grain elevator and the refrigerated railroad car—became paradigmatic technologies of the 19th century (Cronon, 2009). The first financial derivatives originated in the trading needs of grain and cotton merchants, and equipment leasing was invented by a food processor seeking financing for capital goods (Glaeser et al., 1992).

Production of agricultural machinery in the mid-19th century was still characterized by small and medium establishments serving local markets, and was highly specialized by

crop type (Pudup, 1987; Page and Walker, 1991). For example, corn, wheat, beans, and potatoes all used different types of mechanical planters. While improvements in tools and machines were often specific to one crop or group of crops, the enlargement of producers' know-how presumably yielded cross-product spillovers and higher potential for later recombinations. The agricultural machinery sector also had technological complementarities with other industries—the development of the plow, for instance, relied heavily on advances in iron metallurgy (Pudup, 1987).

Agro-processing industries still accounted for a large share of GDP in the late 19th century, and they were technologically progressive. In the 1870s and 1880s, they were among the first industries to adopt the new production technologies of the Second Industrial Revolution (Chandler, 1977). The diversity of ideas acquired from the application of continuous-flow methods in different agro-processing industries was absorbed in a salient historical case of recombination—the famed Ford Motors' assembly line, which incorporated insights from flour mills, meat-packing establishments, breweries and canning factories (Hounshell, 1985). This collection of historical examples and remarks suggests that agricultural diversity may have enhanced cross-fertilization and recombinant innovations.

In my proposed interpretation, the adoption of new ideas and the formation of new skills are interlinked dimensions of the process of structural change fostered by initial diversity. Following the notion of "new work" from Jacobs (1969), Lin (2011) interprets the presence of workers in new occupations as a measure of technological change and related changes in labor markets. One of his findings (using U.S. data for recent decades) is that new work is more prevalent in locations with high initial levels of industrial diversity. The dynamic effects of diversity under skill complementarities are emphasized by Hausmann and Hidalgo (2011) and in the model proposed in Appendix C. Diversity, insofar as it reflects a wider range of available skills, increases the return to acquiring new skills that complement the previously existing ones.

To assess the effects of early agricultural diversity on the acquisition of new ideas, I consider a proxy for technological dynamism over the Second Industrial Revolution: the county-level average number of patents per capita per decade between 1860 and 1940.[17] I interpret patent activity—which has been considered by a number of recent historical studies, including Perlman (2015), Acemoglu et al. (2016), and Akcigit et al. (2017)—as a measure of local technological dynamism broadly defined, since it captures not only new frontier technologies but also micro-innovations adapting technologies to local conditions as well as unsuccessful attempts.

---

[17]I divide patents in each decade by initial population and take the average across decades. The patent data comes from Akcigit et al., 2013 and Petralia et al., 2016.

To assess how diversity affected the formation of new skills, I focus on skills that emerged during the Second Industrial Revolution. Using Census employment data, I identify the high- and middle-skill occupations (as classified by Katz and Margo, 2014) that appear in the manufacturing sector in 1940 but were non-existent in the 1860 IPUMS Census sample, such as electricians, tool makers, and auto mechanics (see Appendix A.1 for details). I then calculate the share of these occupations in industrial employment in 1940, a measure of the presence of new skills that resembles the notion of "new work" in Jacobs (1969) and Lin (2011).

I report OLS and IV estimates for three specifications, controlling for (i) state fixed effects, (ii) state fixed effects and geo-climatic conditions, and (iii) state fixed effects, geo-climatic conditions, crop-specific controls, and socio-economic conditions. The estimates of the effects of early agricultural diversity on the log of average patents per capita per decade in 1860-1940 and on the share of manufacturing employment in new occupations in 1940 are reported in Panel A and Panel B of Table 4, respectively.[18]

The results in all specifications indicate positive significant effects of agricultural diversity on the acquisition of new ideas and new skills. According to the IV estimates in my preferred specification (column (4)), a one standard deviation increase in early agricultural diversity led to increases of 32% (28 log points) in patents per capita per decade and 0.6 pp. in the share of new skills in manufacturing employment.

## 6.3 Knowledge- and skill-intensive industrial activities

The mechanisms emphasized above have implications about the effects of early agricultural diversity across different industrial activities. If the positive effects of diversity are explained by the role of knowledge spillovers and skill complementarities in the acquisition of new ideas and new skills, then we would expect the effects of diversity to be larger in activities where these forces are more prevalent, namely in knowledge- and skill-intensive activities.

A differentially positive association between economic diversity and high-tech or new activities is hinted at by the urban and regional economics literature (e.g., by the empirical studies of Henderson et al., 1995, Greunz, 2004, and Neffke et al., 2011, and the theoretical contributions of Duranton and Puga, 2001, and Helsley and Strange, 2002). In the model presented in Appendix C, the same features that generate positive effects of diversity on industrialization and long-run development—complementarities and cross-sector spillovers—also generate a differentially positive impact on industrial activities with high levels of complexity.

---

[18] The sample in Panel A is reduced to 1,804 as 17 counties with no patents drop out. The sample in Panel B is reduced to 1,818 as 3 counties with no manufacturing workers recorded in 1940 drop out.

TABLE 4. EFFECTS OF AG.DIVERSITY ON NEW IDEAS AND NEW SKILLS

| | Specification 1 | | Specification 2 | | Specification 3 | |
|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) |

**Panel A.** *Dependent variable: Ln (Patents per capita per decade,1860-1940)*

| | | | | | | |
|---|---|---|---|---|---|---|
| Ag.Diversity$_{1860}$ | 0.975*** | 2.357*** | 1.060*** | 2.239*** | 1.215*** | 2.805*** |
| | (0.276) | (0.699) | (0.279) | (0.682) | (0.225) | (0.919) |
| $R^2$ | 0.581 | 0.570 | 0.605 | 0.598 | 0.701 | 0.691 |
| Observations | 1,804 | 1,804 | 1,804 | 1,804 | 1,804 | 1,804 |

**Panel B.** *Dependent variable: Share of New Skills in Manufacturing Employment,1940*

| | | | | | | |
|---|---|---|---|---|---|---|
| Ag.Diversity$_{1860}$ | 0.0201*** | 0.0535*** | 0.0212*** | 0.0492** | 0.0199*** | 0.0657** |
| | (0.00660) | (0.0182) | (0.00684) | (0.0196) | (0.00677) | (0.0306) |
| $R^2$ | 0.140 | 0.126 | 0.149 | 0.140 | 0.182 | 0.164 |
| Observations | 1,818 | 1,818 | 1,818 | 1,818 | 1,818 | 1,818 |
| State FE | Y | Y | Y | Y | Y | Y |
| Geo-climatic controls | Y | Y | Y | Y | Y | Y |
| Crop-specific controls | N | N | N | N | Y | Y |
| Socio-economic controls | N | N | N | N | Y | Y |

*Notes:* See Appendix A.1 for variable definitions and sources. The means of the dependent variables in panels A and B are -5.98 and 0.027, respectively. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

To assess whether early agricultural diversity was conducive to the development of knowledge- and skill-intensive industrial activities, I consider the shares of each of these two types of activities in overall manufacturing employment in 1940, in the aftermath of the Second Industrial Revolution. I define as knowledge-intensive industries those that had above-median percentages of engineers and scientists in total industry employment, and as skill-intensive those that had above-median educational attainment of workers. The correlation between the shares of these two types of activities across counties in the sample is 0.58.

Table 5 presents estimates of the effects of 1860 agricultural diversity on the shares of manufacturing workers in knowledge-intensive activities (Panel A) and skill-intensive activities (Panel B) in 1940. For all specifications, the results indicate positive and significant effects. According to my preferred specification (column (4)), increasing 1860 agricultural diversity by one standard deviation would lead to increases in the shares of knowledge- and skill-intensive activities in total manufacturing employment of 6.8 and 6.6 pp., respectively.

|  | Specification 1 | | Specification 2 | | Specification 3 | |
|---|---|---|---|---|---|---|
|  | OLS | IV | OLS | IV | OLS | IV |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A.** *Dependent variable: Share of Knowledge-Intensive Activities in Manufacturing,1940* | | | | | | |
| Ag.Diversity$_{1860}$ | 0.190*** | 0.559*** | 0.223*** | 0.551*** | 0.227*** | 0.579** |
|  | (0.0518) | (0.167) | (0.0493) | (0.161) | (0.0532) | (0.239) |
| $R^2$ | 0.403 | 0.378 | 0.438 | 0.420 | 0.473 | 0.457 |
| Observations | 1,818 | 1,818 | 1,818 | 1,818 | 1,818 | 1,818 |
| **Panel B.** *Dependent variable: Share of Skill-Intensive Activities in Manufacturing,1940* | | | | | | |
| Ag.Diversity$_{1860}$ | 0.0968* | 0.488*** | 0.156**** | 0.533*** | 0.144*** | 0.622** |
|  | (0.0518) | (0.172) | (0.0505) | (0.164) | (0.0550) | (0.248) |
| $R^2$ | 0.339 | 0.309 | 0.388 | 0.362 | 0.424 | 0.392 |
| Observations | 1,818 | 1,818 | 1,818 | 1,818 | 1,818 | 1,818 |
| State FE | N | N | Y | Y | Y | Y |
| Geo-climatic controls | N | N | Y | Y | Y | Y |
| Crop-specific controls | N | N | N | N | Y | Y |
| Socio-economic controls | N | N | N | N | Y | Y |

*Notes:* See Appendix A.1 for variable definitions and sources. The means of the dependent variables in panels A and B are 0.39 and 0.38, respectively. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

These results show that agricultural diversity had differentially positive effects on activities with high levels of complexity, lending further support to the hypothesis that diversity's long-run impact is explained by the presence of complementarities and cross-sector spillovers.

## 6.4 Other Channels

Besides the channels examined above, there are other ones that could account for the positive effects of agricultural diversity on development: changes in agricultural productivity (which could in turn boost industrialization); reduced exposure to product-specific shocks (which would decrease volatility and possibly foster economic performance); and lower land concentration (which could affect local institutions). My empirical assessment of these channels is briefly summarized in this section, and presented in detail in Appendix D.

To examine whether agricultural diversity affected development through agricultural productivity, I estimate the effect of the agricultural diversity on land productivity, as well as

the effect of the latter on industrialization. The estimated effects (reported in Table D.1) are not significant for either of the two links.

Next, I examine whether agricultural diversity favored development by reducing volatility in the value of agricultural output. I construct a measure of predicted volatility based on the initial county-level production mix and the year-to-year evolution of national prices. Then, I compare the estimated effects of early agricultural diversity on industrialization obtained with and without controlling for this measure. If diversity positively affected industrialization by reducing volatility, this control should have a negative and significant coefficient, and its inclusion would reduce the coefficient on diversity. However, the evidence (displayed in Table D.2) is not in line with these predictions.

Finally, I assess whether agricultural diversity affected development by shaping land concentration and thus activating political economy mechanisms. To do this, I empirically study the relationship between agricultural diversity and land concentration, as well as the effects of agricultural diversity on key institutions that displayed significant variation at the local level—schools and banks. The results (reported in Tables D.3 and D.4) do not support the relevance of this channel.

## 7 Conclusion

This paper shows that early agricultural diversity had positive and persistent effects on development across U.S. counties. My identification strategy relies on climate-based measures of crop-specific potential yields, which I use to construct an index of potential agricultural diversity. Using this index as an IV, I find that a one standard deviation increase in 1860 agricultural diversity led to increases in 2000 levels of population density and income per capita of about 73% and 6%, respectively. The positive effects of early agricultural diversity emerged during the Second Industrial Revolution, when it boosted the process of structural change.

I advance the hypothesis that the positive effects of agricultural diversity are explained by the presence of complementarities and cross-sector spillovers. As illustrated by a collection of historical examples from the American Second Industrial Revolution, and captured formally in the model presented in Appendix C, agricultural diversity can foster entry into new activities and the acquisition of new ideas and new skills. Providing support for this interpretation, I show that early agricultural diversity led to higher levels of industrial diversification, patent activity and employment in new occupations. Moreover, consistent with an additional implication of the model, agricultural diversity positively affected the shares of industrial workers employed in knowledge- and skill-intensive sectors. In contrast, as shown

in Appendix D, the evidence does not support the relevance of other possible channels.

My findings add to a growing body of evidence about deeply-rooted factors that, although no longer directly relevant, have shaped historical paths of economic growth and thus display persistent effects on comparative development. My focus on the role of agricultural diversity sheds new light on the role of agriculture in the process of development, and also on the relevance of diversity's effects on growth though the acquisition of skills and ideas. Given that these effects cannot be captured in standard two- or three-sector models of structural change, the paper suggests that models with finer levels of aggregation can yield new insights about the development process.

The findings from the history of U.S. counties going back to 1860 may be relevant for understanding the contemporary growth trajectories of developing countries, though naturally this cannot be taken for granted. To further our understanding of the role of diversity, and before any policy implications can be firmly established, future research should attempt to pin down as precisely as possible the mechanisms through which diversity affects development in different contexts across space and time.

# Appendix A. Data

## A.1 Definitions and Sources

**Outcome variables**

*Population density.* Population / area. Digitized U.S. Census data on population for 1860-1940 are drawn from Minnesota Population Center (2016). Data for 1940 onwards are drawn from digitized County Data Books compiled by the U.S. Census Bureau, available from Haines and ICPSR (2010). For 2000 I use the data from the Bureau of Economic Analysis (2015). The data on area is the same I use as a geo-climatic control (see below).

*Income per capita.* Total personal income / population. Data on personal income and population in 2000 is obtained from the Bureau of Economic Analysis (2015).

*Share of Population in Manufacturing.* Average yearly number of manufacturing workers over total population. Digitized U.S. Census data for 1860-1940 are drawn from Minnesota Population Center (2016). Data for 1940 onwards are drawn from digitized County Data Books compiled by the U.S. Census Bureau, available from Haines and ICPSR (2010).

*Manufacturing diversity.* Defined as 1 minus the Herfindahl index of manufacturing employment shares of 59 sectors in total manufacturing employment. Constructed with complete count employment Census micro-data for 1880, 1920, 1930, and 1940, 1-in-100 samples for 1860, 1870, 1900, and 1910, and a 5-in-100 sample in 1900 (in all cases, the largest sample available from Ruggles et al., 2010).

*Patents per capita per decade, 1860-1940.* County-level number of patents divided by initial population in each decade, averaged across decades, between 1860 and 1940. I use data kindly shared by Ufuk Akcigit, Tom Nicholas, and William Kerr for 1860-1920 and data from Petralia et al. (2016) for 1920-1940.

*Share of New Skills in Manufacturing Employment, 1940.* Proportion of a county's manufacturing workers in 1940 that are employed in high- and middle-skill occupations (as classified by Katz and Margo, 2014) which had zero respondents in the 1860 Census 1-in-100 sample. Employment data for 1860 (a 1-in-100 sample) and 1940 (complete count) are drawn from Ruggles et al. (2010). Among high- and middle-skill occupational categories with zero respondents in the 1860 Census sample, there are 14 that are present in manufacturing in 1940 for at least one county in my sample: chemical engineers; electrical engineers; testing technicians; purchasing agents and buyers; advertising agents and salesmen; office machine operators; cranemen, derrickmen, and hoistmen; electricians; heat treaters, annealers, temperers; airplane mechanics and repairmen; automobile mechanics and repairmen; railroad

and car shop mechanics and repairmen; motion picture projectionists; professional nurses (the last two with only a handful of occurrences).

*Share of Skill-Intensive Activities in Manufacturing, 1940.* Proportion of a county's manufacturing workers in 1940 that are employed in industrial activities with above-median skill-intensity. A sector's skill-intensity is the average educational attainment of all workers employed in the sector in the complete count 1940 Census data available from Ruggles et al. (2010).

*Share of Knowledge-Intensive Activities in Manufacturing, 1940.* Proportion of a county's manufacturing workers in 1940 that are employed in industrial activities with above-median knowledge-intensity. A sector's knowledge-intensity is the percentage of engineers and scientists among all workers employed in the sector in the complete count 1940 Census data available from Ruggles et al. (2010).

*Farm Productivity, 1870.* Farm Output / Improved Acres. Farm output is measured in value terms using county-level data from the Census and state-level prices. Data source: digitized U.S. Census data drawn from Minnesota Population Center (2016).

*Bank density.* Number of banks/Population. FDIC data from NHGIS (Minnesota Population Center, 2016).

*Share of farmland in top 10% largest farms.* Share of county-level farmland corresponding to the top 10% largest farms. Source: digitized U.S. Census data from NHGIS (Minnesota Population Center, 2016).

*School expenditures per capita, 1890.* Ordinary expenditures on public common schools over total population. Expenditures data, drawn from Rhode and Strumpf (2003), comes from the Eleventh Census of the United States, Vol. 25, Report on Wealth, Debt, and Taxation, Pt. II Valuation and Taxation.

**Geo-climatic controls**

*Land productivity measures.* Maximum and average of normalized attainable yields for alfalfa, barley, buckwheat, cane sugar, carrot, cabbage, cotton, flax, maize, oats, onion, pasture grasses, pasture legumes, potato, pulses, rice, rye, sorghum, sweet potato, tobacco, tomato, and wheat. I normalize each product's values dividing them by the maximum value for that product in the sample. Measures of attainable yields were constructed by the FAO's Global Agro-Ecological Zones project v3.0 (IIASA/FAO, 2012) using climatic data, including precipitation, temperature, wind speed, sunshine hours and relative humidity (based on which

they determine thermal and moisture regimes), together with crop-specific measures of cycle length (i.e. days from sowing to harvest), thermal suitability, water requirements, and growth and development parameters (harvest index, maximum leaf area index, maximum rate of photosynthesis, etc). Combining these data, the GAEZ model determines the maximum attainable yield (measured in tons per hectare per year) for each crop in each grid cell of 0.083x0.083 degrees. I use FAO's measures of agro-climatic yields (based solely on climate, not on soil conditions) for intermediate levels of inputs/technology and rain-fed conditions. In Appendix B.2, I consider three additional measures of land productivity: the first principal component of the 22 crop-specific attainable yields from FAO-GAEZ; *length of growing period*, the number of days during the year when temperature and moisture are conducive to crop growth and development, also from FAO-GAEZ; and *land suitability for cultivation*, the county-level average of an index of land suitability for cultivation (to be interpreted as the probability that a given area is cultivated) from Ramankutty et al. (2002).

*Area.* Surface area in 1000's of square miles, calculated with county shapefiles from NHGIS (Minnesota Population Center, 2016) using GIS software.

*Temperature.* County-level mean annual temperature measured in Celsius degrees. Data source: IIASA/FAO (2012).

*Rainfall.* County-level average annual precipitation measured in mm. Data source: IIASA/FAO (2012).

*Elevation.* County-level average terrain elevation in km. Data source: IIASA/FAO (2012).

*Latitude.* Absolute latitudinal distance from the equator in decimal degrees, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS (Minnesota Population Center, 2016).

*Longitude.* Absolute longitudinal distance from the Greenwhich meridian in decimal degrees, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS (Minnesota Population Center, 2016).

*Distance to the coastline or Great Lakes.* Minimum distance to a point in the Coastline or the Great Lakes in km, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS (Minnesota Population Center, 2016).

*Distance to the fall line.* Minimum distance to a point in the fall line in km, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS (Minnesota Population Center, 2016).

**Crop-specific controls**

*Dummies for dominance of specific products*: dummy variables for each the five major products (corn, cotton, animals slaughtered, hay, wheat) and one for plantation crops other than cotton (sugar, rice, and tobacco, combined) that take a value of 1 when the product (or group of products) has the largest share in a county's agricultural production value. Data sources: agricultural production data from the 1860 Census of Agriculture and state-level prices from Atack and Bateman (1987) and Craig (1993) (both available from Minnesota Population Center, 2016).

**Socio-economic controls**

*Urbanization rate.* Urban population / Total Population. Data source: digitized U.S. Census data drawn from Minnesota Population Center (2016).

*Farm Productivity, 1860.* Farm Output / Improved Acres. Farm output is measured in value terms using county-level data from the Census and state-level prices. Data source: digitized U.S. Census data drawn from Minnesota Population Center (2016).

*Share of slaves in the population.* Slave Population / Total Population. Data source: digitized U.S. Census data drawn from NHGIS (Minnesota Population Center, 2016).

*Share of foreigners in the population.* Foreign Population / Total Population. Data source: digitized U.S. Census data drawn from NHGIS (Minnesota Population Center, 2016).

*Share of the population below 15 years.* Population below 15 years / Total Population. Data source: digitized U.S. Census data drawn from NHGIS (Minnesota Population Center, 2016).

*Share of the population 65+ years.* Population 65+ years / Total Population. Data source: digitized U.S. Census data drawn from NHGIS (Minnesota Population Center, 2016).

*Distance to railroads.* Minimum distance to a point in a pre-1860 railroad line digitized by Atack (2015c) in km, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS (Minnesota Population Center, 2016).

*Distance to steam-boat navigated rivers.* Minimum distance to a point in a steam-boat navigated river digitized by Atack (2015a) in km, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS (Minnesota Population Center, 2016).

*Distance to canals.* Minimum distance to a point in a steam-boat navigated river digitized by Atack (2015b) in km, calculated from the centroid of each county using GIS software and county shapefiles from NHGIS (Minnesota Population Center, 2016).

*Market Potential.* Following the classic definition of Harris (1954), market potential in county $c$ is given by $\text{Market}_c = \sum_{k \neq c} d_{c,k}^{-1} N_k$, where $k$ is the index spanning neighboring counties, $d_{c,k}$ is the distance between county $c$ and county $k$, and $N_k$ is the population of county $k$ (here, in 1860). Distances are calculated from the centroid of each county using GIS software and county shapefiles from NHGIS (Minnesota Population Center, 2016). Population data are drawn from NHGIS (Minnesota Population Center, 2016).

**Agricultural output predicted price index, average rate of change and volatility**

In Appendix D.2, I incorporate two additional control variables: the average and the standard deviation of $\widehat{P}_{ct}^{\text{a}}$, the rate of change of $P_{ct}^{\text{a}}$, a predicted price index for a county's agricultural production. This index is defined as $P_{ct}^{\text{a}} = \sum_i \theta_{ic} p_{it}$, where the $\theta_{ic}$'s reflect the county's agricultural production mix in 1860, and $p_{it}$ is the national price of product $i$ at time $t$ (with $p_{i1867}$ normalized to 1 for all $i$). After calculating the rate of change $\widehat{P}_{ct}^{\text{a}} = (P_{ct}^{\text{a}} - P_{ct-1}^{\text{a}})/P_{ct-1}^{\text{a}}$ for each year, I compute its average and standard deviation over the 1867-1900 period, $Avg.\widehat{P}_{ct}^{\text{a}}$ and $Std.Dev.(\widehat{P}_{ct}^{\text{a}})$. Annual price data from 1867 to 1900 are available for 15 products that account for over 93.5% of total agricultural output in 1860 in my sample (animals, barley, buckwheat, butter, cheese, cotton, corn, hay, oats, potatoes, rye, sweet potatoes, tobacco, wheat, and wool). Only products with available price data are considered in the calculation of $P_{ct}^{\text{a}}$; the $\theta_{ic}$'s are computed with the restricted set of products so that they add up to 1. Prices for all products except cotton are drawn from Olmstead and Rhode (2006). For animals I use the average of the prices of cattle and hogs; the former are available from 1867 onward (a year later than prices of other products), which sets the starting point of my measures of $P_{ct}^{\text{a}}$. I use the price of butter as a proxy for the price of cheese, since the latter is only available (for Wisconsin cheddar) from 1879 onward; the correlation between these prices from 1879 to 1900 is 0.73. For cotton I use prices in Alabama from Cooley et al. (1977) (I am grateful to Paul Rhode for sharing a copy of this dataset).

## A.2    Adjustment for Changes in County Boundaries

The counties in the sample, with their boundaries defined in 1860, are displayed in gray in Figure A.1. Among these counties, over 70% did not experience subsequent changes in boundaries; among the remaining ones, over two thirds have overlaps of over 80% in terms of area with a county defined in 2000. In this sense, boundaries were relatively stable for counties in the sample. Still, boundary changes need to be taken into account to have consistent units of observation.
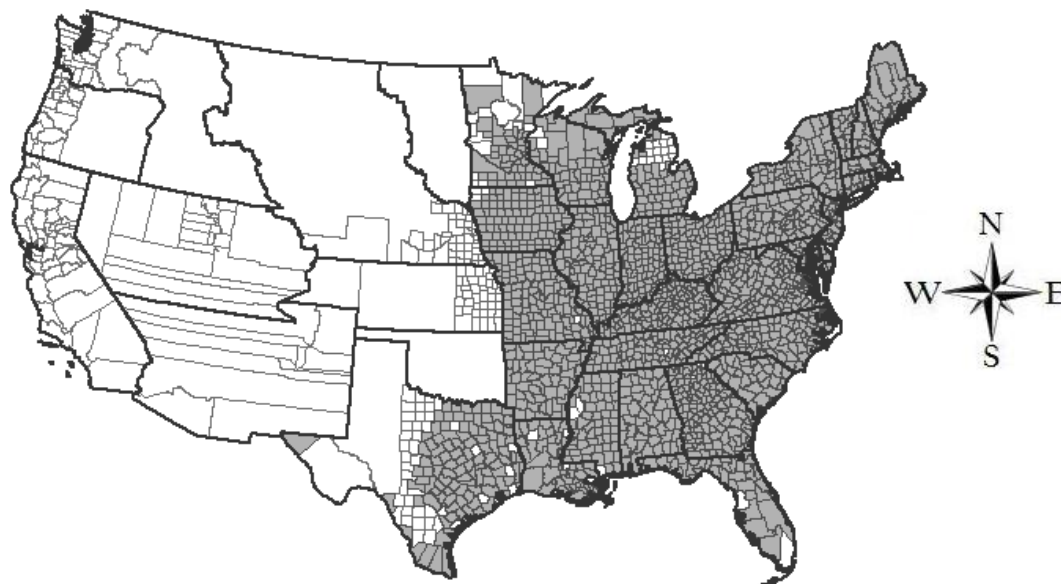
*Notes:* The map shows continental U.S. counties and state boundaries in 1860. Counties in the sample are shaded in gray. Boundary shapefiles are drawn from NHGIS (Minnesota Population Center, 2016).

To adjust data from a period $t$ to 1860 county boundaries, I use geographic shapefiles of county boundaries from NHGIS (Minnesota Population Center, 2016). With the help of GIS software, I determine the county defined in 1860 that contained each county or county fragment defined in period $t$, discarding all fragments with less than 1 square mile. Then I calculate the values of period $t$ variables corresponding to counties defined in 1860 by assigning values from period $t$ variables in proportion to the area of the county defined in period $t$ that each county defined in 1860 represents. This procedure would be fully accurate if the quantities measured in county-level aggregates were uniformly distributed over space.

# Appendix B. Additional Results and Robustness Checks

## B.1    Alternative Measures of Agricultural Diversity

This section considers a number of alternative measures of agricultural diversity. I show that the main results of the paper are robust to adjustments that address potential concerns about measurement error in the agricultural production data, and that they are also robust to considering measures with different functional forms than the standard Hirschman-Herfindahl index.

Measuring agricultural diversity with production data from the 1860 Census of Agriculture is subject to some concerns. First, there is likely mismeasurement in the value of animals produced for the market. The Census collected data on the "value of animals slaughtered." While no printed instructed were issued with reference to the 1860 Census schedule, the collected data presumably refers to the value of animals slaughtered *on the farm.* Thus, the value of livestock sold and slaughtered *off the farm* would be ignored, leading to underestimation. Another issue is that, according to Wright (1979), both demand and yields for cotton were unusually high in 1859. Finally, there is the broad issue that I use gross output rather than value added data.

I address these concerns by considering a number of adjustments and alternative measures. To address measurement error in the value of animals produced for the market, I take advantage of the fact that the 1870 Census collected data on the "value of animals slaughtered or sold for slaughter" rather than "value of animals slaughtered." Both the 1860 and the 1870 Census also collected data on the "value of livestock" on the farm. To construct an alternative measure of animals produced for the market in 1860, I take the 1860 "value of livestock" and multiply it by the ratio of 1870 aggregate "value of animals slaughtered or sold for slaughter" to 1870 "value of livestock" for my sample of counties. Thus, I get county-level estimates of the 1860 "value of animals slaughtered or sold for slaughter" based on the 1860 "value of livestock" on the farm and the ratio between these two variables in 1870 Census. This measure is on average about 25% higher than the "value of animals slaughtered" in 1860, while the correlation coefficient between the two is 0.6. The correlation between my baseline index of agricultural diversity in 1860 and the one obtained when replacing the 1860 "value of animals slaughtered" by the estimates with the "livestock adjustment" just explained is 0.91.

To address the possibly exceptional character of cotton production value in the 1860 Census, I create an alternative measure using 1870 data and considering price and output data for corn (the main agricultural product) as a reference. In particular, I take corn

output in 1860 and multiply it by the ratio of cotton output to corn output in 1870, and by the average ratio of cotton prices to corn prices between 1876 and 1900 (cotton prices are not regularly available before 1876). Thus, I get county-level measures of the 1860 cotton production value without using the 1860 data on cotton output or price.[19] This measure is on average about 5.5% lower than the actual 1860 cotton production value, and the correlation coefficient between the two is 0.93. The correlation between my baseline index of agricultural diversity in 1860 and the one obtained after performing the "cotton adjustment" just explained is also 0.93.

Finally, I consider agricultural variety (i.e., the number of products present at the county level) in 1860 and in 1870 as alternative measures.[20] In these measures, each product is either zero or one, which alleviates concerns about mismeasurement in the value of animals produced for the market, or about abnormal cotton demand or yields in 1859, insofar as these issues mostly affect the intensive margin. In any case, calculating agricultural variety with the "livestock adjustment" or the "cotton adjustment" does not qualitatively change the results.

Considering variety as the relevant measure of diversity provides a check on the implications of using gross output rather than value added data. There is no clear principle indicating whether measuring agricultural diversity with gross output rather or value added data is most appropriate. I use gross output rather than value added data simply because the latter are not available and estimating them would require a large number of questionable assumptions. If using value added data was appropriate, using gross output data would generate measurement error. However, this concern does not apply when agricultural variety is used as a measure of diversity, since the number of products with non-zero value added is generally the same as the number of products with non-zero gross output.

Panel A of Table B.1 provides estimates OLS and IV of the effects of 1860 agricultural diversity, captured with the alternative measures discussed above, on population density in 2000 (in logs), for my preferred specification (controlling for state fixed effects and geo-climatic features). To facilitate comparison, the first two columns reproduce the results using the baseline measure of Ag.Diversity in 1860. In line with the IV strategy from the main analysis, I construct an IV for Ag.Variety using the predicted product shares obtained from the FML model (and considering as present all products with predicted shares above 0.05%).

_____

[19] I also considered measures of cotton output in 1860 in which I average cotton output in 1850 and 1870 and/or continue using the 1860 price or replace it with its average 1876-1900 price (instead of using the ratio to corn prices). All these variations produce qualitatively the same results.

[20] A complete set of prices for 1869/1870 is not available, so I cannot calculate Ag.Diversity in 1870, though I can estimate it with 1870 quantities and 1860 prices. Estimating the long-run effects of diversity using this measure produces qualitatively the same results.

I use the same IV for Ag.Variety in 1860 and 1870. In all cases, the results confirm positive and significant effects of early agricultural diversity on long-run outcomes.

TABLE B.1. ESTIMATED EFFECTS OF AG.DIVERSITY USING ALTERNATIVE MEASURES

| | Dependent variable: Ln Population Density, 2000 |
|---|---|---|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| **Panel A.** *Alternative Measures I* | | | | | | | | | | |
| Ag.Diversity$_{1860}$ | 2.370*** | 4.477*** | | | | | | | | |
| | (0.539) | (1.045) | | | | | | | | |
| Ag.Diversity$_{1860}$ (livestock adjustment) | | | 2.765*** | 5.712*** | | | | | | |
| | | | (0.605) | (1.330) | | | | | | |
| Ag.Diversity$_{1860}$ (cotton adjustment) | | | | | 2.192*** | 4.181*** | | | | |
| | | | | | (0.547) | (1.002) | | | | |
| Ag.Variety$_{1860}$ | | | | | | | 0.0466*** | 0.127*** | | |
| | | | | | | | (0.0103) | (0.0396) | | |
| Ag.Variety$_{1870}$ | | | | | | | | | 0.0510*** | 0.147*** |
| | | | | | | | | | (0.0108) | (0.0493) |
| Observations | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 |
| $R^2$ | 0.354 | 0.329 | 0.358 | 0.317 | 0.349 | 0.327 | 0.343 | 0.282 | 0.347 | 0.258 |
| **Panel B.** *Alternative Measures II (Indexes of Specialization)* | | | | | | | | | | |
| Krugman Specialization Index | -1.034*** | -2.828*** | | | | | | | | |
| | (0.285) | (0.575) | | | | | | | | |
| Gini Coefficient | | | -10.47*** | -21.74*** | | | | | | |
| | | | (0.285) | (0.575) | | | | | | |
| Index of Inequality in Production Structure | | | | | -1.674*** | -4.557** | | | | |
| | | | | | (0.639) | (1.015) | | | | |
| Entropy Index | | | | | | | -4.740*** | -8.542*** | | |
| | | | | | | | (1.078) | (1.865) | | |
| Coefficient of Variation | | | | | | | | | -3.026*** | -5.355*** |
| | | | | | | | | | (0.617) | (1.227) |
| Observations | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 |
| $R^2$ | 0.340 | 0.284 | 0.367 | 0.315 | 0.337 | 0.292 | 0.354 | 0.333 | 0.361 | 0.338 |
| State FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Geo-climatic controls | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

*Notes:* The mean of log population density in 2000 is 4.37. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

I also check the robustness of the main results to adopting alternative measures of diversity. I consider several standard measures of specialization, for which higher values represent lower diversity: the Krugman Index of Specialization, defined as $\sum_i |\theta_{ic} - \theta_i|$, where $\theta_i$ is the share of product $i$ in total agricultural output in the sample; the Gini

coefficient of product shares; an index of Inequality in Production Structure proposed by Cuadrado-Roura et al. (1999), defined as $\sum_i (\theta_{ic} - \theta_i)^2$; an entropy index, defined as $E(\alpha)_c = \frac{1}{N\alpha(\alpha-1)} \sum_i \left[ \left( \theta_{ic}/\bar{\theta}_{ic} \right)^2 - 1 \right]$, where $\bar{\theta}_{ic}$ is the average product share in county $c$ (which is always equal to 1/36), and $\alpha$, the weight given to distances between product shares at different parts of the distribution, is set equal to 2; and finally, the coefficient of variation of product shares, defined as $\sigma_{\theta_{ic}}/\bar{\theta}_{ic}$, where $\sigma_{\theta_{ic}}$ is the standard deviation of product shares in county $c$.

Panel B of Table B.1 provides OLS and IV estimates of the effects of 1860 agricultural diversity (actually, specialization), captured by these alternative measures, on population density in 2000 (in logs). The IV for each measure of specialization is constructed by replacing the actual product shares with the predicted shares from the FML model. In all cases, the results indicate significant negative effects of specialization, that is, positive effects of diversity.

## B.2  Additional Geo-climatic controls

This section shows that the results are robust to controlling for additional geo-climatic features and higher-order terms. Table B.2 reports the coefficient estimates from OLS and IV regressions of 2000 population density (in logs) on 1860 agricultural diversity obtained when I sequentially expand the set of geo-climatic controls. Columns (1)-(2) reproduce the main results controlling for state fixed effects and baseline geo-climatic controls.

Columns (3)-(4) report the results obtained when including cubic polynomials in several land productivity measures as well as in temperature, rainfall, and elevation. Besides the baseline measures of land productivity (i.e., the max and the mean of 22 product-specific attainable yields from FAO-GAEZ), I now include: the first principal component of the FAO-GAEZ measures; the individual FAO-GAEZ measures corresponding to products with shares in output above 1% (corn, cotton, alfalfa, pasture grasses, pasture legumes, wheat, oats, potato, sweet potato, tobacco, cane sugar); the length of growing period, that is, the number of days during the year when temperature and moisture are conducive to crop growth and development (also from FAO-GAEZ); and an index of overall land suitability for cultivation from Ramankutty et al. (2002).

Columns (5)-(6) report the results obtained when adding cubic polynomials in distance to steamboat navigated rivers, distance to canals, and distance to the fall line. The importance of the latter is underscored by the contribution of Bleakley and Lin (2012). While it seems natural to include distance to steamboat navigated rivers and distance to canals in this robustness check, they are more appropriately considered as "socio-economic" controls, since they are not solely determined by geography and climate.

In all specifications, the estimated effect of 1860 agricultural diversity on (the log of) 2000 population density is positive and significant.

TABLE B.2. ESTIMATED EFFECTS OF AG.DIVERSITY WITH ADDED GEO-CLIMATIC CONTROLS

| | Dependent variable: Ln Population Density, 2000 | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | Specification 1 | | Specification 2 | | Specification 3 | |
| | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Ag.Diversity$_{1860}$ | 2.370*** | 4.477*** | 1.831*** | 3.828*** | 1.887*** | 4.035*** |
| | (0.539) | (1.045) | (0.474) | (1.152) | (0.448) | (1.153) |
| State FE | Y | Y | Y | Y | Y | Y |
| Geo-climatic controls | Y | Y | Y | Y | Y | Y |
| Cubic polynomials in: | | | | | | |
| Land productivity measures | N | N | Y | Y | Y | Y |
| Temperature, rainfall, elevation | N | N | Y | Y | Y | Y |
| Latitude, longitude | N | N | N | N | Y | Y |
| Distances to rivers, canals, fall line | N | N | N | N | Y | Y |
| $R^2$ | 0.354 | 0.329 | 0.484 | 0.465 | 0.495 | 0.474 |
| Observations | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 |

*Notes:* The mean of log population density in 2000 is 4.37. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

## B.3    Agricultural Specialization Patterns and Development

The effects of particular specialization patterns have attracted considerable attention from economic historians. Engerman and Sokoloff (1997, 2002) argue that climate and soil quality historically affected crop choice, which in turn led to divergent paths of development; in particular, they emphasize that specialization in cotton, sugar, rice, tobacco, and coffee favored slave plantations and thus generated inequalities that harmed long-run performance (see also Nunn, 2008; Bruhn and Gallego, 2012).

Other contributions highlight the effects of specialization in other crops through different mechanisms. Goldin and Sokoloff (1984) argue that the relative productivity of women and children in hay, wheat, and dairy was relatively low, and thus specialization in these products lowered the costs of hiring workers for industrial firms (see also Goldin and Sokoloff, 1982). Earle and Hoffman (1980) point out that wheat, corn, and livestock had highly

seasonal labor requirements, which also lowered labor costs for the industrial sector. Sokoloff and Dollar (1997) also emphasize the high seasonality of grains, but they argue that the availability of cheap seasonal labor could hinder the adoption of more efficient manufacturing technologies. Vollrath (2011) proposes a model in which specialization in crops with high labor elasticity (e.g., rice) reduces the positive effects of agricultural productivity growth on industrialization.

This paper focuses on the effects of agricultural diversity beyond specific specialization patterns. As discussed in section 4, the diversity index may be highly correlated with the shares of dominant crops. To avoid confounding the effects of agricultural diversity with those of specialization in specific crops, the estimating equation includes crop-specific dummy variables for the 5 major agricultural products (corn, cotton, animals slaughtered, hay, and wheat) and for plantation crops other than cotton (which take a value of 1 when the product, or group of products, has the largest share in a county's agricultural production).

The dummy variables are meant to capture the notion of dominance reflected in terms such as "cotton counties" or "wheat counties." This appendix shows that the results are robust to controlling for specific specialization patterns in different ways. Table B.3 shows OLS and IV estimates of the effects of 1860 agricultural diversity on 2000 population density (in logs) with alternative sets of crop-specific controls.[21] To facilitate comparisons, columns (1)-(2) do not include any crop-specific controls and columns (3)-(4) include the dominance dummies as defined in the main analysis. Columns (5)-(6) include dominance dummies that take a value of 1 when the share for the corresponding crop (or group of crops) in the county's agricultural output is above 25%. Columns (7)-(8) include dominance dummies that take a value of 1 when the share is above the 75th percentile of the distribution of that share in the whole sample. Finally, columns (9)-(10) control for the actual shares.

The results confirm that agricultural diversity had a positive effect on long-run development. The coefficients on crop-specific controls suggest, in line with previous contributions, that early specialization in cotton and other plantation crops is negatively associated with long-run development. Specialization in animal production also appears to have a negative association with long-run development. For the purposes of this paper, the main takeaway from this analysis is that the estimated coefficients for agricultural diversity remain consistently positive and significant across specifications with different crop-specific controls.

---

[21] Only agricultural diversity is instrumented in these IV regressions; the results are qualitatively similar if the crop-specific controls are replaced by the corresponding values based on predicted shares from the FML model.

TABLE B.3. EFFECTS OF AGRICULTURAL DIVERSITY CONTROLLING FOR SPECIFIC PRODUCTS

| | Dependent variable: Ln Population Density, 2000 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No crop-specific controls | | Largest share dummies | | $> 25\%$ dummies | | $> 75^{th}$ percentile dummies | | Exact shares | |
| | OLS | IV | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Ag.Diversity$_{1860}$ | 2.370*** | 4.477*** | 2.206*** | 4.276*** | 2.393*** | 5.107*** | 2.338*** | 4.834*** | 1.527** | 4.593** |
| | (0.539) | (1.045) | (0.521) | (1.190) | (0.606) | (1.359) | (0.563) | (1.234) | (0.635) | (2.131) |
| Corn | | | -0.634* | -0.734* | -0.165 | -0.00500 | -0.219** | -0.277** | -2.299*** | -1.282 |
| | | | (0.353) | (0.411) | (0.103) | (0.114) | (0.106) | (0.110) | (0.767) | (0.916) |
| Cotton | | | -0.991*** | -0.906** | -0.422*** | -0.416*** | -0.532*** | -0.395*** | -2.357*** | -1.258 |
| | | | (0.373) | (0.385) | (0.124) | (0.128) | (0.139) | (0.141) | (0.704) | (0.892) |
| Animals Slaughtered | | | -0.923*** | -1.082*** | -0.0761 | -0.269** | -0.366*** | -0.456*** | -2.470*** | -3.176** |
| | | | (0.329) | (0.392)) | (0.0825) | (0.111) | (0.107) | (0.116) | (0.799) | (1.239) |
| Hay | | | -0.193 | 0.298 | 0.206 | 0.132 | -0.122 | -0.0675 | -0.787 | -0.155 |
| | | | (0.380) | (0.439) | (0.139)) | (0.153) | (0.132) | (0.133) | (0.770) | (0.898) |
| Wheat | | | -0.331 | (-0.473) | 0.176** | 0.0829 | -0.0431 | -0.0738 | 0.256 | 0.173 |
| | | | (0.349) | (0.417) | (0.0772) | (0.0905) | (0.104) | (0.104) | (0.632) | (0.703) |
| Tobacco+Cane+Rice | | | -0.965*** | -0.994** | -0.317*** | -0.448*** | -0.487*** | -0.532*** | -2.328*** | -1.866*** |
| | | | (0.367) | (0.407) | (0.0778) | (0.107) | (0.108) | (0.114) | (0.627) | (0.681) |
| State FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Geo-climatic controls | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Crop-specific controls | N | N | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 |
| $R^2$ | 0.354 | 0.329 | 0.372 | 0.349 | 0.373 | 0.338 | 0.367 | 0.334 | 0.390 | 0.353 |

*Notes:* The mean of log population density in 2000 is 4.37. The crop-specific controls are constructed in different ways across specifications, as indicated at the top of the columns (largest share dummies, $> 25\%$ dummies, etc). Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

## B.4  Input-output Linkages

This section examines whether agro-processing and related industries with known input-output connections with specific agricultural products were more likely to be present the larger the importance of those agricultural products at the local level. To do this, I estimate linear probability models where the outcomes are indicator variables for the presence of specific industrial activities in county $c$ in 1880 (the first year within the period under consideration with available full count Census data) and the key regressors are the shares corresponding to the relevant agricultural products in 1860.

I consider six industrial activities (and the agricultural products to which they are known to be linked): textiles industries (linked to cotton, wool); tobacco manufactures (tobacco); bakery, confectionery, and related products (wheat, rye, cane sugar); beverage industries (rye, barley); meat products (animals); canning and preserving of fruits, vegetables, seafoods (orchards, market gardens).[22]

Table B.4 reports the coefficient estimates from OLS and IV regressions controlling for state fixed effects and geo-climatic features. The IV regressions use the predicted shares obtained from the FML model for each of the relevant agricultural products.

The results suggest that there were a number of input-output linkages through which the composition of agricultural production may have shaped industrial production patterns at the local level. In particular, it is consistent with the idea that agricultural diversity may have fostered manufacturing diversity by favoring entry into various agro-processing industries (which could in turn favor entry to other manufacturing activities).

A salient exception to the relevance of input-output connections among the ones considered here is the case of cotton and textiles, for which the IV estimate is negative (though not significant). Of course, it does not come as a surprise that the presence of cotton did not foster the development of textiles at the local level, as it is well-known that cotton production was concentrated in the South while textile industries were concentrated in the Northeast. In regard to the mechanisms analyzed in the paper, this does not mean that the presence of cotton production did not generate specific skills and knowledge that were potentially relevant for manufacturing at the local level.

---

[22] Textile industries comprise several industry groups in the 1950 Census classification system. I group them together because they have a common set of relevant agricultural inputs. I combine bakery products with confectionery and related products for the same reason. For beverage industries, corn was also a relevant input, but I exclude it because it was mostly used for other purposes (in particular, for feeding animals).

TABLE B.4. INPUT-OUTPUT LINKAGES

| | Textiles industries | | Tobacco manufactures | | Bakery, confectionery | | Beverage industries | | Meat products | | Canning and preserving | |
| | OLS | IV | OLS | IV | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cotton | 0.102 | -0.0549 | | | | | | | | | | |
| | (0.0888) | (0.144) | | | | | | | | | | |
| Wool | 0.335 | 0.964** | | | | | | | | | | |
| | (0.279) | (0.421) | | | | | | | | | | |
| Tobacco | | | 1.155*** | 0.771*** | | | | | | | | |
| | | | (0.168) | (0.290) | | | | | | | | |
| Wheat | | | | | 0.486*** | 0.127 | | | | | | |
| | | | | | (0.129) | (0.491) | | | | | | |
| Rye | | | | | 1.513*** | 4.450* | 0.362 | 6.007** | | | | |
| | | | | | (0.630) | (2.661) | (0.851) | (2.959) | | | | |
| Cane Sugar | | | | | 0.344** | 0.494*** | | | | | | |
| | | | | | (0.167) | (0.187) | | | | | | |
| Barley | | | | | | | 6.452*** | 8.729 | | | | |
| | | | | | | | (1.691) | (5.312) | | | | |
| Animals Slaughtered | | | | | | | | | -0.0408 | 0.271 | | |
| | | | | | | | | | (0.109) | (0.266) | | |
| Orchards | | | | | | | | | | | 1.470** | 9.559*** |
| | | | | | | | | | | | (0.676) | (2.933) |
| Market gardens | | | | | | | | | | | 0.915*** | 0.100 |
| | | | | | | | | | | | (0.278) | (0.485) |
| State FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Geo-climatic controls | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 |
| $R^2$ | 0.228 | 0.222 | 0.349 | 0.346 | 0.345 | 0.335 | 0.380 | 0.359 | 0.183 | 0.180 | 0.325 | 0.206 |

Dependent variable: Indicator variable for

*Notes:* The outcomes are indicator variables for the presence of specific industrial activities in 1880 (as indicated for the columns) and the key regressors are the shares corresponding to the relevant agricultural products in 1860 (as indicated for the rows). Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

# Appendix C. A Model of Agricultural Diversity, Structural Change and Long-run Development

The multi-sector model of growth and structural change presented here offers one possible formal representation of the effects of agricultural diversity on development. A highlight of this representation is that entry into new sectors and the acquisition of new skills and ideas are interconnected dimensions of the process of structural change. The model predicts that diversity of skills positively affects industrial variety and overall industrialization, and has larger effects on sectors that require more skills. Moreover, initial differences in diversity lead to differences in long-run development.

I consider a small local economy, open to trade and labor flows. Free trade implies perfectly elastic demands for all goods. This allows me to abstract from the demand side and focus on the production side; consumption decisions play no role in the model. Meanwhile, under the simplest form of spatial equilibrium (with no amenities or congestion costs), free labor mobility implies that wages are equalized across locations. National wage levels—to which local wages are equalized—are taken as given.

These simplifying assumptions imply that productivity differentials across locations lead to differences in population density, without affecting wages. If labor was not perfectly mobile or if higher density generated congestion and increased living costs, then locations with higher productivity would have not only higher population density but also higher wages; the insights of the model about the effects of diversity on sectoral productivities, structural change and skills formation would be the same.

The key assumption underlying the persistent nature of the effects of diversity is that, in contrast to goods and labor, skills do not flow across locations. This stark assumption makes the model simpler and more transparent, but the implications are qualitatively the same if skills are assumed to be imperfectly immobile rather than completely immobile.

In the model, agricultural diversity is given by exogenous climatic conditions, and it determines the initial variety of skills. Industrial production comprises many sectors, each of which may be active or not, depending on expected profits. Skills required in each sector are complementary, and there are cross-sector spillovers within the local economy. Efficiency for each skill required in a sector depends on the set of established skills in other sectors.

In the presence of these complementarities and spillovers, higher diversity—which implies a wider range of skills—increases expected efficiency and profits, thus favoring entry into new sectors. In turn, expanding the range of active industrial sectors goes together with adoption of new technologies and formation of new skills. The essential source of growth is structural change, not only from agriculture to manufacturing, but most importantly within

manufacturing—this is what drives the formation of new skills.

The model's emphasis on diversity of skills, and the mechanism whereby it fosters the acquisition of additional skills, echo Hausmann and Hidalgo (2011). The characterization of complementarities among skills resembles Kremer (1993), while that of cross-sector spillovers resembles Helsley and Strange (2002). Entry decisions are modelled similarly to innovation decisions in many endogenous growth models.

## C.1  The Structure of Production

Production in the local economy comprises agricultural production and industrial production. Producers of all goods can sell as much as they want in the national market at a price equal to one (units of each good are defined to conform to this normalization), and can hire as much labor as they want at wage level $\bar{w}_t$, which is taken as given in the local economy. To keep notation simpler, the time subindex $t$ and the local economy subindex $c$ are dropped whenever it does not create confusion.

*Agricultural Production*

Agricultural production in a plot of land $p$ is given by $Y_p = (A_p x_p)^{1-\sigma} L_p^\sigma$, where $\sigma \in (0,1)$, $A_p$ is land productivity, $x_p$ is land size, and $L_p$ is labor hired in plot $p$. To maximize profits (given by $\Pi_p = Y_p - \bar{w} L_p$), the owner of plot $p$ hires $L_p^* = \left(\frac{\sigma}{\bar{w}}\right)^{\frac{1}{1-\sigma}} A_p x_p$.

Total agricultural employment in the local economy is $L_a^* = \left(\frac{\sigma}{\bar{w}}\right)^{\frac{1}{1-\sigma}} \bar{A}_a X$, where $X = \sum_p x_p$ is total land, $\bar{A}_a = \sum_p \frac{x_p}{X} A_p$ is the average land productivity in the local economy. The distribution of land ownership does not affect total agricultural employment and production.

Agricultural output comprises a number of different crops. Agricultural diversity is determined by exogenous climatic features; for simplicity, it is taken as a primitive rather than explicitly modeled as the outcome of optimal crop choice.

*Manufacturing Production*

There are several industrial sectors (indexed by $i$), each of which may be active or not. Production in sector $i$ is $Y_i = A_i^{1-\alpha} L_i^\alpha$, where $A_i$ reflects sectoral productivity, $L_i$ is labor employed in sector $i$, and $\alpha \in (0,1)$ reflects decreasing returns to labor (caused by limited entrepreneurial ability or some other fixed factor).

An active producer in sector $i$ maximizes profits ($\Pi_i = Y_i - \bar{w} L_i$), which determines the optimal quantity of labor, $L_i^* = \left(\frac{\alpha}{\bar{w}}\right)^{\frac{1}{1-\alpha}} A_i$, as well as equilibrium output, $Y_i^* = \left(\frac{\alpha}{\bar{w}}\right)^{\frac{\alpha}{1-\alpha}} A_i$, and profits, $\Pi_i^* = (1-\alpha)\left(\frac{\alpha}{\bar{w}}\right)^{\frac{\alpha}{1-\alpha}} A_i$. Sectoral productivities are determined by local skills, as explained below.

*Productivity and Local Skills*

Productivity in sector $i$ in local economy $c$ at time $t$ depends on the level of technology ($T_t$) and on the level of efficiency ($E_{ict}$):

$$A_{ict} = T_t \times E_{ict} \ .$$

The level of technology is the same across local economies (reflecting a common technology frontier) and across sectors (for simplicity), i.e., $T_{ict} = T_t$. Thus, productivity differences across locations and sectors are driven by differences in $E_{ict}$.

Efficiency levels in the different sectors of a local economy are determined by local skills. A key assumption is that skills are not mobile across locations.[23] The diversity of local skills is crucial because of the presence of skill complementarities and cross-sector spillovers, which are explained below.

Each sector $i$ requires a set of specific skills indexed by $j_i = 1, ..., J_i$. Some sectors require only one or two skills, others require several: $J_i$ captures the level of *complexity* of sector $i$. Sectoral skill requirements are the same in all local economies and time periods. A sector-specific skill $j_i$ is performed with efficiency level $e_{j_i c t} \in [0, \bar{e}]$, which may vary across local economies and over time.

The number of established skills at time $t$, denoted by $\Omega_{c,t}$, includes all the skills required by active industrial sectors in the local economy as well as the skills established in the agricultural sector. The number of skills established in agriculture, $\Omega_{A,c}$, is assumed to be time-invariant and strictly increasing in the initial level of agricultural diversity. The initial condition, corresponding to a period before the onset of industrialization, is $\Omega_{c,0} = \Omega_{A,c}$. I use $\Omega_{-i,c,t}$ to denote the number of skills in all sectors other than $i$ ($\Omega_{A,c}$ is counted within $\Omega_{-i,c,t}$ for all $i$). In the rest of this section and in the next one, which focus on a single local economy, I drop the subindex $c$ to simplify notation.

Skills within each sector are complementary. Featuring an "O-ring" property (Kremer, 1993), overall efficiency in sector $i$ is given by

$$E_{it} = \prod_{j=1}^{J_i} e_{j_i t} \ .$$

There are cross-sectoral spillovers within each local economy. Efficiency for a particular skill is given by the distance of that skill to a previously established skill in some other sector:

---

[23]This assumption is reasonable insofar as (i) skills cannot be codified and transmitted as disembodied knowledge, (ii) local skills are an attribute of locations rather than individual workers, i.e., hiring an individual worker from another location is not enough to establish the expertise available in that location, and (iii) hiring a large group of workers from another location is not possible because of coordination problems. The implications of the model are qualitatively the same if skills are assumed to be imperfectly immobile rather than completely immobile.

the local economy has high (low) levels of efficiency for doing things that are close (far) to what the economy was previously doing. More precisely, local efficiency for skill $j$ required by sector $i$ is given by $e_{j_i t} = \bar{e} - d_{j_i}$, where $d_{j_i}$ is the distance between skill $j_i$ and the closest previously established skill in a different sector.

Skills lie on a circle with radius $\bar{e}/\pi$ and their positions are independent draws from a uniform distribution on the circle. Thus, the pdf of $d_{j_i}$ is $f(d) = \Omega_{-i,t\text{-}1}\bar{e}^{-\Omega_{-i,t\text{-}1}}(\bar{e} - d)^{\Omega_{-i,t\text{-}1}-1}$, and $\mathbb{E}\left[d(j_i)\right] = \frac{\bar{e}}{1+\Omega_{-i,t\text{-}1}}$. Distances between any two locations in the circle are between zero and $\bar{e}$, ensuring that $e_{j_i t} \in [0, \bar{e}]$.

Since the draws for the distances of different skills required by a sector and the closest ones in other sectors are independent, expected efficiency in sector $i$ is given by

$$\mathbb{E}(E_{it}) = \mathbb{E}\left(\prod_j^{J_i} e_{j_i t}\right) = [\mathbb{E}(e_{j_i t})]^{J_i} = \bar{e}^{J_i}\left[1 - \frac{1}{1+\Omega_{-i,t\text{-}1}}\right]^{J_i} \;.$$

*Active industrial sectors*

There is one potential entrant in each industrial sector per period, who is assumed to be risk neutral. At the time of making the entry decision, the potential entrant only knows the expected level of efficiency. The minimum distances of required skills with respect to previously established ones become known only after paying the entry cost. Entrants produce for one period and then retire.

Entry is costly. Adopting frontier technologies requires micro-innovations and adaptation to the local economy. Thus, the number of active sectors in a given period is a measure of local technological dynamism—in this model, entry into new sectors is intrinsically linked with the acquisition of new ideas and the formation of skills. The entry cost, $F_t = \psi T_t^{1-\alpha}$, increases with the level of technology, reflecting that more advanced technologies are costlier to implement.

A sector becomes active if expected profits are larger than the entry cost. This defines a threshold in expected efficiency above which entry takes place:

$$\mathbb{E}(\Pi_{it}) \geq F_t \quad \Longleftrightarrow \quad \mathbb{E}(E_{it}) \geq \widetilde{E}_{it} = \frac{\psi}{(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}}\left(\frac{\bar{w}_t}{T_t^{1-\alpha}}\right)^{\frac{\alpha}{1-\alpha}} \;.$$

Assuming that $g_{\bar{w}_t} = g_{T^{1-\alpha}}$, which ensures that wages and productivity grow at the same rate in the steady state, the threshold in expected efficiency is time-invariant, i.e., $\widetilde{E}_{it} = \widetilde{E}_i$.

The threshold for entry into sector $i$ can be expressed as a threshold in the number of established skills in other sectors:

$$\widetilde{\Omega}_i = \frac{1}{\left(\frac{\bar{e}_i^J}{\widetilde{E}_i}\right)^{1/J_i} - 1} \;.$$

The number of active industrial sectors can then be expressed as $\sum_{i=1}^{I} \mathbb{1}_{\left[\Omega_{-i,t\text{-}1} \geq \widetilde{\Omega}_i\right]}$, where $\mathbb{1}_{\left[\Omega_{-i,t\text{-}1} \geq \widetilde{\Omega}_i\right]}$ is an indicator function that takes a value of 1 when $\Omega_{-i,t\text{-}1} \geq \widetilde{\Omega}_i$, i.e., when sector $i$ is active, and 0 otherwise. Sectoral expected employment is given by $\mathbb{E}(L_{it}) = \mathbb{1}_{\left[\Omega_{-i,t\text{-}1} \geq \widetilde{\Omega}_i\right]} \times \mathbb{E}(L_{it}^*)$, where $\mathbb{E}(L_{it}^*) = \left(\frac{\alpha}{\bar{w}}\right)^{\frac{1}{1-\alpha}} \mathbb{E}(A_i)$ is expected optimal employment in an active sector.

## C.2    Structural Change

This subsection explains how the diversity of a local economy affects the evolution of its production structure. I show that entry into any given industrial sector positively depends on the number of previously established skills in other sectors; so does the expected level of employment in the sector once activated, and more acutely so for complex sectors. Agricultural diversity, insofar as it increases the variety of local skills, positively affects entry and employment for all industrial activities. Thus, as I also show, agricultural diversity increases the total number of active industrial sectors and total expected employment in manufacturing.

In this framework, there is a close connection between diversity in production and diversity in skills: while the number of active industrial sectors is given by $\sum_{i=1}^{I} \mathbb{1}_{\left[\Omega_{-i,t\text{-}1} \geq \widetilde{\Omega}_i\right]}$, the number of established skills in the local economy is given by $\sum_{i=1}^{I} \mathbb{1}_{\left[\Omega_{-i,t\text{-}1} \geq \widetilde{\Omega}_i\right]} \times J_i + \Omega_A$. This connection underlies the mechanics of endogenous structural change. The variety of local skills—which reflects the existing variety of production activities—favors entry into new industrial activities in the next period. In turn, increasing industrial variety implies increasing skills variety, which favors further diversification in subsequent periods.

The following two lemmas establish that the number of previously existing skills in sectors other than $i$ (i.e., $\Omega_{-i,t\text{-}1}$) has positive effects on expected efficiency in sector $i$ (i.e., $\mathbb{E}(E_{it})$), and that the magnitude of these effects varies with sectoral levels of complexity.

LEMMA 1. The elasticity of expected efficiency in sector $i$ with respect to the number of previously existing skills in other sectors is positive: $\eta_{\mathbb{E}(E_{it}),\Omega_{-i,t\text{-}1}} \equiv \frac{\Delta \mathbb{E}(E_{it})/\mathbb{E}(E_{it})}{\Delta \Omega_{-i,t\text{-}1}/\Omega_{-i,t\text{-}1}} > 0$ .

PROOF. See section C.4.

The intuition behind Lemma 1 is that in a more diverse economic environment (characterized by a higher number of established skills) there is a higher likelihood of finding existing skills that are close to the ones required by any given sector.

Lemma 2 below captures the differential impact of diversity on expected efficiency in complex sectors, which have production chains with more links and thus are more sensitive

to the links' average weakness.

LEMMA 2. The elasticity of expected efficiency in sector $i$ with respect to the number of previously existing skills in other sectors is increasing in the sector's complexity: $\Delta\eta_{\mathbb{E}(E_{it}),\Omega_{-i,t-1}}/\Delta J_i > 0$ .

PROOF. See section C.4.

The following two propositions describe how entry into a sector and sectoral employment levels depend on the variety of skills previously established in other sectors.

PROPOSITION 1. (a) The number of previously established skills in sectors other than $i$ has a weakly positive effect on entry into sector $i$: $\Delta\mathbb{1}_{[\Omega_{-i,t-1}\geq\widetilde{\Omega}_i]}/\Delta\Omega_{-i,t-1} \geq 0$ . (b) The elasticity of expected employment in active sector $i$ with respect to the number of previously established skills in other sectors is positive: $\eta_{\mathbb{E}(L_{it}^*),\Omega_{-i,t-1}} \equiv \frac{\Delta\mathbb{E}(L_{it}^*)/\mathbb{E}(L_{it}^*)}{\Delta\Omega_{-i,t-1}/\Omega_{-i,t-1}} > 0.$

PROOF. (a) See section C.4. (b) Since $\mathbb{E}(L_{it}^*)$ is a linear function of $\mathbb{E}(E_{it})$, we have that $\eta_{\mathbb{E}(L_{it}^*),\Omega_{-i,t-1}} = \eta_{\mathbb{E}(E_{it}),\Omega_{-i,t-1}}$. Thus, Proposition 1 follows directly from Lemma 1. □

PROPOSITION 2. The elasticity of expected employment in active sector $i$ with respect to the number of previously existing skills in other sectors is increasing in the sector's complexity: $\Delta\eta_{\mathbb{E}(L_{it}^*),\Omega_{-i,t-1}}/\Delta J_i > 0.$

PROOF. Since $\eta_{\mathbb{E}(L_{it}^*),\Omega_{-i,t-1}} = \eta_{\mathbb{E}(E_{it}),\Omega_{-i,t-1}}$, Proposition 2 follows directly from Lemma 2. □

Given that $\Delta\Omega_{-i,t-1}/\Delta\Omega_A > 0$, Proposition 1 implies that agricultural diversity has positive effects on entry and expected employment for any industrial sector $i$ (i.e., $\eta_{\mathbb{E}(L_{it}^*),\Omega_A} \equiv \frac{\Delta\mathbb{E}(L_{it}^*)/\mathbb{E}(L_{it}^*)}{\Delta\Omega_A/\Omega_A} > 0$ and $\Delta\mathbb{1}_{[\Omega_{-i,t-1}\geq\widetilde{\Omega}_i]}/\Delta\Omega_A \geq 0$), as well as differentially positive effects in complex sectors ($\Delta\eta_{\mathbb{E}(L_{it}^*),\Omega_A}/\Delta J_i > 0$). Thus, agricultural diversity, by increasing the variety of local skills, positively affects industrial activities both in the extensive and the intensive margin. Proposition 2 captures the differential effects of diversity in complex activities, which motivates the empirical analysis of section 6.3.

The following proposition establishes the positive effects of agricultural diversity on the number of active industrial sectors and total (expected) industrial employment.

PROPOSITION 3. Consider two local economies, $s$ and $z$, that in a given period $t-1$ are identical in all respects except that $s$ has higher agricultural diversity, so $\Omega_{A,s} > \Omega_{A,z}$. In

period $t$, economy $s$ has a weakly higher number of active industrial sectors and a weakly higher expected value of total industrial employment than economy $z$: $\sum_{i=1}^{I} \mathbb{1}_{\left[\Omega_{-i,s,t-1} \geq \widetilde{\Omega}_i\right]} \geq \sum_{i=1}^{I} \mathbb{1}_{\left[\Omega_{-i,z,t-1} \geq \widetilde{\Omega}_i\right]}$ and $\mathbb{E}\left(\sum_{i=1}^{I} L_{i,s,t}\right) \geq \mathbb{E}\left(\sum_{i=1}^{I} L_{i,z,t}\right)$.

PROOF. See section C.4.

## C.3   Long-run Development

In this subsection, I analyze the process of skills formation and structural change, and establish the effects of initial diversity on long-run economic performance.

First, note that the collection of sector-specific thresholds can be arranged in increasing order; this sequence can then be used to define a new index spanning industrial sectors, $q = 1, ..., I$, where sectors with high $q$ have high entry thresholds. If there are sectors with the same entry threshold $\widetilde{\Omega}_i$, the $q$-index assigns them contiguous integers in arbitrary order. Arranging sectors by their entry thresholds is equivalent to arranging them by timing of entry: sectors with high entry thresholds enter production relatively late, if at all.

The equation for the number of skills at time $t$ can be rewritten using the $q$-index. This equation, together with the initial condition (with $t=0$ preceding the onset of industrialization), fully characterizes the evolution of local skills:

$$\Omega_{ct} = \sum_{q=1}^{I} \mathbb{1}_{\left[\Omega_{-q,t-1} \geq \widetilde{\Omega}_q\right]} \times J_q + \Omega_{A,c}$$

$$\Omega_{c0} = \Omega_{A,c}$$

This characterization of the evolution of skills (which determines the evolution of sectoral levels of employment and production) is instrumental to define the *rest points* and the *steady state* of the local economy.

DEFINITION 1. (*Rest point*) Industrial sector $q = m$ (with $m \geq m'$ for any $m'$ such that $\widetilde{\Omega}_{m'} = \widetilde{\Omega}_m$) is a *rest point* of the local economy if $\sum_{q=1}^{m} J_q + \Omega_{A,c} < \widetilde{\Omega}_{m+1}$. In addition, industrial sector $q = I$ is a *rest point*.

When the local economy reaches a rest point, the process of structural change reaches an end. This follows from the definition. If the economy has entered production in all sectors with $q \leq m$ (with sector $m$ having the highest arbitrarily assigned $q$-index value among sectors with the same entry threshold), then the number of established skills is $\sum_{q=1}^{m} J_q + \Omega^A$. If this is below the threshold required to enter production in sector $m+1$ (i.e., $\widetilde{\Omega}_{m+1}$), then the

process of industrial diversification cannot go beyond sector $m$. In addition, the process of structural change cannot go beyond sector $q = I$ (the one with the highest entry threshold), since once that sector becomes active there are no inactive sectors left to enter.

A local economy may have more than one rest point, but it is the first one (the rest point with lowest $q$) that matters, as it defines the *steady state* of the economy.

DEFINITION 2. (*Steady state*) The rest point with the lowest $q$ defines the *steady state* of the local economy. The sector defining the steady state is denoted by $q_c^*$, and the number of skills in the steady state is $\Omega_c^* = \sum_{q=1}^{q_c^*} J_q + \Omega_{A,c}$.

When the economy enters production in sector $q_c^*$, the process of structural change concludes. While the economy then continues to grow at a rate given by (exogenous) technological progress in the frontier, endogenous growth through diversification is shut down. The steady state is characterized by a constant growth rate and a stable composition of production.

The next proposition establishes the effects of agricultural diversity on comparative economic performance in the long-run.

PROPOSITION 4. Consider two local economies, $s$ and $z$, that are identical in all respects except that $s$ has higher agricultural diversity, so $\Omega_{A,s} > \Omega_{A,z}$. In the steady state, economy $s$ has a weakly higher number of active industrial sectors and a higher number of established skills than economy $z$: $q_s^* \geq q_z^*$ and $\Omega_s^* > \Omega_z^*$.

PROOF. See section C.4.

## C.4 Proofs

PROOF OF LEMMA 1.

Using the definition of $\eta_{\mathbb{E}(E_{it}),\Omega_{-i,t-1}}$ and equation (1), we have that

$$\eta_{\mathbb{E}(E_{it}),\Omega_{-i,t-1}} = \frac{\left(\frac{1+\Delta\Omega_{-i,t-1}/\Omega_{-i,t-1}}{1+\Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}+1)}\right)^{J_i} - 1}{\Delta\Omega_{-i,t-1}/\Omega_{-i,t-1}}.$$

Since $\Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}) > \Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}+1)$ and $J_i \geq 1$, we have $\left(\frac{1+\Delta\Omega_{-i,t-1}/\Omega_{-i,t-1}}{1+\Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}+1)}\right)^{J_i} > 1$, and thus $\eta_{\mathbb{E}(E_{it}),\Omega_{-i,t-1}} > 0$. □

PROOF OF LEMMA 2.

Using the expression for $\eta_{\mathbb{E}(E_{it}),\Omega_{-i,t-1}}$ from the proof of Lemma 1, we have that

$$\frac{\Delta\eta_{\mathbb{E}(E_{it}),\Omega_{-i,t-1}}}{\Delta J_i} = \frac{\left(\frac{1+\Delta\Omega_{-i,t-1}/\Omega_{-i,t-1}}{1+\Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}+1)}\right)^{J_i+\Delta J_i} - \left(\frac{1+\Delta\Omega_{-i,t-1}/\Omega_{-i,t-1}}{1+\Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}+1)}\right)^{J_i}}{\Delta\Omega_{-i,t-1}/\Omega_{-i,t-1}} \cdot \frac{1}{\Delta J_i} \; .$$

Noting that $\Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}) > \Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}+1)$ and $J_i \geq 0$, and also that $\Delta J_i \geq 1$, we have that $\left(\frac{1+\Delta\Omega_{-i,t-1}/\Omega_{-i,t-1}}{1+\Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}+1)}\right)^{J_i+\Delta J_i} > \left(\frac{1+\Delta\Omega_{-i,t-1}/\Omega_{-i,t-1}}{1+\Delta\Omega_{-i,t-1}/(\Omega_{-i,t-1}+1)}\right)^{J_i}$, and thus $\frac{\Delta\eta_{\mathbb{E}(E_{it}),\Omega_{-i,t-1}}}{\Delta J_i} > 0$ . □

PROOF OF PROPOSITION 1(a).

For $\Delta\Omega_{-i,t-1} > 0$, if $\Omega_{-i,t-1} \geq \widetilde{\Omega}_i$, then $\Omega_{-i,t-1} + \Delta\Omega_{-i,t-1} \geq \widetilde{\Omega}_i$; thus, if $\mathbb{1}_{\left[\Omega_{-i,t-1}\geq\widetilde{\Omega}_i\right]} = 1$, then $\mathbb{1}_{\left[\Omega_{-i,t-1}+\Delta\Omega_{-i,t-1}\geq\widetilde{\Omega}_i\right]} = 1$. This implies that $\frac{\Delta\mathbb{1}_{\left[\Omega_{-i,t-1}\geq\widetilde{\Omega}_i\right]}}{\Delta\Omega_{-i,t-1}} = \frac{\mathbb{1}_{\left[\Omega_{-i,t-1}+\Delta\Omega_{-i,t-1}\geq\widetilde{\Omega}_i\right]} - \mathbb{1}_{\left[\Omega_{-i,t-1}\geq\widetilde{\Omega}_i\right]}}{\Delta\Omega_{-i,t-1}} \geq 0$. The inequality is strict in cases where $\Delta\Omega_{-i,t-1} > \widetilde{\Omega}_i - \Omega_{-i,t-1} > 0$. Using the same logic it is straightforward to show that for $\Delta\Omega_{-i,t-1} < 0$, we have $\Delta\mathbb{1}_{\left[\Omega_{-i,t-1}\geq\widetilde{\Omega}_i\right]} \leq 0$. □

PROOF OF PROPOSITION 3.

Since economies $s$ and $z$ are identical in all respects except that $\Omega_{A,s} > \Omega_{A,z}$, then for each sector $i$ we have that $\Omega_{-i,s,t-1} > \Omega_{-i,z,t-1}$ and thus $\mathbb{1}_{\left[\Omega_{-i,s,t-1}\geq\widetilde{\Omega}_i\right]} \geq \mathbb{1}_{\left[\Omega_{-i,z,t-1}\geq\widetilde{\Omega}_i\right]}$. This implies that $\sum_{i=1}^{I} \mathbb{1}_{\left[\Omega_{-i,s,t-1}\geq\widetilde{\Omega}_i\right]} \geq \sum_{i=1}^{I} \mathbb{1}_{\left[\Omega_{-i,z,t-1}\geq\widetilde{\Omega}_i\right]}$. In regards to total expected industrial employment, note first that it can be expressed as $\mathbb{E}\left(\sum_{i=1}^{I} L_{i,c,t}\right) = \sum_{i=1}^{I} \mathbb{E}\left(L_{i,c,t}\right) = \sum_{i=1}^{I} \mathbb{1}_{\left[\Omega_{-i,c,t-1}\geq\widetilde{\Omega}_i\right]} \times \mathbb{E}\left(L_{i,c,t}^*\right)$. Next, note that each of the two factors in every term of this summation is (at least weakly) higher in economy $s$: $\mathbb{1}_{\left[\Omega_{-i,s,t-1}\geq\widetilde{\Omega}_i\right]} \geq \mathbb{1}_{\left[\Omega_{-i,z,t-1}\geq\widetilde{\Omega}_i\right]}$ for all $i$ as shown above; and given that $\eta_{\mathbb{E}(L_{it}^*),\Omega_A} > 0$, the inequality $\Omega_{A,s} > \Omega_{A,z}$ also implies that $\mathbb{E}\left(L_{i,s,t}^*\right) > \mathbb{E}\left(L_{i,z,t}^*\right)$. Thus, we have that $\mathbb{E}\left(\sum_{i=1}^{I} L_{i,s,t}\right) \geq \mathbb{E}\left(\sum_{i=1}^{I} L_{i,z,t}\right)$ □

PROOF OF PROPOSITION 4.

If $m$ is a rest point for economy $s$, this means that $\sum_{q=1}^{m} J_q + \Omega_{A,s} < \widetilde{\Omega}_{m+1}$. And since $\Omega_{A,z} < \Omega_{A,s}$, then $\sum_{q=1}^{m} J_q + \Omega_{A,z} < \widetilde{\Omega}_{m+1}$, which means that $m$ is a also rest point for economy $z$. Since any rest point for economy $s$ is also a rest point for economy $z$ (which may have additional rest points), then the sector defining the steady state of economy $s$ (the rest point with lower $q$) must be at least as high (in terms of $q$) as the one for economy $z$, i.e. $q_s^* \geq q_z^*$. Thus, we have that $\Omega_s^* = \sum_{q=0}^{q_s^*} J_q + \Omega_{A,s} > \sum_{q=0}^{q_z^*} J_q + \Omega_{A,z} = \Omega_z^*$ □

# Appendix D. Other channels

## D.1 Agricultural Productivity

The positive effect of early agricultural diversity on industrialization may have been generated, at least partly, through an effect of agricultural diversity on agricultural productivity. Maybe agricultural diversity increased agricultural productivity, which in turn pushed labor out of agriculture? Or maybe agricultural diversity negatively affected agricultural productivity, and this was conducive to industrialization? For agricultural productivity to be a channel underlying the observed empirical patterns, the effect of agricultural diversity on agricultural productivity and the effect of the latter on industrialization need to be significant and have the same sign. I empirically assess the two links below, and find that the evidence does not support the relevance of this channel.

How would agricultural diversity affect agricultural productivity? A positive effect could operate through economies of scope in agricultural production (Paul and Nehring, 2005; Kim et al., 2012). Complementarities or positive externalities across products may arise from the beneficial use of byproducts (e.g., manure from livestock used as fertilizer) or from more efficient use of labor (e.g., if labor requirements for different crops have heterogeneous seasonal patterns). Diversity may also help to preserve soil quality over time (Russelle et al., 2007). Moreover, it could broaden the knowledge base and thus foster agricultural innovation and technology adoption. On the other hand, if agricultural diversity implies foregoing gains from specialization based on comparative advantage or scale economies, it could have negative effects on productivity.

In turn, agricultural productivity may affect industrialization in either direction, through various mechanisms. On the one hand, higher productivity in agriculture can release labor to be employed in manufacturing; it also means cheaper food for workers and cheaper inputs for firms; it creates resources for investment, which can be channeled towards industrial capital formation; and it means higher purchasing power and thus higher demand for local manufacturing production (Johnston and Mellor, 1961). On the other hand, Matsuyama (1992) demonstrates that in open economies agricultural productivity can have a negative effect on industrial growth by shifting comparative advantage in favor of agriculture (naturally, high transport costs would hold back this effect).

Throughout Appendix D (as in section 6), I report OLS and IV estimates for three specifications, controlling for (i) state fixed effects, (ii) state fixed effects and geo-climatic conditions, and (iii) state fixed effects, geo-climatic conditions, crop-specific controls, and socio-economic conditions. Panel A of Table D.1 shows estimates of the effect of agricultural

diversity in 1860 on farm output per improved acre (in logs) ten years later, right around the onset of the Second Industrial Revolution. Panel B displays estimates of the effect of this measure of agricultural productivity on the share of population employed in manufacturing in 1900, the measure of industrialization used in the main analysis. The IV estimates of the effects of agricultural productivity use the first principal component of the FAO-GAEZ normalized crop-specific attainable yields as an IV (in these estimations I exclude the measures of potential land productivity from the set of geo-climatic controls). The results do not offer support for the relevance of this channel: I do not find robust results indicating an effect of agricultural diversity on agricultural productivity nor an effect of the latter on industrialization.

TABLE D.1. AG.DIVERSITY, AG.PRODUCTIVITY, AND INDUSTRIALIZATION

| | Specification 1 | | Specification 2 | | Specification 3 | |
|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A.** *Dependent variable: Ln Farm Productivity, 1870* | | | | | | |
| Ag.Diversity$_{1860}$ | -0.438* | -0.185 | -0.286 | 0.104 | 0.0261 | 0.275 |
| | (0.229) | (0.315) | (0.219) | (0.363) | (0.136) | (0.514) |
| $R^2$ | 0.377 | 0.374 | 0.409 | 0.403 | 0.535 | 0.533 |
| **Panel B.** *Dependent variable: Share of population in manufacturing, 1900* | | | | | | |
| Ln Farm Productivity$_{1870}$ | 0.0136*** | -0.0478 | 0.0131*** | 0.0255 | -0.00191 | 0.00738 |
| | (0.00240) | (0.0519) | (0.00231) | (0.0172) | (0.00218) | (0.0239) |
| $R^2$ | 0.449 | 0.148 | 0.459 | 0.448 | 0.610 | 0.554 |
| State FE | Y | Y | Y | Y | Y | Y |
| Geo-climatic controls | N | N | Y | Y | Y | Y |
| Crop-specific controls | N | N | N | N | Y | Y |
| Socio-economic controls | N | N | N | N | Y | Y |
| Observations | 1,819 | 1,819 | 1,819 | 1,819 | 1,819 | 1,819 |

*Notes:* See Appendix A.1 for variable definitions and sources. In the regressions reported in Panel B, land productivity measures are not included in the set of geo-climatic controls; the IV estimations use the first principal component of the FAO-GAEZ normalized crop-specific attainable yields is used as IV. The sample is reduced to 1,819 as 2 counties with no farm output in 1870 drop out. The means of the dependent variables in panels A and B are 2.56 and 0.034, respectively. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

## D.2 Volatility, Risk and Local Financial Development

Positive effects of agricultural diversity could arise from reduced risk and volatility, as suggested by some of the contributions reviewed in section 2 (Acemoglu and Zilibotti, 1997; Koren and Tenreyro, 2013). Diversification dampens the direct impact of negative product-specific shocks and also facilitates substitution away from negatively affected products. Olmstead and Rhode (2008) document various negative shocks affecting specific agricultural products during the period under consideration, highlighting the potential relevance of this mechanism. Moreover, if agents are risk averse, being able to reduce risk and volatility by diversifying agricultural production may be a precondition for carrying out risky projects with high returns in the manufacturing sector. Note, however, that reduced agricultural volatility makes industrial investments comparatively less attractive, and thus an effect in the opposite direction is also possible.

Beyond the effects mentioned above, lower levels of risk afforded by diversity could affect local financial development. Insofar as it allows local banks to reduce risk exposure by funding a wide array of imperfectly correlated projects, diversity may increase credit supply (Ramcharan, 2010b). On the other hand, people in places that are well-suited to limit volatility through diversification may have lower demand for financial services. Finally, in addition to affecting local financial institutions (positively through supply or negatively through demand), diversity could influence social norms: higher volatility may foster mutual insurance arrangements and thus help to build trust (Durante and Buggle, 2016). In turn, local financial development and trust may affect economic development through channels other than risk management.

To assess whether agricultural diversity fostered industrialization through reduced volatility, I construct a measure of predicted volatility in the value of agricultural production by combining the initial county-level production mix and the year-to-year evolution of national prices between 1867 and 1900. More precisely, I calculate a predicted price index for a county's agricultural production in year $t$ as $P_{ct}^{\mathrm{a}} = \sum_i \theta_{ic} p_{it}$, where the $\theta_{ic}$'s reflect the county's agricultural production mix in 1860, and $p_{it}$ is the national price of product $i$ in period $t$ (with $p_{i1867}$ normalized to 1 for all $i$). I then calculate the rate of change $\widehat{P}_{ct}^{\mathrm{a}} = (P_{ct}^{\mathrm{a}} - P_{ct-1}^{\mathrm{a}})/P_{ct-1}^{\mathrm{a}}$ for each year, as well as its average and standard deviation over the 1867-1900 period, $Avg.\widehat{P}_{ct}^{\mathrm{a}}$ and $Std.Dev.(\widehat{P}_{ct}^{\mathrm{a}})$.[24]

---

[24] Annual price data from 1867 to 1900 are available for 15 products that account for over 93.5% of total agricultural output in 1860 in my sample (see Appendix A.1 for details). Only products with available price data are considered in the calculation of $P_{ct}^{\mathrm{a}}$; the $\theta_{ic}$'s are computed with the restricted set of products so that they add up to 1. The results are qualitatively similar in a reduced sample of 1,626 counties for which over 90% of agricultural production in each county corresponds to the 15 products with available price

Constructed in this way, $\widehat{P}_{ct}^{\mathrm{a}}$ captures the predicted effect of national-level price changes in a county's agricultural production value given the local production mix, and $Std.Dev.(\widehat{P}_{ct}^{\mathrm{a}})$ captures the volatility of predicted agricultural production value. While county-level volatility naturally depends on the specific local production mix and the covariances of national-level price changes, $Std.Dev.(\widehat{P}_{ct}^{\mathrm{a}})$ has a strong negative association with agricultural diversity (the correlation coefficient in the sample is -0.37).

The relevance of the volatility channel can be assessed by comparing estimates of the effects of agricultural diversity in 1860 on industrialization in 1900 obtained with and without controlling for $Std.Dev.(\widehat{P}_{ct}^{\mathrm{a}})$ (together with $Avg.\widehat{P}_{ct}^{\mathrm{a}}$ to avoid potential omitted variable bias). If diversity positively affected industrialization by reducing volatility, $Std.Dev.(\widehat{P}_{ct}^{\mathrm{a}})$ would have a negative and significant coefficient, and its inclusion would reduce the coefficient on agricultural diversity, possibly making it insignificant (if this channel was the only relevant one). The results of the estimations, displayed in Table D.2, are not in line with these predictions. The estimated effect of $St.Dev.(\widehat{P}_{ct}^{\mathrm{a}})$ is not significant and the estimated coefficient on 1860 agricultural diversity remains significant and of similar magnitude.

I can also examine whether diversity affected financial development. Table D.3 presents estimates of the effects of agricultural diversity in 1860 on county-level bank density (the number of banks per capita) in 1920, the measure considered by Rajan and Ramcharan (2011) in their study of local financial development in this period. The results do no show significant causal effects.

## D.3 Land Concentration and Local Institutions

Agricultural diversity may have affected the manufacturing sector by shaping the distribution of land ownership. If there are increasing returns to scale at the product-level (e.g., fixed costs due to crop-specific capital or skills), then places with low potential diversity (e.g, with very high relative productivity for a single product) may be more favorable for the development of large-scale farms. In addition, if economies of scope (the benefits of diversification) at the farm-level are higher for small farms (Chavas and Aliber, 1993), then places with high potential diversity would be favorable to small farms. On the other hand, if different crops are characterized by different optimal production scales, then diversity could be positively associated with inequality in farm size.

In turn, the presence of large landowners may retard the emergence of human capital promoting institutions (Galor et al., 2009) and/or hinder local financial development (Rajan

---

data. Results are also qualitatively similar if I consider the standard deviation of $P_{ct}^{\mathrm{a}}$ instead of the standard deviation of $\hat{P}_{ct}^{\mathrm{a}}$.

and Ramcharan, 2011), thus negatively affecting the process of industrialization. Galor et al. (2009) provide a panel data analysis at the U.S. state-level from 1880 to 1940 showing that concentration in land ownership had a significant adverse effect on educational expenditures; this was a period characterized by a massive expansion of secondary education, which—as suggested by the model presented in their paper—was key for industrialization. Ramcharan (2010a) and Vollrath (2013) provide evidence in the same direction from U.S. county-level data for the same period.

TABLE D.2. ASSESSING THE RISK AND VOLATILITY CHANNEL

| | Dependent variable: Share of population in manufacturing, 1900 | | | | | |
|---|---|---|---|---|---|---|
| | Specification 1 | | Specification 2 | | Specification 3 | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A.** OLS estimates | | | | | | |
| Ag.Diversity$_{1860}$ | 0.0358*** | 0.0387*** | 0.0404*** | 0.0440*** | 0.0333*** | 0.0341*** |
| | (0.00949) | (0.0107) | (0.00919) | (0.0109) | (0.00825) | (0.0117) |
| $Std.Dev.(\widehat{P}_{ct}^{a})$ | | -0.0285 | | 0.00719 | | 0.0238 |
| | | (0.0817) | | (0.0808) | | (0.0751) |
| $Avg.\widehat{P}_{ct}^{a}$ | | 0.484 | | 0.263 | | -0.143 |
| | | (0.586) | | (0.563) | | (0.638) |
| $R^2$ | 0.439 | 0.440 | 0.461 | 0.461 | 0.618 | 0.618 |
| **Panel B.** IV estimates | | | | | | |
| Ag.Diversity$_{1860}$ | 0.0719*** | 0.0764*** | 0.0830*** | 0.0890*** | 0.0884*** | 0.111** |
| | (0.0219) | (0.0247) | (0.0213) | (0.0237) | (0.0309) | (0.0435) |
| $Std.Dev.(\widehat{P}_{ct}^{a})$ | | 0.0510 | | 0.103 | | 0.206 |
| | | (0.0897) | | (0.0911) | | (0.130) |
| $Avg.\widehat{P}_{ct}^{a}$ | | 0.499 | | 0.249 | | 0.115 |
| | | (0.582) | | (0.570) | | (0.661) |
| $R^2$ | 0.433 | 0.434 | 0.452 | 0.453 | 0.607 | 0.604 |
| State FE | Y | Y | Y | Y | Y | Y |
| Geo-climatic controls | N | N | Y | Y | Y | Y |
| Crop-specific controls | N | N | N | N | Y | Y |
| Socio-economic controls | N | N | N | N | Y | Y |
| Observations | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 |

*Notes:* $\bar{G}p_{c,t}^{a}$ is the annual percentage change of $p_{c,t}^{a} = \sum_i \hat{\theta}_{ic} p_{i,t}$, a predictor of the value of agricultural output constructed with predicted shares for 1860 and subsequent national prices. $Std.Dev.(Gp_{c,t}^{a})$ is its standard deviation. See Appendix A for other variable definitions and sources. The mean of the share of population in manufacturing in 1900 is 0.034. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

| | Dependent variable: Bank density, 1920 | | | | | |
|---|---|---|---|---|---|---|
| | Specification 1 | | Specification 2 | | Specification 3 | |
| | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Ag.Diversity$_{1860}$ | -0.0000424 | 0.000173 | -0.0000815* | 0.0000652 | -0.0000890 | -0.0000143 |
| | (0.0000544) | (0.000118) | (0.0000490) | (0.000109) | (0.0000554) | (0.000166) |
| State FE | Y | Y | Y | Y | Y | Y |
| Geo-climatic controls | N | N | Y | Y | Y | Y |
| Crop-specific controls | N | N | N | N | Y | Y |
| Socio-economic controls | N | N | N | N | Y | Y |
| Observations | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 | 1,821 |
| $R^2$ | 0.622 | 0.616 | 0.649 | 0.646 | 0.691 | 0.691 |

*Notes:* See Appendix A.1 for variable definitions and sources. The mean of bank density in 1920 is 0.00038. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Rajan and Ramcharan (2011) argue that landed elites may hinder the development of local banks to maintain their power. Their paper shows that in the early 20th century U.S., counties with higher land inequality had fewer banks per capita and less access to credit. Even if large landowners did not obstruct financial development, high land inequality could imply that many prospective borrowers had limited access to credit because of insufficient collateral (e.g., Chakraborty and Ray, 2007).[25]

Did agricultural diversity lower land concentration? Panel A of Table D.4 shows estimates of the effects of agricultural diversity in 1860 on the share of farmland corresponding to the top 10% largest farms ten years later, around the onset of the Second Industrial Revolution. The results do not indicate a significant causal effect of agricultural diversity on land concentration.

I can also assess whether agricultural diversity affected local school expenditures and/or financial development, the outcomes that would be influenced by the presence of powerful landed elites according to the theories discussed above. Table D.3 (already discussed in section D.2) shows that there were no significant effects of early agricultural diversity on bank density in 1920. Likewise, using data on educational expenditures in 1890 from Rhode and Strumpf (2003), I find no evidence that these were significantly affected by 1860 agricultural

---

[25]Adamopoulos (2008) also argues that land concentration can hinder industrialization insofar as landed elites can influence policies to protect their rents. In his model, which is used to explain the divergence between Argentina and Canada, the policy that blocks industrialization is an import tariff on intermediate inputs required by manufacturing. Naturally, this mechanism cannot explain divergence across U.S. counties.

diversity, as indicated by the results displayed in Panel B of Table D.4. Thus, I conclude that the evidence does not support the relevance of this channel.

TABLE D.4. EFFECTS OF AG.DIVERSITY ON LAND CONCENTRATION AND PUBLIC SCHOOLING

| | Specification 1 | | Specification 2 | | Specification 3 | |
|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A.** *Dependent variable: Share of farmland in the top 10% largest farms, 1870* | | | | | | |
| Ag.Diversity$_{1860}$ | -0.00964 | -0.0658 | 0.0379 | 0.0301 | 0.0386 | -0.0157 |
| | (0.0271) | (0.0812) | (0.0289) | (0.0884) | (0.0329) | (0.137) |
| $R^2$ | 0.264 | 0.261 | 0.308 | 0.308 | 0.386 | 0.384 |
| Observations | 1,819 | 1,819 | 1,819 | 1,819 | 1,819 | 1,819 |
| **Panel B.** *Dependent variable: Ln School expenditures per capita, 1890* | | | | | | |
| Ag.Diversity$_{1860}$ | 0.150 | -0.289 | 0.0796 | -0.603 | 0.136 | -0.629 |
| | (0.107) | (0.400) | (0.0969) | (0.485) | (0.118) | (0.452) |
| $R^2$ | 0.797 | 0.794 | 0.811 | 0.805 | 0.812 | 0.805 |
| Observations | 1,809 | 1,809 | 1,809 | 1,809 | 1,809 | 1,809 |
| State FE | Y | Y | Y | Y | Y | Y |
| Geo-climatic controls | N | N | Y | Y | Y | Y |
| Crop-specific controls | N | N | N | N | Y | Y |
| Socio-economic controls | N | N | N | N | Y | Y |

*Notes:* See Appendix A.1 for variable definitions and sources. The means of the dependent variables in Panels A and B are 0.83 and 0.23, respectively. Robust standard errors clustered on 60-square-mile grid squares are reported in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

# References

Acemoglu, D., Moscona, J. and Robinson, J. A. (2016). State Capacity and American Technology: Evidence from the Nineteenth Century, *American Economic Review: Papers and Proceedings* **106**(5): 61–67.

Acemoglu, D. and Zilibotti, F. (1997). Was Prometheus unbound by chance? Risk, diversification, and growth, *Journal of Political Economy* **105**(4): 709–751.

Adamopoulos, T. (2008). Land inequality and the transition to modern growth, *Review of Economic Dynamics* **11**(2): 257–282.

Akcigit, U., Grigsby, J. and Nicholas, T. (2017). The Rise of American Ingenuity: Innovation and Inventors of the Golden Age, *NBER Working Paper No. 23047* .

Akcigit, U., Kerr, W. R. and Nicholas, T. (2013). The Mechanics of Endogenous Innovation and Growth: Evidence from Historical US Patents.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*, Princeton University Press.

Atack, J. (2015a). Historical Geographic Information Systems (GIS) database of Steamboat-Navigated Rivers During the Nineteenth Century in the United States.

Atack, J. (2015b). Historical Geographic Information Systems (GIS) database of U.S. Canals.

Atack, J. (2015c). Historical Geographic Information Systems (GIS) database of U.S. railroads.

Atack, J. and Bateman, F. (1987). *To their own soil: agriculture in the antebellum North*, Iowa State University Press.

Beaudry, C. and Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate, *Research Policy* **38**(2): 318–337.

Berliant, M. and Fujita, M. (2008). Knowledge Creation as a Square Dance on the Hilbert Cube, *International Economic Review* **49**(4): 1251–1295.

Berliant, M. and Fujita, M. (2011). The dynamics of knowledge diversity and economic growth, *Southern Economic Journal* **77**(4): 856–884.

Bester, C. A., Conley, T. G. and Hansen, C. B. (2011). Inference with dependent data using cluster covariance estimators, *Journal of Econometrics* **165**(2): 137–151.

Bleakley, H. and Lin, J. (2012). Portage and path dependence, *Quarterly Journal of Economics* **127**(2): 587–644.

Broadberry, S. N. and Irwin, D. A. (2006). Labor productivity in the United States and the United Kingdom during the nineteenth century, *Explorations in Economic History* **43**(2): 257–279.

Bruhn, M. and Gallego, F. A. (2012). Good, bad, and ugly colonial activities: do they matter for economic development?, *Review of Economics and Statistics* **94**(2): 433–461.

Bureau of Economic Analysis (2015). Table CA1 Personal Income Summary: Personal Income, Population, Per Capita Personal Income, http://www.bea.gov/iTable/iTableHtml.cfm?reqid=70&step=1&isuri=1.

Bustos, P., Caprettini, B. and Ponticelli, J. (2016). Agricultural productivity and structural trans-formation: Evidence from brazil, *American Economic Review* **106**(6): 1320–1365.

Chakraborty, S. and Ray, T. (2007). The development and structure of financial systems, *Journal of Economic Dynamics and Control* **31**(9): 2920–2956.

Chandler, A. (1977). *The Visible Hand: The Managerial Revolution in American Business*, Harvard University Press.

Chavas, J.-P. and Aliber, M. (1993). An analysis of economic efficiency in agriculture: a nonpara-metric approach, *Journal of Agricultural and Resource Economics* **18**(1): 1–16.

Conley, T. G. (1999). GMM estimation with cross sectional dependence, *Journal of econometrics* **92**(1): 1–45.

Cooley, T. F., DeCanio, S. J. and Matthews, M. S. (1977). An agricultural time series-cross section data set, *NBER Working Paper No. 197* .

Craig, L. A. (1993). *To sow one acre more: childbearing and farm productivity in the antebellum North*, Johns Hopkins University Press.

Cronon, W. (2009). *Nature's metropolis: Chicago and the Great West*, WW Norton & Company.

Cuadrado-Roura, J. R., García-Greciano, B. and Raymond, J. L. (1999). Regional convergence in productivity and productive structure: The Spanish case, *International Regional Science Review* **22**(1): 35–53.

David, P. A. (1990). The dynamo and the computer: an historical perspective on the modern productivity paradox, *American Economic Review* **80**(2): 355–361.

Desrochers, P. (2001). Local diversity, human creativity, and technological innovation, *Growth and change* **32**(3): 369–394.

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

Durante, R. and Buggle, J. (2016). Climate Risk, Cooperation and the Co-Evolution of Culture and Institutions.

Duranton, G. and Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products, *American Economic Review* **91**(5): 1454–1477.

Earle, C. and Hoffman, R. (1980). The Foundation of the Modern Economy: Agriculture and the Costs of Labor in the United States and England, 1800-60, *American Historical Review* **85**(5): 1055–1094.

Ellison, G., Glaeser, E. L. and Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns, *American Economic Review* **100**: 1195–1213.

Engerman, S. L. and Sokoloff, K. L. (2002). Factor Endowments, Inequality, and Paths of Devel-opment Among New World Economics, *NBER Working Paper No. 9259* .

Engerman, S. and Sokoloff, K. (1997). Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies, *in S. Haber (Ed), How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800-1914, pp. 260–340* pp. 260–340.

Fenske, J. (2014). Ecology, Trade, and States in Pre-Colonial Africa, *Journal of the European Economic Association* **12**(3): 612–640.

Galor, O. (2011). *Unified growth theory*, Princeton University Press.

Galor, O., Moav, O. and Vollrath, D. (2009). Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence, *Review of Economic Studies* **76**(1): 143–179.

Glaeser, E., Kallal, H. D., Scheinkman, J. and Shleifer, A. (1992). Growth in cities, *Journal of Political Economy* **100**(6): 1126–52.

Glaeser, E. L., Kerr, S. P. and Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines, *Review of Economics and Statistics* **97**(2): 498–520.

Goldin, C. D. and Katz, L. F. (2009). *The race between education and technology*, Harvard University Press.

Goldin, C. and Katz, L. F. (1998). The origins of technology-skill complementarity, *Quarterly Journal of Economics* **113**(3): 693–732.

Goldin, C. and Sokoloff, K. (1982). Women, children, and industrialization in the early republic: Evidence from the manufacturing censuses, *Journal of Economic History* **42**(04): 741–774.

Goldin, C. and Sokoloff, K. (1984). The relative productivity hypothesis of industrialization: The American case, 1820 to 1850, *Quarterly Journal of Economics* **99**(3): 461–487.

Gollin, D. (2010). Agricultural productivity and economic growth, *Handbook of agricultural economics* **4**: 3825–3866.

Gordon, R. J. (2016). *The rise and fall of American growth: The US standard of living since the civil war*, Princeton University Press.

Greunz, L. (2004). Industrial structure and innovation-evidence from European regions, *Journal of Evolutionary Economics* **14**(5): 563–592.

Haines, M. R. and ICPSR (2010). *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002*, Interuniversity Consortium for Political and Social Research.

Hanlon, W. W. and Miscio, A. (2017). Agglomeration: A long-run panel data approach, *Journal of Urban Economics* **99**: 1 – 14.
  **URL:** *http://www.sciencedirect.com/science/article/pii/S0094119017300013*

Harris, C. D. (1954). The Market as a Factor in the Localization of Industry in the United States, *Annals of the Association of American Geographers* **44**(4): 315–348.

Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output, *Journal of Economic Growth* **16**(4): 309–342.

Helsley, R. W. and Strange, W. C. (2002). Innovation and input sharing, *Journal of Urban Economics* **51**(1): 25–45.

Henderson, V., Kuncoro, A. and Turner, M. (1995). Industrial development in cities, *Journal of Political Economy* **103**(5): 1067–1090.

Hirschman, A. O. (1981). *Essays in trespassing*, Cambridge University Press.

Hornbeck, R. and Keskin, P. (2015). Does Agriculture Generate Local Economic Spillovers? Short-Run and Long-Run Evidence from the Ogallala Aquifer, *American Economic Journal: Economic Policy* **7**(2): 192–213.

Hornbeck, R. and Naidu, S. (2014). When the levee breaks: black migration and economic development in the American South, *American Economic Review* **104**(3): 963–990.

Hounshell, D. (1985). *From the American system to mass production, 1800-1932: The development of manufacturing technology in the United States*, Johns Hopkings University Press.

IIASA/FAO (2012). Global Agro-ecological Zones (GAEZ v3.0), IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Imbs, J., Montenegro, C. and Wacziarg, R. (2012). Economic Integration and Structural Change, *mimeo* .

Imbs, J. and Wacziarg, R. (2003). Stages of diversification, *American Economic Review* **93**(1): 63–86.

Jacobs, J. (1969). *The economies of cities*, Random House.

Johnston, B. F. and Mellor, J. W. (1961). The role of agriculture in economic development, *American Economic Review* **51**(4): 566–593.

Jones, C. I. (2011). Intermediate goods and weak links in the theory of economic development, *American Economic Journal: Macroeconomics* **3**(2): 1–28.

Katz, L. F. and Margo, R. A. (2014). Technical change and the relative demand for skilled labor: The United States in historical perspective, *in L. Platt Boustan, C. Frydman and R. Margo (Eds.), Human capital in history: The American record, pp. 15-57.* .

Kim, K., Chavas, J.-P., Barham, B. and Foltz, J. (2012). Specialization, diversification, and productivity: a panel data analysis of rice farms in korea, *Agricultural Economics* **43**(6): 687–700.

Koren, M. and Tenreyro, S. (2013). Technological diversification, *American Economic Review* **103**(1): 378–414.

Kremer, M. (1993). The O-ring theory of economic development, *Quarterly Journal of Economics* **108**(3): 551–575.

Kuhlmann, C. (1944). The processing of agricultural products after 1860, *in H. F. Williamson (Ed.), The growth of the American economy; an introduction to the economic history of the United States,* pp. 105–142.

Li, N. (2013). An Engel curve for variety.

Lin, J. (2011). Technological adaptation, cities, and new work, *Review of Economics and Statistics* **93**(2): 554–574.

Matsuyama, K. (1992). Agricultural productivity, comparative advantage, and economic growth, *Journal of Economic Theory* **58**(2): 317–334.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, *in P. Zarembka, Frontiers in econometrics* pp. 105–142.

Meyer, D. R. (1989). Midwestern industrialization and the American manufacturing belt in the

nineteenth century, *Journal of Economic History* **49**(4): 921–937.

Michaels, G. (2011). The Long Term Consequences of Resource-Based Specialisation, *Economic Journal* **121**(551): 31–57.

Michalopoulos, S. (2012). The Origins of Ethnolinguistic Diversity, *American Economic Review* **102**(4): 1508–39.

Minnesota Population Center (2016). National historical geographic information system: Version 11.0 [database], Minneapolis: University of Minnesota.
**URL:** *http://doi.org/10.18128/D050.V11.0.*

Mullahy, J. (2015). Multivariate fractional regression estimation of econometric share models, *Journal of Econometric Methods* **4**(1): 71–100.

Neffke, F., Henning, M., Boschma, R., Lundquist, K.-J. and Olander, L.-O. (2011). The dynamics of agglomeration externalities along the life cycle of industries, *Regional Studies* **45**(1): 49–65.

Nunn, N. (2008). Slavery, Inequality, and Economic Development in the Americas, *in E. Helpman, Institutions and economic performance, Harvard University Press, pp. 148–180.* .

Nunn, N. (2014). Historical Development, *The Handbook of Economic Growth* **2**: 347–402.

Nunn, N. and Qian, N. (2011). The potato's contribution to population and urbanization: evidence from a historical experiment., *Quarterly Journal of Economics* **126**(2): 593–650.

Olmstead, A. L. and Rhode, P. W. (2006). Tables Da693-706, Da717-729, Da733-745, Da755-765, Da755-765, Da768-773, Da968-982, Da1020-1038, *in Carter, Gartner, Haines, Olmstead, Sutch, and Wright (eds.), Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition* .

Olmstead, A. L. and Rhode, P. W. (2008). *Creating Abundance: Biological Innovation and American Agricultural Development*, Cambridge University Press.

Olsson, O. (2000). Knowledge as a set in idea space: An epistemological view on growth, *Journal of economic growth* **5**(3): 253–275.

Olsson, O. (2005). Technological opportunity and growth, *Journal of economic growth* **10**(1): 31–53.

Page, B. and Walker, R. (1991). From settlement to Fordism: the agro-industrial revolution in the American Midwest, *Economic Geography* **67**(4): 281–315.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric Methods for Fractional Response Variables With an Application to 401 (K) Plan Participation Rates, *Journal of Applied Econometrics* **11**(6): 619–632.

Paul, C. J. M. and Nehring, R. (2005). Product diversification, production systems, and economic performance in U.S. agricultural production, *Journal of Econometrics* **126**(2): 525–548.

Perlman, E. R. (2015). Dense Enough To Be Brilliant: Patents, Urbanization, and Transportation in Nineteenth Century America.

Petralia, S., Balland, P.-A. and Rigby, D. (2016). Histpat dataset.
**URL:** *http://dx.doi.org/10.7910/DVN/BPC15W*

Pudup, M. B. (1987). From farm to factory: structuring and location of the US farm machinery

industry, *Economic Geography* **63**(3): 203–222.

Rajan, R. G. and Ramcharan, R. (2011). Land and credit: A study of the political economy of banking in the United States in the early 20th century, *Journal of Finance* **66**(6): 1895–1931.

Ramalho, E. A., Ramalho, J. J. and Murteira, J. M. (2011). Alternative estimating and testing empirical strategies for fractional regression models, *Journal of Economic Surveys* **25**(1): 19–68.

Ramankutty, N., Foley, J. A., Norman, J. and McSweeney, K. (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change, *Global Ecology and Biogeography* **11**(5): 377–392.

Ramcharan, R. (2010a). Inequality and redistribution: Evidence from US counties and states, 1890-1930, *Review of Economics and Statistics* **92**(4): 729–744.

Ramcharan, R. (2010b). The link between the economic structure and financial development, *BE Journal of Macroeconomics* **10**(1).

Restuccia, D., Yang, D. T. and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis, *Journal of Monetary Economics* **55**(2): 234–250.

Rhode, P. W. and Strumpf, K. S. (2003). Assessing the importance of Tiebout sorting: Local heterogeneity from 1850 to 1990, *American Economic Review* **93**(5): 1648–1677.

Romer, P. M. (1990). Endogenous Technological Change, *Journal of Political Economy* **98**(5 part 2): S71–S102.

Rosenberg, N. (1972). *Technology and American economic growth*, Harper & Row.

Rosenberg, N. (1979). Technological interdependence in the American economy, *Technology and Culture* **20**(1): 25–50.

Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R., Schroeder, M. B. and Sobek, M. (2010). *Integrated Public Use Microdata Series: Version 5.0*, University of Minnesota.

Russelle, M. P., Entz, M. H. and Franzluebbers, A. J. (2007). Reconsidering integrated crop–livestock systems in North America, *Agronomy Journal* **99**(2): 325–334.

Schmookler, J. (1966). *Invention and economic growth.*, Harvard University Press.

Schumpeter, J. A. (1934). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle*, Harvard University Press.

Sivakumar, A. and Bhat, C. (2002). Fractional split-distribution model for statewide commodity-flow analysis, *Journal of the Transportation Research Board* **1790**(1): 80–88.

Sokoloff, K. L. and Dollar, D. (1997). Agricultural Seasonalily and the Organization of Manufacturing in Early Industrial Economies: The Contrast Between England and the United States, *Journal of Economic History* **57**(02): 288–321.

Spolaore, E. and Wacziarg, R. (2013). How Deep Are the Roots of Economic Development?, *Journal of Economic Literature* **51**: 325–69.

Usher, A. P. (1929). *A History of Mechanical Inventions*, McGraw-Hill.

Van den Bergh, J. C. (2008). Optimal diversity: increasing returns versus recombinant innovation, *Journal of Economic Behavior & Organization* **68**(3): 565–580.

Vollrath, D. (2011). The agricultural basis of comparative development, *Journal of Economic Growth* **16**(4): 343–370.

Vollrath, D. (2013). Inequality and school funding in the rural United States, 1890, *Explorations in Economic History* **50**(2): 267–284.

Weiss, T. J. (1992). US Labor Force Estimates and Economic Growth, 1800-1860, *in R.E. Gallman and J.J. Wallis, American economic growth and standards of living before the Civil War*, University of Chicago Press, pp. 19–78.

Weitzman, M. L. (1998). Recombinant growth, *Quarterly Journal of Economics* **113**(2): 331–360.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*, MIT press.

Wright, G. (1979). The efficiency of slavery: another interpretation, *American Economic Review* **69**(1): 219–226.

Zeppini, P. and Van den Bergh, J. C. (2013). Optimal diversity in investments with recombinant innovation, *Structural Change and Economic Dynamics* **24**: 141–156.