For Amber Waves of Grain: Commodity Booms and Structural Transformation in Nineteenth Century America

Jeff Chan*

Wilfrid Laurier University

March 30, 2024

Abstract

This paper examines the effect that the export boom in wheat had on US counties from 1870 to 1900. I find that increased exports in wheat, exogenously driven by declines in British wheat production, led to an unequal effect on wheat production across counties. Specifically, counties that were less well-suited to wheat production increased their wheat output relative to more productive counties. These low-yield counties also reduced their urban population share and increased the amount of agricultural activity occurring within their borders. Taken together, my results suggest that the wheat boom in nineteenth century America led to more wheat production in counties which were less well-suited for growing it, which slowed these counties' transition from rural to urban.

JEL codes: F14, N51, N91, Q17, R11
Keywords: wheat, export boom, nineteenth century, United States

*Email: jchan@wlu.ca. I thank the editor, Ryan Cardwell, and two anonymous referees for making suggestions that substantially improved the quality of the paper.

1 Introduction

In the second half of the 1800s the United States went from exporting very little wheat to becoming the world's leading exporter of the agricultural good by the end of the century, driven in large part by British imports. Multiple explanations for this phenomenon have been proposed, ranging from the expansion of food production during the Civil War to feed soldiers to technological improvements in shipping and communications such as steamships and the telegraph. By the end of the century, the United States would become the largest wheat exporter in the world.[1] At the same time, the United States was undergoing a series of rapid changes that would shape its economy for the coming 20th century. The shift from agriculture from manufacturing was one of the most important of these changes, accompanied by a corresponding movement of people from rural to urban. Despite the large amount of work investigating the US export of grains such as wheat during this period, there has been no work to my knowledge that has studied how the wheat export boom in the nineteenth century affected the structural transformation of the United States from a country of farmers to its eventual position as a factory for the world.

This paper aims to fill this gap by studying how the wheat export boom induced counties to change their agricultural and manufacturing activities, as well as its population distribution across rural and urban locations. I combine time series variation in US wheat exports as well as British wheat production to obtain exogenous variation in US wheat exports. I then pair these time series with predicted wheat yields by county, calculated from FAO-GAEZ; I use these predicted wheat yields as a measure of how productive a given county is in producing wheat based on its climate and geographic conditions.

I first find that counties with lower wheat productivity increased their wheat production relative to counties with higher productivity during the wheat export increase; this increase was accompanied by an increase in farms, farmland, and farming output. I find some weak evidence that this increase in aqricultural output was mirrored by an offsetting decrease in manufacturing activity, especially in capital and labour (as measured by total wages paid). Finally, I show that counties with lower wheat productivity saw a decrease to their urban population, resulting in an overall increase in the rural share of the population.

[1] This is a position which it continues to occupy as of the writing of this paper.

A sizeable literature has examined the growth in the trade in grain between the United States and Britain in the nineteenth century. Sharp and Weisdorf (2013), for example, study the integration of the wheat markets in the US and UK using price data. O'Rourke (1997) instead examines the effects that the increase in cheap US grain had on different European countries' economies, including the UK. Pederson et al. (2023) also study the UK-US wheat markets' integration but incorporate Canada as an additional component of the cross-Atlantic wheat market, finding that Canada faced similar barriers in accessing the UK wheat market despite being a part of the British empire. Ejrnaes et al. (2008) also study the degree of convergence in the grain market in the nineteenth century, finding that in the earlier periods shocks in Britain were the most important price shocks, while in the later period the United States became the most important source of shocks. This last finding, in particular, is consistent with the identification strategy in my paper which uses UK wheat production levels as exogenous variation for US wheat exports. More generally, my paper adds to the literature on the US and UK grain markets by further highlighting the effect that the integration and growth of these markets had on American counties, some of which responded by increasing their production of wheat and moving further towards agriculture at the potential expense of other economic activities.

This paper also relates closely to other work examining the role of commodity-based trade shocks on local labor markets. The most relevant work in this area is Brauer and Kersting (2023), who study the effects of the American grain export boom on Prussia during the nineteenth century, finding that wheat-growing regions did not experience any adverse effects; the authors attribute this lack of response to out-migration, which mitigated the impacts of the grain trade shock. In contrast, I find strong effects of the wheat export boom in America, with regions that had lower predicted wheat yields disproportionately increasing their farming activities to produce more wheat. In that sense, my findings are more consistent with contemporary work in trade and commodities booms like Costa et al. (2016), who find that the commodities export boom in Brazil brought on by the demand for raw materials in China during the 2000s saw faster wage growth from 2000 to 2010 for affected regions in Brazil.[2] Another paper highly relevant to my work is Williamson (1980), who links wheat export growth in the late 19th century US to changes in economic outcomes using a simple model and some quantitative counterfactual exercises. My paper tests a similar hypothesis to that in Williamson (1980), but estimates the causal relationship between wheat exports and local outcomes using

[2]More generally, my work is also related to studies of how trade shocks affect local labor markets in modern settings, such as Autor et al. (2013), Topalova (2010), and McCaig (2011).

an instrumental variables approach and exploiting data on innate land suitability for wheat production. My work is also in line with the results in Chan (2023), who shows that port-level trade shocks induced more manufacturing activity and population increases in affected ports. In contrast to this work, however, my paper shows that an export shock led to a re-orientation towards agriculture due to increased wheat cultivation, but principally for counties that have lower predicted wheat yields.

In addition to the preceding paragraph's discussion of work on commodity booms in general, there is a long-running literature studying the wheat boom in the Canadian context. Much of this literature finds that the wheat boom that occurred in Canada during the 19th and early 20th centuries had a positive effect on economic growth and local outcomes. Bologna Pavlik et al. (2023), in a recent example, find that the wheat boom-induced structural break in growth during the late 1800s accounted for approximately 40% of observed economic growth in Canada during the study period. This is mostly in line with other work in the field such as Chambers and Gordon (1966), Lewis (1975), McInnis (2007), Bertram (1973), Watkins (1977), Ankli (1980), and Dick (1980) which use various methods to calculate the impact of the wheat boom on economic growth. More generally, Keay (2007) studies the role of natural resources in Canadian economic growth in the 20th century and finds that natural resources did not constrain GNP per capita growth and had a positive impact on GNP per capita levels. In line with this body of work, my paper provides complementary evidence of the positive growth-inducing effects of a wheat boom on local outcomes in a neighboring country, the United States. My methodology, which makes use of predicted crop suitability data as a proxy for underlying productivity in wheat by US county, has also been an unexplored avenue in studies of wheat booms during this period (to my knowledge); I therefore also add to the plurality of methodologies that have been used in the study of wheat booms' effects, and also confirm the general finding of this literature that wheat booms have a growth-enhancing effect.

This paper also belongs to a broader, ongoing research agenda seeking to understand the drivers of economic activity and structural transformation in the United States during the nineteenth century. Donaldson and Hornbeck (2016), for example, take a market access approach and show that increased market access induced by the US railroad network expansion had substantial impacts on agricultural land value. Similarly, Chan (2022) finds that railroad-induced market access led to more agricultural output and population increases in the US. Nunn et al. (2020) argue that the large inflows of immigrants to the US during this era also had a transformative effect on the country, with counties which received more immigrants during this time period having better modern-day economic outcomes such as higher incomes and lower unemployment, which the authors attribute to the persistence of shorter-run advantages accrued by these counties such as higher productivity and innovation. My paper will also focus on a large, external shock but will instead argue that in my case the shock led to a stunted industrialization and rural-to-urban transition process for the counties which were induced to grow more wheat.

The rest of the paper proceeds as follows. Section 2 provides some background on the setting, Section 3 describes the data and methodology used in the analysis, Section 4 goes over the results, and Section 5 concludes.

2 Background

2.1 US Wheat Production

The United States produced over 287 million bushels of wheat in 1870. By 1900, wheat production had ballooned to 658 million bushels. A variety of factors contributed to the large increase in wheat output, which catapulted the United States to becoming one of the world's leading wheat suppliers. One major contributor was the westward expansion of the United States. During this period, a significant number of settlers moved onto lands west of the Mississippi River. Many such settlers took advantage of legislation such as the 1862 Homestead Act, which provided free land to settlers provided that they improved the lot and lived on it for a minimum period of time. The movement of settlers was facilitated greatly by the rapid expansion of the railroad network. In 1870 there were approximately 53,000 miles of track laid in the US; by 1890, this figure had more than doubled to almost 130,000 miles. More crucially, over 50,000 miles of this new track was laid in the west. In addition to facilitating the movement of people, improved rail access also allowed farmers to more easily ship their goods to distant markets, both domestic and foreign. This combination of factors resulted in an increase in the overall farming output of the US, including in wheat.

2.2 US Wheat Exports

In 1870, the United States exported approximately 37 million bushels, at a value of 47 million USD. By 1900, wheat exports had grown to 102 million bushels valued at 73 million USD. As a point of comparison, US total exports of merchandise (excluding re-exports) were 377 million USD in 1870 and 1.371 billion USD in 1900. In this context, the growth of wheat exports may not have had as tremendous of an impact on overall US exports as other goods. There are two important caveats to this observation, however. First, an increase of 37 to 102 million bushels is still a tremendous increase in exports over a relatively short period of time, constituting an almost tripling of quantity exported. Additionally, 1870 and 1900 were relatively anomalous years in terms of the overall trend occurring during this period; this is an artifact of the large year-to-year fluctuations in wheat exports. This can be seen in Figure 2, which shows US wheat export quantities by year. The figure shows that in 1899 and 1898, US wheat export quantities were even higher than in 1900, at 139 and 148 million bushels, respectively. In contrast, in 1869 and 1868, US wheat exports were only 18 and 16 million bushels, respectively. Wheat exports, therefore, exhibited astonishingly rapid growth during the period from 1870 to 1900.

Various explanations have been proposed for why US exports of wheat increased during the late nineteenth century. One such explanation was the improvements in communications and shipping technology that occurred during this era, which led to greater market integration at both national and international levels. The increasing use of telegraphs both across and within countries allowed sellers and farmers to coordinate more effectively, and for markets across countries to become more integrated. Yates (1986), for example, argues that the telegraph allowed markets in the US to develop and grow to a national level due to the easing of coordination issues. Similar evidence from China by Hao et al. (2022) supports the idea that telegraph expansion in the nineteenth century led to better grain market integration.

Similar arguments also suggest that the adoption of transoceanic steamships facilitated the US-Europe trade in grains such as wheat. North (1958) collects data and finds that freight rates for a variety of goods, including wheat, fell dramatically from 1740 to 1913; he finds, for example, that the freight rate for wheat from New York to Liverpool fell from over 21 cents per bushel in 1873 to less than 3 cents a bushel in 1901. North (1958) argues that this decrease in freight rates facilitated drops in the cost of food thanks to imported foodstuffs and spurred economic development.

2.3 British Wheat Production

Spurred in part by the increasingly available and cheap imported wheat from abroad, British production of wheat declined steadily; this decline is exhibited clearly in the British wheat production statistics, as seen in Figure 3. While domestic supply of wheat was continually decreasing, demand for foodstuffs did not. The UK's population climbed from 31 million in 1871 to over 38 million by 1901. More and more of the British population was moving towards urban areas, however, and fewer remained each year in rural areas to engage in agricultural activity. In 1850, the urban population share for the UK was 39.6 percent; by 1900, this figure had increased to 67.4 percent (Bairoch and Goertz (1986)).[3] This increasing gap between domestic supply and demand was filled by a corresponding rise in imports of wheat, principally from the United States.

The decline in British wheat production and its substitution with imported wheat was caused in large part by the repeal of the Corn Laws in 1846.[4] The Corn Laws were a series of tariffs and other protections against imported grains. The protections were enacted in part due to the end of the Napoleonic Wars in 1815 and the corresponding decrease in corn prices that followed. The protections allowed British farmers to continue serving the domestic market. These protections were crucial to the viability of British domestic production; Williamson (1990) notes that contemporary proponents of the Laws stated that European farmers produced wheat for a cost of 40 shillings per quarter while their English counterparts needed prices to rise to above 80 shillings per quarter to make a profit. Despite the clear incentive for farmers to argue for the continuation of these protections, by the 1840s a growing opposition to the Corn Laws was arguing for their repeal. One principal argument of this group was that the Corn Laws led to higher food prices; this argument became especially important in 1845 and 1846 as the Irish Potato Famine began. As a consequence, Sir Robert Peel, the Prime Minister at the time, pushed for a repeal of the Corn Laws. The Corn Laws were successfully repealed in the summer of 1846. The effects of the repeal of the Corn Laws remains an open question in the literature.[5] What is almost certainly true, however, is that the removal of the trade barriers exposed British farmers to foreign competition and likely contributed to the shift from agriculture to manufacturing. This shift, coupled with the lowered barriers to imported wheat, led to an increasing reliance on foreign wheat; this became especially true as Canada and the US began producing more wheat due to their westward

[3] For comparison, the comparable figures from the same source state that the US had an urban population share of 5.3 percent in 1850 and 35.9 percent in 1900.

[4] "Corn" in this context refers not only to what we know as corn, but also other grains such as wheat and barley.

[5] For an example of recent work, see Irwin and Chepeliev (2021).

expansions.

2.4 US-British Trade in the Nineteenth Century

More generally, the trade relationship between the US and the British during this period was very strong. The UK was the United States' largest trading partner from 1870 to 1900. This trade linkage was long-standing and up to the mid-19th century reflected ta commodities-for-manufactures relationship, where the US would ship mainly agricultural and other natural resource-based products to the UK in exchange for the manufactured goods that the British, by virtue of the Second Industrial Revolution, had a large advantage producing. In particular, US cotton, produced with the use of slaves in the South, was a key export through the whole of the 1800s; this cotton would be shipped mainly to the UK, where its textile factories would produce finished products that would then be shipped around the world (including back to America). Near the end of the century, the nature of this relationship would begin to change, owing to the US's increasing capabilities in manufacturing. Even by 1900, however, cotton and other agricultural outputs would remain a key export for America.[6]

Trade policy in the context of the US-UK trade relationship did not play a central role, at least on the part of the United States. Trade policy during this period for the US was characterized by high tariffs. As Irwin (2017) notes in chapters 5 and 6, tariffs were raised during the US Civil War (1861-1865) and remained high for the remainder of the 19th century due to Republican opposition to their removal and Democrats' inability to coalesce enough political capital to lower tariffs. In contrast, the UK was undergoing a period of removal of tariffs and other trade barriers. In addition to the Corn Laws, which enabled American producers of grains to eventually penetrate and dominate the British market (as discussed above), the UK also engaged in a variety of policies and agreements to lower trade barriers with many other countries. An exemplar of this activity is the British-French Cobden-Chevalier Treaty of 1860, which reduced trade barriers in a range of goods. Perhaps most importantly, it introduced the concept of the most-favored-nation (MFN) clause to trade negotiations; MFN status meant that if a country grants another country low tariffs, it is also obligated to grant that low tariff to any country with which it has granted MFN status. This mechanism ensured that trade agreements would systematically reduce tariffs across all trading partners in such agreements and con-

[6]For an overview of US trade and for an investigation into how trade affects manufacturing during this period, see Chan (2023).

tributed to the flourishing of trade during the mid to late 19th century.[7] In summary, the British-US trade relationship certainly benefited from a loosening of trade restrictions, particularly on the British side for US exports; the US did not, however, reciprocate and maintained its high tariffs throughout the period.[8]

3 Data and Methodology

3.1 Data

3.1.1 County-Level Outcomes

The outcome variables and all other county-level data used in this paper (unless otherwise specified) come from Haines and ICPSR (2010), which collates county-level summary tables produced using data from the US Decennial Census, the Census of Agriculture, and the Census of Manufactures and published by the US Census Bureau. Specifically, I make use of data on population (both aggregate and separated by subgroups), farms, land use, agricultural output, and manufacturing output. I also supplement the Haines and ICPSR (2010) data with additional manufacturing variables concerning establishments, total wages paid, and capital from Manson et al. (2021).

3.1.2 Predicted Crop Yields

Data on potential yields of wheat are obtained from FAO-GAEZ. In some supplementary analysis, I also make use of potential yields of another important crops during this period: cotton. The GAEZ data use an agroclimatic model which uses the climate, local conditions, and other geographical inputs to predict what the potential yields of various crops would be at very fine resolutions (0.083 degree by 0.083 degree cells) for the landmass of the Earth. Following Chan (2022) and Fiszbein (2022), I set potential yields to be produced using an intermediate level of technology and rain-fed conditions. The GAEZ data are then converted to the county level by taking the mean potential yield of cells that fall within each county, which provides me with a measure of wheat productivity for a given county. This measure for wheat productivity is shown in Figure 6, which maps the GAEZ measure into the counties in my sample; for comparison, the change in wheat production is shown in Figure 5. The maps show that there is considerable variation in both of the

[7] For a more in-depth discussion of these types of agreements, as well as the trade environment of the 19th century, see Meissner (2024).

[8] Some scholars, such as O'Rourke (2000), attribute American manufacturing growth during this time to the high tariffs, although this remains a source of academic debate.

key wheat variables used in this paper.

The GAEZ data have been widely used across many fields in economics because it provides an arguably time-invariant measure of productivity by crop at small-scale resolutions suitable for econometric analysis. According to FAO-GAEZ, the intent of the potential yields was to ascertain the "agronomically possible upper limit to produce individual crops under given agro-climatic, soil and terrain conditions and applying specific management assumptions and agronomic input levels."[9] In other words, in principle the GAEZ data use only factors that do not change over time in its calculations for potential yields, such as terrain and climate, and not as a result of any human investments in land improvement or other behaviors. Because of this design feature, economists have used the GAEZ data as proxies for land suitability in a variety of historical settings. Galor and Ozak (2016), for example, use the GAEZ crop suitability measures instead of actual crop output to circumvent issues of endogenous crop choice in order to show that parts of the world that had higher potential yields (and thus higher potential gains from investing in agriculture) had a longer-run mindset culturally. Similarly, Dickens (2022) uses the GAEZ data as a proxy for ancient land productivity to show that neighbouring ethnic groups with large differences in their land productivity had more similar languages which suggests that they likely engaged in trade to offset the differences in their lands' endowments. These examples, as well as many other papers, rely on the explicit design objective of the GAEZ data that it provides estimates of potential crop yields free from any human intervention and using only geographical factors.

The use of GAEZ data in work by economists, including in this paper, does not necessarily imply that such work presumes that the actual wheat productivity in different land areas is perfectly proxied by the GAEZ predicted yields or even that the true maximum potential yields are time-invariant. wheat productivity almost certainly varied over time for lands; Olmstead and Rhode (2002) show, for example, that American wheat production underwent significant productivity changes from 1800 to 1940, a large share of which they attribute to biological innovations. Instead, the purpose of using this data in my paper and in many other such works is to obtain a measure of what the underlying productivity for wheat production would be in different areas absent any human intervention such as land improvements, fertilizers, infrastructure, and the like. While a hypothetical set of predicted yields that incorporates these things would produce more realistic

[9] See https://gaez.fao.org/pages/theme-details-theme-3. Accessed March 28, 2024.

potential yields, it would also introduce multiple sources of endogeneity. Land improvements, for example, are an endogenous choice on the part of farmers and could respond to the wheat export shock. Similarly, the decision on whether to mechanize or use more advanced farming equipment could itself be a response to the increased attractiveness of the international wheat market for US farmers. There is a substantial literature in international trade that shows that being able to export more easily can incentivize firms to increase their productivity by investing in more machinery or doing more innovative activities (see Lileeva and Trefler (2010), for example). I therefore employ a time-invariant measure of land productivity in wheat to abstract away from these endogenous productivity factors and to focus purely on geographical attributes that can affect wheat production in different areas.

Of course, there could be issues with using the GAEZ data as proxies for underlying crop suitability if there are issues with the data itself, as highlighted in a review article on historical data by Giuliano and Matranga (2021). One such issue is that the FAO calculated potential yields using modern-day crops as the basis; if crops were substantially different in their historical iterations then this could lead to differences in how reasonable a proxy the predicted crop yields could be for historical land productivity for different crops. In addition, there could be concerns that the technologies and cultivation methods presumed by FAO-GAEZ in their calculations were not available historically.

In the context of my paper, I have taken precautions to mitigate the above concerns in the use of the GAEZ data. First, the GAEZ data allow you to select the level of technology in the calculation of potential yields; I select the intermediate level of technology, which Fiszbein (2022) argues is the best proxy for the technology level in 19th century America. Second, I do not allow for man-made irrigation to be used in the calculation of potential yields and instead use rain-fed irrigation only; this eliminates issues of improvements to irrigation technologies over time. Finally, the proximity of the time period in this paper (1870-1900) to the time period that the GAEZ data is based on (1961-2010). I therefore argue that issue relating to crop differences over time is relatively minor compared to studies where the GAEZ data is used to proxy for land suitability from many hundreds or thousands of years in the past. Even if there are any such differences in my setting, the research design does not require that the GAEZ data is a perfect proxy for potential yields in 19th century America; as long as the GAEZ data is strongly correlated with the true potential yields at the time, their use as the main source of geographical variation in wheat-affectedness should remain valid.

3.1.3 Wheat Production and Exports Time Series

Time series on wheat production and trade flows for the US and UK come from various sources. Annual US wheat production data (in millions of bushels) was sourced from the NBER's macrohistory database and downloaded via FRED. I also obtain UK annual wheat production data from the NBER macrohistory database (National Bureau of Economic Research (2023)). Finally, I obtain annual US wheat exports and UK imports of wheat from the US from U.S. Bureau of the Census (1960).

3.1.4 1890 Boundaries

County level data are converted to 1890 county definitions using the methodology from Hornbeck (2010). This method takes county boundaries over time and concords them by overlaying them using GIS. These overlaid boundaries are then broken up into their component pieces using the Intersect tool. The resulting pieces, and their areas, then determine how much of a county should be mapped to any given 1890 county with which it overlaps. For more information on this approach, see Hornbeck (2010).[10]

3.1.5 Other Data, Boundaries, and Sample Restrictions

In addition to the aforementioned data, I also employ some GIS data. County boundary definitions are obtained from NHGIS (Manson et al. (2021)), while data on the spatial extent of the US railroad network over time was obtained from Atack (2016). To maintain as much consistency as possible, I keep only counties that are reported in the baseline data and that can be mapped to 1890 county boundaries for all 4 sample years. In practice, there will still be missing observations for many given counties due to missing data for different variables. For this estimation sample, I report summary statistics for the key variables used in the analysis in Table A1 in the Appendix.

3.2 Specification

In my empirical analysis, I estimate the following equation:

$$a(y_{ct}) = \beta_0 + \beta_1 a(GAEZ_c) * a(exp_t) + \gamma_c + \delta_{st} + \epsilon_{ct} \tag{1}$$

[10]This approach is commonly used and has been employed in papers such as Donaldson and Hornbeck (2016) and Chan (2022).

$a()$ is the inverse hyperbolic sine transformation, which I implement for most variables in my analysis unless otherwise stated. This transformation is very similar to the log transformation,[11] and similarly helps address skewness in the data; this is a particularly significant issue for variables such as population and output, both of which are used in this paper. In contrast to the log transformation, however, the inverse hyperbolic sine has the benefit of being defined at zero.

y_{ct} represents various county-level outcomes of interest in this paper, such as population, wheat output, and agricultural activity. The key right-hand side variable is the interaction between US wheat exports from the previous year (and averaged with the previous and subsequent year), exp_t, as well as a given county's predicted wheat yields from GAEZ, $GAEZ_c$. The coefficient on this interaction term yields the relative change in the outcome between high and low yield counties due to changes in aggregate wheat export changes. This approach is very similar to that of Brauer and Kersting (2023), who look at the effects of grain imports from 1870 to 1913 on Prussian counties with higher initial concentrations of output in directly affected crops. My specification is also similar to that of Fernihough and O'Rourke (2020), who look at European cities and how their proximity to coal fields, instrumented by proximity to Carboniferous-era rock strata, affected their eventual growth during the Industrial Revolution. I similarly use a fundamental characteristic of a county, its suitability for wheat, to examine whether the wheat export boom had differential impacts on US counties.

In all specifications, unless otherwise noted, I include county fixed effects as well as state-year fixed effects. I cluster standard errors by state. In all specifications, observations will be weighted by that county's 1870 population.

3.3 Instrumental Variable

One concern with the right hand side variable in equation 1 is that wheat exports may be endogenous. For example, US wheat exports may increase due to expanded production of wheat in US counties due to technology shocks or improvements in aqricultural techniques. In order to mitigate such concerns, I employ an instrumental variable of the following form:

$$IV_{ct} = a(GAEZ_c) * a(UKprod_t) \tag{2}$$

[11] Specifically, the inverse hyperbolic sine transformation is $a(x) = ln(x + \sqrt{x^2 + 1})$.

The IV replaces US wheat exports with UK wheat production. Historically, British imports of wheat were often increased in years in which there were issues with the domestic supply. Figures 2 and 3 present time series plots of US wheat exports and UK wheat production, respectively. The figures illustrate that US wheat exports seem to move in the opposite direction to that of British wheat production. In particular, as US wheat exports have trended upward from 1870 to 1900, the opposite has occurred for UK production. The visual evidence is supported by the correlation between the (transformed) UK production and US wheat exports time series, using data from 1866 to 1900; the correlation is -0.5089, implying that UK wheat production and US wheat exports are strongly negatively correlated.

In addition, the instrument is very strong and predicts the endogenous regressor well. The first stage is presented in Table A2 in the Appendix. I also present the reduced form estimates in all main tables as well (in Panel B).

The instrumental variable can be interpreted as a variant of the shift-share instruments widely used in a range of fields within economics. Examples of uses of the shift-share methodology include instrumental variables used for immigration inflows (Card (2001)) and Chinese import exposure (Autor et al. (2013)). Like these examples, my IV assigns aggregate shocks, in my case wheat exports, to different regions according to differences in regional exposure, which in my paper is determined by predicted wheat yields. A recent literature that has further elaborated on the application and drawbacks of the shift-share methodology. Goldsmith-Pinkham et al. (2020), for example, argue that the use of shift-share IVs is econometrically equivalent to using the shares as instruments with a weighting matrix based on the national-level trends (the "shifts"); they follow by stating that the identification assumptions for shift-share IVs boil down to the exogeneity of the shares. In my setting, the shares would be equivalent to counties' predicted wheat yields. Given that the predicted yields are based on exogenously determined geographical characteristics, I argue that the predicted yields are plausibly exogenously determined. Goldsmith-Pinkham et al. (2020) also suggest a specification check where using only those shares that have the highest Rotemberg weights (i.e. those shares that have the highest predictive contribution) produces similar results; in my setting, since there is only one "share", this specification check and concerns surrounding it are trivially satisfied. While my IV is not a canonical case of a shift-share instrument, I nonetheless conclude that it satisfies the criteria described by Goldsmith-Pinkham et al. (2020) for it to be a valid research design.

Another parallel approach described in Borusyak et al. (2022) takes the stance that the shifts/shocks are exogenous and the shares are endogenous. In their approach, if one satisfies some conditions then their methodology can correct for correct inference and ensure consistency of estimates; the implication, of course, is that absent these methods inference will be incorrect and estimates may be inconsistent. In my setting, however, it is more plausible that the shares (i.e. predicted yields based on geographical characteristics) are exogenous. I therefore opt for the approach outlined in Goldsmith-Pinkham et al. (2020) instead.[12]

4 Results

4.1 Main Analysis

Before presenting the results of the regression analysis, I first provide an overview of the changes to wheat production occurring within counties during the sample period. Table 1 illustrates the total wheat output per sample year for two sets of counties: 1) the full set of counties and 2) the sample set of counties used in my regression analysis. Across both samples, I find that the number of counties producing wheat increases over time. Across both samples, wheat output (as measured in millions of bushels) and farming output (as measured in millions of 1890 USD) increase, approximately doubling. The key implication from Table 1 is therefore that wheat production was increasing over the sample period.

I then focus on whether counties in my regression sample differed in the changes to their wheat output based on their predicted wheat yields. I split the counties into those with above and below median wheat suitability, as measured by the FAO-GAEZ data. I then take each decadal change (1870-1880, 1880-1890, 1890-1900) and examine the changes to wheat output. The statistics are presented in Table 2. I first show that, across all periods, counties with above-median wheat suitability produced more wheat. In contrast, however, counties with below-median suitability tended to increase their wheat production disproportionately more. Across all 3 periods, the share of below-median counties producing more wheat between the start and end of the period was not much lower than that of above-median counties and in fact overtook above-median counties by 1890-1900; this implied that the number of below-median counties increasing

[12]Both papers recognize the other and separately describe that the optimal choice of which of the two to follow depends on the plausibility of whether the shocks or shares are more likely to be exogenous in the setting.

their wheat yields was similar to the number of above-median counties doing the same. The results are even more stark when examining the total change in wheat output across the two county types. Although in 1870-1880, the initial wheat production of below-median counties was less than a quarter of above-median counties, their increase in wheat output was over two-thirds the increase in wheat output for above-median counties. This trend continues in the later periods, with below-median counties' increase in wheat output actually becoming much higher than that for above-median counties. The takeaway from Table 2 is that, despite starting with less wheat output, it is the counties with lower suitability that have been driving the increase in wheat output occurring in America during this time. It is worthwhile to note, however, that except during 1880-1890 for above-median counties, all county types have been increasing their wheat output. This insight is important to keep in mind when I discuss the subsequent regression analysis, which looks at the relative change in outcomes between counties with lower versus higher wheat suitability.

I now turn to the regression results in the paper. One note before discussing the regression results is that all main tables contain two panels. The first panel, Panel A, typically contains the IV results (and also the OLS results, for the population table). Panel B contains the reduced form estimates using the instrumental variable directly in regressions. For the sake of brevity, I have opted to discuss only the Panel A results in the text but interested readers looking for reduced form estimates are directed to the tables.

First, in Table 3, I present results from specifications which examine the effects of the wheat export shock on county populations. In Panel A of this table, I present both OLS (in the first 4 columns) and IV results (in the latter 4 columns). In both the OLS and IV estimates, I do not find a large or statistically significant effect on total population. I then turn to examining urban and rural populations separately. In contrast to the overall lack of effect on population, the coefficients in both OLS and IV suggest that lower predicted yield counties decreased their urban populations relative to high-yield counties. This is correspondingly offset by relative increases to their rural populations, although the rural specification's coefficient is significant only in the IV specification. Further confirming this result, I also show that the rural population share (in columns 4 and 8) increases in low-yield counties relative to high-yield counties. Taken together, the findings in Table 3 suggest that the wheat export boom induced low-yield counties' population to reallocate internally from urban to rural areas. To get a sense of the magnitude, I calculate the difference between a county at the 25th percentile of wheat productivity versus another county at the 75th percentile, in terms of the effect on the urban population variable. Going from the 25th to the 75th percentile in wheat productivity implies a 0.32 higher change in (transformed) urban population for a one unit increase in transformed US wheat exports. Transformed US wheat exports increased by 1.7 from 1870 to 1900. Putting this all together, this implies that, relative to a 75th percentile productivity county, a county with wheat productivity at the 25th percentile would have a 0.544 unit lower increase in its transformed urban population, which is equivalent to a 0.12 standard deviation difference. I therefore conclude that the effects estimated for urban population are moderate in effect size.

To further understand the cause of this reallocation, I next examine agricultural outcomes' response to the wheat export shock in Table 4. In this table, and in all subsequent tables, I focus only on presenting and discussing the IV estimates, although an additional table in the Appendix (Table A3) contains the OLS estimates for some key main results. Table 4's estimates are consistent with low-yield counties increasing their agricultural activity. More specifically, low-yield counties respond to wheat export shocks by increasing the number of farms, increasing farmland acreage (both overall and in improved acreage), and increase their wheat production. Two other agricultural variables, farm output and a dummy variable for that county reporting wheat production, are also negative, which is consistent with low-yield counties increasing in these areas as well; the estimates, however, are insignificant at conventional levels. The coefficients in this table clearly indicate that the main mechanism behind the urban-to-rural reallocation from Table 3 is that low-yield counties, relative to high-yield counties, increase agricultural activity to produce more wheat because of the increased export demand for wheat. Again, focusing on magnitudes, a county with wheat productivity at the 25th percentile would have a 0.21 standard deviation higher amount of (transformed by inverse hyperbolic sine) improved farmland relative to a 75th percentile county, if both faced the actual change in US wheat exports from 1870 to 1900. This, once again, suggests a moderately sized effect size.

The previous two tables suggest that low-yield counties moved towards agricultural production. One potential cost of this movement is that these counties may have offset this increase in agricultural activity by substituting away from manufacturing. To test this, I examine whether the wheat export shocks had an effect on manufacturing outcomes. The resulting coefficients are reported in Table 5. None of the estimated coefficients are statistically significant, although some are non-trivially large in magnitude. The signs of the coefficients suggest that establishments, capital, and total wages paid decrease in low-yield counties, while output value increases. The bulk of the evidence in Table 5 therefore suggests that manufacturing activity may have declined in response to the increase in export demand for wheat. Again, however, the evidence in this table is weak and therefore suggestive at best.

4.2 Robustness Checks

In this subsection, I test for the robustness of my main analysis by performing a series of robustness checks. For these checks, I focus on several of the key outcome variables from the prior subsection. These variables are the urban population, the rural population share, the number of farms, the acreage of improved farmland, and wheat output. The tables for the robustness checks are presented in the Appendix.

One concern from the main analysis could be that the measure of potential wheat yields could be correlated with suitability in other crops that could also be undergoing booms. In particular, wheat productivity might be high in locations that are also very productive in cotton; this would be an issue since cotton was also a crop that comprised a very large proportion of US exports during this period. In Table A4, I therefore include a flexible control for cotton suitability by taking the inverse hyperbolic sine transformed cotton suitability measure (similarly obtained from FAO-GAEZ) and interact it with year dummies. The estimates in Table A4 are very similar to the main estimates, although the coefficients are slightly less precisely estimated. Almost all, however, with the exception of the coefficient for the urban population specification, remain statistically significant at conventional levels. The magnitudes of the coefficients also remain relatively similar to those from the main tables. I therefore conclude that the main analysis does not suffer from serious concerns about wheat suitability being correlated with cotton suitability.

In my next robustness check, I investigate whether the results from the main analysis differ between eastern and western states. An important ongoing change during this period in the US was the colonization of the American West. One potential explanation for my results could, in this context, be that this western migration could be driving the increase in wheat production and therefore the other changes occurring. In order to investigate this possibility, I split the sample of counties into western and eastern states, using the current definition of western states as defined by the Census Bureau. I present these heterogeneity results in Table A5. In contrast to this alternate explanation, I find that the main results seem to be driven in fact by eastern states. One important caveat, however, is that there are far fewer western counties than eastern counties, with western counties comprising less than 10% of all observations in my sample. Nonetheless, the evidence shown does not suggest that the magnitude of the estimates are larger for western states. For farms and improved farmland, the coefficient magnitudes are similar to the main results, while the urban population and wheat coefficients are also of meaningful size. Although caution must be used in interpretation of these results due to the lack of precision in the western estimates, the results shown in Table A5 do not support an explanation where the west was responsible for the main results in this paper.

In addition, one concern could be that trade infrastructure could be an important driver of my results. For example, Canada-US trade in the early to mid-1800s was largely free of trade barriers, although this ended by 1843. If proximity to pre-existing trade infrastructure, either in service of this prior Canada-US trade or more generally, was important one would expect the eastern counties (where most of the important trade infrastructure existed at the time, including New York[13]) to be the principal drivers of the effects found in this paper. The robustness exercise that splits counties into east and west in Table A5 does show that results may have been slightly stronger in the east although the differences are relatively minor. The results are not purely driven by east or west, however, nor are the coefficients very different in magnitude for most columns. I therefore conclude that proximity to trade infrastructure may have played a role but likely was not a main determinant of whether a county was affected by the wheat export shock.

One concern in the main analysis could be that my findings are driven by idiosyncratic differences between counties in the South and those in the rest of the US. Institutional quality, for one, might be worse in a region where oppression of Black Americans is systemic and the norm; this might suggest that the South might not see the same gains from the wheat boom as the rest of the country. This would also be in line with a large literature that argues that natural resource endowments may not have positive effects on economic growth. On the other hand, one could also argue that the South may have been more primed to benefit from a wheat boom than the rest of the US for historical reasons. For example, Majewski and Tchakerian (2007) argue that farmers in the South traditionally held more land in an unimproved state due to their practice of shifting cultivation; this practice would mean that southern farmers had large plots of land abandoned for years for recovery of soil nutrients. If this additional reserve of land allowed farmers in the South to respond

[13] New York was by far the most important port in the US at the time. In 1870, New York accounted for over half of the US's trade flows. See Chan (2023) for more details on port-level trade during this time.

more quickly to the wheat export boom by allowing them to plant more wheat, then the South may have disproportionately benefited from the mechanisms proposed in my paper.

I test for the above possibilities of the South having differential effects by conducting an additional robustness check. Specifically, in addition to separately estimating the main results by splitting the sample into east and west counties (Table A5) I also do the same but with south and non-south counties.[14] This exercise is reported in Table A6 in the Appendix. Some of the main results are stronger in the southern states, such as those for urban population, while some of the results are stronger in non-southern states, such as that for farms. Overall, however, the results are broadly similar across both regions. This conclusion adds to the debate on whether natural resource abundance helps or hinders economic growth. If the American South, with its history of slavery and continued institutionalized oppression of Black Americans well into the 20th century, can be seen as having poorer institutions than their non-south counterparts, then this paper's setting coupled with a natural resource shock in the form of the wheat export boom could be seen as a natural experiment in the testing of this hypothesis. The estimates in Table A6 provide some evidence that, at least in the US context, institutions do not seem to matter heavily in the transmission of natural resources' benefits to local economic outcomes. This could also be because while the institutional apparatus to oppress Black Americans may have been extractive in nature, in most respects states in the South have somewhat similar institutions in other ways to their non-south counterparts; both types of states, for example, are subject to federal oversight and operate under the same national-level institutions (such as the US Supreme Court). Despite this caveat, I conclude that the evidence shown in Table A6 is not consistent with institutional quality working to affect gains from natural resource booms within-country.

In Table A7, I drop each sample year one by one, reporting the resulting coefficients in Panels A to D, to test for whether my results are dependent on any particular year. The results from Table A7, when dropping the years 1880, 1890, or 1900, are very similar to the main results presented in the prior subsection. When 1870 is dropped in Panel A, the estimated coefficients become smaller and lose statistical significance although the magnitudes are still meaningful and the signs remain unchanged. The fact that 1870 is a key year in my analysis is not surprising since it is the first year in my sample and the period from 1870 to 1880 represents some of the largest growth in wheat exports and production in the United States during the late

[14] I use the definition of South states as defined by the US Census Bureau.

nineteenth century. Removing 1870 from my sample therefore constitutes the elimination of an important source of variation.

In Table A8, I check whether the predicted yields from GAEZ are consistent with counties producing more wheat in the cross section.[15] In column 1, I regress the transformed GAEZ measure for wheat against wheat output in 1870. The results are significant and positive, suggesting that counties with higher predicted yields produce more wheat. Columns 2,3, and 4 repeat this exercise with the later years (1880, 1890, and 1900) and find that this correlation diminishes in later years. This is entirely consistent with the main findings, which argue that the wheat boom led to lower-yield counties disproportionately increasing their wheat output; this mechanically would induce the weaker correlation seen in columns 2-4. Finally, in column 5 I instead regress wheat output on predicted yields for cotton, another leading export crop during this period. Reassuringly, the coefficient in this specification is negative and significant, suggesting that counties with higher predicted yields in cotton do not produce more wheat; instead, they likely substitute towards producing more cotton instead. Taken together, the estimates in Table A8 strongly support the validity of the use of the GAEZ predicted yield data as a valid proxy for crop productivity.

In Table A9, I test whether or not the wheat export boom led to changes in the proportion of farmland improved or in mechanization. The earlier analysis in 4 suggests that the share of improved farmland did not change, since the magnitude of the coefficients on farmland and improved farmland are very similar. Column 1 of Table A9 confirms this, since the share of farmland that is improved is insignificant. In column 2, I use the (inverse hyperbolic sine of) value of farming implements and machinery as a proxy for whether or not mechanization increased or decreased due to the wheat boom. The coefficient, again, is statistically insignificant, suggesting that mechanization was not a major component of the change in wheat production in less suitable counties nor of any adjustments that suitable counties engaged in.

The predicted wheat yield data from FAO-GAEZ are built using a series of geographical characteristics such as soil, climate, and terrain. The nature of this measure suggests that the inclusion of such geographical factors as control variables would be inappropriate. For one illustration of this point, suppose that rainfall (interacted with year dummies) were to be included as a control variable. Even if one were to find that the

[15] Note that the descriptive statistics in Table 2 already provide some evidence in support of this.

main treatment variable became smaller or less significant, this would not necessarily imply that omitted variables were at play. This is because it would not be clear whether the rainfall control was simply capturing the fact that higher rainfall counties are likely higher wheat yield counties for that very reason; the additional control could therefore simply be absorbing some of the main effect of interest. Nonetheless, to test for whether the predicted yield measure is simply capturing some omitted geographical characteristics I include the latitude and longitude of each county's centroid (interacted with year) as an additional control; these results are reported in Table A10. The results are very similar in magnitude to the main estimates, although precision is somewhat affected (particularly for rural share and wheat output). The table therefore implies that omitted geographical characteristics, as proxied by the inclusion of latitude and longitude, are not an important concern in my analysis.

In Table A11, I explore the use of an alternate source for the construction of my IV. Specifically, I make use of British Historical Statistics (Mitchell (1988)). Unfortunately, Mitchell (1988) does not contain a time series for British wheat production that covers a significant proportion of my sample period. I therefore make use of British imports of wheat (as measured in quantities with thousands of hundredweights as the unit). I simply substitute the time series used in the main IV for wheat production in the UK for this new import time series. The results, reported in Table A11, are very similar to the main analysis. In particular, all results remain quantitatively similar in magnitude and in sign. I therefore conclude that the main results in the analysis are not purely the artefact of the particular data source chosen.

4.3 Further Discussion

One interesting issue that has not been discussed in this paper yet is whether one should view the use of more peripheral lands, which may have a lower underlying wheat productivity, as "inefficient" if they could also produce wheat with the aid of mechanical equipment or land improvements. In one sense, it is inefficient because it induces farmers to use land for wheat cultivation when it may have been relatively more productive in producing other crops (or better suited for other uses altogether). In particular, if the wheat boom slowed the rural-to-urban transition and in so doing stymied the manufacturing boom that would propel American economic growth in the early 20th century, then the increase in wheat production could be seen as a hindrance to US growth during this period. On the other hand, as the referee implies, farmers clearly responding to incentives and could reasonably be assumed to be making an informed choice based on all available uses of their land. The fact that they chose to produce more wheat with their land, despite the lower underlying suitability for it, does not necessarily imply that there was misallocation in the sense that macroeconomists might use the word in modern connotations (see Adamopoulos et al. (2022) for a recent example). In this paper, and in some of the exposition in particular, I have described the results as showing that the wheat export shock may have slowed America's transition from agriculture to manufacturing, which was a key component of the US rise to economic dominance in the 20th century, and as such may have been suboptimal from that perspective. One should keep the above caveats and alternative interpretations in mind, however.

5 Conclusion

This paper finds that the wheat export boom of the late nineteenth century had an uneven effect across the United States. Counties with a low predicted yield in wheat increased their rural population shares and increased agricultural activity to produce more wheat relative to higher-yield counties. I also find some weak evidence that manufacturing activity may have declined as well. These changes suggest that low-yield counties, induced to produce more wheat, may have slowed or even reversed the ongoing transition from rural to urban living during this period. Given the importance of this transition in the structural transformation of the US economy from a resource-based economy to one that would become a globally leading manufacturing powerhouse, my results suggest that low-yield counties may have been left behind during this transformation because of the commodities boom in wheat.

My findings also echo the modern debates about whether a reliance on natural resources are a hindrance to economic development. Those that believe in the existence of the so-called natural resource curse[16] argue that the over-development of the natural resources sector can lead to an atrophied manufacturing sector, slowing the transformation of developing countries. The wheat export boom and its subsequent effect on US counties shows that this mechanism may also have been operating in the United States in the nineteenth century via a slowed transition to manufacturing. Subsequent studies on the longer-run implications of this stunted structural transformation in low-yield counties may therefore yield important insights into how modern-era countries overly relying on their primary sectors can move faster along their paths to economic

[16] Also known as "Dutch Disease".

development.

References

Adamopoulos, T., L. Brandt, J. Leight, and D. Restuccia, "Misallocation, Selection and Productivity: A Quantitative Analysis with Micro Data from China," *Econometrica*, 2022, *90* (3), 1261–1282.

Ankli, R., "The growth of the Canadian economy, 1896-1920: Export led and/or neoclassical growth," *Explorations in Economic History*, 1980, *17* (3), 251–274.

Atack, J., "Historical Geographic Information Systems (GIS) database of U.S. Railroads for 1826-1911," 2016.

Autor, D., D. Dorn, and G. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–2168.

Bairoch, P. and G. Goertz, "Factors of Urbanisation in the Nineteenth Century Developed Countries: A Descriptive and Econometric Analysis," *Urban Studies*, 1986, *23*, 285–305.

Bertram, G., "The Relevance of the Wheat Boom in Canadian Economic Growth," *Canadian Journal of Economics*, 1973, *6* (4), 545–566.

Bologna Pavlik, J., V. Geloso, and C. Pender, "Reaffirming the Wheat Boom: New Evidence for Structural Breaks in Canadian Economic Growth," 2023.

Borusyak, K., P. Hull, and X. Jaravel, "Quasi-Experimental Shift-Share Research Designs," *Review of Economic Studies*, 2022, *89* (1), 181–213.

Brauer, R. and F. Kersting, "Trade shocks, labour markets and migration in the First Globalisation," *Economic Journal*, 2023.

Card, D., "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

Chambers, E. and D. Gordon, "Primary products and economic growth: An empirical measurement," *Journal of Political Economy*, 1966, *74* (4), 315–322.

Chan, J., "The Local Effects of the First Golden Age of Globalization: Evidence from American Ports, 1870-1900," 2023.

Chan, Jeff, "Farming Output, Concentration, and Market Access: Evidence from the Nineteenth Century American Railroad Expansion," *Journal of Development Economics*, 2022, *157.*

Costa, F., J. Garred, and J. Pessoa, "Winners and Losers from a Commodities-for-Manufactures Trade Boom," *Journal of International Economics*, 2016, *102*, 50–69.

Dick, T., "Canadian wheat production and trade 1896-1930," *Explorations in Economic History*, 1980, *17* (3), 275–302.

Dickens, A., "Understanding Ethnolinguistic Differences: The Roles of Geography and Trade," *Economic Journal*, 2022, *132* (643), 953–980.

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

Ejrnaes, M., K. Persson, and S. Rich, "Feeding the British: convergence and market efficiency in the nineteenth-century grain trade," *Economic History Review*, 2008, *61* (s1), 140–171.

Fernihough, A. and K. O'Rourke, "Coal and the European Industrial Revolution," *Economic Journal*, 2020, *131*, 1135–1149.

Fiszbein, M., "Agricultural Diversity, Structural Change and Long-Run Development: Evidence from the U.S.," *American Economic Journal: Macroeconomics*, 2022, *14* (2).

Galor, O. and O. Ozak, "The Agricultural Origins of Time Preference," *American Economic Review*, 2016, *106* (10), 3064–3103.

Giuliano, P. and A. Matranga, *Handbook of Historical Economics*, Academic Press, 2021.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

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

Hao, Y., Y. Li, and J. Nye, "Wiring China: The impact of telegraph construction on grain market integration in late imperial China, 1870–1911," *Economic History Review*, 2022, *75* (3), 857–880.

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

Irwin, D., *Clashing Over Commerce: A History of US Trade Policy*, University of Chicago Press, 2017.

_ **and M. Chepeliev**, "The Economic Consequences of Sir Robert Peel: A Quantitative Assessment of the Repeal of the Corn Laws," *Economic Journal*, 2021, *131* (640), 3322–3337.

Keay, I., "The Engine or the Caboose? Resource Industries and Twentieth-Century Canadian Economic Performance," *Journal of Economic History*, 2007, *67* (1), 1–32.

Lewis, F., "The Canadian Wheat Boom and Per Capita Income: New Estimates," *Journal of Political Economy*, 1975, *83* (6), 1249–1257.

Lileeva, A. and D. Trefler, "Improved Access to Foreign Markets Raises Plant-level Productivity…For Some Plants," *Quarterly Journal of Economics*, 2010, *125* (3), 1051–1099.

Majewski, J. and V. Tchakerian, "The Environmental Origins of Shifting Cultivation: Climate, Soils, and Disease in the Nineteenth-Century US South," *Agricultural History*, 2007, *81* (4), 522–549.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles, "IPUMS National Historical Geographic Information System: Version 16.0 [dataset]," 2021.

McCaig, B., "Exporting out of poverty: Provincial poverty in Vietnam and U.S. market access," *Journal of International Economics*, 2011, *85* (1), 102–113.

McInnis, M., "Canadian economic development in the wheat boom era: A reassessment.," 2007.

Meissner, C., *One from the Many: The Global Economy Since 1850*, Oxford University Press, 2024.

Mitchell, B., *British Historical Statistics*, Cambridge University Press, 1988.

National Bureau of Economic Research, "NBER Macrohistory Database," 2023. retrieved on Sept 28 2023.

North, D., "Ocean Freight Rates and Economic Development 1750-1913," *Journal of Economic History*, 1958, *18* (4), 537–555.

Nunn, N., N. Qian, and S. Sequeira, "Immigrants and the Making of America," *The Review of Economic Studies*, 2020, *87* (1), 382–419.

Olmstead, A. and P. Rhode, "The red queen and the hard reds: Productivity growth in American wheat, 1800-1940," *Journal of Economic History*, 2002.

O'Rourke, K., "The European Grain Invasion, 1870–1913," *Journal of Economic History*, 1997, *57* (4), 775–801.

_ , "Tariffs and Growth During the Late 19th Century," *Economic Journal*, 2000, *110* (463), 456–483.

Pederson, M., V. Geloso, and P. Sharp, "Globalization and Empire: Market Integration and International Trade among Canada, the US, and Britain, 1750–1870," *Social Science History*, 2023, *48* (1), 19–40.

Sharp, P. and J. Weisdorf, "Globalization revisited: Market integration and the wheat trade between North America and Britain from the eighteenth century," *Explorations in Economic History*, 2013, *50* (1), 88–98.

Topalova, P., "Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India," *American Economic Journal: Applied Economics*, 2010, *2* (4), 1–41.

U.S. Bureau of the Census, "Historical Statistics of the United States, Colonial Times to 1957," 1960. Washington D.C., Library of Congress Card No. A 60-9150.

Watkins, M., "The staple theory revisited," *Journal of Canadian Studies*, 1977, *12* (5), 83–95.

Williamson, J., "Greasing the wheels of sputtering export engines: Midwestern grains and American growth," *Explorations in Economic History*, 1980, *17* (3), 189–217.

_ , "The Impact of the Corn Laws Just Prior to Repeal," *Explorations in Economic History*, 1990, *27*, 123–156.

Yates, J., "The Telegraph's Effect on Nineteenth Century Markets and Firms," *Business and Economic History*, 1986, *15*, 149.

6 Figures and Tables

Figure 1: UK imports of wheat from US

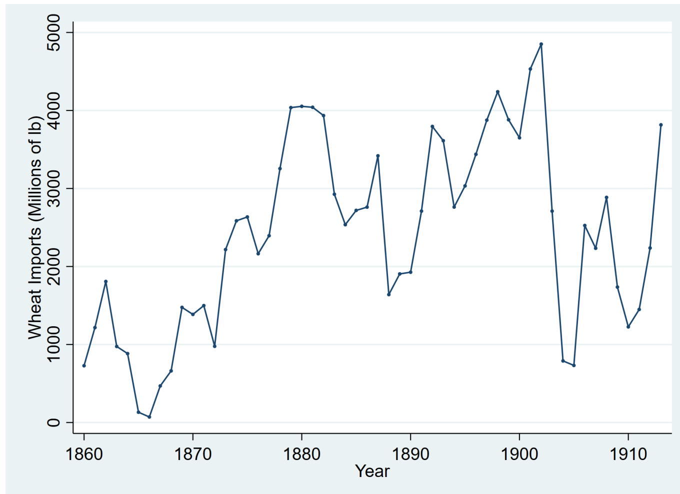


Figure 2: US exports of wheat

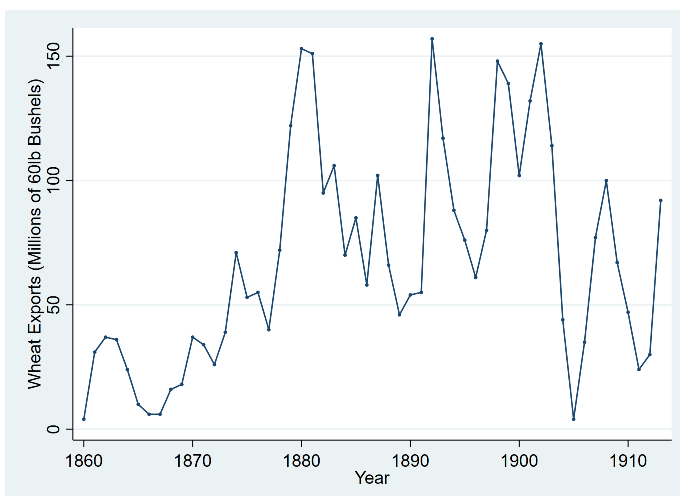


Figure 3: UK wheat production

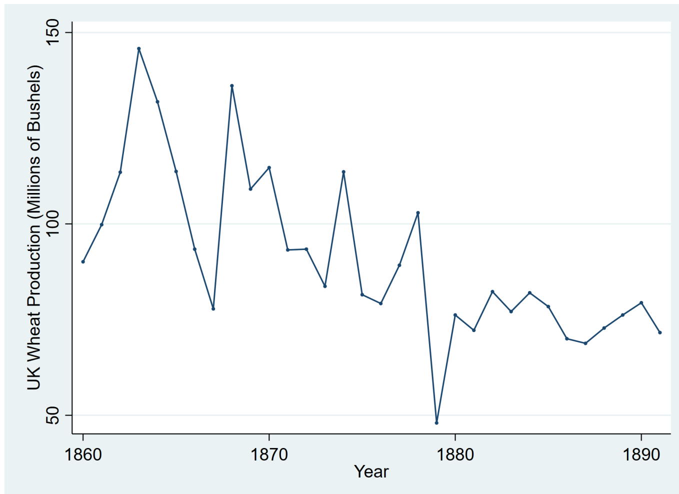


Figure 4: US wheat production

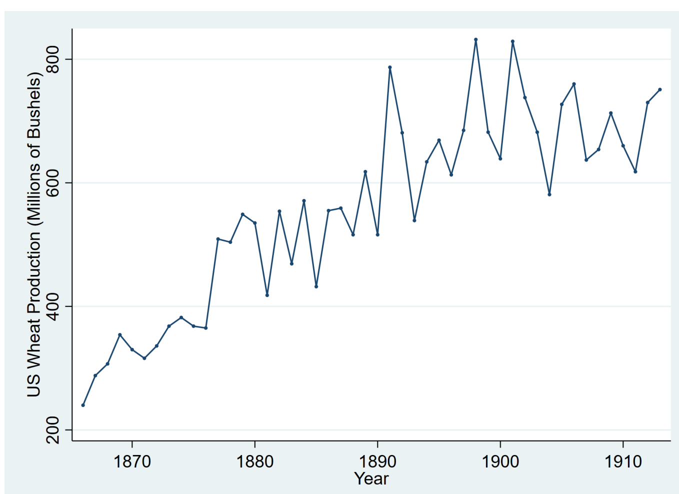
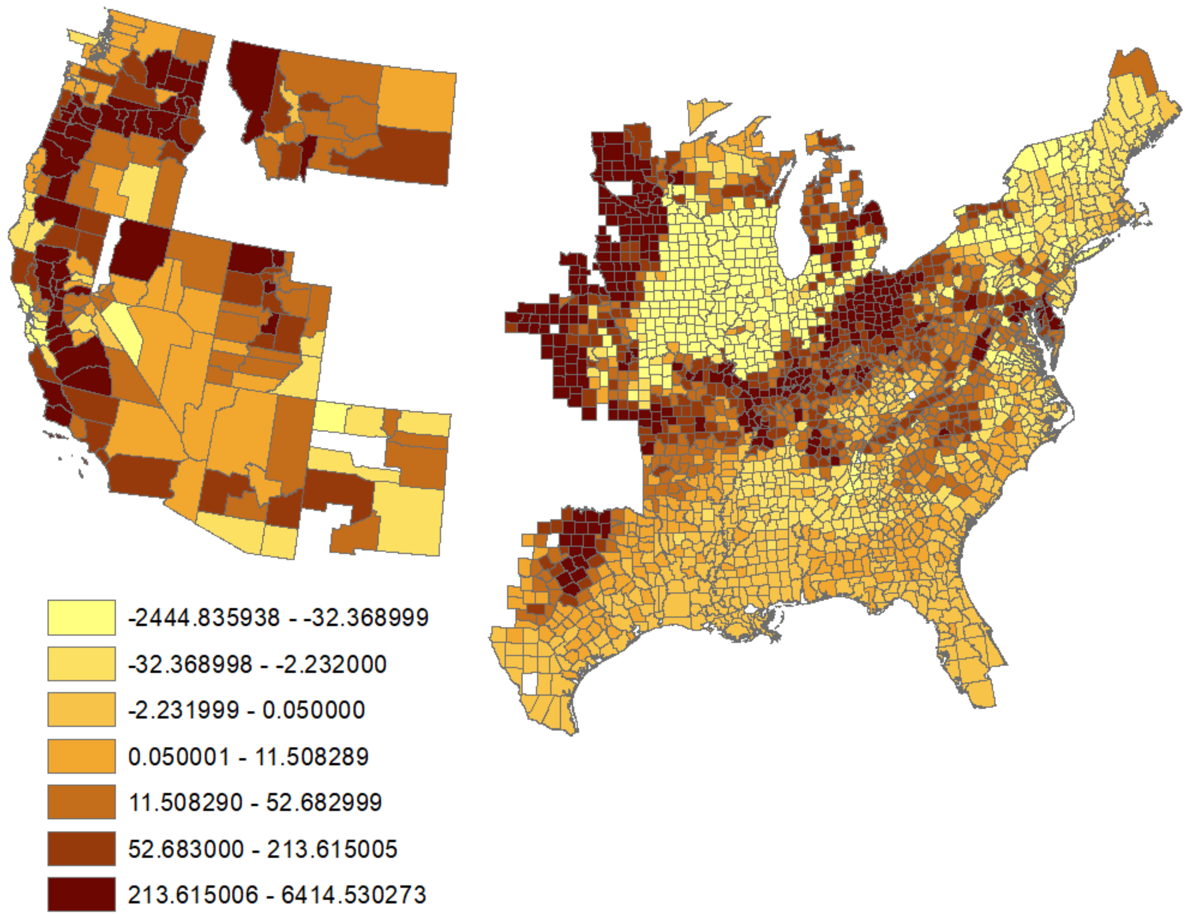
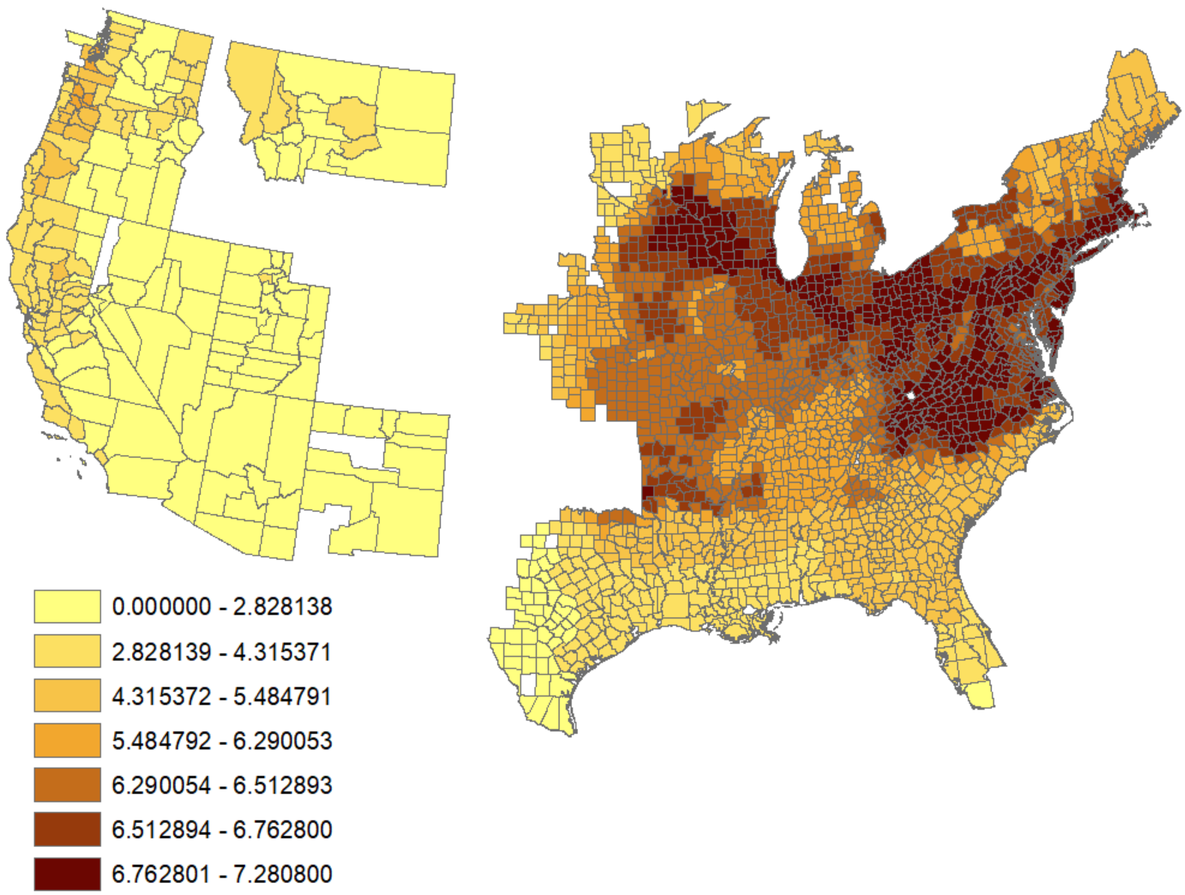


Figure 5: Change in Wheat Production (1870-1900), Thousands of Bushels



Notes: each color in the map represents one of 7 quantiles for the change in wheat production, with darker colors representing more positive changes.

Figure 6: Predicted Wheat Yields



Notes: each color in the map represents one of 7 quantiles for predicted wheat yields, with darker colors representing higher predicted yields.

Table 1: Wheat and Farming Production by Year

year	no. of counties producing wheat	wheat output (millions of bushels)	farming output (millions of 1890 USD)
Panel A: All Counties			
1870	2068	286.8	1916
1880	2250	456.0	1928
1890	2377	467.9	3253
1900	2526	636.8	3576
Panel B: Sample Counties			
1870	2066	286.8	1916
1880	2165	451.9	1918
1890	2089	410.8	3135
1900	2218	506.3	3360

Notes: Panel A uses the full set of counties that can be mapped to 1890 boundaries. Panel B uses only the regression sample of counties which are consistently reported across all 4 decadal years in the sample period.

Table 2: Changes to Wheat Production

above or below median predicted yield	sh. counties growing more wheat	wheat output, start of period	change in wheat output	no. counties
Panel A: 1870-1880				
below	0.628	48.94	66.75	1197
above	0.740	237.8	95.60	1196
Panel B: 1880-1890				
below	0.270	118.5	44.05	1197
above	0.284	333.4	-84.69	1196
Panel C: 1890-1900				
below	0.702	162.5	86.36	1197
above	0.589	248.3	8.520	1196

Notes: each panel presents statistics for the decadal change mentioned within the sample period. sh. counties growing more wheat refers to the proportion of counties that report growing more wheat at the end of the period, relative to the start of the period. no. counties refers to the total number of counties in the given category. above or below median predicted yield refers to whether a given county is above or below the median wheat predicted yield, as measured by the FAO-GAEZ data on predicted wheat yields. All counties analyzed in this table are in the regression sample.

Table 3: Population

Dep. Var.:	(1) a(pop)	(2) a(urban pop)	(3) a(rural pop)	(4) sh. rural	(5) a(pop)	(6) a(urban pop)	(7) a(rural pop)	(8) sh. rural
Panel A: OLS and IV								
a(GAEZ$_c$) x a(exp$_t$)	-0.0344	0.920**	-0.475	-0.0480**	-0.0521	1.228*	-0.578*	-0.0601**
	(0.0816)	(0.445)	(0.287)	(0.0190)	(0.112)	(0.647)	(0.333)	(0.0277)
Observations	9,564	9,564	9,564	9,564	9,564	9,564	9,564	9,564
R-squared	0.982	0.896	0.901	0.969	0.000	0.001	0.001	0.002
Panel B: Reduced Form								
IV$_{ct}$	0.148	-3.494*	1.644*	0.171**				
	(0.318)	(1.842)	(0.949)	(0.0787)				
Observations	9,564	9,564	9,564	9,564				
R-squared	0.982	0.896	0.901	0.969				

Notes: *** p$<$ 0.01, **p$<$ 0.05, *p$<$ 0.1. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. In Panel A, Columns 1-4 are estimated by OLS while columns 5-8 are estimated using IV; for a description of the instrument, see the main text. All estimates in Panel B estimated using OLS.

Table 4: Agricultural Outcomes

Dep. Var.:	(1) a(farms)	(2) a(farmland)	(3) a(farmland, improved)	(4) a(farm output)	(5) a(wheat output)	(6) $\mathbb{1}$(wheat)
Panel A: IV						
a(GAEZ$_c$) x a(exp$_t$)	-0.460**	-0.513***	-0.495**	-0.240	-0.998**	-0.0770
	(0.198)	(0.181)	(0.186)	(0.182)	(0.489)	(0.0621)
Observations	9,479	9,481	9,481	9,482	9,497	9,564
R-squared	0.004	0.011	0.008	-0.001	0.004	0.002
Panel B: Reduced Form						
IV$_{ct}$	1.312**	1.461***	1.412**	0.685	2.844**	0.219
	(0.564)	(0.518)	(0.530)	(0.518)	(1.394)	(0.177)
Observations	9,479	9,481	9,481	9,482	9,497	9,564
R-squared	0.942	0.967	0.963	0.939	0.945	0.765

Notes: *** p$<$ 0.01, **p$<$ 0.05, *p$<$ 0.1. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns in Panel A estimated using IV; for a description of the instrument, see the main text. All columns in Panel B estimated using OLS.

Table 5: Manufacturing Outcomes

	(1)	(2)	(3)	(4)
Dep. Var.:	a(mfg estab.)	a(mfg capital)	a(mfg total wages)	a(mfg output)
Panel A: IV				
a(GAEZ$_c$) x a(exp$_t$)	0.0460	0.508	0.329	-0.227
	(0.194)	(0.522)	(0.866)	(0.316)
Observations	9,543	9,447	9,446	9,456
R-squared	0.000	0.001	0.000	0.000
Panel B: Reduced Form				
IV$_{ct}$	-0.131	-1.452	-0.942	0.649
	(0.553)	(1.494)	(2.476)	(0.905)
Observations	9,543	9,447	9,446	9,456
R-squared	0.949	0.985	0.958	0.928

Notes: *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns in Panel A estimated using IV; for a description of the instrument, see the main text. All columns in Panel B estimated using OLS.

A Additional Figures and Tables

Table A1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
a(pop)	10.126	1.201	0	15.118	9584
a(urban pop)	3.135	4.584	0	15.079	9584
a(rural pop)	9.946	1.185	0	12.673	9584
sh. rural	0.895	0.195	0	1	9582
a(farms)	7.729	1.178	0	9.846	9493
a(farmland)	12.706	1.119	0	15.278	9495
a(farmland, improved)	11.851	1.405	0	14.571	9495
a(farm output)	14.022	1.344	0	16.824	9496
a(wheat output)	9.975	4.128	0	16.367	9511
$\mathbb{1}$(wheat)	0.891	0.312	0	1	9584
a(mfg estab.)	4.336	1.657	0	10.836	9561
a(mfg capital)	12.519	6.309	0	27.694	9463
a(mfg total wages)	9.101	4.821	0	20.227	9462
a(mfg output)	12.894	3.198	0	21.444	9474
a(exp$_t$)	4.892	0.681	3.858	5.558	9584
a(GAEZ$_c$)	2.362	0.395	0	2.683	9572
a(GAEZ$_c$) x a(exp$_t$)	11.555	2.529	0	14.913	9572

Notes: Summary statistics are for counties in main estimation sample.

Table A2: First Stage

	(1)
Dep. Var.:	a($GAEZ_c$) x a(exp_t)
IV_{ct}	-2.845***
	(5.53e-07)
Observations	9,564
R-squared	1.000

Notes: *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$. County and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's 1870 population.

Table A3: Main Results, OLS

	(1)	(2)	(3)	(4)	(5)
Dep. Var.:	a(urban pop)	sh. rural	a(farms)	a(farmland, improved)	a(wheat output)
a($GAEZ_c$) x a(exp_t)	0.920**	-0.0480**	-0.307*	-0.372**	-0.976*
	(0.445)	(0.0190)	(0.154)	(0.142)	(0.486)
Observations	9,564	9,564	9,479	9,481	9,497
R-squared	0.896	0.969	0.942	0.963	0.945

Notes: *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns estimated using OLS.

Table A4: Main Results, Controlling for Cotton Suitability

	(1)	(2)	(3)	(4)	(5)
Dep. Var.:	a(urban pop)	sh. rural	a(farms)	a(farmland, improved)	a(wheat output)
a($GAEZ_c$) x a(exp_t)	0.997	-0.0455**	-0.555**	-0.621***	-1.351*
	(0.648)	(0.0225)	(0.228)	(0.188)	(0.700)
Observations	9,564	9,564	9,479	9,481	9,497
R-squared	0.008	0.029	0.070	0.077	0.033

Notes: *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns estimated using IV; for a description of the instrument, see the main text.

Table A5: Main Results, Eastern and Western States

Dep. Var.:	(1) a(urban pop)	(2) sh. rural	(3) a(farms)	(4) a(farmland, improved)	(5) a(wheat output)
Panel A: Eastern States					
a(GAEZ$_c$) x a(exp$_t$)	1.476*	-0.0822**	-0.427**	-0.510**	-1.700***
	(0.796)	(0.0341)	(0.193)	(0.206)	(0.588)
Observations	8,796	8,796	8,715	8,717	8,733
R-squared	0.001	0.003	0.003	0.008	0.009
Panel B: Western States					
a(GAEZ$_c$) x a(exp$_t$)	0.528	0.00218	-0.548	-0.457	0.832**
	(0.778)	(0.0456)	(0.490)	(0.438)	(0.318)
Observations	768	768	764	764	764
R-squared	0.005	-0.000	0.019	0.012	0.017

Notes: *** p< 0.01, **p< 0.05, *p< 0.1. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns estimated using IV; for a description of the instrument, see the main text.

Table A6: Main Results, South and Non-South States

Dep. Var.:	(1) a(urban pop)	(2) sh. rural	(3) a(farms)	(4) a(farmland, improved)	(5) a(wheat output)
Panel A: South States					
a(GAEZ$_c$) x a(exp$_t$)	2.497**	-0.0610*	-0.216	-0.155	-1.098*
	(0.860)	(0.0322)	(0.149)	(0.227)	(0.539)
Observations	4,548	4,548	4,536	4,535	4,532
R-squared	0.003	0.003	0.002	0.003	0.008
Panel B: Non-South States					
a(GAEZ$_c$) x a(exp$_t$)	-0.414	-0.0589	-0.752**	-0.902***	-0.879
	(0.838)	(0.0491)	(0.287)	(0.286)	(0.867)
Observations	5,016	5,016	4,943	4,946	4,965
R-squared	-0.000	0.002	0.007	0.014	0.001

Notes: *** p< 0.01, **p< 0.05, *p< 0.1. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns estimated using IV; for a description of the instrument, see the main text.

Table A7: Main Results, Dropping Years

Dep. Var.:	(1) a(urban pop)	(2) sh. rural	(3) a(farms)	(4) a(farmland, improved)	(5) a(wheat output)
Panel A: Dropping 1870					
a(GAEZ$_c$) x a(exp$_t$)	0.552	-0.0791	-0.260	-0.345	-0.432
	(1.035)	(0.0584)	(0.219)	(0.487)	(2.180)
Observations	7,173	7,173	7,163	7,163	7,162
R-squared	-0.000	-0.002	-0.007	-0.007	0.001
Panel B: Dropping 1880					
a(GAEZ$_c$) x a(exp$_t$)	1.379*	-0.0690*	-0.541**	-0.565***	-0.975
	(0.784)	(0.0348)	(0.217)	(0.205)	(0.602)
Observations	7,173	7,173	7,093	7,095	7,103
R-squared	0.002	0.004	0.010	0.013	0.004
Panel C: Dropping 1890					
a(GAEZ$_c$) x a(exp$_t$)	1.064*	-0.0548**	-0.387**	-0.438***	-0.991*
	(0.547)	(0.0240)	(0.175)	(0.162)	(0.513)
Observations	7,173	7,173	7,087	7,089	7,115
R-squared	0.002	0.004	0.014	0.017	0.005
Panel D: Dropping 1900					
a(GAEZ$_c$) x a(exp$_t$)	1.355**	-0.0547**	-0.491**	-0.512**	-1.107***
	(0.662)	(0.0225)	(0.229)	(0.250)	(0.356)
Observations	7,173	7,173	7,082	7,084	7,101
R-squared	0.001	0.001	0.001	0.004	0.006

Notes: *** $p< 0.01$, **$p< 0.05$, *$p< 0.1$. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns estimated using IV; for a description of the instrument, see the main text.

Table A8: GAEZ Predicted Yields and Actual Wheat Output

Dep. Var.:	(1) a(wheat output)	(2) a(wheat output)	(3) a(wheat output)	(4) a(wheat output)	(5) a(wheat output)
a(GAEZ$_c$)	5.399**	3.394	4.117	3.554**	
	(2.660)	(2.521)	(2.720)	(1.694)	
a(GAEZcotton$_c$)					-3.007*
					(1.717)
Observations	2,335	2,390	2,381	2,391	2,335
R-squared	0.645	0.552	0.599	0.577	0.641

Notes: *** p< 0.01, **p< 0.05, *p< 0.1. State FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's 1870 population. All columns estimated using OLS.

Table A9: Land Improvement and Mechanization

Dep. Var.	(1) sh. improved	(2) a(implements and machinery)
a(GAEZ$_c$) x a(exp$_t$)	0.0369	-0.314
	(0.0331)	(0.201)
Observations	9,477	9,481
R-squared	0.001	0.001

Notes: *** p< 0.01, **p< 0.05, *p< 0.1. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns estimated using IV; for a description of the instrument, see the main text.

Table A10: Main Results, Latitude and Longitude x Year Controls

Dep. Var.:	(1) a(urban pop)	(2) sh. rural	(3) a(farms)	(4) a(farmland, improved)	(5) a(wheat output)
a(GAEZ$_c$) x a(exp$_t$)	1.064*	-0.0424	-0.610***	-0.653***	-1.186
	(0.631)	(0.0285)	(0.220)	(0.192)	(0.734)
Observations	9,564	9,564	9,479	9,481	9,497
R-squared	0.010	0.017	0.061	0.095	0.028

Notes: *** p< 0.01, **p< 0.05, *p< 0.1. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns estimated using IV.

Table A11: Main Results, Alternative Source for IV

Dep. Var.:	(1) a(urban pop)	(2) sh. rural	(3) a(farms)	(4) a(farmland, improved)	(5) a(wheat output)
a(GAEZ$_c$) x a(exp$_t$)	1.064*	-0.0424	-0.610***	-0.653***	-1.186
	(0.631)	(0.0285)	(0.220)	(0.192)	(0.734)
Observations	9,564	9,564	9,479	9,481	9,497
R-squared	0.010	0.017	0.061	0.095	0.028

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. County FE and state-year FE are included. a() refers to the inverse hyperbolic sine transformation. Standard errors are clustered by state. Observations are weighted by each county's initial, 1870 population. All columns estimated using IV.