Title
Essays on Economic History

Author
Xu, Meng

Publication Date
2018

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Essays on Economic History

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Meng Xu

2018

Abstract of the Dissertation

Essays on Economic History

by

Meng Xu

Doctor of Philosophy in Economics

University of California, Los Angeles, 2018

Professor Dora Luisa Costa, Chair

The first chapter of my dissertation uses a competition framework to reexamine the U.S. railway expansion from 1870 to 1890. Do reductions in trade costs increase the income levels of trading regions? The answer of most theoretical trade models is yes, and these models are used by policy-makers to justify investments in new transportation infrastructure to reduce trade costs. I reexamine this question by investigating the impact of the U.S. railway expansion from 1870 to 1890 on county economic growth. My main insight is that while the construction of a new transportation network can benefit a location by increasing producers' access to markets, it also can hurt a location by giving other locations a competitive edge. I derive and test empirically testable implications of the joint effects of market access and competition on economic outcomes from a multisector trade model. I find that on average more than 50% of the benefit of market access is offset by the negative effects of competition. Overall, I find that the expansion of railways increased agricultural land values by 20% to 40%, thus implying that railways explain 1% to 2% of 1890 GNP. My estimates of the impact of railways on American economic growth are lower than those of Donaldson and Hornbeck (DH16) and closer to those of Fogel (Fog64).

In the second chapter I study the effects of sector-specific productivity shocks on the distribution of production. The general-equilibrium trade model implies that the production in a certain location depends not only on its own productivity shocks but also on productivity shocks in competing locations. I derive several testable implications of the joint effects of

one location's own productivity shock and productivity shocks in competing locations on the location's production. I test these implications by investigating the impact of the boll weevil in the early 20th century United States on the cotton belt's cotton acreage. I find (1) when few lands in the cotton belt were infested by the boll weevil, the presence of the boll weevil decreased the infested county's cotton acreage; (2) as the boll weevil expanded to more lands, the infested county's cotton acreage would gradually recover; (3) as the boll weevil expanded to more lands, cotton acreage in uninfested counties would increase. The movement of other economic outcomes which closely associated with the cotton production (such as corn acreage, total farm acreage, agricultural land value, population, and the tenancy system) mirrored the movement of the cotton acreage. My findings are in line with the argument that the boll weevil caused the internal shift in the cotton belt but rarely changed the whole cotton belt.

In the third chapter, coauthored with Ming Gu, we discuss the long-run effects of the propaganda by the central government on people's preference formation. This link is taken as given but never proved in the previous literature. This paper examines the impact of government propaganda on people's attitudes towards gender equality in China, which imposes strict media censorship, advocates gender equality, and experiences growth in female labor participation. We find the high propaganda intensity would lead to progressive gender-related attitudes in female labor participation, but also conservative attitudes in housework share. It is consistent with the context of gender equality propaganda in China. Propaganda has larger positive effects on women than men. Women were affected in the age range between 0-25, while men were affected in the age range between 0-17. Propaganda also reduced boy preference for both men and women at the age range between 18-25, the marriage age and childbearing age of most Chinese.

The dissertation of Meng Xu is approved.

James Tong

WIlliam Walker Hanlon

Leah Michelle Boustan

Dora Luisa Costa, Committee Chair

University of California, Los Angeles

2018

Table of Contents

List of Figures

List of Tables

Acknowledgments

Completion of this doctoral dissertation is possible with the support of several people. I would like to express my sincere gratitude to all of them. First and foremost, I offer my sincerest gratitude to my committee chair, Professor Dora Costa, who has supported me throughout my dissertation, from the ways of thinking to every sense of the word, with her valuable guidance, scholarly inputs, consistent encouragement, and extreme patience; while allowing me the room to work in my own way. She offered me uncountable scholarly inputs, even though we do not share the same research interests. She always made herself available to answer my questions and read and edit my dissertation despite her very busy schedules. She always encouraged me instead of blaming me when I had a bottleneck and got stuck in research. Without her, this dissertation would not have been completed. One could not wish for a better advisor.

Besides my main advisor, I would like to thank the rest of my doctoral committee: Professor Walker Hanlon, Professor Leah Boustan, Professor James Tong. I thank Professor Walker Hanlon. I got the initial inspiration of the first and the most important chapter from him. I always felt excited when I talked to him and got many insightful ideas. It is my highest honor to be the first few students of his academic career. I thank Professor Leah Boustan for her hard questions which incentivized me to widen my research from various perspectives. I thank Professor James Tong for his excellent knowledge about China's historical and contemporary data. I could not forget that he often invited all Chinese Ph.D. students in the Division of Social Science UCLA to dinner together. It gave me opportunities to know many students in other department and exchange my ideas with them.

I also thank my undergraduate advisor Professor Se Yan at Peking University. The seed of the first and the most important chapter of this dissertation was planted in 2011 with him. I own my thanks to him, who not only first drew my attention to economic history but also inspired me to think the impact of railways on the regional competition. With his guidance, I found that many prosperous towns which were located along the traditional trade road declined after the construction of the parallel railways in adjacent counties. The idea of the first chapter of my doctoral dissertation is based on this finding.

I thank my family and friends for supporting me throughout all hard times in my Ph.D. studies at UCLA. My parents, Hong Xu and Qiongfang Zhou, are delighted that I am finally finishing my dissertation. They supported me in every possible way to see the completion of this work. I feel lucky that I met Jen-Wen Chang, who became my husband in 2017, in the Ph.D. program. Although we are in different fields of economics — he is a theorist in microeconomics while I am an economic historian, he worked as an editor for me. He read this entire draft, discussed the logical details, provided valuable comments, and helped me to organize all three chapters as one dissertation. He also unconditionally supported me when I felt frustrated in life and research. I thank my cousin, Yifan Zhou, for helping me digitize some data. I appreciate my grandmother, Shulan Dan, who is an illiterate rural woman and no longer with us but believed in my ability to get a Ph.D. degree. I thank my best friends Xiyan Wang, Yi Jin, and Airan Liu for their friendship since college. I spent several fun weekends with Xiyan Wang, who was a Ph.D. student in transportation economics at UCI, in the Greater Los Angeles Area. I traveled with Yi Jin, who is a Ph.D. student in Human Geography Urban Studies at LSE in Morocco and Scotland. I made phone calls with Airan Liu, who was a Ph.D. student in sociology at the University of Michigan to complain about the difficulties in Ph.D. life. I also talked my research with them, and they gave me some inspirations from their perspectives.

Last, I must thank the era that I am living in and the country in which I grew up. The environment that I grew up has formed my ways of thinking. The rapid economic growth in China not only makes it possible for me to study in the United States but also provide fantastic insights for me to think of the economy developing process. It is also an era that many great brains have built a beautiful knowledge system in economics. Without the theoretical framework in economics, it is impossible to think of history in the way of my dissertation.

Vita

2012	B.A. (Economics) and B.A. (Sociology), Peking University.
2013	M.A. (Economics), UCLA, Los Angeles, California.
2014	Ph.D. Candidate (Economics), UCLA, Los Angeles, California.

CHAPTER 1

Regional Competition and Economic Growth: Winners and Losers from Market Integration

1.1 Introduction

Models of international and interregional trade posit that reductions in trade costs will increase income level in trading regions. Such reductions can be achieved through investments in transportation infrastructure that give producers greater market access. Whether an infrastructure investment is worth it depends on the increase in total welfare and is an empirical issue. One difficulty in calculating total welfare is that, to date, researchers have failed to account for the negative impact of reduced trade costs. Some locations suffer because they face increased competition from other locations which produce similar products.

My paper is the first to account for both the positive effects of market access and the negative effects of competition on locations connected by a new transportation network. I examine the impact of the U.S. railway expansion of 1870 to 1890 on the economic outcomes of counties using insights from a multisector general-equilibrium trade model. I find that the railway expansion increased agricultural land values by 20% to 40%, equivalent to 1% to 2% of 1890 GNP. On average, more than 50% of the positive effect of market access was offset by the negative effects of competition.

My findings reconcile the long-standing controversy over the impact of new transportation networks on economic growth. (Fog64) used the "social saving" approach and found that the overall increase in transportation costs attributable to railroads was no more than 4.7% of 1890 GNP for all economic sectors and 2.7% for agriculture (2.1% for intraregional trade and 0.6% for interregional trade). If all 1890 railroads were replaced by proposed canals and improved roads, the increase in transportation cost in agriculture may have been as low as 1.8% of 1890 GNP(0.8% for intraregional trade and 1% for interregional trade). Estimates of the impact of railroads on economic growth in counties with fewer water transportation options are larger ((Cra04)). (DH16) use a "market access" approach and find that by removing all railways in 1890, the reduction in agricultural land values generates annual economic losses equal to 3.2% of GNP. I argue that following the market-access approach without considering the effects of structural competition substantially overestimates the effect of railways on total agricultural land values. My estimates of the impact of the railways are one-third to two-thirds of those asserted by (DH16). By removing all railways in 1890, the loss of agricultural land values amounts to only 1% to 2% of 1890 GNP.

My findings can explain why some locations were winners and others were losers when market access increased. For example, counties in New Mexico experienced a high increase in market access from 1870 to 1890 but below-average economic growth. The positive effects of market access were fully offset by the negative effects of competition. As I show in my theoretical framework, those offsets are more likely when neighboring counties have similar endowments and industrial structures. My results are consistent with (CT00)'s study of highway expansion, which finds that neighboring counties experienced a decline.

My findings can help elucidate the roles of productivity and comparative advantage in economic growth in three ways. First, my model implies that highly productive counties will suffer less from trade shocks that increase competition. Economies of scale made urbanized counties with large cities highly productive and my empirical work shows that these counties experienced higher growth than other urbanized counties. Second, I show that trade shocks can change a location's comparative advantage. I find that greater agricultural competition reduced agricultural land and output values, while greater nonagricultural competition positively affected these values. Third, I find that a trade network that favored agriculture over nonagriculture sectors helped economic growth. I find that increasing agricultural competition reduced economic growth. My results contrast with (Mat92)'s prediction of a negative link between agricultural productivity and economic growth for a small, open economy.[1]

My findings also have implications for regional development. Southern economic growth lagged that of the Midwest after the Civil War. Some Southern cities, such as New Orleans, La., and Mobile, Ala., experienced stagnant or declining growth. Railways built by Northerners connected small Southern towns to Northern cities, drawing large Southern cities into fierce competition. Railroads brought greater competition to the Southern cities than to Midwestern cities, which explains the growth gap between the Southern and Midwestern cities.

Recent research has highlighted the negative effects of competition on trade integration. (GLZ14) find that countries similar in industrial structure to China experienced losses after China's accession to the WTO in 2001. (CPR14) find that total-factor productivity gains in the computer and electronics industry in California adversely affected neighboring states competing in the same industry. (CF16) also focuses on China's trade integration. They find that during trade integration, the interior region of China lost out because workers emigrated to the coastal regions. These papers use structural estimation and focus on the net gain of trade integration. In contrast, my paper not only discusses the aggregate gain of trade integration but also decomposes the positive and negative effects of trade integration. My historical setting also enables me to answer broader questions about the early stages of industrialization, above and beyond assessments of trade integration.

In Section 2, I model the impact of the expansion of the railway network on economic outcomes using the multisector Eaton-Kortum model. Each location has its unique endowments and industrial structure and thus experiences different competitive shocks. Even though a new transportation network increases market access, the negative effects of competition can offset the positive effects of market access and lead to flat or even negative economic growth. I derive two implications from this model: (1) holding market access constant, higher competition will lead to worse economic performance, and (2) because resources are allocated to sectors facing fewer competitive pressures, competition in one sector will have a negative

[1] (Mat92) assumes that workers are immobile across locations. It is a strong assumption.

impact on economic outcomes in that sector but a positive economic impact on outcomes in a sector facing fewer competitive pressures.

In Section 3, I demonstrate how I bring my model to the data. I show that the positive and negative effects of trade integration alike can be measured in a simple reduced-form manner without price and trade-flow data. My methodology thus can be used in other historical and developing-country contexts, where the data are often limited. I also construct a measure of changes in competition that can be used by other researchers. I first measure the changes in sector-specific competition by fixing the trade cost between producers and markets at the initial level and allowing the trade cost between competitors (producers in other locations who produce similar goods) and markets to decrease after the advent of a new transportation network. The intuition is that if competing locations are better connected to markets, sectors in each location will experience increases in competition. Because increases in sector-specific competition depend on relative productivity, if a location has relatively higher productivity in a sector, it will experience less competition when competitors connect to its markets, and vice versa. Finally, I measure total competition in a location as the weighted summation of sector-specific competition in that location, where sectors are weighted by their importance in the whole economy. The aggregate competition changes in a location is thus affected by industrial structure and productivity in all locations in the same trade network.

In Section 4, I describe the empirical strategy and examine the effects of competition on economic performance. I find that (1) an increase in competition reduced economic growth measured by several indicators, including total population, urban population, city size, agricultural land values, the value of agricultural output, the number of workers in manufacturing, and value added in manufacturing; (2) an increase in agricultural competition reduced agricultural land and output value, while an increase in competition in all sectors positively affected agricultural outcomes; and (3) an increase in agricultural competition negatively affected the economy as a whole, while an increase in competition in all sectors positively affected the whole economy. In Section 5, I discuss several strategies to mitigate potential endogeneity problems.

In Section 6, I find that my model can explain several empirical observations. The effects of competition help explain the development gap between counties with new rail tracks and their adjacent counties, the growth gap between urbanized counties with large cities and other urbanized counties, and the growth gap between large Southern cities and large Midwestern cities after the Civil War.

In Section 7, I undertake a counterfactual analysis to show that in the scenario of no railways, agricultural land values would lose about 20% to 40% of its value. I also show that this estimate varies widely across locations (the standard deviation is about 10% to 20%). As previously noted, my estimates are equivalent to about 1% to 2% of 1890 GNP and are one-third to two-thirds of the ones in (DH16). My estimates are smaller than Fogel (1964)'s estimates for agriculture, a 2.7% loss in 1890 GNP. In an extreme case, with an extended canal network and improved country roads, the decline agricultural land values was as low as 16.1%, or 0.8% of 1890 GNP.

In Section 8, I offer concluding remarks.

1.2 Theoretical Framework

I apply a simple version of the multisector EK model in (CPR14) to my context. This economy comprises several sectors (j) and locations (production locations i, markets n, competing locations m).

Consumer preference is Cobb-Douglass, where the consumption share of each sector j is α^j. The utility of a representative consumer who lives in location i is

$$U_i = \prod_j (c_i^j)^{\alpha^j}, \sum_j \alpha^j = 1, \tag{1.1}$$

where c_i^j is the amount of goods of sector j consumed by the representative consumer in location i.

In equilibrium, utility is equalized among locations. Thus we have

$$\bar{U} = \frac{Y_i}{L_i P_i}, \forall i, \tag{1.2}$$

where Y_i is the total income in location i, L_i is the number of workers in location i, and P_i is price index in location i.

To simplify to (CPR14), I assume that goods are produced by labor and land (including house structure). Materials are not used in production.[2]

Income in location i is expressed as

$$Y_i = w_i L_i + r_i H_i, \tag{1.3}$$

[2] If a sector uses materials produced in other sectors for its production, there will be input-output linkages across sectors. I exclude the input-output linkages between sectors for several reasons. First, it is hard to incorporate input-output linkage into a simple, applicable reduced-form strategy. The biases caused by excluding materials from production can be partially corrected by incorporating the demand from other producers into the model (Appendix D.1). With this correction, the biases in estimating competition effects on agricultural outcomes are minimized, since in the process of agricultural production, materials produced by other sectors play a very weak role, and agriculture provides a substantial portion of inputs for manufacturing production. This simplification makes more biases in estimating competition effects on manufacturing outcomes, since for manufacturing producers, competition not only exists in markets but also in the supply of materials. Excluding materials from production prevents me from estimating the effects of competition in the supply of materials.

where w_i is the nominal wage in location i, r_i is the nominal land and house structure rent rate in location i, and H_i is the total amount of land and house structure in location i.

Price index P_i in location i is defined as

$$P_i = \prod_j (\frac{P_i^j}{\alpha^j})^{\alpha^j}, \tag{1.4}$$

where P_i^j is the expected price of goods in sector j location i, derived below.

Production functions are also Cobb-Douglass, with constant labor share β. Each location-sector pair (i, j) draws an idiosyncratic productivity z_i^j independently from a Frechet distribution $F_i^j(z) = \exp(z^{-\theta})$ with a shape parameter θ. Productivity in sector j and location i is also affected by exogenous endowments of location i, which generates a location-sector-specific productivity level T_i^j.

Therefore, the marginal cost of production in sector j and location i is

$$MC_i^j = \frac{r_i^\beta w_i^{1-\beta}}{z_i^j T_i^j}. \tag{1.5}$$

Goods are tradable with iceberg costs. One unit of any good in any sector j shipped from location i to location n requires producing $\kappa_{ni} > 1$ unit of goods in location i. The exogenous productivities T_i^j, as well as shipping costs, construct the geographical variation in competition.

Consumers shop around and find the cheapest location to buy from. Therefore, under perfect competition, location n can purchase sector j goods at the following price:

$$p_n^j(z^j) = \min_i \{\kappa_{ni} MC_i^j\} = \min_i \left\{ \frac{\kappa_{ni} r_i^\beta w_i^{1-\beta}}{z_i^j T_i^j} \right\}. \tag{1.6}$$

With the Frechet distribution of z_i^j, the expected price of goods consumed in sector j location n is

$$P_n^j = \delta_1 [\sum_i (r_i^\beta w_i^{1-\beta} \kappa_{ni})^{-\theta} T_i^j]^{-\frac{1}{\theta}}. \tag{1.7}$$

The share of location n's total expenditure on the goods of sector j purchased from location i is given as

$$\pi^j_{ni} = \frac{X^j_{ni}}{X^j_n} = \frac{(r_i^\beta w_i^{1-\beta}\kappa_{ni})^{-\theta} T_i^j}{\sum_m (r_m^\beta w_m^{1-\beta}\kappa_{nm})^{-\theta} T_m^j}. \tag{1.8}$$

I define consumer market access for sector j in market n as

$$CMA^j_n \equiv (P^j_n)^{-\theta} = \delta_2 \sum_i (r_i^\beta w_i^{1-\beta})^{-\theta} \kappa_{ni}^{-\theta} T_i^j, \tag{1.9}$$

which measures how cheaply a consumer who lives in market n can access goods in sector j. The index is increasing in productivity but decreasing in shipping cost and input price.

I define the aggregate consumer market access in market n as

$$CMA_n \equiv P_n^{-\theta} = \prod_j \left(\frac{CMA^j_n}{\alpha^j} \right)^{\alpha^j}, \tag{1.10}$$

which measures how cheaply a consumer who lives in market n can purchase a basket of consumer goods.

Sector-specific firm market access is defined as

$$FMA^j_i \equiv T^j_i \sum_n Y_n \left(\frac{P^j_n}{\kappa_{ni}} \right)^{\theta}. \tag{1.11}$$

A location i firm's market access in sector j is the weighted sum of the ratios of expected price and shipping cost weighted by market size. It is also increasing in productivity.

Aggregate firm market access in location i is equal to the summation of FMA^j_i, as follows:

$$FMA_i \equiv \sum_j \alpha^j FMA^j_i = \delta_3 \sum_j \alpha^j \sum_n \frac{\kappa_{ni}^{-\theta} Y_n}{\sum_m (r_m^\beta w_m^{1-\beta})^{-\theta} \kappa_{nm}^{-\theta} \frac{T^j_m}{T^j_i}}. \tag{1.12}$$

The expression of FMA_i is interesting. The numerator $\kappa_{ni}^{-\theta} Y_n$ is the same as the expression of market access from market n in (DH16). It is increasing in market size of n, Y_n, but decreasing in trade cost between location i and market n. The denominator is the summation of competition from competing location m. $(r_m^\beta w_m^{1-\beta})^{-\theta} \kappa_{nm}^{-\theta} \frac{T^j_m}{T^j_i}$ is the location i sector

j's competition from competing location m in market n. It is increasing in the productivity of location m to location i in sector j $\frac{T_m^j}{T_m^j}$, but decreasing to the input price in location m and the trade cost between location m to market n.

Under the above framework, the logarithm land rental rate, the number of workers, and total output can be expressed as linear functions of aggregate consumer market access and aggregate firm market access.

The number of workers (population) can be expressed as

$$(1+\beta\theta)\ln L_i = \delta_4 + \beta\theta \ln H_i + (1+\frac{1}{\theta})\ln CMA_i + \ln FMA_i. \tag{1.13}$$

Total land rental income in sector j can be expressed as

$$(1+\beta\theta)\ln(r_i H_i^j) = \delta_5 + \beta\theta\ln H_i + (1-\beta)\ln CMA_i + (1+\beta\theta)\ln FMA_i^j - \beta\theta\ln FMA_i. \tag{1.14}$$

The number of workers in sector j can be expressed as

$$(1+\beta\theta)\ln L_i^j = \delta_6 + \beta\theta\ln H_i + (\beta+\frac{1}{\theta})\ln CMA_i + (1+\beta\theta)\ln FMA_i^j - \ln FMA_i. \tag{1.15}$$

The value of output is sector j location i can be expressed as

$$(1+\beta\theta)\ln Y_i^j = \delta_7 + \beta\theta\ln H_i + (1-\beta)\ln CMA_i + (1+\beta\theta)\ln FMA_i^j - \beta\theta\ln(FMA_i). \tag{1.16}$$

These equations have a number of implications.

Implication 1: Holding other variables constant, increasing the aggregate consumer market access CMA_i will cause both economic outcome of the whole economy (total population) and sector-specific economic outcomes (the number of workers hired in the sector, land rental income in the sector, and output value in the sector) to increase in location i.

Implication 2: Holding other variables constant, increasing the aggregate firm market access FMA_i will cause the number of workers L_i to increase in location i.

Implication 3: Holding other variables constant, increasing firm market access in sector j will cause sector j–related economic outcomes (the number of workers hired in the sector, total land rental income in the sector, and output value in the sector) to increase, while increasing aggregate firm market access will cause sector j–related economic outcomes to decrease.

My main contribution to the literature, however, is accounting for the effects of competition on economic outcomes. In particular, I single out the effects of structural competition faced by location i to FMA_i. To do so, I fix the connections between location i and all markets n κ_{ni} at the initial level, but allow input prices and connections between all competing locations m and all markets n κ_{nm} to change with time. The changes in competition between time t and time $t+1$ for location i is defined as

$$\Delta Competition_i \equiv \ln\left(\sum_j \alpha^j \sum_n \frac{\kappa_{ni,t}^{-\theta} Y_{n,t}}{\sum_m (r_{m,t}^{\beta} w_{m,t}^{1-\beta})^{-\theta} \kappa_{nm,t}^{-\theta} \frac{T_m^j}{T_i^j}}\right) - \ln\left(\sum_j \alpha^j \sum_n \frac{\kappa_{ni,t}^{-\theta} Y_{n,t+1}}{\sum_m (r_{m,t+1}^{\beta} w_{m,t+1}^{1-\beta})^{-\theta} \kappa_{nm,t+1}^{-\theta} \frac{T_m^j}{T_i^j}}\right). \tag{1.17}$$

In particular, the larger the trade cost decline from competing location m to market n is, the larger $\Delta Competition_i$ is, and the smaller the increase in FMA_i is. This relationship between competition and FMA_i, together with Implications 2 and 3, yields two additional implications:

Implication 4: Holding other variables constant, a larger competition increase in all sectors implies less aggregate firm market access increase, and then less population increase.

Implication 5: Holding other variables constant, a larger competition increase in sector j implies less increase in firm market access in sector j, less increase in sector-related economic outcomes (the number of workers hired in the sector, total land rental income in the sector, and output value in the sector); meanwhile, a larger competition increase in all sectors implies less aggregate firm market access increase, and then a larger increase in sector-related economic outcomes.

Implication 4 captures the negative effects of competition. Implication 5 states that trade could affect relative competition among sectors and then a location's comparative advantage.

The structure of (1.17) helps us understand regional development in history. It solves the discrepancy between the market-access approach in (DH16) and empirical findings in previous literature that some locations experienced large increases in market access but later declined. (CT00) finds that counties that were not directly linked by new highway routes but were located *near* highway routes declined. (Fab13) finds that periphery counties linked by new highway routes declined. The two stories can be explained by the framework of my paper. In (CT00), neighboring counties usually share similar natural resources. For a nearby location i, which is not linked by new highways, and its competing location m, which is directly connected by highways, the increase of κ_{nm} is larger than the increase of κ_{ni}, and the competition effects offset the market-access effects. In (Fab13), because of economies of scale, central locations have higher productivity than their nearby periphery locations; thus, when both central locations and periphery locations are connected by highways, periphery locations will face a larger increase in competition.

I empirically test Implications 4 and 5, and I show that the negative effects of competition explain the growth gap between three groups of locations from 1870 to 1890: counties with new rail tracks and their adjacent counties without new rail tracks, urbanized counties with a large city and other urbanized counties, and Midwestern and Southern counties. Further, I apply Implication 2 to the counterfactual analysis to estimate the aggregate railway effects on agricultural land values.

1.3 From Theory to Empirical Strategy

1.3.1 Market Access and Competition

The sector-specific consumer market access in equation (1.9) captures how cheaply a consumer who lives in location i can access goods in sector j. The cheapness can be divided into three parts: the cheapness of factor price in producer n $(r_n^{\beta} w_n^{1-\beta})^{-\theta}$, the cheapness of trade cost $\kappa_{ni}^{-\theta}$ between location i and producer n, and the location-sector-specific productivity in producer n T_n^j. I exclude the endogenously co-determined factor price from the empirical measurement of the term CMA_i^j, since input price is endogenously determined in the model. Including it in the measurement of CMA and FMA will raise endogeneity bias.

I use the following first-order approximation to measure the sector-specific consumer market access CMA_{it}^j:

$$CMA_{it}^j \approx \sum_n \kappa_{nit}^{-\theta} T_n^j, \tag{1.18}$$

which focuses on how cheaply consumers in location i can access producers with high productivity in sector j.[3]

The aggregate consumer market access in location i CMA_i is approximated by Equation (1.10).

The aggregate firm market access can be approximated as

$$FMA_{it} \approx \sum_j \alpha^j \sum_{n \neq i} \frac{\kappa_{nit}^{-\theta} Y_{nt}}{\sum_{m \neq i} \kappa_{nmt}^{-\theta} \frac{T_m^j}{T_i^j}}. \tag{1.19}$$

Similarly, the aggregate firm market access in agriculture can be approximated as

$$FMA_{it}^A \approx \sum_{j \in A} \alpha^j \sum_{n \neq i} \frac{\kappa_{nit}^{-\theta} Y_{nt}}{\sum_{m \neq i} \kappa_{nmt}^{-\theta} \frac{T_m^j}{T_i^j}}, \tag{1.20}$$

[3] It can be proved that the first-order approximation is highly correlated with the CMA_i^j, which contains the observed land rental rate and manufacturing wages from data.

and the aggregate firm market access in manufacturing can be approximated as

$$FMA_{it}^{M} \approx \sum_{j\in M} \alpha^j \sum_{n\neq i} \frac{\kappa_{nit}^{-\theta} Y_{nt}}{\sum_{m\neq i} \kappa_{nmt}^{-\theta} \frac{T_m^j}{T_i^j}}. \tag{1.21}$$

I also exclude the competition from location i itself in the denominator, since the simplified structure can help us better understand the competition effects. The difference between including and excluding self-competition is minimal, because the contribution of self-competition to overall competition is tiny. I also exclude the market access from location i to itself, since market size and other economic outcomes are co-determined in the model. We can ignore the distortion caused by the exclusion, since the trade cost is assumed to be time-invariant, and any market is only a small portion of all markets for most producers.

1.3.2 Market Access and Changes in Competition

As I have discussed, aggregate firm market access FMA_i can be decomposed into two parts: one is market access, market size weighted by trade cost; the other is competition, the connection between competitors who produce similar goods. Following (DH16), I define market access of location i at year t as $MA_{it} \approx \sum_{n\neq i} \kappa_{nit} Y_{nt}$.

Following the strategies in measure FMA, the changes in competition is:

$$\Delta Competition_i \approx \ln\left(\sum_{j} \alpha^j \sum_{n\neq i} \frac{\kappa_{ni,t}^{-\theta} Y_{n,t}}{\sum_{m\neq i} \kappa_{nm,t}^{-\theta} \frac{T_m^j}{T_i^j}}\right) - \ln\left(\sum_{j} \alpha^j \sum_{n\neq i} \frac{\kappa_{ni,t}^{-\theta} Y_{n,t+1}}{\sum_{m\neq i} \kappa_{nm,t+1}^{-\theta} \frac{T_m^j}{T_i^j}}\right) \tag{1.22}$$

In particular, holding other variables constant, if the trade cost between competing location m and market n decreases between period t and $t+1$ ($\kappa_{nm,t+1} < \kappa_{nm,t}$, $\kappa_{nm,t+1}^{-\theta} > \kappa_{nm,t}^{-\theta}$), producer i will face fiercer competition from competitor m in the new transportation network ($\Delta Competition_i$ will be positive).

Changes in competition in sector $J \in \{\text{Agriculture, Manufacturing}\}$ is expressed as

$$\Delta Competition_i^J = \ln\left(\sum_{j\in J} \alpha^j \sum_{n\neq i} \frac{\kappa_{ni,t}^{-\theta} Y_{n,t}}{\sum_{m\neq i} \kappa_{nm,t}^{-\theta} \frac{T_m^j}{T_i^j}}\right) - \ln\left(\sum_{j\in J} \alpha^j \sum_{n\neq i} \frac{\kappa_{ni,t}^{-\theta} Y_{n,t+1}}{\sum_{m\neq i} \kappa_{nm,t+1}^{-\theta} \frac{T_m^j}{T_i^j}}\right) \tag{1.23}$$

Next, I linearly decompose $\Delta \ln FMA_i$ into the positive parts $\Delta \ln MA_i$ and negative parts $\Delta Competition_i$, as follows:

$$\Delta \ln FMA_i = \lambda_1 \Delta \ln MA_i + \lambda_2 \Delta Competition_i + \varepsilon_i. \quad (1.24)$$

One concern is that the decomposition is arbitrary, and the relationship between $\Delta \ln FMA_i$ and $\Delta \ln MA_i$ is not linear conditional on $\Delta Competition_i$, and the relationship between $\Delta \ln FMA_i$ and $\Delta Competition_i$ is not linear conditional on $\Delta \ln MA_i$. In Appendix A, I show that the relationship is indeed linear, by fitting a polynomial to residual changes in log firm market access and residual changes in competition, conditional on residual changes in log market access and other control variables.

1.3.3 Parameters and Data

In order to measure the aggregate consumer market access CMA_i, aggregate firm market access FMA_i, aggregate firm market access in agriculture FMA_i^A, aggregate firm market access in manufacturing FMA_i^M, market access in previous literature MA_i, changes in competition $\Delta Competition_i$, changes in competition in agriculture $\Delta Competition_i^A$, and changes in competition in manufacturing $\Delta Competition_i^M$, I need to know the value of a series of parameters: consumption share of subsector α^j, transportation cost κ_{nit}, location-sector-specific productivity T_i^j, market size Y_{it}, and trade elasticity θ.

1.3.3.1 Consumption Share α_j

Historical consumption data are not available. Following the method in (CPR14), I use the value of domestic products plus the value of imports minus the value of exports as a proxy for value of consumed goods. The consumption share of sector j is calculated by the following equation:

$$\alpha^j = \frac{Y^j + IMP^j - EXP^j}{\sum_j (Y^j + IMP^j - EXP^j)}, \quad (1.25)$$

where Y^j is the total value of product in sector j, IMP^j is the value of imports in sector j, and EXP^j is the value of exports in sector j. I obtain the data from (HPR10) and the (Tre70).

I do not consider international markets in the baseline estimation. However, because international trade involves U.S. production, I have to consider them. I discuss this concern in Appendix D.1.

1.3.3.2 Transportation Cost κ_{nit}

Transportation cost κ_{nit} is measured as the least pairwise transportation cost in 1870 and 1890. I obtained the data from (DH16).

1.3.3.3 Location-Sector-Specific Productivity T_i^j

Location-sector-specific productivity T_i^j describes the natural resources that deterministically affect the productivity of a specific industry in a county. The whole economy is made up of three major sectors: agriculture, mining, and manufacturing. Based on similarity of goods and availability of data, each sector is divided into several subsectors.[4]

I use the agro-climatic crop suitability data ((IIA12)) as a proxy for T_i^j in crop-planting subsectors. For livestock production, I use modern livestock productivity data ((Rob11)) as a proxy. One concern is whether producers will face competition only from competitors who produce similar goods. If sector j did not exist in location i, the changes in trade

[4] Subsectors of agriculture are barley, buckwheat, beans and peas, cotton, flax, maize, oats, rice, rye, sweet potatoes, tobacco, wheat, white potatoes, and livestock. Subsectors of manufacturing are food and kindred products; tobacco products; textile mill products; apparel and other finished products made from fabrics and similar materials; lumber and wood products except furniture; furniture and fixtures; paper and allied products; printing, publishing, and allied industries; chemicals and allied products; petroleum refining and related industries; rubber and miscellaneous plastics products; leather and leather products; stone, clay, glass, and concrete products; primary metal industries; fabricated metal products, except machinery and transportation equipment; industrial and commercial machinery and computer equipment; electronic and other electrical equipment and components, except computer equipment; transportation equipment; measuring, analyzing, and controlling instruments, photographic, medical and optical goods, watches and clocks; and miscellaneous manufacturing industries. Subsectors of mining are stone, coal, copper, silver, gold, slate, lead, petroleum, zinc, nickel, asphalt, and iron.

cost between producers in sector j in other locations and markets would not affect economic performance in location i. Therefore, if sector j did not exist in location i in 1870, in my OLS regression I set $T_i^j = 0$.

For nonagricultural subsectors, T_i^j is measured as value added per worker in 1870. I obtained the data source from (HPR10). The division of manufacturing subsectors is based on the two-digit code in the 1987 Standard Industrial Classification (SIC) System ((DOL87)).

To make all sectors comparable, I normalize T_i^j to a range of 0-1 in each subsector.

One concern is whether subsector choice is not exogenous in agriculture and value added per worker is not exogenous in manufacturing. I will undertake an IV strategy to deal with this problem in Section 5.1.

1.3.3.4 Market Size Y_{nt}

I use population as a proxy for market size Y_{nt} in the baseline analysis. In Appendix C, I also use the total value of agricultural output plus value added in nonagricultural sectors as a robustness check.

1.3.3.5 Trade Elasticity θ

It is impossible to accurately estimate the shape parameter under the framework of multisector EK model with available historical data from 1870 to 1890. Two earlier research papers provide reasonable values of θ for my context. (CP15) compute θ for 20 tradable sectors with data at the two-digit level of the third revision of the International Standard Industrial Classification (ISIC Rev. 3). I match their classification to my classification ((DOL87)). Weighting their sectoral θ by consumption share in 1870, I calculate the value of aggregate θ across all sectors as 7.35. The value of aggregate θ across all manufacturing sectors is 6.10. The matching table and weights of each sector are in Appendix B. (DH16) also compute a reasonable value of θ in this context. They estimate the value of θ in agriculture using historical data of the same period. Their value is 8.22, very close to the (CP15) estimation, 8.11. It is possible to generate biases by using parameter estimates from other literature

—the estimates of trade elasticity vary by researcher. I also discuss other possible values of θ in Appendix D.2 as a robustness check.

Since for simplicity I assume θ is location-sector-invariant, I believe it's reasonable to apply my weighted average value 7.35 when I discuss economic outcomes related to all sectors, to apply (DH16) value 8.22 when I discuss economic outcomes related to agriculture, and to apply my weighted average value 6.10 when I discuss economic outcomes related to manufacturing.[5]

[5] I discuss an extended model, with the assumption that θ can vary by sector, and its estimation results in Appendix F. The results are very similar to the results in the body of the paper.

1.4 Empirical Strategy and Results

In this section, I test Implication 4 and Implication 5. I reconsider the question in (Mat92) of whether the link between comparative advantage in agriculture and economic growth is positive or negative.

1.4.1 Negative Effects of Competition on Economic Growth

I use the data described in the previous section to estimate the effects of the changes in competition between 1870 and 1890 on changes in economic outcomes during the same period. My baseline model is as follows:

$$\Delta \ln EconomicOutcome_i = \delta + \eta_1 \Delta \ln(MA_i) + \eta_2 \Delta Competition_i + \delta_s + f(x_i, y_i) + \varepsilon_i \quad (1.26)$$

where δ_s refers to state fixed effects and $f(x_i, y_i)$ refers to latitude and longitude and their quadratic and cubic terms of a location. I exclude changes in log consumer market access from the regression, since changes in log consumer market access and changes in log market access are highly correlated.[6]

I run the model (1.26) on the whole economy and separately on the agricultural and manufacturing sectors; I consider several variables representing the economic outcome.

Table 1.1 reports the results. In Panel A, I allow the market size to change between 1870 and 1890. One concern is whether changes in log market access and changes in competition depend not only on reduced trade cost but also on reallocation and growth of population. To better understand the positive market-access effects and negative effects of competition caused by reduced trade cost, in Panel B I fix the market size at the 1870 level. In general, conditional on changes in log market access, the effects of competition are negative and substantial for all measurements of economic growth. The effects are slightly smaller in Panel B than in Panel A. In most columns, the negative effects of competition are statistically

[6] The correlation coefficient is larger than 0.99 regardless of the value of θ. The results would clearly show positive effects and negative effects by dropping one of the two variables that cause multicollinearity.

Table 1.1: Positive and Negative Effects of Railway Expansion, 1870–1890

	(1) Changes in Log Total Population	(2) Changes in Log Farm Land Values	(3) Changes in Log Value of Farm Output	(4) Changes in Log Manufacturing Labor	(5) Changes in Log Manufacturing Value Added	(6) Changes in Log Urban Population	(7) Changes in Log Urban Population (NE, MW)	(8) Changes in Log City Size	(9) Changes in Log City Size (NE, MW)
Panel A: Allow Market Size to Change between 1870 and 1890									
Changes in log market access	0.610^{***}	0.811^{***}	0.680^{***}	0.380^{***}	0.397^{***}	0.766^{***}	0.635^{**}	1.063^{*}	-0.216
	(0.0927)	(0.116)	(0.118)	(0.109)	(0.120)	(0.217)	(0.253)	(0.572)	(0.555)
Changes in competition	-0.909^{***}	-0.970^{***}	-0.817^{***}	-0.905^{***}	-0.823	-1.494^{**}	-8.091^{***}	-2.437^{*}	-16.18^{***}
	(0.176)	(0.176)	(0.195)	(0.325)	(0.589)	(0.652)	(2.357)	(1.202)	(5.374)
% Market Access Effects Offset by Competition Effects	55.59%	53.66%	53.91%	74.56%	64.78%	121.67%	115.59%	88.08%	NA
Panel B: Fix Market Size at the 1870 Level									
Changes in log market access	0.597^{***}	0.839^{***}	0.688^{***}	0.376^{***}	0.389^{***}	0.619^{***}	0.595^{**}	0.693	-0.271
	(0.102)	(0.129)	(0.140)	(0.117)	(0.129)	(0.201)	(0.270)	(0.581)	(0.654)
Changes in competition	-0.711^{***}	-0.879^{***}	-0.698^{***}	-0.780^{***}	-0.684	-0.930^{*}	-7.268^{**}	-1.495	-11.61^{*}
	(0.204)	(0.184)	(0.240)	(0.278)	(0.542)	(0.477)	(2.840)	(1.145)	(6.180)
% Market Access Effects Offset by Competition Effects	55.79%	59.14%	57.29%	83.17%	70.49%	116.11%	146.47%	108.95%	NA
Panel C: Model-dependent Market Access in Donaldson and Hornbeck (2016), Locations Equally Compete with Each Other									
Changes in log market access	0.676^{***}	0.863^{***}	0.733^{***}	0.406^{***}	0.428^{***}	0.692^{***}	0.643^{**}	0.908	-0.213
	(0.0986)	(0.127)	(0.133)	(0.114)	(0.123)	(0.227)	(0.246)	(0.592)	(0.551)
Changes in competition	-0.505^{***}	-0.430^{***}	-0.368^{***}	-0.634^{**}	-0.541	-0.851	-8.148^{***}	-1.443^{*}	-16.18^{***}
	(0.133)	(0.0989)	(0.120)	(0.269)	(0.523)	(0.578)	(2.355)	(0.803)	(5.372)
Value of θ	7.35	8.22	8.22	6.10	6.10	7.35	6.10	7.35	6.10
Competition Is Measured in	All Sectors	Agriculture	Agriculture	Manufacturing	Manufacturing	All Sectors	Manufacturing	All Sectors	Manufacturing
N	2327	2327	2327	1775	1775	456	350	162	130

Notes: Market size is measured by population. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

significant, except for manufacturing value added in both panels and city size of all cities in Panel B.

(DH16) also use a specification of market access $MA_i = \sum_n \kappa_{ni}^{-\theta} MA_n^{\frac{-(1+\theta)}{\theta}} L_n$, which considers competition between locations but also assumes that there is no productivity difference in each sector across locations and that all locations equally compete with each other. I put the changes in log market access based on this specification and changes in competition defined in my paper into the same regression and report the results in Panel C of Table 1.1. It shows that competition effects become smaller and implies that competition effects are partially absorbed into market-access effects. However, the effects are still substantial and significant. Ignoring the different sector-specific productivity across locations, the complex specification of market access in (DH16) does not fully consider competition between locations.

Since $\Delta \ln MA$ and $\Delta Competition$ are positively correlated in all specifications, I calculate the percentage of positive market-access effects that were offset by negative effects of competition using the following equation:

$$\%\text{offset} = \frac{d\Delta Competition}{d\Delta \ln(MA)} \frac{\eta_2}{\eta_1}. \tag{1.27}$$

The results are also in Table 1.1. For all measures of economic outcomes, the percentage of positive market-access effects that were offset by negative effects is higher than 50% on average. The percentage is even higher in manufacturing, urban, and city-related economic outcomes. In the sample of urbanized counties and cities, positive market-access effects were fully offset by negative effects of competition.

One interesting question is whether a county benefited or suffered from 1870–1890 railway expansion. The question can be answered by combining the effects of market access and competition $\eta_1 \Delta \ln(MA_i) + \eta_2 \Delta Competition_i$ (in Equation 1.26). If the combined effects are positive, then I can conclude that the county benefited from the railway expansion; otherwise, it suffered. I plot the combined effects on the map using Panel B, since it shows a pure effect caused by reduced trade cost instead of by population redistribution and growth.

Figure 1.1 depicts railway expansion winners and losers in terms of population growth. In the upper map, all counties are classified to seven quantiles. Darker colors indicate higher combined effects. In the lower map, darker colors indicate positive combined effects while lighter colors indicate negative combined effects.

In general, the new transportation network benefited counties in northern Michigan, northern Wisconsin, northern Minnesota, northern Montana, southern California, and coastal Oregon, as well as most counties in Utah and Washington, counties in the southern Appalachian Mountains, and almost all counties in Texas and Maine. In general, other regions suffered.

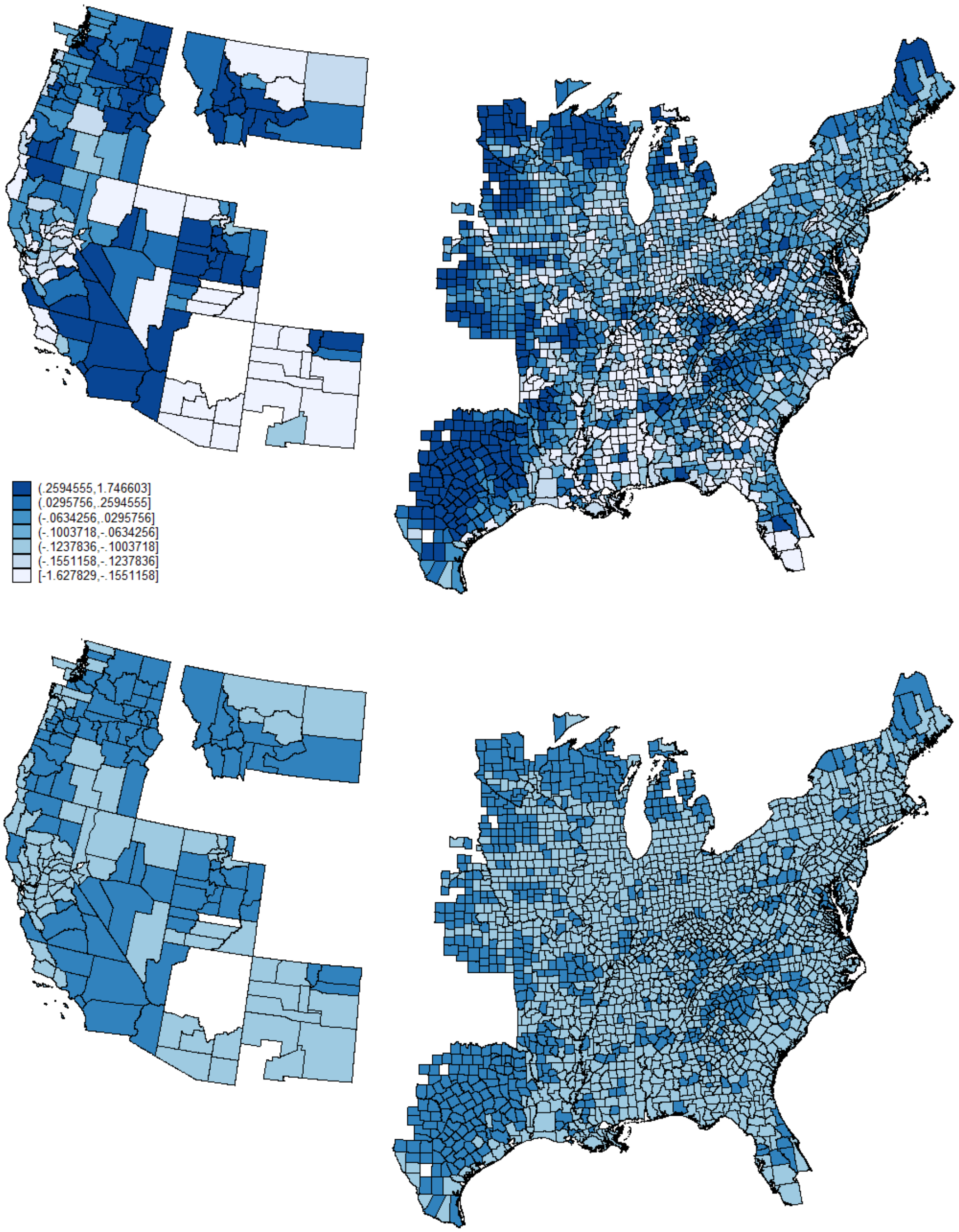
It is interesting to compare the maps of winners and losers with the maps of changes in log market access (Figure 1.2). The two maps look similar, but interesting differences emerge. For example, the changes in log market access in most New Mexico counties was very high while the combined effects are negative. It's easy to understand, since competition also increased a lot in New Mexico then. In fact, if we look at the Figure 1.3, New Mexico counties performed poorly. Meanwhile, in Kansas and Nebraska, if we look only at the changes in log market access, it was not as high as in New Mexico counties. However, the changes in competition was also much lower than in New Mexico counties (Figure 1.2). Unlike in New Mexico, the combined effects on the population in most of Kansas and Nebraska were positive (Figure 1.1). And if we look at the changes in log population (Figure 1.3), many counties in Kansas and Nebraska experienced a huge population gain during this period.

1.4.2 Role of Comparative Advantage

I use the following reduced-form model to test Implication 5:

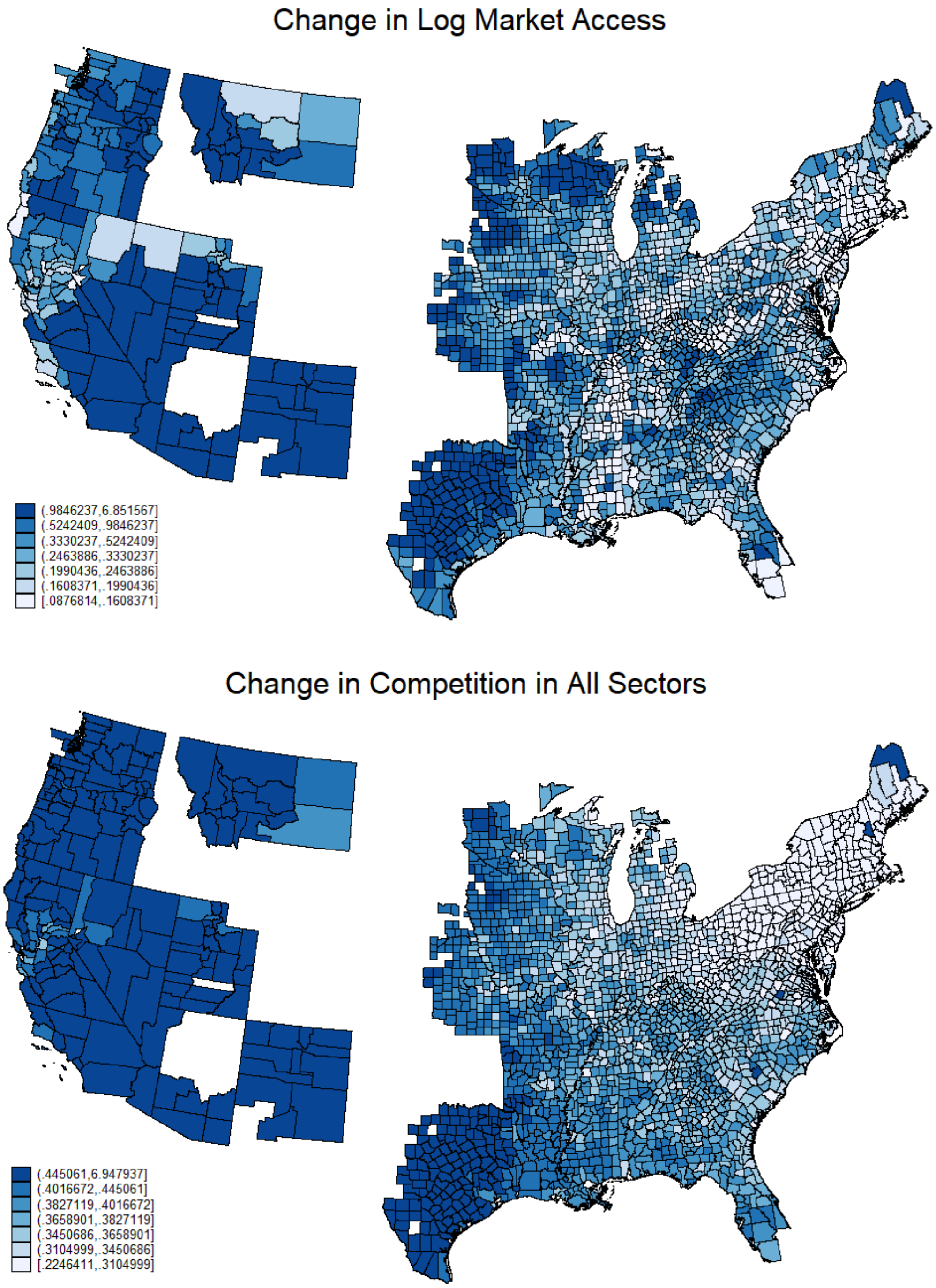
$$\begin{aligned}\Delta \ln EconomicOutcome_i^J = \delta + \eta_1 \Delta \ln(MA_i) + \eta_2 \Delta Competition_i^J + \eta_3 \Delta Competition_i \\ + \delta_s + f(x_i, y_i) + \varepsilon_i, \end{aligned} \tag{1.28}$$

Figure 1.1: Winners and Losers of Railway Expansion in Terms of Population, 1870–1890



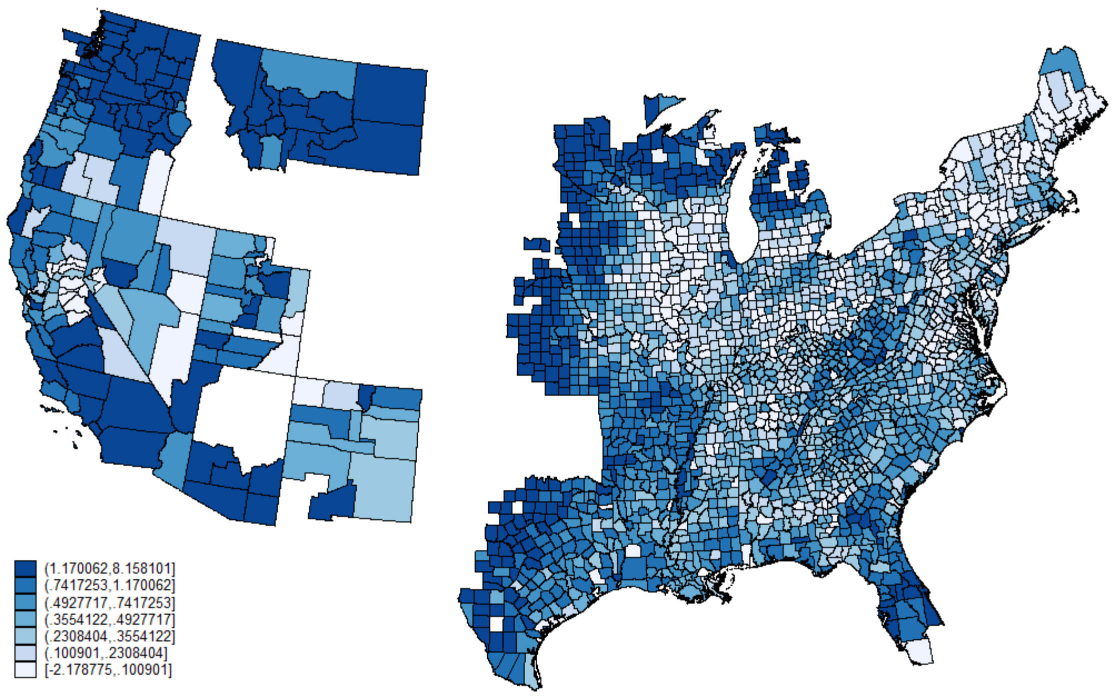
Notes: In the upper map, all counties are classified to seven quantiles. Darker colors indicate higher combined effects. In the lower map, darker colors indicate positive combined effects while lighter colors indicate negative combined effects. $\theta = 7.35$. Market size is measured by population and fixed at the 1870 level.

Figure 1.2: Changes in Log Market Access and Competition, 1870–1890



Notes: The upper map indicates changes in log market access from 1870 to 1890. The lower map indicates changes in competition from 1870 to 1890. All counties are classified to seven quantiles. Darker colors indicate higher effects. $\theta = 7.35$. Market size is measured by population and fixed at the 1870 level.

Figure 1.3: Changes in Log Population, 1870–1890



Notes: All counties are classified to seven quantiles. Darker colors indicate higher population growth.

where $J \in$ {Agriculture, Manufacturing}. According to Implication 5, η_2 is negative while η_3 is positive. The results are reported in Table 1.2, where columns (1) and (2) focus on the agriculture sector, (3) and (4) on the manufacturing sector.

The results for agriculture-related economic outcomes are exactly the same as the theory predicts. It means if trade shocks favor agriculture—that is, competition increases in agriculture less than it does in other sectors—the shocks will positively affect economic outcome in agriculture.

In contrast, the results for manufacturing-related economic outcomes are not the same as the theory prediction. The results can still be explained, since the assumption that there is no linkage between sectors is arbitrary and may create bias. The bias is much more severe for manufacturing than for agriculture, since agriculture provides a substantial portion of manufacturing input, but not vice versa. As a result, the development of manufacturing is not only negatively affected by competition from agricultural development but also promoted by cost reduction resulting from agricultural development. If the positive effects of agriculture

Table 1.2: Role of Comparative Advantage

	(1)	(2)	(3)	(4)
	Changes in Log Farm Land Values	Changes in Log Value of Farm Output	Changes in Log Manufacturing Labor	Changes in Log Manufacturing Value Added
Panel A: Allow Market Size to Change between 1870 and 1890				
Changes in competition in agriculture/manufacturing	-2.502***	-2.228***	1.989	0.309
	(0.574)	(0.578)	(1.452)	(1.418)
Changes in competition in all sectors	2.216***	1.786**	-2.136***	-1.553**
	(0.718)	(0.665)	(0.702)	(0.719)
Panel B: Fix Market Size at the 1870 Level				
Changes in competition in agriculture/manufacturing	-2.221***	-2.073***	0.918	-1.095
	(0.456)	(0.533)	(1.685)	(1.542)
Changes in competition in all sectors	2.044***	1.910***	-1.215	-0.450
	(0.550)	(0.536)	(0.848)	(0.779)
Value of θ in Competition in Agriculture/Manufacturing	8.22	8.22	6.10	6.10
Value of θ in Competition in All Sectors	7.35	7.35	7.35	7.35
N	2327	2327	2327	2327

Notes: Changes in log market access with θ=8.22 and θ=7.35 are controlled in columns (1) and (2). Changes in log market access with θ=6.10 and θ=7.35 are controlled in columns (3) and (4). State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

dominate, it is easy to understand that changes in competition in all sectors negatively affects manufacturing.

1.4.3 Link between Comparative Advantage in Agriculture and Economic Growth

(Mat92) predicts that comparative advantage in agriculture and economic growth are negatively linked in a small, open economy. I reexamine the prediction under the framework of competition to see whether trade integration that relatively favors agriculture will generate positive or negative economic growth. I undertake the following empirical strategy:

$$\begin{aligned}\Delta \ln EconomicOutcome_i = \delta + \eta_1 \Delta \ln(MA_i) + \eta_2 \Delta Competition_i^{Agriculture} + \eta_3 \Delta Competition_i \\ + \delta_s + f(x_i, y_i) + \varepsilon_i, \qquad (1.29)\end{aligned}$$

where $EconomicOutcome_i$ refers to the economic outcome of the whole economy. I discuss total population and the value of agricultural output plus value added in manufacturing here. If (Mat92)'s prediction is true, I will expect that η_2 is positive.

In Table 1.3, the coefficient of changes in competition in agriculture is significantly negative, contradicting (Mat92)'s prediction. The results imply a positive link between comparative advantage in agriculture and economic growth. (Mat92) assumes workers are immobile across locations. Under this assumption, with comparative advantage in agriculture, worker are locked in agriculture. The productive agriculture squeezes out manufacturing. However, the assumption is too strong. If workers are mobile across locations, low-productivity agriculture cannot provide cheap inputs for manufacturing. In the early stages of development, production in many manufacturing sectors relied highly on materials produced by agriculture. Manufacturing suffered from low-productivity agriculture. As a result, workers moved to other locations.

Table 1.3: Effects of Agricultural Competition on the Whole Economy, Holding Constant the Changes in Competition in All Sectors

	(1)	(2)
	Changes in Log Total Population	Changes in Log Value of Agricultural Output Plus Manufacturing Value Added
Panel A: Allow Market Size to Change between 1870 and 1890		
Changes in competition in agriculture	-1.889***	-2.032***
	(0.406)	(0.562)
Changes in competition in all Sectors	1.165**	1.102*
	(0.506)	(0.633)
Panel B: Fix Market Size at the 1870 Level		
Changes in competition in agriculture	-1.603***	-1.844***
	(0.328)	(0.535)
Changes in competition in all sectors	1.244***	1.388**
	(0.358)	(0.580)
Value of θ in Competition in Agriculture	8.22	8.22
Value of θ in Competition in All Sectors	7.35	7.35
N	2327	2327

Notes: Changes in log market access with θ=8.22 and θ=7.35 are controlled. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.5 Problem of Endogeneity

In the previous section, I show an interesting picture of the effects of negative competition on economic outcomes. Productivity variation caused by distribution of natural endowments and reduced trade cost between competing locations m and markets n are two main sources of geographical variation of changes in competition.

Two problems make their measurement endogenous. First, in the baseline estimation, I set the value of T_i^j as 0 for subsectors in agriculture if subsector j goods were not produced in location i in 1870. However, the crop choice is endogenous. In manufacturing subsectors, I use value added per worker to measure T_i^j, which is by default 0 if subsector j goods were not produced in location i in 1870. Second, trade cost reduction between competing locations and markets κ_{nm} is endogenous, since most of the new railways were built by private companies whose major concern was profit.

To solve the problem of endogeneity and identify the causal effects of competition, I undertake the following strategies: First, I use an IV strategy to eliminate the endogeneity of T_i^j. Second, I inspect whether a spurious relationship exists between changes in 1870–1890 competition and earlier economic growth. Third, I control railway variables. Since the main source of endogeneity of κ_{nm} is rail-route selection, κ_{nm} is exogenous conditional on route selection. I also control other variables that were likely to affect route selection. These variables include pre-1870 growth and productivity in 1870.

1.5.1 IV Strategies for T_i^j

In the agricultural sector, a concern is that crop choice is not exogenous. In my IV regression, I use the original value of land suitability without considering whether subsector j goods were produced in location i in 1870. Then the endogeneity of subsector choice is instrumented by low land suitability.

In manufacturing subsectors, I use value added per worker to measure productivity, but the measure is not exogenous. To solve the problem of endogeneity, I instrument value added per worker by input accessibility, calculated by the following equation:

$$T_{i,IV}^{j} = \sum_{k \in Input_j} \gamma^k \sum_{p \neq i} \kappa_{ip,1870}^{-\theta} T_p^k, j \in \text{Manufacturing} \tag{1.30}$$

which is the weighted sum of accessibility of all inputs. k indicates subsectors that provide input for subsector j. γ^k is the ratio of value of input k in the total value of all inputs of subsector j.[7] p is the location where producers produce input from subsector k. θ is the trade elasticity. For the weighted average of all subsectors, I use the 7.35 in (CP15). $\sum_{p \neq i} \kappa_{ip}^{-\theta} T_p^k$ measures how cheaply a producer in location i sector j can access inputs in subsector k produced in other locations. The intuition is that if producers of goods in sector j location i have better access to high-productivity input-providing locations, the value added per worker will be high. For inputs in agriculture and mining, the input accessibility highly depends on natural resources. For inputs in manufacturing sectors, input accessibility mainly depends on productivity in locations other than location i. In general, compared to T_i^j, the values of $T_{i,IV}^j$ mainly rely on distribution of natural resources and characteristics of other locations.

Table 1.4, Table 1.5, and Table 1.6 compare the results of baseline OLS (Panel A) and IV (Panel B) to the exercises in Section 4.1 to 4.3. Results remain the same under IV regressions.

1.5.2 Spurious Effects on Previous Economic Growth

If counties or cities that faced fiercer competition between 1870 and 1890 tended to grow more slowly before the new transportation network was built (i.e., slow growth caused fierce competition but not vice versa), then the estimations in Section 4 would be biased. A typical strategy to test the problem of endogeneity is the visual inspection of the pretreatment trends.

Table 1.7 shows the correlation between changes in competition from 1870 to 1890 and previous economic growth. Panel A shows the baseline results for comparison. Panel B shows the results for economic growth from 1850 to 1870. The correlation is positive for

[7] I calculate the input-output linkage based on (Bat04).

Table 1.4: Positive and Negative Effects of Railway Expansion, 1870–1890, OLS and IV Strategy

	(1) Changes in Log Total Population	(2) Changes in Log Farm Land Values	(3) Changes in Log Value of Farm Output	(4) Changes in Log Manufacturing Labor	(5) Changes in Log Manufacturing Value Added	(6) Changes in Log Urban Population	(7) Changes in Log Urban Population (NE, MW)	(8) Changes in Log City Size	(9) Changes in Log City Size (NE, MW)
Panel A: Baseline OLS Regressions									
Changes in log market access	0.610^{***}	0.811^{***}	0.680^{***}	0.380^{***}	0.397^{***}	0.766^{***}	0.635^{**}	1.063^{*}	-0.216
	(0.0927)	(0.116)	(0.118)	(0.109)	(0.120)	(0.217)	(0.253)	(0.572)	(0.555)
Changes in competition	-0.909^{***}	-0.970^{***}	-0.817^{***}	-0.905^{***}	-0.823	-1.494^{**}	-8.091^{***}	-2.437^{*}	-16.18^{***}
	(0.176)	(0.176)	(0.195)	(0.325)	(0.589)	(0.652)	(2.357)	(1.202)	(5.374)
Panel B: IV Regressions									
Changes in log market access	0.609^{***}	0.816^{***}	0.684^{***}	0.475^{***}	0.511^{***}	0.737^{**}	1.001^{***}	0.940^{*}	0.000466
	(0.0907)	(0.114)	(0.116)	(0.108)	(0.124)	(0.295)	(0.247)	(0.534)	(0.476)
Changes in competition	-0.920^{***}	-0.989^{***}	-0.832^{***}	-1.519^{***}	-1.554^{**}	-1.403^{*}	-17.36^{***}	-2.168^{*}	-27.60^{***}
	(0.180)	(0.179)	(0.194)	(0.413)	(0.641)	(0.835)	(3.711)	(1.127)	(6.052)
Value of θ	7.35	8.22	8.22	6.10	6.10	7.35	6.10	7.35	6.10
Competition Is Measured in	All Sectors	Agriculture	Agriculture	Manufacturing	Manufacturing	All Sectors	Manufacturing	All Sectors	Manufacturing
N	2327	2327	2327	1775	1775	456	350	162	130

Notes: Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: Role of Comparative Advantage, OLS and IV Strategy

	(1)	(2)	(3)	(4)
	Changes in Log Farm Land Values	Changes in Log Value of Farm Output	Changes in Log Manufacturing Labor	Changes in Log Manufacturing Value Added
Panel A: Baseline OLS Regressions				
Changes in competition in agriculture/manufacturing	-2.502***	-2.228***	1.989	0.309
	(0.574)	(0.578)	(1.452)	(1.418)
Changes in competition in all sectors	2.216***	1.786**	-2.136***	-1.553**
	(0.718)	(0.665)	(0.702)	(0.719)
Panel B: IV Regressions				
Changes in competition in agriculture/manufacturing	-2.264***	-2.128***	0.221	18.21
	(0.619)	(0.578)	(26.84)	(61.32)
Changes in competition in all sectors	1.742**	1.570**	-1.342	-9.847
	(0.850)	(0.785)	(12.29)	(28.00)
Value of θ in Competition in Agriculture/Manufacturing	8.22	8.22	6.10	6.10
Value of θ in Competition in All Sectors	7.35	7.35	7.35	7.35
N	2327	2327	2327	2327

Notes: Changes in log market access with θ=8.22 and θ=7.35 are controlled in columns (1) and (2). Changes in log market access with θ=6.10 and θ=7.35 are controlled in columns (3) and (4). State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.6: Effects of Agricultural Competition on the Whole Economy, Holding Constant the Changes in Competition in All Sectors, OLS and IV Strategy

	(1)	(2)
	Changes in Log Total Population	Changes in Log Value of Agricultural Output Plus Manufacturing Value Added
Panel A: Baseline OLS Regressions		
Changes in competition in agriculture	-1.889***	-2.032***
	(0.406)	(0.562)
Changes in competition in all sectors	1.165**	1.102*
	(0.506)	(0.633)
Panel B: IV Regressions		
Changes in competition in agriculture	-1.730***	-1.616***
	(0.501)	(0.573)
Changes in competition in all sectors	0.871	0.359
	(0.727)	(0.792)
Value of θ in Competition in Agriculture	8.22	8.22
Value of θ in Competition in All Sectors	7.35	7.35
N	2327	2327

Notes: Changes in log market access with θ=8.22 and θ=7.35 are controlled. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: Correlation between Changes in Competition, 1870–1890, and Previous Economic Growth

	(1) Changes in Log Total Population	(2) Changes in Log Farm Land Values	(3) Changes in Log Value of Farm Output	(4) Changes in Log Manufacturing Labor	(5) Changes in Log Manufacturing Value Added	(6) Changes in Log Urban Population	(7) Changes in Log Urban Population (NE, MW)	(8) Changes in Log City Size	(9) Changes in Log City Size (NE, MW)
Panel A: Baseline Effects, 1870–1890									
Changes in log market access	0.610^{***}	0.811^{***}	0.680^{***}	0.380^{***}	0.397^{***}	0.766^{***}	0.635^{**}	1.063^{*}	-0.216
	(0.0927)	(0.116)	(0.118)	(0.109)	(0.120)	(0.217)	(0.253)	(0.572)	(0.555)
Changes in competition	-0.909^{***}	-0.970^{***}	-0.817^{***}	-0.905^{***}	-0.823	-1.494^{**}	-8.091^{***}	-2.437^{*}	-16.18^{***}
	(0.176)	(0.176)	(0.195)	(0.325)	(0.589)	(0.652)	(2.357)	(1.202)	(5.374)
Panel B: Pretreatment Effects, 1850–1870									
Changes in log market access	-0.207	-0.468^{*}	-0.337	-0.505^{**}	-0.454	1.462	0.243	1.803	1.972
	(0.179)	(0.257)	(0.244)	(0.187)	(0.296)	(1.316)	(2.725)	(1.572)	(1.549)
Changes in competition	0.630^{***}	0.899^{***}	0.713^{**}	0.441	0.519	-2.882	86.79^{***}	-5.757	4.697
	(0.201)	(0.310)	(0.329)	(0.291)	(0.666)	(5.168)	(19.03)	(3.840)	(27.45)
Panel C: Pretreatment Effects, 1850–1860 (Before the Civil War)									
Changes in log market access	-0.366^{**}	-0.788^{***}	-0.728^{***}	-0.462^{**}	-0.915^{*}	-0.800	-7.580^{***}	1.007	2.051
	(0.171)	(0.264)	(0.267)	(0.221)	(0.516)	(2.140)	(1.619)	(1.445)	(1.722)
Changes in competition	0.719^{***}	1.351^{***}	1.168^{***}	0.414	1.139	0.420	-19.05	-3.422	10.56
	(0.237)	(0.362)	(0.329)	(0.414)	(1.052)	(4.289)	(23.62)	(3.675)	(21.91)
Panel D: Pretreatment Effects, 1860–1870 (During the Civil War)									
Changes in log market access	0.159	0.321	0.392^{**}	-0.0423	0.384	2.262	7.823^{***}	0.796	-0.0795
	(0.126)	(0.223)	(0.183)	(0.0914)	(0.347)	(1.961)	(1.452)	(0.544)	(0.671)
Changes in competition	-0.0886	-0.452	-0.456^{*}	0.0271	-0.566	-3.301	105.8^{***}	-2.335^{**}	-5.860
	(0.216)	(0.307)	(0.268)	(0.307)	(1.240)	(6.430)	(26.26)	(1.136)	(7.330)
Value of θ	7.35	8.22	8.22	6.10	6.10	7.35	6.10	7.35	6.10
Competition Is Measured in	All Sectors	Agriculture	Agriculture	Manufacturing	Manufacturing	All Sectors	Manufacturing	All Sectors	Manufacturing
N	2327	2327	2327	1775	1775	456	350	162	130

Notes: Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

most economic outcomes. And for those negative values (column 6 and column 8, effects of competition in all sectors on urban population and city size), the pretreatment effects are statistically insignificant.

One argument is that economic development was interrupted by the Civil War. And the 1850–1870 pretreatment period may not reflect economic development in a normal environment. Panel C of Table 1.7 shows the effects of competition from 1850 to 1860. Conditional on all other variables, the changes in competition from 1870 to 1890 is positively correlated with most 1850–1860 economic outcomes. The correlation is negative in columns (7) and (8), but it is never statistically significant.

Did counties that faced larger increases in competition from 1870 to 1890 suffer the most during the Civil War? Panel D of Table 1.7 shows the results. Competition from 1860 to 1870 is negatively correlated with economic growth in some terms.[8] For most cases, the pretreatment effects are statistically insignificant. The pretreatment effects of competition on the value of farm output are significantly negative. The effects (-0.456) between 1860 and 1870 are still much smaller than the effects (-0.817) between 1870 and 1890. The pretreatment effects of competition in all sectors on city size for all cities with a population larger than 10,000 in 1870 are negative, substantial, and significant. The effects are only slightly smaller than the effects from 1870 to 1890 (-2.335 vs. -2.437). In general, for most cases, I could not observe the spurious pretreatment effects of changes in competition between 1870 and 1890.

Table 1.8 shows the pretreatment effects in terms of the role of comparative advantage. In contrast to the baseline results, the pretreatment trends were different in 1850–1870 and 1850–1860. In 1860–1870, the pretreatment trend has the same sign as the 1870–1890 trend. However, the values of pretreatment effects from 1860 to 1870 were much smaller.

Table 1.9 shows the relationship between changes in competition in agriculture from 1870 to 1890 and economic growth in earlier periods conditional on changes in market access and

[8]Total population, agricultural land and output value, value added in manufacturing, urban population in the sample of all urbanized counties, city size of all cities with a population larger than 10,000 in 1870, and city size of Northeastern and Midwestern cities with a population larger than 10,000 in 1870.

Table 1.8: Role of Comparative Advantage, Pretreatment Effects

	(1)	(2)
	Changes in Log Farm Land Values	Changes in Log Value of Farm Output
Panel A: Baseline Effects, 1870–1890		
Changes in competition in agriculture	-2.502***	-2.228***
	(0.574)	(0.578)
Changes in competition in all sectors	2.216***	1.786**
	(0.718)	(0.665)
Panel B: Pretreatment Effects, 1850–1870		
Changes in competition in agriculture	2.219	1.606
	(1.436)	(1.118)
Changes in competition in all sectors	-2.064	-1.429
	(1.935)	(1.373)
Panel C: Pretreatment Effects, 1850–1860 (before the Civil War)		
Changes in competition in agriculture	3.769***	2.787**
	(1.354)	(1.161)
Changes in competition in all sectors	-3.611**	-2.382*
	(1.601)	(1.361)
Panel D: Pretreatment Effects, 1860–1870 (during the Civil War)		
Changes in competition in agriculture	-1.550*	-1.180
	(0.872)	(0.711)
Changes in competition in all sectors	1.546	0.953
	(1.071)	(0.911)
Value of θ in Competition in Agriculture	8.22	8.22
Value of θ in Competition in All Sectors	7.35	7.35
N	2327	2327

Notes: Changes in log market access with θ=8.22 and θ=7.35 are controlled. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.9: Effects of Agricultural Competition on the Whole Economy, Holding Constant the Changes in Competition in All Sectors, Pretreatment Effects

	(1)	(2)
	Changes in Log Total Population	Changes in Log Value of Agricultural Output Plus Manufacturing Value Added
Panel A: Baseline Effects, 1870–1890		
Changes in competition in agriculture	-1.889***	-2.032***
	(0.406)	(0.562)
Changes in competition in all sectors	1.165**	1.102*
	(0.506)	(0.633)
Panel B: Pretreatment Effects, 1850–1870		
Changes in competition in agriculture	1.665*	2.113*
	(0.840)	(1.102)
Changes in competition in all sectors	-1.772	-2.440*
	(1.129)	(1.291)
Panel C: Pretreatment Effects, 1850–1860 (before the Civil War)		
Changes in competition in agriculture	2.117***	3.081**
	(0.752)	(1.307)
Changes in competition in all sectors	-2.077**	-2.932*
	(0.857)	(1.471)
Panel D: Pretreatment Effects, 1860–1870 (during the Civil War)		
Changes in competition in agriculture	-0.451	-1.100
	(0.459)	(0.700)
Changes in competition in all sectors	0.305	0.611
	(0.617)	(0.917)
Value of θ in Competition in Agriculture	8.22	8.22
Value of θ in Competition in All Sectors	7.35	7.35
N	2327	2327

Notes: Changes in log market access with θ=8.22 and θ=7.35 are controlled. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

changes in competition in all sectors. In contrast to the baseline results, the 1850–1870 and 1850–1860 pretreatment trends were exactly opposite. The pretreatment trend 1860–1870 has the same sign as the period 1870–1890. However, the spurious effects were neither substantial nor statistically significant.

1.5.3 Control Rail-route Selection

Most railways were built by private companies to pursue economic profit. They were constructed to connect locations with economic prosperity (high previous economic growth) to locations with economic potential (low previous economic growth). Thus the construction of railway networks is endogenous.

To mitigate this bias, following the strategy in (DH16), I control for whether a county is directly linked by any new railway, whether a county is near any railway (within a 10-mile, 20-mile, 30-mile, and 40-mile buffer of the county), a county's mileage of new railways, the mileage of new railways nearby (within a 10-mile, 20-mile, 30-mile, and 40-mile buffer of the county), and quadratic and cubic terms of all mileage variables. The major risk of endogeneity in changes in competition caused by rail-route selection is eliminated by controlling all railway variables.

Another strategy is to control variables that affected rail-route selection. First, I control previous economic growth and the quadratic terms in the 1850–1860 and 1860–1870 periods. Second, I control the initial productivity of the most productive subsectors,[9] since profit-driven companies were more likely to link locations with productive sectors to markets.

Table 1.10 shows the results for effects of competition on economic growth with additional control variables related to rail-route selection. In general, negative effects are robust with additional control variables but they become slightly smaller than the ones in the baseline regression. In small subsamples of urbanized counties and cities, the significance level is sensitive to additional control variables, since an over-fitting problem might be serious in

[9] To make productivity in different sectors comparable, I standardize the productivity of each sector to be within 0 to 1.

Table 1.10: Negative Effects of Competition, Railway Expansion 1870–1890, Control Rail-Route Selection

	(1) Changes in Log Total Population	(2) Changes in Log Farm Land Values	(3) Changes in Log Value of Farm Output	(4) Changes in Log Manufacturing Labor	(5) Changes in Log Manufacturing Value Added	(6) Changes in Log Urban Population	(7) Changes in Log Urban Population (NE, MW)	(8) Changes in Log City Size	(9) Changes in Log City Size (NE, MW)
Panel A: Baseline, No Additional Controls									
Changes in competition	-0.909^{***}	-0.970^{***}	-0.817^{***}	-0.905^{***}	-0.823	-1.494^{**}	-8.091^{***}	-2.437^{*}	-16.18^{***}
	(0.176)	(0.176)	(0.195)	(0.325)	(0.589)	(0.652)	(2.357)	(1.202)	(5.374)
Panel B: Control Railway Variables									
Changes in competition	-0.746^{***}	-0.912^{***}	-0.721^{***}	-0.632^{**}	-0.502	-1.118	-7.956^{**}	-4.070	-12.27^{*}
	(0.167)	(0.232)	(0.185)	(0.274)	(0.476)	(0.926)	(3.550)	(2.431)	(6.792)
Panel C: Control Pretreatment Growth, 1850–1860 and 1860–1870, and Quadratic Terms									
Changes in competition	-0.811^{***}	-0.849^{***}	-0.718^{***}	-0.851^{**}	-0.791	-1.396^{*}	-11.50^{***}	-1.298	-14.26^{***}
	(0.168)	(0.156)	(0.172)	(0.356)	(0.614)	(0.743)	(2.739)	(0.933)	(4.408)
Panel D: Control Initial Maximum Productivity among All Subsectors									
Changes in competition	-0.954^{***}	-1.016^{***}	-0.844^{***}	-0.922^{***}	-0.865	-1.570^{*}	-8.193^{***}	-2.723^{**}	-16.04^{***}
	(0.182)	(0.178)	(0.197)	(0.335)	(0.583)	(0.803)	(2.310)	(1.238)	(5.020)
Panel E: Control All above Variables									
Changes in competition	-0.645^{***}	-0.751^{***}	-0.595^{***}	-0.572^{*}	-0.510	-1.466	-10.84^{***}	-3.041	-10.10^{*}
	(0.150)	(0.168)	(0.151)	(0.311)	(0.502)	(1.023)	(3.398)	(2.105)	(5.542)
Panel F: IV Regressions and Control All Above Variables									
Changes in competition	-0.636^{***}	-0.767^{***}	-0.610^{***}	-0.988^{**}	-1.058^{*}	-1.866^{*}	-21.81^{***}	-3.103^{*}	-9.450
	(0.153)	(0.166)	(0.148)	(0.400)	(0.593)	(1.036)	(4.386)	(1.869)	(5.811)
Value of θ	7.35	8.22	8.22	6.10	6.10	7.35	6.10	7.35	6.10
Competition Is Measured in	All Sectors	Agriculture	Agriculture	Manufacturing	Manufacturing	All Sectors	Manufacturing	All Sectors	Manufacturing
N	2327	2327	2327	1775	1775	456	350	162	130

Notes: Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses. Railway variables include whether a county is directly linked by any new railway, whether a county is near any railway (within a 10-mile, 20-mile, 30-mile, and 40-mile buffer of the county), a county's mileage of new railways, the mileage of new railways nearby (within a 10-mile, 20-mile, 30-mile, and 40-mile buffer of the county), and quadratic and cubic terms of all mileage variables.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

small samples. Besides, when I combine the IV strategies with additional control variables, the results are still robust.

Table 1.11 shows the results for the role of comparative advantage. Table 1.12 shows the link between comparative advantage in agriculture and economic growth. In both tables, with additional control variables, comparative advantage has smaller effects but signs are still consistent with the model prediction.

In summary, my results are robust to the strategies that eliminate endogeneity of rail-route selection.

Table 1.11: Role of Comparative Advantage, Control Rail-Route Selection

	(1)	(2)
	Changes in Log Farm Land Values	Changes in Log Value of Farm Output
Panel A: Baseline, No Additional Controls		
Changes in competition in agriculture	-2.502***	-2.228***
	(0.574)	(0.578)
Changes in competition in all sectors	2.216***	1.786**
	(0.718)	(0.665)
Panel B: Control Railway Variables		
Changes in competition in agriculture	-2.404***	-2.082***
	(0.556)	(0.511)
Changes in competition in all sectors	2.120***	1.732***
	(0.645)	(0.606)
Panel C: Control Pretreatment Growth, 1850–1860 and 1860–1870, and Quadratic Terms		
Changes in competition in agriculture	-2.224***	-2.010***
	(0.538)	(0.542)
Changes in competition in all sectors	1.895**	1.576**
	(0.715)	(0.684)
Panel D: Control Initial Maximum Productivity among All Subsectors		
Changes in competition in agriculture	-2.404***	-2.367***
	(0.556)	(0.499)
Changes in competition in all sectors	2.120***	1.914***
	(0.645)	(0.617)
Panel E: Control All Above Variables		
Changes in competition in agriculture	-1.968***	-1.760***
	(0.448)	(0.442)
Changes in competition in all sectors	1.692***	1.448**
	(0.615)	(0.587)
Panel F: IV Regressions and Control All Above Variables		
Changes in competition in agriculture	-1.989***	-1.809***
	(0.495)	(0.505)
Changes in competition in all sectors	1.680**	1.487*
	(0.707)	(0.771)
Value of θ in Competition in Agriculture	8.22	8.22
Value of θ in Competition in All Sectors	7.35	7.35
N	2327	2327

Notes: Changes in log market access with θ=8.22 and θ=7.35 are controlled. Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.12: Effects of Agricultural Competition on the Whole Economy, Holding Constant the Changes in Competition in All Sectors, Control Rail-Route Selection

	(1)	(2)
	Changes in Log Total Population	Changes in Log Value of Agricultural Output Plus Manufacturing Value Added
Panel A: Baseline, No Additional Controls		
Changes in competition in agriculture	-1.889***	-2.032***
	(0.406)	(0.562)
Changes in competition in all sectors	1.165**	1.102*
	(0.506)	(0.633)
Panel B: Control Railway Variables		
Changes in competition in agriculture	-1.787***	-1.982***
	(0.369)	(0.515)
Changes in competition in all sectors	1.191**	1.167*
	(0.459)	(0.592)
Panel C: Control Pretreatment Growth, 1850–1860 and 1860–1870, and Quadratic Terms		
Changes in competition in agriculture	-1.656***	-1.746***
	(0.365)	(0.504)
Changes in competition in all sectors	0.937*	0.775
	(0.471)	(0.614)
Panel D: Control Initial Maximum Productivity among All Subsectors		
Changes in competition in agriculture	-1.806***	-1.959***
	(0.348)	(0.522)
Changes in competition in all sectors	0.979**	0.939*
	(0.384)	(0.548)
Panel E: Control All Above Variables		
Changes in competition in agriculture	-1.463***	-1.605***
	(0.279)	(0.425)
Changes in competition in all sectors	0.850**	0.753
	(0.359)	(0.533)
Panel F: IV Regressions and Control All Above Variables		
Changes in competition in agriculture	-1.383***	-1.365***
	(0.408)	(0.490)
Changes in competition in all sectors	0.695	0.310
	(0.628)	(0.734)
Value of θ in Competition in Agriculture	8.22	8.22
Value of θ in Competition in All Sectors	7.35	7.35
N	2327	2327

Notes: Changes in log market access with θ=8.22 and θ=7.35 are controlled. Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.6 Competition and Regional Development

In this section, I show that changes in competition shed lights on several interesting stories of regional development. First, competition explains the growth gap between counties that were directly linked by new railways and counties that had no new rail track, but were adjacent to a county that was linked by new railways. Second, competition explains agglomeration effects, that urban population grew faster in counties with a large city, even though these counties experienced less growth in market access. Third, competition explains why, compared with Midwestern cities, Southern cities grew more slowly after the Civil War, even though their increase in market access was high.

The general strategy I employ in this section is based on (BK86). First, I use the following regression to show the growth gap:

$$\Delta \ln EconomicOutcome_i = \delta + \eta_1 D_i + \delta_s + f(x_i, y_i) + \varepsilon_i, \tag{1.31}$$

where D_i refers to the dummy variables that categorize locations into several groups. The coefficient η_1 refers to growth gap between groups.

Second, I use the following regression to see whether a difference exists between location groups in changes in log market access, and whether a difference exists between location groups in changes in competition conditional on changes in log market access:

$$\Delta \ln(MA_i) = \delta + \eta_1 D_i + \delta_s + f(x_i, y_i) + \varepsilon_i \tag{1.32}$$

$$\Delta Competition_i = \delta + \eta_1 D_i + \eta_2 \Delta \ln(MA_i) + \delta_s + f(x_i, y_i) + \varepsilon_i. \tag{1.33}$$

Third, I add changes in log market access into Equation 1.31 to see whether the absolute values of coefficients of dummy variables shrink. If so, changes in log market access can explain the growth gap, as follows:

$$\Delta \ln EconomicOutcome_i = \delta + \eta_1 D_i + \eta_2 \Delta \ln(MA_i) + \delta_s + f(x_i, y_i) + \varepsilon_i. \qquad (1.34)$$

Fourth, I add changes in competition into the regression to see whether the absolute values of coefficients of dummy variables further shrink (Equation 1.35). If so, conditional on changes in log market access, changes in competition is a mechanism to explain the growth gap.

$$\Delta \ln EconomicOutcome_i = \delta + \eta_1 D_i + \eta_2 \Delta \ln(MA_i) + \eta_3 \Delta Competition_i + \delta_s + f(x_i, y_i) + \varepsilon_i \qquad (1.35)$$

1.6.1 Counties With New Railway Tracks vs Counties Near New Railway Tracks

(CT00) find that highways positively affected counties on the routes but negatively affected adjacent counties. This finding opens up an interesting question: Why did neighboring counties experience worse economic performance despite enjoying a larger increase in market access?

One hypothesis is that neighboring counties had similar endowments. When counties were directly linked by new transportation routes, producers in their neighboring counties faced much fiercer competition, since competitors who produced similar goods were better connected to markets. As a result, the significant growth gap between counties directly linked by highways and adjacent counties was not only attributed to the difference in market access but also caused by the difference in competition.

To test the hypothesis, I investigate the performance of counties with new rail track, adjacent counties (baseline group in the regression), and counties far from the new railways. Unlike the story in (CT00), neighboring counties did not experience negative growth compared to counties far from new railways between 1870 and 1890, but I can still observe a big growth gap between counties with new rail tracks and their adjacent counties without new rail tracks (Panel A of Table 1.14). The gap between counties with new rail tracks and their adjacent counties is substantial for the total population, agriculture-related economic

Table 1.13: Changes in Log Market Access and Changes in Competition (Conditional on Changes in Log Market Access) by Location

	(1) Changes in Market Access	(2) Changes in Market Access	(3) Changes in Market Access	(4) Changes in Competition in All Sectors	(5) Changes in Competition in Agriculture	(6) Changes in Competition in Manufacturing
Panel A: With, near, without New Rail Tracks						
Directly linked by new railways	0.458***	0.514***	0.385***	-0.116***	-0.158***	-0.0591***
	(0.0434)	(0.0486)	(0.0444)	(0.0311)	(0.0352)	(0.0190)
Not adjacent to counties with new rail tracks	-0.00464	-0.00621	0.000160	0.00472	0.0207**	-0.00125
	(0.0247)	(0.0271)	(0.0146)	(0.00630)	(0.00863)	(0.00294)
N	2327	2327	1775	2327	2327	1775
Panel B: Urbanized Counties with Towns, Small Cities, and Large Cities in the United States						
County with large city (city size > 25,000)	-0.0265*			-0.0253***		
	(0.0142)			(0.00698)		
County with small city (city size = 10,000–24,999)	-0.0100			-0.00286		
	(0.0184)			(0.0106)		
N	456			456		
Panel C: Urbanized Counties with Towns, Small Cities, and Large Cities in the Northeast and the Midwest						
County with large city (city size > 25,000)			-0.0234***			-0.0116***
			(0.00626)			(0.00166)
County with small city (city size = 10,000–24,999)			-0.0232***			-0.00587***
			(0.00663)			(0.00100)
N			350			350
Panel D: Southern Cities and Midwestern Cities						
Southern cities			-0.0149			0.0149***
			(0.0192)			(0.00230)
N			68			68
Value of θ	7.35	8.22	6.10	7.35	8.22	6.10

Notes: Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Changes in market access is controlled in columns (4) to (6). Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

outcomes, and manufacturing-related economic outcomes. If we compare the group of neighboring counties and counties located far from new railways, there is a significant difference in total population and agriculture-related economic outcomes, but the difference is much narrower than the gap between counties with new rail tracks and adjacent counties. In addition, the gap in manufacturing-related economic outcomes is not significantly different from zero.

What causes the wide gap between counties with new rail tracks and adjacent counties? Panel A of Table 1.13 shows that counties directly connected by new railways had a larger increase in market access and faced less competition than adjacent counties with the same changes in log market access.

In panel B of Table 1.14, I control changes in log market access; I find that the effects of market access on all economic outcomes are significantly positive. The gaps in all economic outcomes between counties directly linked by new railways and adjacent counties shrink after controlling market access. The growth gap measured by agriculture-related economic outcomes and the number of manufacturing workers becomes insignificant, while the growth gap measured by total population and manufacturing value added remains significant. In panel C of Table 1.14, I further control for changes in competition. Compared to the results in panel B, the growth gap between two groups of counties further reduces, becoming insignificant in all economic outcomes. The competition effects on all economic outcomes are negative. And the effects are all statistically significant, except for value added in manufacturing, where $\theta = 6.10$.

I can conclude that both the difference in changes in market access and the difference in changes in competition help explain the growth gaps between counties directly linked by new railways and their adjacent counties.

1.6.2 Agglomeration Effects in Urbanized Counties with Large Cities

Another interesting story that can be bolted onto my framework is agglomeration effects in urbanized counties with a large city. From 1870 to 1890, urban population grew faster

Table 1.14: Counties with New Rail Tracks, Adjacent Counties, and Counties Far from New Rail Tracks

	(1) Changes in Log Total Population	(2) Changes in Log Farm Land Values	(3) Changes in Log Value of Farm Output	(4) Changes in Log Manufacturing Labor	(5) Changes in Log Manufacturing Value Added
Panel A: Baseline without Additional Controls					
Directly linked by new railways	0.294***	0.307***	0.290***	0.192**	0.220***
	(0.0817)	(0.110)	(0.0979)	(0.0714)	(0.0739)
Not adjacent to counties with new rail tracks	-0.110**	-0.239***	-0.155**	0.0710	0.0651
	(0.0442)	(0.0620)	(0.0607)	(0.0674)	(0.0745)
Panel B: Control for Changes in Log Market Access					
Directly linked by new railways	0.141**	0.0497	0.0798	0.121	0.142*
	(0.0657)	(0.109)	(0.0881)	(0.0752)	(0.0842)
Not adjacent to counties with new rail tracks	-0.109**	-0.236***	-0.152**	0.0710	0.0651
	(0.0435)	(0.0584)	(0.0592)	(0.0672)	(0.0746)
Changes in log market access	0.334***	0.501***	0.409***	0.186	0.204
	(0.0862)	(0.132)	(0.123)	(0.113)	(0.127)
Panel C: Control for Changes in Log Market Access and Changes in Competition					
Directly linked by new railways	0.0401	-0.105	-0.0487	0.0694	0.0966
	(0.0636)	(0.118)	(0.0835)	(0.0716)	(0.0786)
Not adjacent to counties with new rail tracks	-0.105**	-0.215***	-0.135**	0.0699	0.0641
	(0.0421)	(0.0561)	(0.0574)	(0.0669)	(0.0743)
Changes in log market access	0.574***	0.828***	0.682***	0.349***	0.347**
	(0.0879)	(0.141)	(0.130)	(0.119)	(0.133)
Changes in competition	-0.871***	-0.979***	-0.814***	-0.869**	-0.765
	(0.174)	(0.206)	(0.202)	(0.322)	(0.582)
Value of θ	7.35	8.22	8.22	6.10	6.10
Competition Is Measured in	All Sectors	Agriculture	Agriculture	Manufacturing	Manufacturing
N	2327	2327	2327	1775	1775

Notes: Baseline is counties without new rail tracks that are adjacent to counties with new rail tracks. Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in counties with large cities (Table 1.15).[10] What drove the persistent growth in urbanized counties with large cities? Panel B (sample of all urbanized counties) and Panel C (sample of urbanized counties in the Northeast and the Midwest) in Table 1.13 show that counties with a large city experienced smaller increases in log market access but also smaller increases in competition conditional on changes in log market access.

Can the difference in changes in competition explain why these counties experienced smaller increases in market access but greater urban population growth? The answer is yes. Holding market access constant, counties with a large city still had much higher urban population growth than other urbanized counties (Panel B of Table 1.15), but the growth gap becomes very small and insignificant when further controlling for changes in competition in manufacturing (Panel C of Table 1.15). The difference in changes in competition (conditional on changes in log market access) is the channel at work to explain the growth gap between counties with a large city and other urbanized counties.

Why did counties with large cities experience smaller increases in competition? The theoretical measurement of competition in market n faced by producers who produce goods in subsector j in location i from producers who produce in the same subsector in location m can be expressed as $\frac{T_m^j}{T_i^j}(\frac{r_m^\beta w_m^{1-\beta}}{r_i^\beta w_i^{1-\beta}})^{-\theta}\kappa_{mn}^{-\theta}$. If the productivity of subsector j in location i, T_i^j, is high, as the trade cost between competing location m and market n κ_{mn} decreases, the increase in competition $\frac{T_m^j}{T_i^j}(\frac{r_m^\beta w_m^{1-\beta}}{r_i^\beta w_i^{1-\beta}})^{-\theta}\kappa_{mn}^{-\theta}$ caused by decrease in κ_{mn} will be mitigated. If we compare productivity in manufacturing subsectors, there is almost no subsector in which counties with large cities have significantly lower productivity than other urbanized counties. The only exception is the subsector Miscellaneous Manufacturing Industries, in which large cities have significantly lower productivity in the sample of all urbanized counties. And in both the sample of all urbanized counties and urbanized counties in industrialized regions in the Northeast and the Midwest, in 9 out of 20 manufacturing subsectors, counties with large cities have significantly higher productivity than other urbanized counties (Table 1.16). Because of the relatively high productivity, counties with large cities face less competition

[10] I define large cities as those with a population of least 25,000.

Table 1.15: Urbanized Counties with Large Cities vs. Other Urbanized Counties

	(1) Changes in Log Urban Population	(2) Changes in Log Urban Population (NE, MW)
Panel A: Baseline without Additional Controls		
County with large city (city size > 25,000)	0.103**	0.168***
	(0.0502)	(0.0517)
County with small city (city size = 10,000—24,999)	0.0444	0.0518
	(0.0543)	(0.0603)
Panel B: Control for Changes in Log Market Access		
County with large city (city size > 25,000)	0.111**	0.177***
	(0.0494)	(0.0534)
County with small city (city size = 10,000—24,999)	0.0475	0.0612
	(0.0541)	(0.0599)
Changes in log market access	0.304**	0.407
	(0.123)	(0.274)
Panel C: Control for Changes in Log Market Access and Changes in Competition		
County with large city (city size > 25,000)	0.0752	0.0994
	(0.0494)	(0.0643)
County with small city (city size = 10,000—24,999)	0.0434	0.0221
	(0.0570)	(0.0569)
Changes in log market access	0.750***	0.622**
	(0.217)	(0.248)
Changes in competition	-1.420**	-6.667**
	(0.615)	(2.627)
Value of θ	7.35	6.10
Competition Is Measured in	All Sectors	Manufacturing
N	456	350

Notes: Baseline is urbanized counties without a city (city size larger than 10,000 in 1870). Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

than other urbanized counties, conditional on the changes in market access.

Competition plays a role of self-reinforcement in agglomeration. Counties with a large city enjoyed the advantage of economies of scale and had relatively higher productivity. And the persistence of growth is not because these counties experienced a larger increase in market access but because their high productivity meant they faced less fierce competition in the new transportation network.

1.6.3 Poor Performance of Southern Cities

The framework of competition also helps explain the slow growth of Southern cities relative to Midwestern cities after the Civil War.

Though Southern cities grew more slowly than their Midwestern counterparts (Table 1.17, column 1), no significant difference exists in the change of market access between the two groups of cities (Table 1.13, Panel D, column 3). The significant gap in growth between Southern and Midwestern cities remains almost the same after controlling for changes in log market access (Table 1.17), but the gap disappears when further controlling for changes in competition in manufacturing (Table 1.17, columns 4 and 5). Holding changes in market access constant, Southern cities experienced a significantly larger increase in competition in manufacturing than Midwestern cities (Table 1.13, panel D, column 6). The difference in competition is the channel at work to explain the growth gap between Southern cities and Midwestern cities. It helps explain the relatively slow growth of Southern cities after the Civil War.

Southern cities were much more integrated into the railway network after the Civil War than before, but trade integration is not always good news. Before the Civil War, almost all Southern cities were connected to the North by railways. After the Civil War, Northerners built railways to connect to small Southern towns. This created fierce competition for traditional Southern cities. As a result, many small Southern towns (e.g., Charlotte, Birmingham) grew faster after the Civil War, while many large cities were stagnant or declined. The typical examples are New Orleans and Mobile, which experienced higher than average

Table 1.16: Manufacturing Productivity among Urbanized Counties

	All Urbanized Counties			Urbanized Counties in the Northeast and Midwest		
Productivity of Subsectors	Large City	Small City	N	Large City	Small City	N
Food and Kindred Products	0.00995	-0.00969	438	0.0148	-0.00839	346
	(0.0100)	(0.00745)		(0.0107)	(0.00588)	
Tobacco Products	0.0164**	0.00501	219	0.0120*	-0.00104	187
	(0.00628)	(0.00859)		(0.00662)	(0.00927)	
Textile Mill Products	0.00195	-0.000399	336	0.00438	0.00207	289
	(0.00920)	(0.00926)		(0.00912)	(0.00947)	
Apparel and Other Finished Products Made from Fabrics and Similar Materials	0.0302*	0.0251*	350	0.0327*	0.0212*	298
	(0.0161)	(0.0128)		(0.0169)	(0.0116)	
Lumber and Wood Products, Except Furniture	0.00393	0.00509	433	0.00331	0.00154	342
	(0.00267)	(0.00370)		(0.00303)	(0.00316)	
Furniture and Fixtures	0.0203***	0.00500	294	0.0278***	0.00389	260
	(0.00676)	(0.00765)		(0.00602)	(0.00590)	
Paper and Allied Products	-0.0249	-0.00607	155	-0.0309	-0.00975	144
	(0.0216)	(0.0278)		(0.0229)	(0.0289)	
Printing, Publishing, and Allied Industries	0.0612***	0.0169	232	0.0550***	0.0203	186
	(0.0220)	(0.0194)		(0.0117)	(0.0165)	
Chemicals and Allied Products	0.0301	-0.00381	146	0.0334	0.00873	122
	(0.0254)	(0.0164)		(0.0285)	(0.0163)	
Petroleum Refining and Related Industries	0.0252	0.0211	46	0.0138	-0.0156	43
	(0.0389)	(0.114)		(0.0395)	(0.107)	
Rubber and Miscellaneous Plastics Products	0.0213	0.0105	36	0.0721	-0.00310	30
	(0.0646)	(0.0643)		(0.0646)	(0.0191)	
Leather and Leather Products	0.0454***	0.0222**	411	0.0436***	0.0184**	336
	(0.0157)	(0.00898)		(0.0143)	(0.00833)	
Stone, Clay, Glass, and Concrete Products	0.0233**	-0.00499	303	0.0331**	-0.00138	258
	(0.0102)	(0.00768)		(0.0122)	(0.00837)	
Primary Metal Industries	0.00841	0.00180	329	0.00622	0.00181	277
	(0.00651)	(0.00699)		(0.00697)	(0.00745)	
Fabricated Metal Products, Except Machinery and Transportation Equipment	0.0158***	0.0136**	377	0.0203***	0.0138**	310
	(0.00315)	(0.00539)		(0.00233)	(0.00553)	
Industrial and Commercial Machinery and Computer Equipment	0.0188*	0.00257	364	0.0169*	-0.00228	305
	(0.00930)	(0.00693)		(0.00940)	(0.00711)	
Electronic and Other Electrical Equipment and Components, Except Computer Equipment	-0.0508	-0.152	25	-0.0508	-0.187	24
	(0.166)	(0.168)		(0.167)	(0.168)	
Transportation Equipment	0.0246***	0.0105***	402	0.0225***	0.0112***	331
	(0.00408)	(0.00308)		(0.00347)	(0.00318)	
Measuring, Analyzing, and Controlling Instruments; Photographic, Medical and Optical Goods; Watches and Clocks	0.0436	-0.0310	31	0.0176	-0.0310	29
	(0.0804)	(0.0755)		(0.0828)	(0.0763)	
Miscellaneous Manufacturing Industries	-0.0525**	-0.0175	157	-0.0349	-0.0155	137
	(0.0228)	(0.0242)		(0.0212)	(0.0143)	

Notes: Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

increase in market access compared to urbanized counties in the South and the Midwest, but also higher increase in manufacturing competition. As a result, New Orleans slipped in the rankings of largest cities from ninth to twelfth largest, and Mobile slipped from 39th to 97th.

Table 1.17: Growth Gap between Southern Cities and Midwestern Cities

	(1) Changes in Log City Population
Panel A: Baseline without Additional Controls	
Southern cities	-0.222*
	(0.116)
Panel B: Control for Changes in Log Market Access	
Southern cities	-0.188*
	(0.0983)
Changes in log market access	2.263
	(1.756)
Panel C: Control for Changes in Log Market Access and Changes in Competition	
Southern cities	0.157
	(0.135)
Changes in log market access	0.852
	(1.227)
Changes in competition	-23.20**
	(8.306)
Value of θ	6.10
Competition Is Measured in	Manufacturing
N	68

Notes: Baseline is Midwestern cities. Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.7 Counterfactual Analysis

1.7.1 Remove Railways: Compare with Estimation in (DH16)

Whether a county benefited from railway construction depends on which effects dominate, the positive market-access effects or the negative effects of competition. Indeed, many counties suffered from new railway construction. A new question arises: Does the model without structural competition overestimate positive railway effects on the overall economy? In this section, I show that (DH16) overestimate railway effects on agricultural land values. Following their strategy, I discuss the counterfactual scenario: If all railways disappeared, how large would the decrease in agricultural land values be?

In this section, I focus neither on the how large the effects of competition are anymore nor on decomposing competition and market access. Instead, I focus on Equation 1.14, in which aggregate consumer market access and aggregate firm market access in agriculture positively affect agricultural land values, while aggregate firm market access in all sectors negatively affects agricultural land values. For the best comparison, I focus on the θ value 8.22 from the estimation in (DH16).[11]

The regression specification is

$$\begin{aligned}\Delta \ln LandValue_i^{Agriculture} =& \delta + \eta_1 \Delta \ln(CMA_i) + \eta_2 \Delta \ln(FMA_i^{Agriculture}) + \eta_3 \Delta \ln(FMA_i) \\ &+ \delta_s + f(x_i, y_i) + \varepsilon_i. \end{aligned} \tag{1.36}$$

Table 1.18 reports the results of Equation 1.14. As implied by the theoretical model, changes in log aggregate consumer market access and changes in log aggregate firm market access in agriculture positively affect changes in log agricultural land values, while changes in log aggregate firm market access in all sectors negatively affect changes in log agricultural land values. The results are consistent in both the unweighted regression and the weighted

[11] In our baseline model, for simplicity, I assume the value of θ is consistent across all subsectors; however, it would generate biases. As an extension, I develop a model that allows the value of θ to vary by subsectors, assuming that all lands are used in agriculture. I estimate the model as a robustness check.

Table 1.18: Effects of Consumer Market Access and Firm Market Access on Agricultural Land Values

	Changes in Log Farm Land Values	
	(1)	(2)
Changes in log consumer market access	0.233***	0.238*
	(0.0857)	(0.136)
Changes in log firm market access of agriculture	1.736***	2.374***
	(0.404)	(0.272)
Changes in log firm market access of all sectors	-1.201***	-2.116***
	(0.365)	(0.351)
Value of θ	8.22	8.22
Weighted by Farm Land Values in 1870	No	Yes
N	2327	2327
R^2	0.630	0.641
adj. R^2	0.622	0.633

Notes: State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

regression.[12] This model provides more data fitness and explains more than the model used to analyze the counterfactual scenario in (DH16), (see Appendix E.4 for details.) Therefore, I need to apply the model to the counterfactual analysis and reestimate the aggregate railway effects.

Based on the estimates with 1870–1890 data, I reestimate the aggregate railway effects on total land values in 1890. In the baseline, I assume the population distribution is the same as the one I observed in 1890. I am also holding the total population constant, and I discuss the effects of population distribution in the years with fewer railways: 1870, 1850, and 1830. The estimation is shown in Panel A of Table 1.19. Under the framework of the multisector EK model with structural competition, the aggregate railway effects were much smaller and more varied than the ones in (DH16). In the baseline estimation (with the 1890 population distribution), the percentage decline in agricultural land values without railways is 23.1% (1.2% of 1890 GNP). It is only 38.4% of the estimation in (DH16). Unlike the estimation in

[12] In (DH16), to minimize the influence of outliers and to estimate the appropriate average effect for the counterfactual analysis, I weight the regression by 1870 land values. I follow this strategy in the counterfactual analysis.

Table 1.19: Estimation of Aggregate Railway Effects on Agricultural Land Values—Comparison between This Paper and (DH16) in Counterfactual Analysis

	(1) Estimated by This Paper	(2) Estimated by Donaldson and Hornbeck (2016)
Panel A: Percentage Decline in Land Value without Railways		
Assuming the population distribution from 1890	23.1	60.2
	(18.6)	(4.2)
Assuming the population distribution from 1870	19.3	59.1
	(20.4)	(4.1)
Assuming the population distribution from 1850	26.7	59.3
	(15.0)	(4.1)
Assuming the population distribution from 1830	39.8	60.1
	(12.8)	(4.0)
Panel B: Decrease in Agricultural Land Values by Regions (Assuming the Population Distribution from 1890)		
Northeast	0.05 billion	0.5 billion
South	0.2 billion	0.5 billion
Midwest	1.4 billion	2.5 billion
Plains	0.4 billion	0.9 billion
Far West	-0.3 billion	0.5 billion
Panel C: Percentage Decline in Land Values without Railways, with Improved Traditional Transportation Methods (Assuming the Population Distribution from 1890)		
With proposed canals	21.6	52.4
	(15.9)	(4.2)
With improved country roads, wagon cost of 14 cents (40% cost reduction)	17.0 (13.3)	47.5 (3.9)
With proposed canals and improved country roads	16.1 (11.1)	40.0 (3.7)

(DH16), my estimation of railway effects is sensitive to the change of population distribution. The largest loss is estimated under the population distribution in 1830, that the percentage decline in agricultural land values without railways is 39.8% (2.0% of 1890 GNP). It is still less than two-thirds of the effect in (DH16).

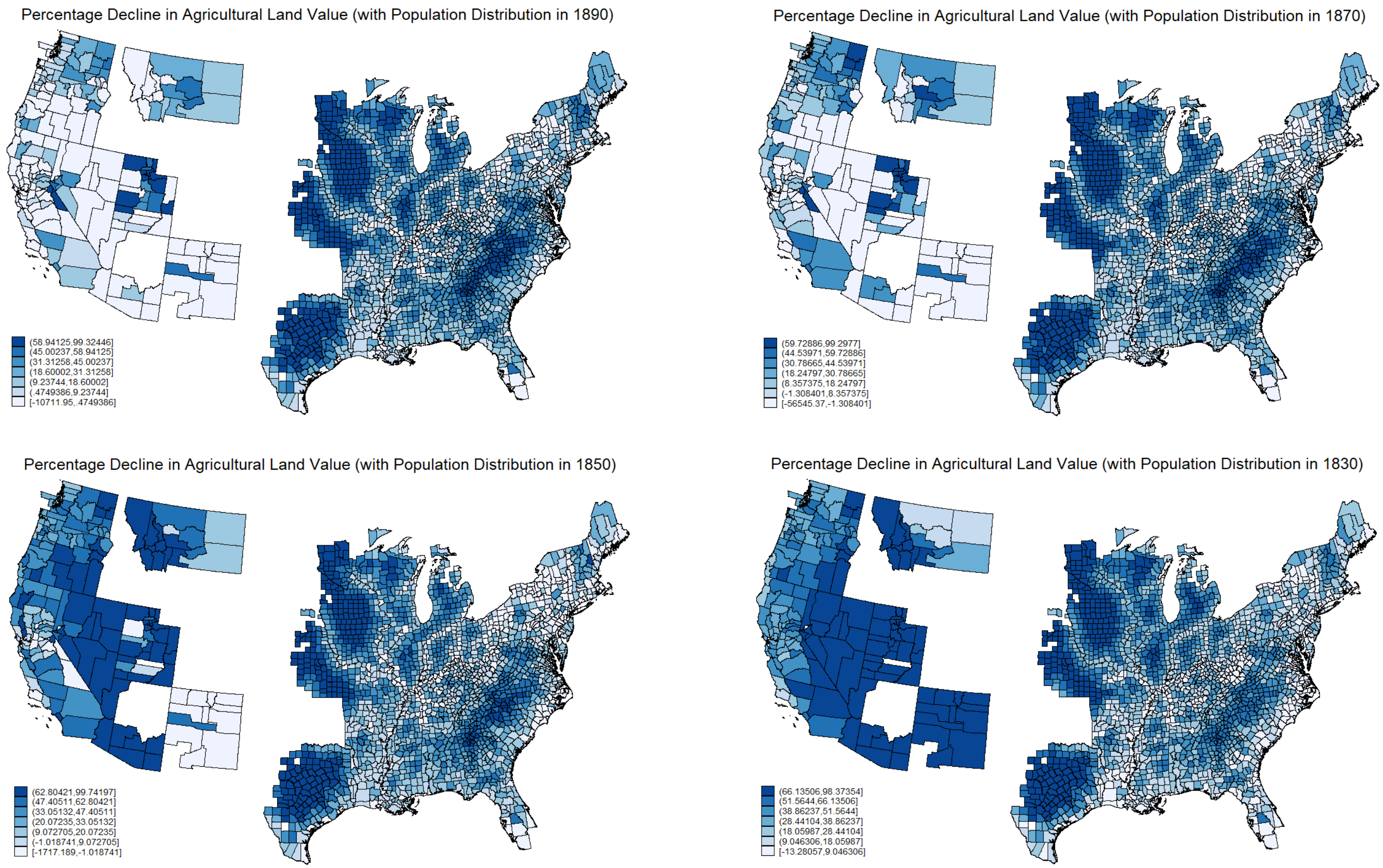
Panel B of Table 1.19 shows the decrease in agricultural land values by regions (assuming the population distribution from 1890). The land value losses estimated by this paper are much smaller than the ones in (DH16) in all regions. An extreme example is the Far West, which would experience increased land values in the counterfactual scenario.

Panel C of Table 1.19 shows that improving traditional transportation methods would slightly further reduce the percentage decline in land values. In the extreme case, with an extended canal system and improved country roads (proposed by (Fog64)), the percentage decline is as low as 16.1% (0.8% of 1890 GNP).

Figure 1.4 shows the percentage decline in agricultural land values. Darker colors indicate a larger percentage of land value losses. From the population distribution in 1890 to the population distribution in 1830, Eastern counties show a similar trend. Counties with convenient water transportation were less affected without railways. Western counties show a different story. If I assume that population distribution is the same in 1870 and 1890, most Western counties lose little land values without railways. Some Western counties would experience an increase in land values by removing all railways. The first transcontinental railway, the Pacific Railroad, from St. Louis to San Francisco, was completed in 1869. Among all census years, 1870 population distribution should be the one closest to counterfactual population distribution in the western United States. The result with 1870 population distribution might be the most reasonable in terms of distribution of percentage decline in agricultural land values for the Western United States.

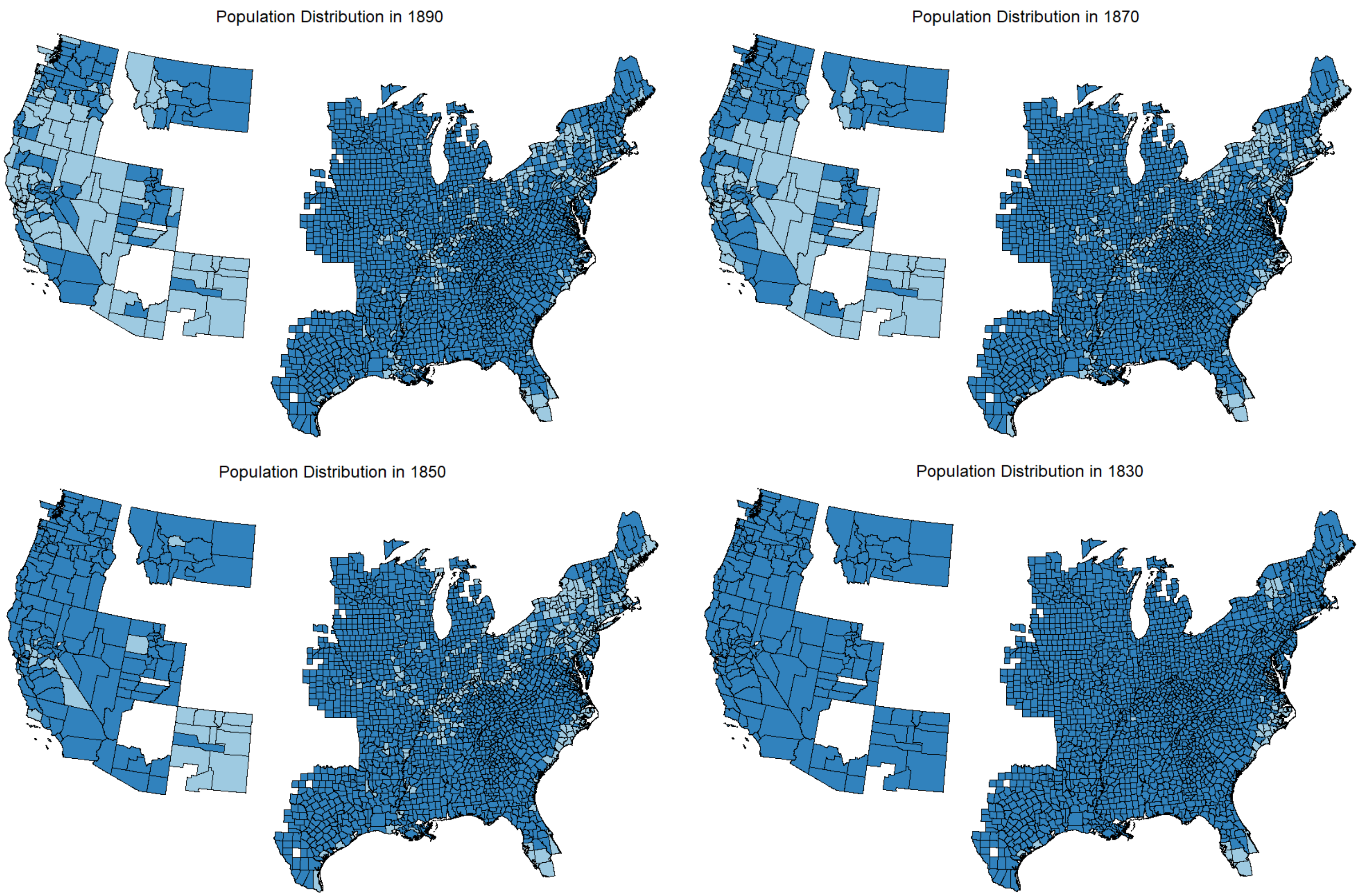
In Figure 1.5, I recategorize counties into two groups. Darker colors indicate counties that would experience land value losses by removing all railways in 1890. They were the winners in the 1890 rail system. Lighter colors indicate counties that would experience land values gain by removing all railways in 1890. They were the losers in 1890 rail system.

Figure 1.4: Percentage Decline in Agricultural Land Values



Notes: All counties are classified to seven quantiles. Darker colors indicate larger percentages of land value loss.

Figure 1.5: Winners and Losers in the 1890 Rail System



Notes: Darker colors indicate counties would experience lower land values by removing all railways in 1890. They are winners in the 1890 rail system. Lighter colors indicate counties that would experience higher land values by removing all railways in 1890. They are losers in the 1890 rail system.

More counties would become losers with later population distribution. Only a few counties become losers with the 1830 population distribution. With the 1870 and 1890 population distributions, most losers are in the western United States, where many counties featured a harsh natural environment, low agricultural productivity, and were vulnerable to trade shocks that increased competition. These counties covered most of New Mexico, Arizona, Nevada, and inland Oregon.

The counterfactual analysis still suffers some shortages under the framework of the multi-sector EK model. First, unlike the simple model without structural competition in (DH16), adding information about industrial structure and productivity helps me better interpret the data. However, this complication makes estimating the counterfactual population distribution impossible from the model itself. Therefore, I have to use the observed historical data as a proxy for the counterfactual population distribution. Second, I lack sufficient data to be able to observe industrial structure in both the early periods (1830, 1850) and the later period (1890); I have to assume it fixed at the 1870 level. In fact, in a world without railways, the distribution of industry and productivity would be different from the one that existed in 1870. In general, I cannot conclude that I accurately estimate the aggregate railway effects on agricultural land values. Despite the model's shortcomings, it still provides evidence to show that it is very likely to overestimate the aggregate railway effects if structural competition is deleted.

1.8 Conclusion

This paper develops a new "market access-competition" approach for analyzing the effects of a new transportation network on economic growth. The multisector EK model highlights that producers in a given location faced structural-related competition from producers of similar goods in other locations. The geographical variation of changes in competition comes from the distribution of natural resources for all location-sector pairs, as well as the reduced trade cost resulting from the new transportation network.

I develop a reduced-form method for decomposing the positive market-access effects and negative structural-related effects of competition with limited historical data. I find that the competition effects are substantial and crucial. On the aggregate level, I find that the expansion of railways increased agricultural land values by 20% to 40%, thus implying that railways contributed to 1% to 2% of 1890 GNP. My estimate is one-third to two-thirds of (DH16)'s estimates and lower than (Fog64)'s. It also explains why so many counties grew slowly despite a huge increase in market access.

This paper contributes to the literature by virtue of my methodological measurement of the economic benefits of a new transportation infrastructure. Other researchers have focused either on comparing locations that gained a new transportation network with those that did not or on estimating the aggregate effects of a new transportation network. The disadvantage of the former is that the methodology assumes that all locations connected by the new network are equally affected and that new routes have no effect on locations far from the new routes. The disadvantage of the social savings methodology employed by (Fog64) is that it does not consider competition between production locations. Although (DH16)'s market-access approach allows for competition between production locations, it assumes that all locations compete equally with each other. Thus, there is not much difference between their approach with and without competition. In this paper, I develop a new method that considers productivity in each location-pair, derive a competition index from a multisector general-equilibrium model, and identify the negative effects of trade integration. My method contributes to the estimation of both aggregate effects and individual effects, and it combines the complementary strengths of two earlier contributions to the literature.

The policy implications of my research are profound—the World Bank promotes infrastructure spending as an engine for development and spends heavily on transportation to connect underdeveloped regions. My research shows the pitfalls of this approach—low productivity and increased competition are likely to make these regions suffer—and encourages policymakers to rethink the unintended, often deleterious, consequences of these projects. Fortunately, my reduced-form methodology can be easily applied in developing countries, where data are limited.

This paper suggests some avenues for future research. First, my current work focuses on competition in production sectors. But competition in nonproduction sectors such as trade, commerce, and transshipment also interests me. For example, I am intrigued by how so many traditional U.S. and Chinese transportation hubs declined after the completion of expanded transportation networks, which better connected them to markets but forced them to compete with new hubs. I want to develop a method to measure this competition and further identify the causality between competition and economic growth.

Second, like trade shocks, productivity shocks can also affect the intensity of competition between locations. Previous literature ignores competition and undertakes difference-in-differences strategies. In fact, the counterfactual trend of affected regions should be different from the trend that we observed on unaffected regions. The results of traditional difference-in-differences strategy is biased. On the one hand, unaffected regions did not experience negative productivity shock, but was still affected by the change in competition. They faced less competition after the shock and experienced a large increase in production. In the counterfactual scenario without the negative productivity shocks, the affected regions would not experience a reduction in competition. Difference-in-differences strategy overestimates the effects. On the other hand, if a sector in a given location experiences a negative productivity shock, the location's productive resources will shift to other sectors, retarding technological progress in the affected sector. Affected regions would improve their productivity, and thus face less competition from unaffected regions; meanwhile, unaffected regions would slow down technological progress in the sector and face more competition. As a result, traditional difference-in-differences strategies underestimate productivity shocks.

Third, the effects of trade shocks and productivity shocks on competition vary by the geographic similarity of economic activities, thus naturally, the pace of development varied from region to region. I observe more rise and fall of locations in the preindustrial period and the early stage of industrialization than I observe today. One hypothesis about the observation is that in the early stages of industrialization, the distribution of economic activities highly relied on the distribution of natural resources, and production in a specific sector was geographically dispersed. As I discussed in this paper, nearby locations fiercely competed with each other. Trade shocks or productivity shocks could easily cause economic activities to shift from large cities to nearby smaller, fast-growing cities. As specialization and agglomeration increased, the disparity in industrial structure and productivity became wider among nearby locations. As a result, same level shocks had larger effects on reshaping the distribution of economic activities in the preindustrial period and the early stage of industrialization than they do nowadays. I will test this hypothesis by constructing a panel data set containing trade shocks, productivity shocks, distribution of production, and distribution of economic activities from the 19th century to the present.

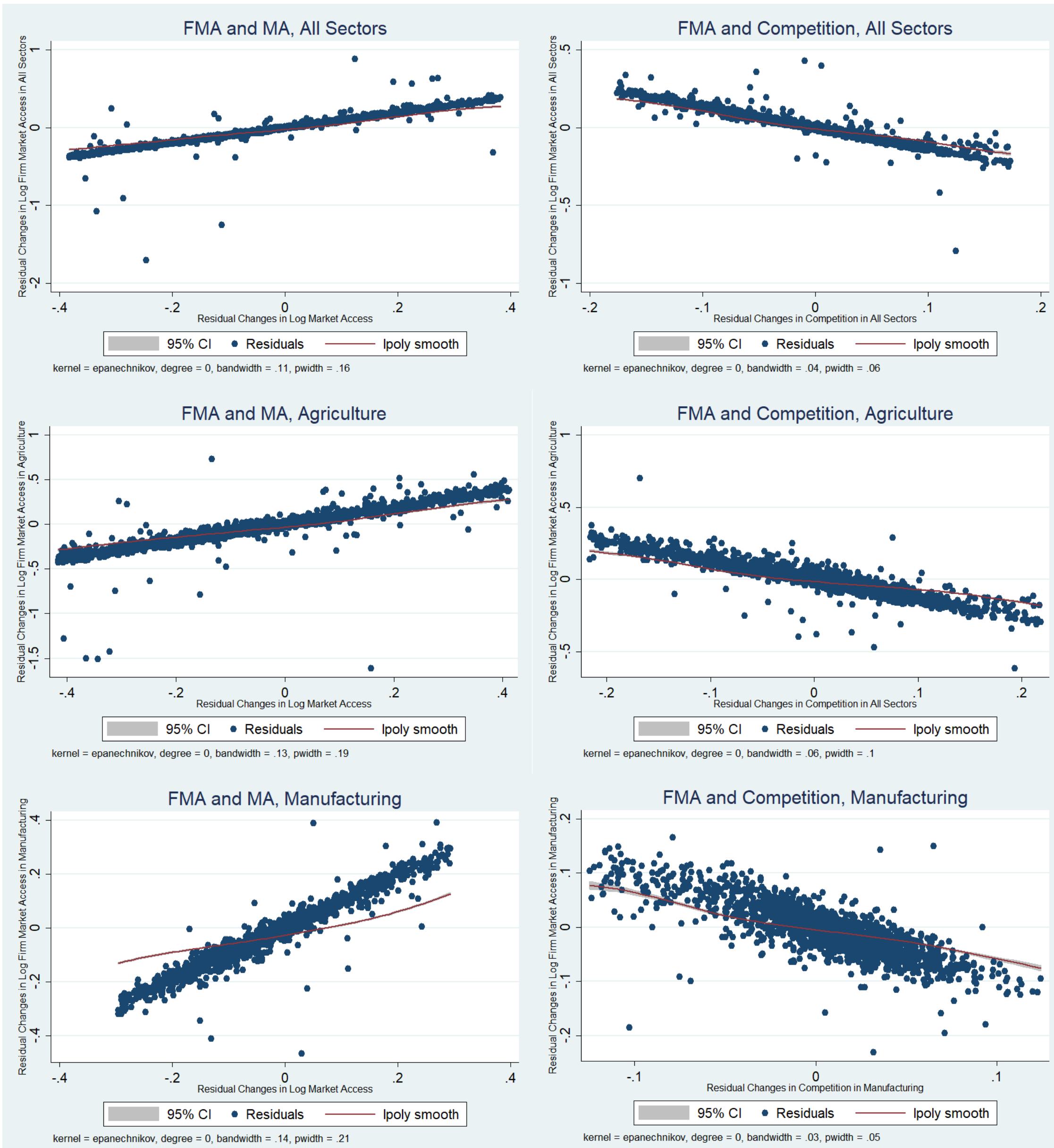
1.A Linear Decomposition

In this appendix, I show that the linear decomposition of $\Delta \ln FMA_i$ to $\Delta \ln MA_i$ and $\Delta Competition_i$ is not arbitrary.

I calculate residual changes in log firm market access and log market access, after conditioning on changes in competition and other control variables in the baseline regression,[13] and residual changes in log firm market access and changes in competition, after conditioning on changes in log market access and other control variables in the baseline regression. I limit the sample to residual changes in market access or changes in competition within one standard deviation. Figure 1.6 shows scatter points, a kernel-weighted local polynomial fit line, and the 95% confidence interval of the fit line. The left three subfigures plot the local polynomial relationship between residual changes in log firm market access and residual changes in log market access, conditional on changes in competition and other control variables. The right three subfigures plot the local polynomial relationship between residual changes in log firm market access and residual changes in competition, conditional on residual changes in log market access and other control variables. The top two subfigures show the relationships of variables for all sectors. The middle two subfigures show the relationships of agriculture-related variables. The bottom two subfigures show the relationships of manufacturing-related variables. It is obvious that the relationship between changes in log firm market access and log market access is positively linear, while the relationship between changes in log firm market access and changes in competition is negatively linear.

[13] Including state fixed effects and latitude and longitude variables.

Figure 1.6: Local Polynomial Relationship Between Changes in Log Firm Market Access and Log Market Access, Between Changes in Log Firm Market Access and Changes in Competition, 1870–1890



Notes: Residual changes are calculated by regressing changes in the indicated variables on state fixed effects and longitude-latitude variables. The left three figures plot the local polynomial relationship between residual changes in log firm market access and residual changes in log market access, conditional on changes in competition. The right three figures plot the local polynomial relationship between residual changes in log firm market access and residual changes in competition, conditional on residual changes in log market access. All figures are based on the Epanechnikov kernel function, with default bandwidth and pwidth.

1.B Sector-Matching Table to Calculate θ

Table 1.20: Sector-Matching Table

Sectors in This Paper (Manufacturing is Grouped by SIC 1987)	Sectors in (CP15) (Manufacturinging is Grouped by ISIC Rev.3)	θ Estimated by (CP15)	Weight
Food and Kindred Products	Food Products, Beverages, and Tobacco	2.55	0.0491
Tobacco Products	Food Products, Beverages, and Tobacco	2.55	0.0063
Textile Mill Products	Textiles, Textile Products, Leather, and Footwear	5.56	0.0480
Apparel and Other Finished Products Made from Fabrics, and Similar Materials	Textiles, Textile Products, Leather, and Footwear	5.56	0.0279
Lumber and Wood Products, Except Furniture	Wood and Products of Wood and Cork	10.83	0.0396
Furniture and Fixtures	Other	5	0.0100
Paper and Allied Products	Pulp, Paper, Paper Products, Printing and Publishing	9.07	0.0057
Printing, Publishing, and Allied Industries	Pulp, Paper, Paper Products, Printing and Publishing	9.07	0.0128
Chemicals and Allied Products	Chemicals	4.75	0.0112

Continued on next page

Table 1.20 – *Continued from previous page*

Sectors in this paper (Manufacturing is grouped by SIC 1987)	Sectors in (CP15) (Manufacturing is grouped by ISIC Rev.3)	θ estimated by (CP15)	Weight
Petroleum Refining and Related Industries	Coke, Refined Petroleum, and Nuclear Fuel	51.08	0.0015
Rubber and Miscellaneous Plastics Products	Rubber and Plastics Products	1.66	0.0029
Leather and Leather Products	Textiles, Textile Products, Leather and Footwear	5.56	0.0326
Stone, Clay, Glass, And Concrete Products	Other Nonmetallic Mineral Products	2.76	0.0155
Primary Metal Industries	Basic Metals	7.99	0.0242
Fabricated Metal Products, Except Machinery and Transportation Equipment	Fabricated Metal Products, Except Machinery and Equipment	4.30	0.0341
Industrial and Commercial Machinery and Computer Equipment	Machinery and Equipment n.e.c	4.75	0.0303
Electronic and Other Electrical Equipment and Components, Except Computer Equipment	Electrical Machinery and Apparatus, n.e.c.	10.60	0.0010
Transportation Equipment	Other Transport Equipment (Other Than Motor Vehicles Trailers and Semitrailers)	0.37	0.0163

Continued on next page

Table 1.20 – *Continued from previous page*

Sectors in this paper (Manufacturing is grouped by SIC 1987)	Sectors in (CP15) (Manufacturing is grouped by ISIC Rev.3)	θ estimated by (CP15)	Weight
Measuring, Analyzing, And Controlling Instruments; Photographic, Medical and Optical Goods; Watches and Clocks	Medical, Precision and Optical Instruments, Watches and Clocks	8.98	0.0027
Miscellaneous Manufacturing Industries	Manufacturing n.e.c. and Recycling	5.00	0.0162
Barley, Buckwheat, Corn, Cotton, Flax, Livestock, Oats, Peas and Beans, Sweet Potatoes, Rice, Rye, Tobacco, Wheat, White Potatoes	Agriculture	8.11	0.5833
Asphalt, Cinnabar, Coal, Copper, Gold, Iron, Nickel, Lead, Peat, Petroleum, Silver, Stone, Zinc	Mining	15.72	0.0288

1.C Results of Market Size Measured by Income

In this appendix, I measure market size by value of agricultural output plus manufacturing value added. Table 1.21 is comparable with Table 1.1. It discusses the negative effects of competition on economic growth. Table 1.22 is comparable with Table 1.2. It shows results about the role of comparative advantage. Table 1.23 is comparable with Table 1.3. It shows the positive link between comparative advantage in agriculture and economic growth. All results are consistent with the ones where market size is measured by population.

Table 1.21: Positive and Negative Effects of Railway Expansion, 1870–1890, Market Size Is Measured by Income

	(1) Changes in Log Total Population	(2) Changes in Log Farm Land Values	(3) Changes in Log Value of Farm Output	(4) Changes in Log Manufacturing Labor	(5) Changes in Log Manufacturing Value Added	(6) Changes in Log Urban Population	(7) Changes in Log Urban Population (NE, MW)	(8) Changes in Log City Size	(9) Changes in Log City Size (NE, MW)
Panel A: Allow Market Size to Change between 1870 and 1890									
Changes in log market access	0.569***	0.815***	0.696***	0.391***	0.431***	0.845***	0.748***	0.773	-0.241
	(0.112)	(0.110)	(0.114)	(0.133)	(0.136)	(0.176)	(0.254)	(0.516)	(0.465)
Changes in competition	-0.757***	-1.024***	-0.893***	-0.961***	-0.989*	-3.217***	-11.12***	-3.584	-16.10**
	(0.162)	(0.146)	(0.158)	(0.309)	(0.531)	(0.924)	(3.027)	(2.206)	(6.866)
% Market Access Effects Offset by Competition Effects	50.86%	59.84%	61.07%	74.47%	69.50%	240.54%	125.88%	103.68%	NA
Panel B: Fix Market Size at the 1870 Level									
Changes in log market access	0.572***	0.840***	0.694***	0.369***	0.396***	0.692***	0.545*	0.777	-0.324
	(0.113)	(0.120)	(0.128)	(0.133)	(0.138)	(0.199)	(0.272)	(0.506)	(0.647)
Changes in competition	-0.711***	-1.002***	-0.803***	-0.799***	-0.754	-1.820**	-7.033**	-2.839	-11.39
	(0.192)	(0.161)	(0.216)	(0.259)	(0.482)	(0.793)	(2.879)	(1.706)	(6.599)
% Market Access Effects Offset by Competition Effects	52.30%	62.39%	60.49%	73.24%	64.45%	175.93%	146.68%	111.97%	NA
Value of θ	7.35	8.22	8.22	6.10	6.10	7.35	6.10	7.35	6.10
Competition Is Measured in	All Sectors	Agriculture	Agriculture	Manufacturing	Manufacturing	All Sectors	Manufacturing	All Sectors	Manufacturing
N	2327	2327	2327	1775	1775	456	350	162	130

Notes: Market size is measured by agricultural output value plus manufacturing value added. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.22: Role of Comparative Advantage, Market Size Is Measured by Income

	(1)	(2)	(3)	(4)
	Changes in Log Farm Land Values	Changes in Log Value of Farm Output	Changes in Log Manufacturing Labor	Changes in Log Manufacturing Value Added
Panel A: Allow Market Size to Change Between 1870 and 1890				
Changes in competition in agriculture/manufacturing	-2.274***	-2.206***	2.951*	1.824
	(0.511)	(0.490)	(1.472)	(1.515)
Changes in competition in all sectors	1.753**	1.663***	-2.240***	-2.065***
	(0.698)	(0.605)	(0.612)	(0.624)
Panel B: Fix Market Size at the 1870 Level				
Changes in competition in agriculture/manufacturing	-2.325***	-2.239***	2.292	0.938
	(0.459)	(0.521)	(1.680)	(1.600)
Changes in competition in all sectors	1.918***	1.977***	-1.585**	-1.194
	(0.536)	(0.524)	(0.731)	(0.727)
Value of θ in Competition in Agriculture/Manufacturing	8.22	8.22	6.10	6.10
Value of θ in Competition in All Sectors	7.35	7.35	7.35	7.35
N	2327	2327	2327	2327

Notes: Market size is measured by agricultural output value plus manufacturing value added. Changes in log market access with θ=8.22 and θ=7.35 are controlled in columns (1) and (2). Changes in log market access with θ=6.10 and θ=7.35 are controlled in columns (3) and (4). State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.23: Effects of Agricultural Competition on the Whole Economy, Holding Constant the Changes in Competition in All Sectors, Market Size Is Measured by Income

	(1)	(2)
	Changes in Log Total Population	Changes in Log Value of Agricultural Output Plus Manufacturing Value Added
Panel A: Allow Market Size to Change between 1870 and 1890		
Changes in competition in agriculture	-1.529***	-2.115***
	(0.349)	(0.484)
Changes in competition in all sectors	1.010*	1.236*
	(0.528)	(0.624)
Panel B: Fix Market Size at the 1870 Level		
Changes in competition in agriculture	-1.491***	-2.047***
	(0.337)	(0.538)
Changes in competition in all sectors	1.065***	1.561**
	(0.369)	(0.609)
Value of θ in Competition in Agriculture	8.22	8.22
Value of θ in Competition in All Sectors	7.35	7.35
N	2327	2327

Notes: Market size is measured by agricultural output value plus manufacturing value added. Changes in log market access with θ=8.22 and θ=7.35 are controlled. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.D More Robustness Checks

1.D.1 Demand from International Trade and Other Producers

One concern is that demand for a specified product is not proportional to population or income across locations. There are three possibilities. First, consumer preference is not homogeneous across locations. Second, domestic products are not demanded by domestic and international consumers. Third, for simplicity and data availability, I use a simple model without material in production. However, producers in many sectors sell goods not only to consumers but also to other producers, who use intermediate goods in production.

It is hard to measure the heterogeneity in consumer preference, but we can incorporate demand from international and other producers into the model. Net export of goods in a certain sector can be incorporated into the demand from the related international trade port. The first-order approximation of aggregate firm market access (Equation 1.19) can be rewritten as:

$$FMA_{it} \approx \sum_j \alpha^j \sum_{n \neq i} \frac{\kappa_{nit}^{-\theta} Y_{nt}}{\sum_{m \neq i} \kappa_{nmt}^{-\theta} \frac{T_m^j}{T_i^j}} + \sum_j \sum_p \frac{\kappa_{pit}^{-\theta} NEXP_{p,t}^j}{\sum_{m \neq i} \kappa_{pmt}^{-\theta} \frac{T_m^j}{T_i^j}}, \tag{1.37}$$

in which p is an index for ports and $NEXP_{pt}^j$ is subsector j's net exports value at port p in census year t. I obtain the statistics of imports and exports in each main port from (Tre70).

Following the strategy in the baseline analysis, changes in FMA_i can divide into two parts, changes in market access and the changes in competition, as the following equations:

$$\begin{aligned} \Delta \ln(MA_i) = & \ln(\sum_j \alpha^j \sum_{n \neq i} \kappa_{ni,t+1}^{-\theta} Y_{n,t+1} + \sum_j \sum_p \kappa_{pi,t+1}^{-\theta} NEXP_{p,t+1}^j) \\ & - \ln(\sum_j \alpha^j \sum_{n \neq i} \kappa_{ni,t}^{-\theta} Y_{n,t} + \sum_j \sum_p \kappa_{pi,t}^{-\theta} NEXP_{p,t}^j) \end{aligned} \tag{1.38}$$

$$\Delta Competition_i = \ln(\sum_j \alpha^j \sum_{n\neq i} \frac{\kappa_{ni,t}^{-\theta} Y_{n,t}}{\sum_{m\neq i} \kappa_{nm,t}^{-\theta} \frac{T_m^j}{T_i^j}} + \sum_j \sum_p \frac{\kappa_{pit}^{-\theta} NEXP_{p,t}^j}{\sum_{m\neq i} \kappa_{pmt}^{-\theta} \frac{T_m^j}{T_i^j}})$$
$$- \ln(\sum_j \alpha^j \sum_{n\neq i} \frac{\kappa_{ni,t}^{-\theta} Y_{n,t}}{\sum_{m\neq i} \kappa_{nm,t+1}^{-\theta} \frac{T_m^j}{T_i^j}} + \sum_j \sum_p \frac{\kappa_{pit}^{-\theta} NEXP_{p,t}^j}{\sum_{m\neq i} \kappa_{pm,t+1}^{-\theta} \frac{T_m^j}{T_i^j}}). \quad (1.39)$$

To deal with the issue of demand from other producers,[14] I need to build up an input-output matrix for the late 19th century. Based on the dataset in (Bat04), I build up a rough input-output matrix: In order to produce one dollar of products from sector j, how much did an average producer need to spend on material from sector k? And then, combining the input-output matrix with county-level manufacturing data in the 1870 census, I can calculate the demand for intermediate goods of each sector from each county. The changes in market access and changes in competition can be further modified as the following equations:

$$\Delta \ln(MA_i) = \ln(\sum_j \alpha^j \sum_{n\neq i} \kappa_{ni,t+1}^{-\theta} Y_{n,t+1} + \sum_j \sum_p \kappa_{pi,t+1}^{-\theta} NEXP_{p,t+1}^j + \sum_j \sum_o \kappa_{oi,t+1}^{-\theta} INPUT_{o,t+1}^j)$$
$$- \ln(\sum_j \alpha^j \sum_{n\neq i} \kappa_{ni,t}^{-\theta} Y_{n,t} + \sum_j \sum_p \kappa_{pi,t}^{-\theta} NEXP_{p,t}^j + \sum_j \sum_o \kappa_{oi,t}^{-\theta} INPUT_{o,t}^j) \quad (1.40)$$

$$\Delta Competition_i = \ln(\sum_j \alpha^j \sum_{n\neq i} \frac{\kappa_{ni,t}^{-\theta} Y_{n,t}}{\sum_{m\neq i} \kappa_{nm,t}^{-\theta} \frac{T_m^j}{T_i^j}} + \sum_j \sum_p \frac{\kappa_{pit}^{-\theta} NEXP_{p,t}^j}{\sum_{m\neq i} \kappa_{pmt}^{-\theta} \frac{T_m^j}{T_i^j}} + \sum_j \sum_o \frac{\kappa_{oit}^{-\theta} INPUT_{o,t}^j}{\sum_{m\neq i} \kappa_{omt}^{-\theta} \frac{T_m^j}{T_i^j}})$$
$$- \ln(\sum_j \alpha^j \sum_{n\neq i} \frac{\kappa_{ni,t}^{-\theta} Y_{n,t}}{\sum_{m\neq i} \kappa_{nm,t+1}^{-\theta} \frac{T_m^j}{T_i^j}} + \sum_j \sum_p \frac{\kappa_{pit}^{-\theta} NEXP_{p,t}^j}{\sum_{m\neq i} \kappa_{pm,t+1}^{-\theta} \frac{T_m^j}{T_i^j}}$$
$$+ \sum_j \sum_o \frac{\kappa_{oit}^{-\theta} INPUT_{o,t}^j}{\sum_{m\neq i} \kappa_{om,t+1}^{-\theta} \frac{T_m^j}{T_i^j}}), \quad (1.41)$$

where $INPUT_o^j$ is the value of materials in sector j demanded by producers in location o.

Table 1.24 shows the effects of negative competition considering demand from foreign countries and demand from other producers. To be able to add the market access and

[14] Because of data availability, I assume that only manufacturing producers demand materials from other sectors. It will not substantially affect the results since for primary sectors, agriculture and mining, the cost of input from other sectors is only a tiny portion of total cost.

competition in consumer markets, foreign markets, and other intermediate goods markets, I measure market size by the value of agricultural output and value added in manufacturing and mining and fixed at the 1870 level, since county-sector-level data in manufacturing and mining is available only in 1870. In general, results are robust. Competition effects are always substantial and negative. The significance level is sensitive in the sample of urbanized counties and cities because of the small sample size.

Table 1.25 shows the results for the role of comparative advantage. Table 1.26 shows the positive link between comparative advantage in agriculture and economic growth. All results are consistent with the ones that don't consider demand from other producers and world markets.

1.D.2 Other Value of θ

Data availability restrains me from accurately estimating the value of θ across sectors. In the baseline estimation, I focus on the weighted average value in (CP15) of 7.35 to discuss the competition in all sectors. I also use the value in (DH16) of 8.22 to discuss the competition effects in agriculture and the weighted average of manufacturing sectors in (CP15) of 6.10 to discuss the competition effects in manufacturing. In this section, I check other values.

An abundance of literature estimates the value of θ. In addition to (CP15) and (DH16), (EK02) estimates the value as 8.28. (CDK12) estimates the value across all sectors as 6.53. (Don10) estimates the average value across 13 agricultural categories as 3.80. (SW14)'s estimation is 4.10. (HM14) review all estimates in the literature and find the mean value to be 6.74 and the median value to be 5.03. All the above estimates fall into the range between the two extreme estimates in (EK02), 3.60 and 12.86.

Table 1.27 shows results of several possible θ values. In addition to the values in the baseline (6.10, 7.35, 8.22) and the two extreme values (3.60 and 12.86) in (EK02), I also discuss the value in (HM14), whose mean value is 4.10 and median value is 5.03. In general, within the wide range of 3.60 and 12.86, competition effects are still negative and substantial. For some extreme values, the negative effects are not significant. To make the results clearer,

Table 1.24: Effects of Negative Competition, Considering International Trade Demand and Demand from Other Producers

	(1) Changes in Log Total Population	(2) Changes in Log Farm Land Values	(3) Changes in Log Value of Farm Output	(4) Changes in Log Manufacturing Labor	(5) Changes in Log Manufacturing Value Added	(6) Changes in Log Urban Population	(7) Changes in Log Urban Population (NE, MW)	(8) Changes in Log City Size	(9) Changes in Log City Size (NE, MW)
Panel A: Baseline									
Changes in log market access	0.572^{***}	0.840^{***}	0.694^{***}	0.369^{***}	0.396^{***}	0.692^{***}	0.545^{*}	0.777	-0.324
	(0.113)	(0.120)	(0.128)	(0.133)	(0.138)	(0.199)	(0.272)	(0.506)	(0.647)
Changes in competition	-0.711^{***}	-1.002^{***}	-0.803^{***}	-0.799^{***}	-0.754	-1.820^{**}	-7.033^{**}	-2.839	-11.39
	(0.192)	(0.161)	(0.216)	(0.259)	(0.482)	(0.793)	(2.879)	(1.706)	(6.599)
Panel B: Consider International Trade									
Changes in log market access	0.549^{***}	0.807^{***}	0.674^{***}	0.374^{***}	0.402^{***}	0.675^{***}	0.524^{*}	0.832^{*}	-0.360
	(0.114)	(0.123)	(0.127)	(0.133)	(0.139)	(0.199)	(0.263)	(0.462)	(0.658)
Changes in competition	-0.663^{***}	-0.935^{***}	-0.770^{***}	-0.812^{***}	-0.773	-1.965^{**}	-6.794^{**}	-3.554^{*}	-10.21^{**}
	(0.195)	(0.150)	(0.208)	(0.257)	(0.482)	(0.900)	(2.757)	(1.813)	(4.558)
Panel C: Consider Both International Trade and Input Demand in Manufacturing Production									
Changes in log market Access	0.522^{***}	0.798^{***}	0.665^{***}	0.373^{***}	0.398^{***}	0.636^{***}	0.409	0.747^{**}	-0.505
	(0.112)	(0.123)	(0.126)	(0.135)	(0.137)	(0.179)	(0.262)	(0.334)	(0.635)
Changes in competition	-0.581^{***}	-0.924^{***}	-0.758^{***}	-0.877^{***}	-0.822^{*}	-2.050^{**}	-3.977	-3.841^{**}	-7.701
	(0.196)	(0.149)	(0.206)	(0.247)	(0.445)	(0.866)	(2.343)	(1.576)	(4.463)
Value of θ	7.35	8.22	8.22	6.10	6.10	7.35	6.10	7.35	6.10
Competition Is Measured in	All Sectors	Agriculture	Agriculture	Manufacturing	Manufacturing	All Sectors	Manufacturing	All Sectors	Manufacturing
N	2327	2327	2327	1775	1775	456	350	162	130

Notes: Market size is measured by value of agricultural output and value added in manufacturing and mining and fixed at the 1870 level. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.25: Role of Comparative Advantage, Considering International Trade and Demand from Other Producers

	(1)	(2)
	Changes in Log Farm Land Values	Changes in Log Value of Farm Output
Panel A: Baseline		
Changes in competition in agriculture	-2.325***	-2.239***
	(0.459)	(0.521)
Changes in competition in all sectors	1.918***	1.977***
	(0.536)	(0.524)
Panel B: Consider International Trade		
Changes in competition in agriculture	-1.640***	-1.612***
	(0.316)	(0.417)
Changes in competition in all sectors	1.064***	1.089**
	(0.335)	(0.438)
Panel C: Consider Both International Trade and Input Demand in Manufacturing Production		
Changes in competition in agriculture	-1.450***	-1.387***
	(0.298)	(0.373)
Changes in competition in all sectors	0.877**	0.813*
	(0.349)	(0.423)
Value of θ in Competition in Agriculture	8.22	8.22
Value of θ in Competition in All Sectors	7.35	7.35
N	2327	2327

Notes: Market size is measured by value of agricultural output and value added in manufacturing and mining and fixed at the 1870 level. Changes in log market access with θ=8.22 and θ=7.35 are controlled. Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust

Table 1.26: Link between Comparative Advantage in Agriculture and Economic Growth, Considering International Trade and Demand from Other Producers

	(1)	(2)
	Changes in Log Total Population	Changes in Log Value of Agricultural Output Plus Manufacturing Value Added
Panel A: Baseline		
Changes in competition in agriculture	-1.491***	-2.047***
	(0.337)	(0.538)
Changes in competition in all sectors	1.065***	1.561**
	(0.369)	(0.609)
Panel B: Consider International Trade		
Changes in competition in agriculture	-1.046***	-1.462***
	(0.262)	(0.452)
Changes in competition in all sectors	0.373	0.679
	(0.287)	(0.581)
Panel C: Consider Both International Trade and Input Demand in Manufacturing Production		
Changes in competition in agriculture	-0.979***	-1.287***
	(0.255)	(0.391)
Changes in competition in all sectors	0.346	0.447
	(0.318)	(0.526)
Value of θ in Competition in Agriculture	8.22	8.22
Value of θ in Competition in All Sectors	7.35	7.35
N	2327	2327

Notes: Market size is measured by value of agricultural output and value added in manufacturing and mining and fixed at the 1870 level. Changes in log market access with θ=8.22 and θ=7.35 are controlled. Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.27: Results of Several Possible θ Values

	(1) Changes in Log Total Population	(2) Changes in Log Farm Land Values	(3) Changes in Log Value of Farm Output	(4) Changes in Log Manufacturing Labor	(5) Changes in Log Manufacturing Value Added	(6) Changes in Log Urban Population	(7) Changes in Log Urban Population (NE, MW)	(8) Changes in Log City Size	(9) Changes in Log City Size (NE, MW)
Panel A: Eaton and Kortum (2002) Lower Bound (θ=3.60)									
Changes in competition	-0.751	-4.083***	-2.635***	-0.166	-0.117	-5.261***	-7.119**	-7.270**	-15.81**
	(0.860)	(1.031)	(0.879)	(1.088)	(1.434)	(1.521)	(2.676)	(2.996)	(6.001)
Panel B: Simonovska and Waugh (2014) (θ=4.10)									
Changes in competition	-1.122***	-2.773***	-1.789***	-0.612	-0.302	-4.962***	-7.439***	-6.723**	-16.78**
	(0.316)	(0.703)	(0.633)	(0.585)	(0.976)	(1.397)	(2.582)	(2.739)	(6.329)
Panel C: Median Value in Head and Mayer (2014) (θ=5.03)									
Changes in competition	-1.026***	-1.702***	-1.160**	-0.798**	-0.574	-4.468***	-7.679***	-5.856**	-16.34**
	(0.236)	(0.398)	(0.447)	(0.391)	(0.714)	(1.167)	(2.448)	(2.332)	(5.855)
Panel D: Weighted Average in Manufacturing in Caliendo and Parro (2015) (θ=6.10)									
Changes in competition	-0.950***	-1.319***	-0.999***	-0.905***	-0.823	-3.107***	-8.091***	-4.401**	-16.18***
	(0.222)	(0.264)	(0.319)	(0.325)	(0.589)	(0.934)	(2.357)	(1.814)	(5.374)
Panel E: Baseline, Weighted Average in Caliendo and Parro (2015) (θ=7.35)									
Changes in competition	-0.909***	-1.093***	-0.897***	-0.817***	-0.815**	-1.494**	-7.812***	-2.437*	-13.54***
	(0.176)	(0.200)	(0.223)	(0.148)	(0.312)	(0.652)	(2.269)	(1.202)	(4.147)
Panel F: Agriculture in Donaldson and Hornbeck (2016) (θ=8.22)									
Changes in competition	-0.862***	-0.970***	-0.817***	-0.692***	-0.701***	-0.969*	-8.095***	-1.632*	-13.37***
	(0.149)	(0.176)	(0.195)	(0.0981)	(0.211)	(0.482)	(2.236)	(0.897)	(3.812)
Panel G: Eaton and Kortum (2002) Upper Bound (θ=12.86)									
Changes in competition	-0.574***	-0.604***	-0.524***	-0.338***	-0.370***	-0.237	-9.155***	-0.516	-11.63***
	(0.0683)	(0.126)	(0.102)	(0.0488)	(0.0658)	(0.218)	(1.909)	(0.337)	(2.197)
N	2327	2327	2327	1775	1775	456	350	162	130

Notes: Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I graph the results in Figure 1.7.

Column 1 of Table 1.27 shows the effects of competition in all sectors on total population. The effects are negative and significant for all values except for the extreme lower bound 3.60, which the negative effects of competition are not significant. From 4.10 to 12.86, the absolute value of negative effects reduces, and the estimation becomes more precise as the value of θ increases.

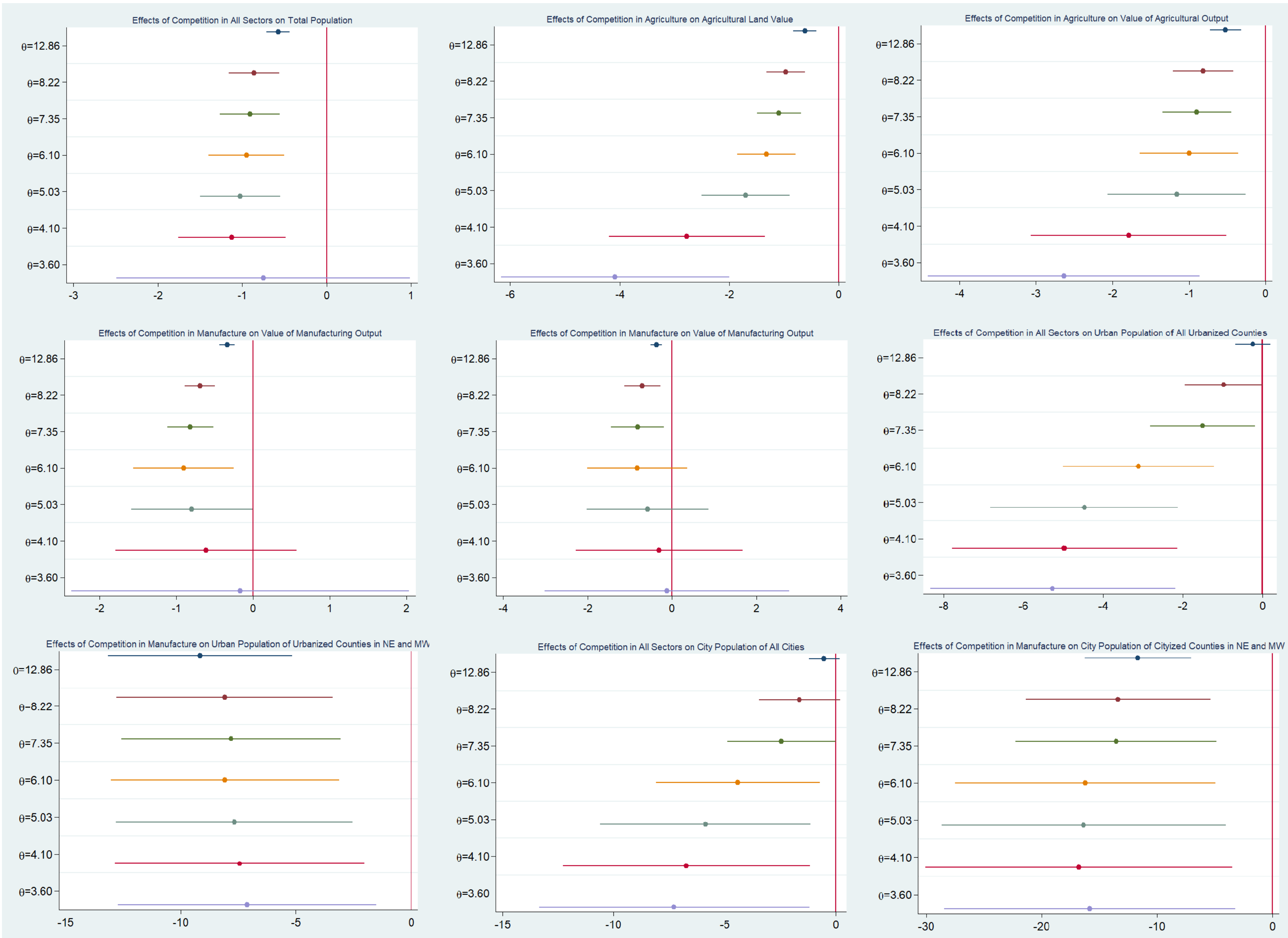
Column 2 of Table 1.27 shows the effects of competition in agriculture on agricultural land values. The effects are negative and significant for all values of θ. From 3.60 to 12.86, the absolute value of negative effects of competition reduces, and the estimation becomes more precise as the value of θ increases. Column 3 of Table 1.27 shows the effects of competition in agriculture on the value of agricultural output. The trend is similar to the one of competition effects on agricultural land values.

Column 4 of Table 1.27 shows the effects of competition in manufacturing on workers hired in manufacturing. Competition effects are negative for all values of θ. The effects are not significant for $\theta = 3.60$ or $\theta = 4.10$. The absolute value of negative effects of competition increases from 3.60 to 6.10 and then decreases from 6.10 to 12.86. As the value of θ increases, the estimation becomes more precise. Column 5 shows the effects of competition in manufacturing on manufacturing value added. The trend is similar to the one of competition effects on manufacturing workers. The negative effects are not significant for θ =3.60, 4.10, 5.03, or 6.10.

Column 6 of Table 1.27 shows the effects of competition in all sectors on the urban population of all urbanized counties. The effects are negative for all values of θ discussed, and significant for all values except for the upper-bound extreme value $\theta = 12.86$. Similar to the trend of competition effects on total population and agriculture-related economic outcomes, the absolute value of negative effects reduces, and the estimation becomes more precise as the value of θ increases. The effects of competition in all sectors on city size for all cities larger than 10,000 in 1870 show a similar trend (column 8).

Column 7 of Table 1.27 shows the effects of competition in manufacturing on the urban

Figure 1.7: Results of Several Possible θ Values



population of urbanized counties in the two most industrialized regions, the Northeast and the Midwest. The effects are negative, substantial, and significant for all values of θ discussed here. There is no substantial variation in values of θ. A 1% increase in competition in manufacturing leads to a 7.12%–9.16% decrease in urban population of urbanized counties in the Northeast and the Midwest. The estimation is not so precise for all values of θ, since the sample size is limited for this case. Column 9 shows the same effects on city size of all cities larger than 10,000 in 1870 in the Northeast and the Midwest. The effects are also negative, substantial, and significant for all values of θ. A 1% increase in competition in manufacturing leads to an 11.63%–16.78% decrease in city size. The estimations are not precise.

Table 1.28 shows the results for the role of comparative advantage. Table 1.29 shows the positive link between comparative advantage in agriculture and economic growth. All results are consistent with results in the body of the paper.

Table 1.28: The Role of Comparative Advantage, Different Value of θ

	(1)	(2)
	Changes in Log Farm Land Value	Changes in Log Value of Farm Output
Panel A: Eaton and Kortum (2002) Lower Bound (θ=3.60)		
Changes in competition in agriculture	-10.11***	-6.487**
	(2.419)	(2.580)
Changes in competition in all sectors	6.910***	4.414*
	(2.373)	(2.497)
Panel B: Simonovska and Waugh (2014) (θ=4.10)		
Changes in competition in agriculture	-8.411***	-5.528**
	(2.088)	(2.244)
Changes in competition in all sectors	6.231***	4.131*
	(2.086)	(2.188)
Panel C: Median Value in Head and Mayer (2014) (θ=5.03)		
Changes in competition in agriculture	-6.221***	-4.562***
	(1.623)	(1.529)
Changes in competition in all sectors	4.867***	3.664**
	(1.566)	(1.517)
Panel D: Weighted Average in Manufacturing in Caliendo and Parro (2015) (θ=6.10)		
Changes in competition in agriculture	-4.275***	-3.622***
	(1.172)	(1.068)
Changes in competition in all sectors	3.158***	2.802**
	(1.069)	(1.069)
Panel E: Baseline, Weighted Average in Caliendo and Parro (2015) (θ=7.35)		
Changes in competition in agriculture	-2.714***	-2.443***
	(0.672)	(0.704)
Changes in competition in all sectors	1.760***	1.678**
	(0.601)	(0.685)
Panel F: Agriculture in Donaldson and Hornbeck (2016) (θ=8.22)		
Changes in competition in agriculture	-2.306***	-2.036***
	(0.541)	(0.585)
Changes in competition in all sectors	1.459***	1.331**
	(0.499)	(0.571)
Panel G: Eaton and Kortum (2002) Upper Bound (θ=12.86)		
Changes in competition in agriculture	-1.479***	-1.268***
	(0.328)	(0.355)
Changes in competition in all Sectors	0.904**	0.783**
	(0.355)	(0.344)
N	2327	2327

Notes: Changes in log market access are controlled. Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.29: Link between Comparative Advantage in Agriculture and Economic Growth, Different Value of θ

	(1)	(2)
	Changes in Log Total Population	Changes in Log Value of Agricultural Output Plus Manufacturing Value Added
Panel A: Eaton and Kortum (2002) Lower Bound (θ=3.60)		
Changes in competition in agriculture	-6.574***	-5.371**
	(1.664)	(2.618)
Changes in competition in all sectors	3.563**	2.277
	(1.706)	(2.375)
Panel B: Simonovska and Waugh (2014) (θ=4.10)		
Changes in competition in agriculture	-5.235***	-4.087*
	(1.474)	(2.228)
Changes in competition in all sectors	3.370**	2.284
	(1.521)	(2.076)
Panel C: Median Value in Head and Mayer (2014) (θ=5.03)		
Changes in competition in agriculture	-3.996***	-3.415**
	(1.094)	(1.459)
Changes in competition in all sectors	2.864**	2.355*
	(1.097)	(1.399)
Panel D: Weighted Average in Manufacturing in Caliendo and Parro (2015) (θ=6.10)		
Changes in competition in agriculture	-2.922***	-3.020***
	(0.734)	(1.007)
Changes in competition in all sectors	1.942***	2.081*
	(0.691)	(1.037)
Panel E: Baseline, Weighted Average in Caliendo and Parro (2015) (θ=7.35)		
Changes in competition in agriculture	-1.819***	-2.143***
	(0.393)	(0.711)
Changes in competition in all sectors	0.919**	1.263*
	(0.366)	(0.729)
Panel F: Agriculture in Donaldson and Hornbeck (2016) (θ=8.22)		
Changes in competition in agriculture	-1.439***	-1.779***
	(0.301)	(0.583)
Changes in competition in all sectors	0.602*	0.964
	(0.299)	(0.592)
Panel G: Eaton and Kortum (2002) upper bound (θ=12.86)		
Changes in competition in agriculture	-0.616***	-1.141***
	(0.194)	(0.311)
Changes in competition in all sectors	0.0467	0.599**
	(0.173)	(0.287)
N	2327	2327

Notes: Changes in log market access are controlled. Market size is measured by population and allowed to change between 1870 and 1890. State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.E Model Comparison

In the the body of the paper, I proved that structural competition has negative effects on economic growth. In this appendix, I further show that adding structural competition can increase the model's explanatory power. I compare my specification, which considers both changes in market access and changes in competition, with regressions derived from the single-sector Eaton-Kortum model in (DH16)(Equation 1.42), as follows:

$$\Delta \ln y_i = \delta + \eta_1 \Delta \ln(MA_i) + \delta_s + f(x_i, y_i) + \varepsilon_i, \tag{1.42}$$

where the term of market access can be defined in two ways: the less model-dependent version, market access without competition, $MA_i = \sum_{n \neq i} \kappa_{ni}^{-\theta} L_n$; and the more model-dependent version, market access with competition that does not contain any information about the industry structure of any locations, $MA_i = \sum_n \kappa_{ni}^{-\theta} MA_n^{-\frac{1+\theta}{\theta}} L_n$.

Here is an interesting question: Is the empirical model derived from the multisector EK model better than the model derived from the single-sector EK model in previous literature? Particularly, due to data availability and simplification, I exclude factor price from the measurement of changes in competition. Thus, the empirical model I discuss in this paper is not exactly theory-dependent and does not depend on running the data through the particular structure of the model. Is the empirical model in this work still more powerful than the one in (DH16) in explaining the 1870–1890 data? In the following subsections, I compare the two empirical models by several model-selection strategies: goodness-of-fit(adjusted R^2), information criteria (AIC, BIC), and non-nested model tests (encompassing approach, comprehensive approach). Goodness-of-fit and information criteria help us to know which model suffers less information loss. Non-nested model tests help us to know whether one model provides all the information that the other model can provide, not the other way around. In other words, one model is the true model, while the other is false model.

I focus on two economic outcomes discussed in (DH16): population and agricultural land values. I focus on equation 1.26 in discussing population and equation 1.28 in discussing agricultural land value.

1.E.1 Goodness-of-Fit and Information Criteria

Since the models differ in degree of freedom, Adjusted R^2, Akaike information criterion (AIC), and Bayesian information criterion (BIC) are better choices than R^2 for model selection.

Panel A of Table 1.8 displays the value of adjusted R^2, AIC, and BIC of my model and the model in (DH16). For both population and agricultural land values, my model has higher adjusted R^2 and lower AIC and BIC.

1.E.2 Non-Nested Model Tests: Encompassing Approach

(MR86) introduces what they called the "encompassing approach" to test the ability of one model to explain the features of another model. I use this approach to compare two non-nested empirical models that include information about competition: mine and the one discussed in (DH16). For population, the encompassing model is

$$\begin{aligned} \Delta \ln Population_i =& \delta + \eta_1 \Delta \ln(MAsimple_i) + \eta_2 \Delta Competition_i + \eta_3 \Delta \ln(MAcomplex_i) \\ &+ \delta_s + f(x_i, y_i) + \varepsilon_i, \end{aligned} \tag{1.43}$$

where $MAsimple_i = \sum_{n \neq i} \kappa_{ni}^{-\theta} L_n$ and $MAcomplex_i = \sum_n \kappa_{ni}^{-\theta} MA_n^{-\frac{1+\theta}{\theta}} L_n$. I test separately the hypotheses (1) $\eta_1 = \eta_2 = 0$ and (2) $\eta_3 = 0$. If I can reject hypothesis (1) but cannot reject hypothesis (2), then I can conclude that my model can explain all the features of the model in (DH16) but not the reverse. If both hypotheses can be rejected or neither hypothesis can be rejected, then I cannot conclude that one model can explain all the features of the other model.

Similarly, for agricultural land values, the encompassing model is:

Figure 1.8: Model Comparison: MA vs. MA+Competition

Model Comparison						
Dependent Variable	Changes in Log Population			Change in Log Farm Land Value		
	(1) This Paper	(2) Donaldson and Hornbeck (2016) Simple MA	(3) Donaldson and Hornbeck (2016) Complex MA	(4) This Paper	(5) Donaldson and Hornbeck (2016) Simple MA	(6) Donaldson and Hornbeck (2016) Complex MA
Panel A: Goodness-of-Fit and Information Criteria						
Adjusted R^2	0.5509	0.5166	0.5484	0.6345	0.6018	0.6179
AIC	3891.08	4061.505	3903.175	5628.555	5824.862	5728.973
BIC	3937.098	4101.771	3943.441	5686.078	5865.128	5769.239
Panel B: Encompassing						
F statistic	10.06***		8.26***	3.37**		0.32
Panel C: Comprehensive						
J statistic	2.76***		2.87***	3.56***		-0.57
Cox statistic	-3.88***		-7.77***	-20.06***		1.63*

$$
\begin{aligned}
\Delta \ln LandValue_i^{Agriculture} =& \delta + \eta_1 \Delta \ln(MAsimple_i) + \eta_2 \Delta Competition_i^{Agriculture} \\
&+ \eta_3 \Delta Competition_i + \eta_4 \Delta \ln(MAcomplex_i) \\
&+ \delta_s + f(x_i, y_i) + \varepsilon_i,
\end{aligned}
$$

where I separately test the hypotheses (1) $\eta_1 = \eta_2 = \eta_3 = 0$ and (2) $\eta_4 = 0$.

Panel B of Table 1.8 shows the results of the encompassing approach. For agricultural land values, hypothesis (1) is rejected, but I do not have enough evidence to reject hypothesis (2). Thus, my model is preferred. All features in the (DH16) model can be explained by my model, but the reverse is not true. For population, both hypothesis (1) and (2) are rejected. I do not have enough evidence to prove that my model captures all the features in the (DH16) model.

1.E.3 Non-Nested Model Tests: Comprehensive Approach

The encompassing approach has some problems. First, coefficients of variables other than $\Delta \ln MAsimple_i$, $\Delta \ln MAcomplex_i$, $\Delta Competition_i$, and $\Delta Competition_i^{Agriculture}$ remain a mixture of parts of η_1, η_2, and η_3, and the F test does not establish that any of these parts is zero. Second, since $\Delta \ln MAsimple_i$, $\Delta \ln MAcomplex_i$, $\Delta Competition_i$, and $\Delta Competition_i^{Agriculture}$ are correlated, the collinearity may be a problem ((Gre03)).

To overcome these problems, (DM81) propose the J-test. The idea of the J-test is that if one model is true, then the fitted value from the other model, when added to the model, should be insignificant, and vice versa.

Cox (1961, 1962) argues that the underlying requirements of J-test are not met in non-nested models. (Cox61), (Cox62) (Pes74), and (PD78) propose a reformulated test statistic, the Cox statistic.

Panel C of Table 1.8 reports the results of the J-test and the Cox-Pesaran test. The results are similar to the ones in the encompassing approach. For agricultural land values, the J-test shows that my model is the true model. The Cox-test does not prefer my model, but my model is preferred with a high absolute value of Cox-statistic and a high significance level, while the (DH16) model is preferred with a low absolute value of Cox-statistic and a low significance level. For population, we cannot tell which model is preferred.

In summary, in terms of agricultural land values, my model can explain all the information explained by the (DH16) model, but not vice versa. In terms of population, we cannot tell which model is preferred, but my model loses less information.

1.E.4 Counterfactual Model Comparison

I also compare models used in counterfactual analysis. Table 1.30 shows that the model derived from the multisector EK model not only has information loss (higher adjusted R^2, lower AIC and BIC in Panel A) but also explains more than just the information explained by the model without consideration of competition.

Table 1.30: Model Comparison between This Paper and (DH16) in Counterfactual Analysis

	(1) Estimated by This Paper	(2) Estimated by Donaldson and Hornbeck (2016)
Panel A: Goodness-of-Fit and Information Criteria		
Adjusted R^2 (unweighted)	0.6217	0.5981
Adjusted R^2 (weighted)	0.6327	0.6169
AIC (unweighted)	5709.924	5846.353
AIC (weighted)	1299.411	1395.273
BIC (unweighted)	5759.695	5886.619
BIC (weighted)	1351.182	1435.539
Panel B: Encompassing		
F statistic (unweighted)	9.09***	1.75
F statistic (weighted)	31.33***	0.91
Panel C: Comprehensive		
J statistic (unweighted)	5.43***	1.32
J statistic (weighted)	8.27***	0.96
Cox statistic (unweighted)	-66.62***	-3.52***
Cox statistic (weighted)	-19.10***	-227.77***

1.F An Extended Model: Value of θ Varies across Sectors

1.F.1 Theoretical Settings

I now discuss an extended model that allows trade elasticity to vary across agriculture, manufacturing, and mining. Following (CP15), I assume that all subsectors have the same θ ($\theta^j = \theta^A$, $\forall j \in Agriculture$). Besides, for simplicity, I assume that producers use both land and labor to produce agricultural goods, but use only labor to produce nonagricultural goods.[15] This assumption does not deviate far from the truth. (CI01) found that the input share of land in agriculture is 0.19, while the input share of land in nonagricultural sectors is only 0.06. In this section, I focus on the linear expression of agricultural outcomes (total agricultural land values and value of agricultural output).[16]

Similar to the trade share in the baseline model (Equation 1.8), trade share in the agricultural subsector can be expressed as

$$\pi_{ni}^{j} = \frac{X_{ni}^{j}}{X_{n}^{j}} = \frac{(r_i^{\beta^A} w_i^{1-\beta^A} \kappa_{ni})^{-\theta^A} T_i^j}{\sum_m (r_m^{\beta^A} w_m^{1-\beta} \kappa_{nm})^{-\theta^A} T_m^j}, \forall j \in Agriculture. \tag{1.44}$$

Consumer market access of sector j ($j \in Agriculture$) in location i can be expressed as

$$CMA_i^j \equiv (P_i^j)^{-\theta^A} = \delta_{A2} \sum_n (r_n^{\beta^A} w_n^{1-\beta^A})^{-\theta^A} \kappa_{ni}^{-\theta^A} T_n^j. \tag{1.45}$$

Consumer market access of sector j ($j \in nonagriculturalsectors$) in location i can be

[15](DH16) discuss a similar model with two sectors (agriculture and manufacturing) and input-output linkage. In their model setting, they assume that producers use labor and land to produce agricultural goods and manufacturing goods. Based on the model, they derive a reduced-form expression of log agricultural land rental rate as a linear function of log total lands, log consumer market access in agriculture, log consumer market access in manufacturing, log firm market access in agriculture, and log input market access in agriculture from manufacturing. However, they make a mistake that they can only derive the linear expression of log agricultural land rental rate by assuming that all lands are used in agricultural production and that manufacturing production does not use land at all. Besides, the key difference between my model and their model is that I assume that the deterministic productivity varies across sectors as well as locations (T_i^j). This assumption helps us discuss the effects of structural competition.

[16]It is impossible to write log economic outcomes in nonagricultural sectors with the assumption that value of θ varies across subsectors.

expressed as

$$CMA_i^j \equiv (P_i^j)^{-\theta^j} = \delta_{j2} \sum_n (w_n)^{-\theta^j} \kappa_{ni}^{-\theta^j} T_n^j. \tag{1.46}$$

Price index P_i in location i can be expressed as

$$P_i = \prod_j (\frac{P_i^j}{\alpha^j})^{\alpha^j} = \prod_j (\frac{(CMA_i^j)^{-\frac{1}{\theta^j}}}{\alpha^j})^{\alpha^j}. \tag{1.47}$$

Aggregate consumer market access in location i is defined as

$$CMA_i \equiv \prod_j (CMA_i^j)^{\frac{\alpha^j}{\theta^j}}. \tag{1.48}$$

Firm market access of sector j ($j \in Agriculture$) in location i can be expressed as

$$FMA_i^j \equiv T_i^j \sum_n \kappa_{ni}^{-\theta^A} (CMA_n^j)^{-1} X_n^j, \forall j \in Agriculture, \tag{1.49}$$

where X_i^j is the total demand in sector j, location n. If we assume that nonagricultural production does not require inputs produced by agriculture, $X_i^j = \alpha^j Y_n$, then all demands come from consumers. If we assume that nonagricultural production uses both labor and materials produced by agriculture, $X_i^j = \alpha^j Y_n + INPUT_n^j$, where $INPUT_n^j$ is demand for goods in sector j from nonagricultural producers in location n.

Aggregate firm market access in agriculture in location i is defined as

$$FMA_i^A = \sum_{j \in Agriculture} FMA_i^j. \tag{1.50}$$

Total agricultural land rental rental rate in location i can be expressed as

$$(1 + \beta^A \theta^A) \ln(r_i H_i) = \delta_8 + \beta^A \theta^A \ln H_i + (1 - \beta^A)\theta^A \ln(CMA_i) + (1 + \beta^A \theta^A) \ln FMA_i^A \tag{1.51}$$

Total value of agricultural output in location i can be expressed as

$$(1 + \beta^A \theta^A) \ln Y_i^A = \delta_9 + \beta^A \theta^A \ln H_i + (1 - \beta^A)\theta^A \ln(CMA_i) + \ln FMA_i^A. \tag{1.52}$$

1.F.2 From Theory to Empirical Works

Similar to the application of the baseline model, location-sector-specific consumer market access is approximately measured by Equation 1.18. Aggregate consumer market access is calculated by Equation 1.48. Aggregate firm market access in agriculture is measured as

$$FMA_{it}^{A} \approx \sum_{j \in Agriculture} \alpha^{j} \sum_{n \neq i} \frac{\kappa_{nit}^{-\theta^{A}} Y_{nt}}{\sum_{m \neq i} \kappa_{nmt}^{-\theta^{A}} \frac{T_{m}^{j}}{T_{i}^{j}}}. \tag{1.53}$$

Changes in log aggregate firm market access in agriculture ΔFMA_{i}^{A} can be decomposed into changes in competition and changes in log market access, as follows:

$$\Delta Competition_{i}^{A} = \ln\Big(\sum_{j \in Agri} \alpha^{j} \sum_{n \neq i} \frac{\kappa_{ni,t}^{-\theta^{A}} Y_{n,t}}{\sum_{m \neq i} \kappa_{nm,t}^{-\theta^{A}} \frac{T_{m}^{j}}{T_{i}^{j}}}\Big) - \ln\Big(\sum_{j \in Agriculture} \alpha^{j} \sum_{n \neq i} \frac{\kappa_{ni,t}^{-\theta^{A}} Y_{n,t+1}}{\sum_{m \neq i} \kappa_{nm,t+1}^{-\theta^{A}} \frac{T_{m}^{j}}{T_{i}^{j}}}\Big) \tag{1.54}$$

$$\Delta \ln MA_{i} = \ln\Big(\sum_{n} \kappa_{ni,t+1}^{-\theta^{A}} Y_{n,t+1}\Big) - \ln\Big(\sum_{n} \kappa_{ni,t}^{-\theta^{A}} Y_{n,t}\Big). \tag{1.55}$$

I get the value of trade elasticity $\theta^{A} = 8.22$ from (DH16). (CP15) provide the estimation of θ^{j}, $\forall j \in ManufacureandMining$ (see Table 1.20).

1.F.3 Baseline Regression and Results

The following specification describes the estimation strategy:

$$\Delta \ln y_{i}^{A} = \delta + +\eta_{1} \Delta \ln(CMA_{i}) + \eta_{2} \Delta \ln(MA_{i}) + \eta_{3} \Delta Competition_{i}^{A} + \delta_{s} + f(x_{i}, y_{i}) + \epsilon_{i} \tag{1.56}$$

where y_{i}^{A} is total agricultural land values and value of agricultural output.

Table 1.31 shows the results. Because $\Delta \ln CMA_{i}$ and $\Delta \ln MA_{i}$ are highly correlated (the value correlation coefficient is higher than 0.99 for all specifications), the coefficients of the two variables are not as predicted by theoretical model. The competition effects are negative and significant for all for columns and two dependent variables.

Table 1.31: Effects of Competition on Agricultural Outcomes

	(1)	(2)	(3)	(4)
Panel A	Changes in Log Farm Land Values			
Changes in log consumer market access	-1.521	-11.70	-6.450	-16.02*
	(7.439)	(8.523)	(8.473)	(9.038)
Changes in log market access	0.985	2.173**	1.579	2.705**
	(0.855)	(0.968)	(0.970)	(1.036)
Changes in competition in agriculture	-0.950***	-0.977***	-0.816***	-0.964***
	(0.184)	(0.150)	(0.216)	(0.173)
Panel B	Changes in Log Value of Agricultural Output			
Changes in log consumer market access	4.113	-5.917	-0.141	-7.882
	(4.742)	(7.051)	(4.757)	(6.894)
Changes in log market access	0.209	1.383*	0.704	1.612**
	(0.497)	(0.788)	(0.501)	(0.772)
Changes in competition in agriculture	-0.870***	-0.869***	-0.697**	-0.784***
	(0.211)	(0.173)	(0.261)	(0.224)
Market Size Is Measured by Population	X		X	
Market Size Is Measured by Value of Agricultural Output and Value Added in Manufacturing and Mining		X		X
Allow Changes in Market Size	X	X		
Fix Market Size at the 1870 Level			X	X
N	2327	2327	2327	2327

Notes: State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.F.4 Counterfactual Analysis

In the counterfactual analysis, I do not decompose firm market access into market access and competition; instead, I focus on the original theoretical-model-derived specification, as follows:

$$\Delta \ln LandValue_i^A = \delta + \eta_1 \Delta \ln(CMA_i) + \eta_2 \Delta \ln(FMA_i^A) + \delta_s + f(x_i, y_i) + \varepsilon_i. \quad (1.57)$$

Table 1.32 shows the results. Column 1 shows the results of Equation 1.57 without any weights. Column 2 shows the results of Equation 1.57 weighted by agricultural land values in 1870. The weighted regression reflects the aggregate railway effects. Neither result violates the model implications that consumer market access and firm market access in agriculture causes total land values to rise.

Table 1.32: Effects of Consumer Market Access and Firm Market Access on Agricultural Land Values

	Changes in Log Farm Land Values	
	(1)	(2)
Changes in log consumer market access	1.408**	1.430*
	(0.642)	(0.719)
Changes in log firm market access of agriculture	0.612***	0.406***
	(0.107)	(0.103)
Weighted by Farm Land Value in 1870	No	Yes
N	2327	2327
R^2	0.626	0.627
adj. R^2	0.618	0.619

Notes: State fixed effects and cubic polynomials in county latitude and longitude are controlled. Robust standard errors clustered by state are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column 2 of Table 1.33 Panel A shows the percentage decline in land values in the counterfactual scenario without railways. The results are very similar to the estimation in the body of the paper. Under all assumptions of population distributions, the aggregate percentage decline in land values without railways is around 30%. It is half the estimation in (DH16). Panel B shows the decrease in agricultural land values by regions (assuming the

Table 1.33: Estimation of Aggregate Railway Effects on Agricultural Land Values Comparison between This Paper and (DH16) in Counterfactual Analysis

	(1) Estimated under Baseline Model	(2) Estimated under Appendix Model	(3) Estimated by Donaldson and Hornbeck (2016)
Panel A: Percentage Decline in Land Values without Railways			
Assuming the population distribution from 1890	23.1 (18.6)	28.3 (11.7)	60.2 (4.2)
Assuming the population distribution from 1870	19.3 (20.4)	29.6 (11.1)	59.1 (4.1)
Assuming the population distribution from 1850	26.7 (15.0)	32.5 (10.0)	59.3 (4.1)
Assuming the population distribution from 1830	39.8 (12.8)	34.3 (9.4)	60.1 (4.0)
Panel B: Decrease in Agricultural Land Values by Regions (Assuming the Population Distribution from 1890)			
Northeast	0.05 billion	0.1 billion	0.5 billion
South	0.2 billion	0.2 billion	0.5 billion
Midwest	1.4 billion	1.4 billion	2.5 billion
Plains	0.4 billion	0.4 billion	0.9 billion
Far West	-0.3 billion	0.02 billion	0.5 billion
Panel C: Percentage Decline in Land Values without Railways, with Improved Traditional Transportation Methods (Assuming the Population Distribution from 1890)			
With proposed canals	21.6 (15.9)	22.8 (10.2)	52.4 (4.2)
With improved country roads, wagons cost of 14 cents (40% cost reduction)	17.0 (13.3)	20.4 (9.0)	47.5 (3.9)
With both proposed canals and improved country roads	16.1 (11.1)	16.0 (7.6)	40.0 (3.7)

population distribution from 1890). When compared with results in the body of the paper, the decreased land values in northeast changes from 0.05 billion to 0.1 billion. The decreased land values in the Far West changes from -0.3 billion to 0.02 billion.

Results are very similar in other regions. Compared to the estimation with in (DH16) (column 3), land value losses are much lower in all regions. Panel C shows the results of land value losses in the counterfactual scenario without railways but with improved traditional transportation methods. The results are similar to those in the main body of the paper.

Bibliography

[Bat04] Thomas J. Weiss with Jeremy Atack Bateman, Fred. "STATE SAMPLES FROM THE CENSUS OF MANUFACTURING: 1850, 1860, AND 1870." https://doi.org/10.3886/ICPSR04071.v1, 2004.

[BK86] R. M. Baron and D. A Kenny. "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *Journal of Personality and Social Psychology*, **51**(6):1173–1182, 1986.

[CDK12] Arnaud Costinot, Dave Donaldson, and Ivana Komunjer. "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *Review of Economic Studies*, **79**(2):581–608, 2012.

[CF16] A. Kerem Cosar and Pablo D. Fajgelbaum. "Internal Geography, International Trade, and Regional Specialization." *American Economic Journal: Microeconomics*, **8**(1):24–56, February 2016.

[CI01] Francesco Caselli and Wilbur John Coleman II. "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation." *Journal of Political Economy*, **109**(3):584–616, 2001.

[Cox61] D. R. Cox. *Tests of Separate Families of Hypotheses*, volume 1 of *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, chapter 0, pp. 105–123. University of California Press, 1961.

[Cox62] D. R. Cox. "Further Results on Tests of Separate Families of Hypotheses." *Journal of the Royal Statistical Society. Series B (Methodological)*, **24**(2):406–424, 1962.

[CP15] Lorenzo Caliendo and Fernando Parro. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies*, **82**(1):1–44, 2015.

[CPR14] Lorenzo Caliendo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel G. Sarte. "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy." International Finance Discussion Papers 1119, Board of Governors of the Federal Reserve System (U.S.), August 2014.

[Cra04] Nicholas Crafts. "Social savings as a measure of the contribution of a new technology to economic growth." Economic History Working Papers 22554, London School of Economics and Political Science, Department of Economic History, July 2004.

[CT00] Amitabh Chandra and Eric Thompson. "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system." *Regional Science and Urban Economics*, **30**(4):457–490, July 2000.

[DH16] Dave Donaldson and Richard Hornbeck. "Railroads and American Economic Growth: A "Market Access" Approach *." *The Quarterly Journal of Economics*, **131**(2):799–858, 2016.

[DM81] Russell Davidson and James G MacKinnon. "Several Tests for Model Specification in the Presence of Alternative Hypotheses." *Econometrica*, **49**(3):781–793, May 1981.

[DOL87] DOL. "Standard Industrial Classification (SIC) System.", 1987.

[Don10] Dave Donaldson. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." NBER Working Papers 16487, National Bureau of Economic Research, Inc, October 2010.

[EK02] Jonathan Eaton and Samuel Kortum. "Technology, Geography, and Trade." *Econometrica*, **70**(5):1741–1779, September 2002.

[Fab13] Benjamin Faber. "Trade Integration, Market Size and Industrialization: Evidence from China's National Trunk Highway System." CEP Discussion Papers dp1244, Centre for Economic Performance, LSE, October 2013.

[Fog64] Robert Fogel. *Railroads and American Economic Growth: Essays in Econometric History*. Johns Hopkins Press, 1964.

[GLZ14] Julian di Giovanni, Andrei A. Levchenko, and Jing Zhang. "The Global Welfare Impact of China: Trade Integration and Technological Change." *American Economic Journal: Macroeconomics*, **6**(3):153–183, July 2014.

[Gre03] William H. Greene. *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, 5. edition, 2003.

[HM14] Keith Head and Thierry Mayer. *Gravity Equations: Workhorse,Toolkit, and Cookbook*, volume 4 of *Handbook of International Economics*, chapter 0, pp. 131–195. Elsevier, January 2014.

[HPR10] Michael R. Haines, Inter university Consortium for Political, and Social Research. "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002." https://doi.org/10.3886/ICPSR02896.v3, 2010.

[IIA12] IIASA/FAO. "Global Agro-ecological Zones (GAEZ v3.0)." http://gaez.fao.org/Main.html#, 2012.

[Mat92] Kiminori Matsuyama. "Agricultural productivity, comparative advantage, and economic growth." *Journal of Economic Theory*, **58**(2):317–334, December 1992.

[MR86] Grayham E Mizon and Jean-Francois Richard. "The Encompassing Principle and Its Application to Testing Non-nested Hypotheses." *Econometrica*, **54**(3):657–678, May 1986.

[PD78] M Pesaran and Angus Deaton. "Testing Non-Nested Nonlinear Regression Models." *Econometrica*, **46**(3):677–94, 1978.

[Pes74] M. H. Pesaran. "On the General Problem of Model Selection." *Review of Economic Studies*, **41**(2):153–171, 1974.

[Rob11] T.P et al. Robinson. "Global Livestock Production Systems.", 2011.

[SW14] Ina Simonovska and Michael E. Waugh. "Trade Models, Trade Elasticities, and the Gains from Trade." NBER Working Papers 20495, National Bureau of Economic Research, Inc, September 2014.

[Tre70] Treasury Department, Bureau of Statistics. "Annual Report of The Chief of The Bureau of Statistics on The Commerce and Navigation of The United States for The Fiscal Year Ended June 30, 1870.", 1870.

CHAPTER 2

The Boll Weevil Cannot Kill King Cotton: Productivity Shock Under A Regional Competition Framework 1889-1929

2.1 Introduction

How large are the effects of sector-specific exogenous shock on the distribution of production? This question is related to one of the central debates in applied economics about the effects of economic shocks. The standard approach in empirical studies treats each location as an isolated island and assumes that the production in one location is not affected by productivity shocks in another location. However, in an open economy, all locations are connected by a trade network and compete with other locations which produce similar products. As a result, once some locations experienced exogenous productivity shocks, other locations, which are free from negative productivity shocks, would face less competition and expand their production. Moreover, labor and capital mobility amplifies the network effects.

In this project, I derive several testable implications of a general-equilibrium trade model, which takes competition effects into account. This model is based on the multi-sector Eaton-Kortum model ((EK02))[1]. I then estimate the effects of the negative cotton productivity shock from the boll weevil plague on the cotton production in the early 20th century southern

[1] In the multi-sector Eaton-Kortum model, the producer market access of the certain crop j in location i not only positively correlates with the productivity of crop j in location i but also negatively correlates with the productivity of the same crop in other locations which produce crop j. Therefore, the crop-specific negative shock in affected locations will attenuate the competition faced by other locations. It thus provides us a good framework.

United States. My results provide a clear picture of the dynamic effects of the boll weevil and a perspective to explain why the boll weevil did not bring the agricultural revolution to the Cotton Belt.

The boll weevil caused large declines in cotton productivity since it arrived in Texas from Mexico in 1892. It gradually expanded to the whole Cotton Belt from then on and infested all the cotton-growing areas in the United States by the 1920s. At the same time, we witnessed the dynamic rise and fall of the cotton acreage in different regions of the Cotton Belt. Initially, the boll weevil infested the western Cotton Belt; we observed declining cotton acreage in many locations in the western Cotton Belt as well as increasing cotton acreage in the eastern Cotton Belt. Once the boll weevil expanded to the eastern Cotton Belt, the cotton acreage in the eastern Cotton Belt declined while the cotton acreage in the western Cotton Belt recovered. Since cotton was the most important crop in the southern United States, the rise and fall of cotton acreage correlated with a series of economic outcomes. Cotton production was closely related to the migration of population since cotton is a labor-intensive crop. Cotton production also correlated with the tenancy system in the southern United States.

My findings are consistent with the implications of the general-equilibrium trade model. I find that (1) when few areas in the Cotton Belt was infested by the boll weevil, the boll weevil had negative effects on the cotton acreage; (2) as more and more areas were infested by the boll weevil, the cotton acreage in the infested locations gradually recovered; (3) the expansion of the boll weevil would increased the cotton acreage in locations which were temporarily not infested by the boll weevil. The boll weevil had similar effects on all economic outcomes, which were closely associated with the cotton acreage (such as corn acreage, total farm acreage, agricultural land values, population, rural population, black population, number of farms, and the tenancy system).

One concern is that many other factors, which coincided with the expansion of the boll weevil, affected the redistribution of the cotton production in the same period. There might be alternative explanations for the hypothesized cause-effect relationship. For example, railway expansion, advantages in mechanization and soil conversation, and benefits from developed irrigation system also contribute to the recovery of the cotton production in the western Cotton Belt. The serious racial violence, which coincided with the boll weevil's worst year in the eastern Cotton Belt, is also believed to be one reason for the declining cotton acreage in the west Cotton Belt. Previous literature uses the distance from the location to the entry point of the boll weevil as an instrumental variable. However, this variable correlates with other factors which might lead to the redistribution of the cotton production. As a result, the instrumental strategy violates the exclusion restriction. To eliminate the effects of alternative explanations, I control a series of factors which might affect the redistribution of the cotton production.

The discussion of the boll weevil's effects on the cotton acreage, the crop choice, and the following economic consequences (agricultural labor arrangements, land value, migration, tenancy system, etc.) is not new. Early literature exploits state-level variation of boll weevil ((Str55); (Hig76); (Osb85); (Wri86); (RS01)) while the most recent studies move on to county-level data ((LOR09); (Bak15); (ABH17)). However, there is still room for more consideration, especially the effects of the boll weevil on the cotton acreage and the economic outcomes. Previous literature focuses on the effects of the boll weevil on the cotton yields. When they discuss the effects of the boll weevil on the cotton acreage, they do not consider the effects of competition in a trade network and do not control other factors which might affect the rise and fall of the cotton acreage. Consequently, there may be estimation biases.

Indeed, some of my findings are different from previous studies. First, I find the corn acreage changed with the cotton acreage instead of opposite to the cotton acreage. Corn acreage declined with the arrival of the boll weevil, but recovered as more and more lands were infested by the boll weevil. It is in line with (JD54)'s remark that corn was complementary to the cotton plantation. Second, diversification and land abandonment coexisted with the arrival of the boll weevil. When a county was infested by the boll weevil, but most lands in the Cotton Belt were still uninfested, total land acreage, as well as the share of cotton in all farmlands, decreased. As more and more lands were infested in the Cotton Belt, both total farm acreage and cotton share in all farmlands recovered. Third, contrary to the previous literature, I find that with the arrival of the boll weevil, the number of fixed-rent tenant farms did not significantly change, while the number of share tenant farms decreased. As the boll weevil infested more lands in the Cotton Belt, the number of share tenant recovered. As the boll weevil expanded, the number of share tenant also increased in non-infested locations. This finding shows that, as the bottom class in the tenancy system, share tenants were more vulnerable to the negative shocks and had higher incentives to move to locations with better economic opportunities.

This paper makes two main contributions. First, it provides a new framework to discuss the effects of the productivity shocks on the crop choice in the agriculture production when regions compete with each other. This framework can also be applied to discuss the effects of any other sector-specific productivity shocks on the change of industrial structure and the following economic outcomes. Second, in economic history, it is the first paper that controls confounding factors which might also affect the distribution of the cotton production in the Cotton Belt. It helps us better understand the effects of the expansion of the boll weevil on the dynamic shift of the cotton production and a series of associated economic outcomes in the early 20th century southern United States.

I organize this paper as follows. Section 2 documents the expansion of the boll weevil and the rise and fall of the cotton acreage in each county of the Cotton Belt. Section 3 is the theoretical framework and its implications. In section 4, I introduce the empirical strategies, potential problems in the previous literature, and data used for estimation. In section 5, I discuss how the presence and the expansion of the boll weevil affected the cotton acreage. In Section 6, I discuss how the presence of expansion of the boll weevil affected other economic outcomes which are closely associated with the cotton acreage. Section 7 concludes.

2.2 The Rise and Fall of the Cotton Acreage with the Expansion of the Boll Weevil

Historical observations are consistent with the regional competition framework. Figure 2.1 shows the progress of the boll weevil across the Cotton Belt in 1892-1924; and Figure 2.2 shows the ebbs and flows of cotton production during this period. During 1889-1899, there was little to no growth of the cotton production in most counties due to the declining cotton prices in the market. During 1899-1909, the boll weevil struck east Texas and Delta states (Arkansas, Louisiana, Mississippi), and cotton production shifted from these regions to the north-western Cotton Belt and eastern Cotton Belt. During 1909-1919, a period with sky-rocketing cotton prices, the boll weevil spread to the eastern states. Cotton production again shifted from the eastern states back to the west. Finally, when the boll weevil took over the entire Cotton Belt during 1919-1924, cotton production in Texas continued to grow.

(Gie04) indicates that when the boll weevil came, tenants moved out from the infested areas. Most moved tenants went to either West Texas where the dry weather limited boll weevil, or to the eastern and northern Cotton Belt ahead of the boll weevil throng. The story implies that in the first couple of years of the boll weevil expansion, cotton production in unaffected northwest Texas and the eastern and northern Cotton Belt benefited from tenants outmigration from the southeastern Cotton Belt. The dynamic redistribution of the cotton production in the sample period reflects the observation in (Gie04). Once the boll weevil spread to the eastern states, the western states regained its competitive edge against the eastern states. Coupled with increasing cotton market prices, cotton production in the western Cotton Belt recovered and prospered again.

2.3 Theoretical Framework

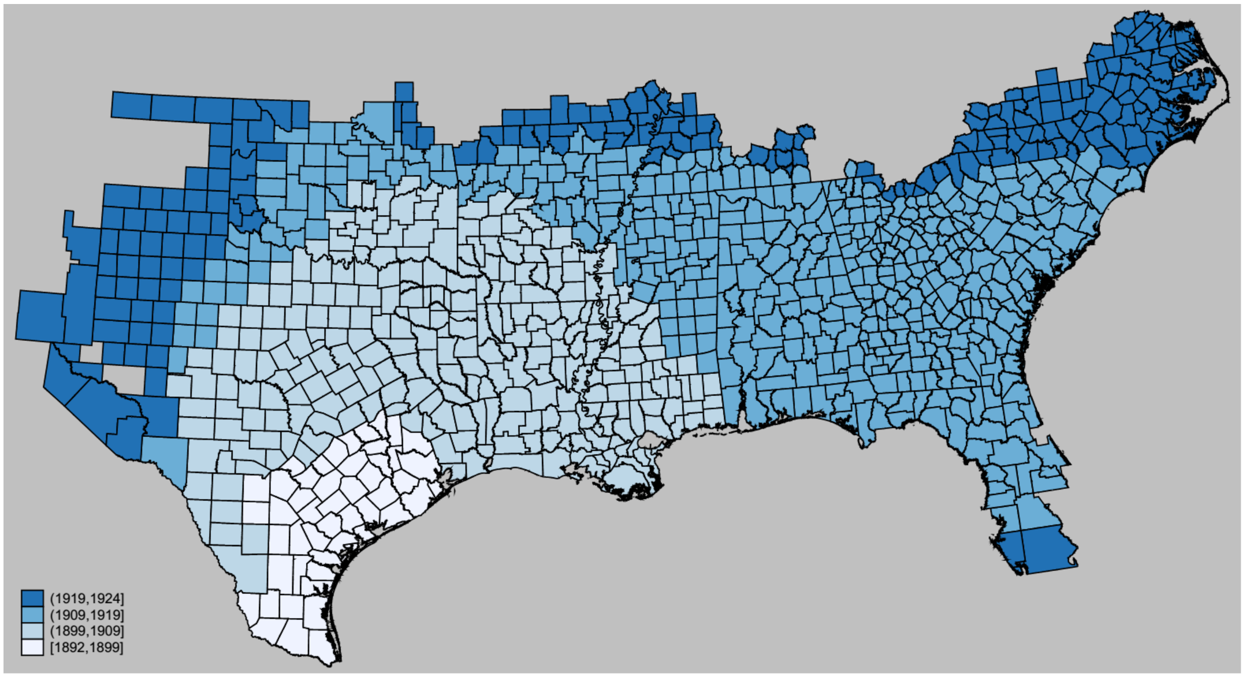
Following multi-sector Eaton-Kortum model in (CPR14), I apply a modified version to my context. I divide agriculture into several crop-based sub-sectors. Each sub-sector of agriculture has its own location-sector specific productivity.

Consumer preference has the Cobb-Douglass structure, consumption share of each sub-sector j is α^j. The utility of the consumer who lives in location i is expressed as the following:

$$U_n = \prod_j (c_i^j)^{\alpha^j}, \tag{2.1}$$

in which c_i^j is the amount of goods of sub-sector j consumed by consumers in location i and

Figure 2.1: The Boll Weevil Expansion 1892–1924



$\sum_j \alpha^j = 1$.

Goods are produced by land, labor, and capital. Income in location i is expressed as following:

$$Y_i = w_i L_i + q_i H_i + r K_i, \tag{2.2}$$

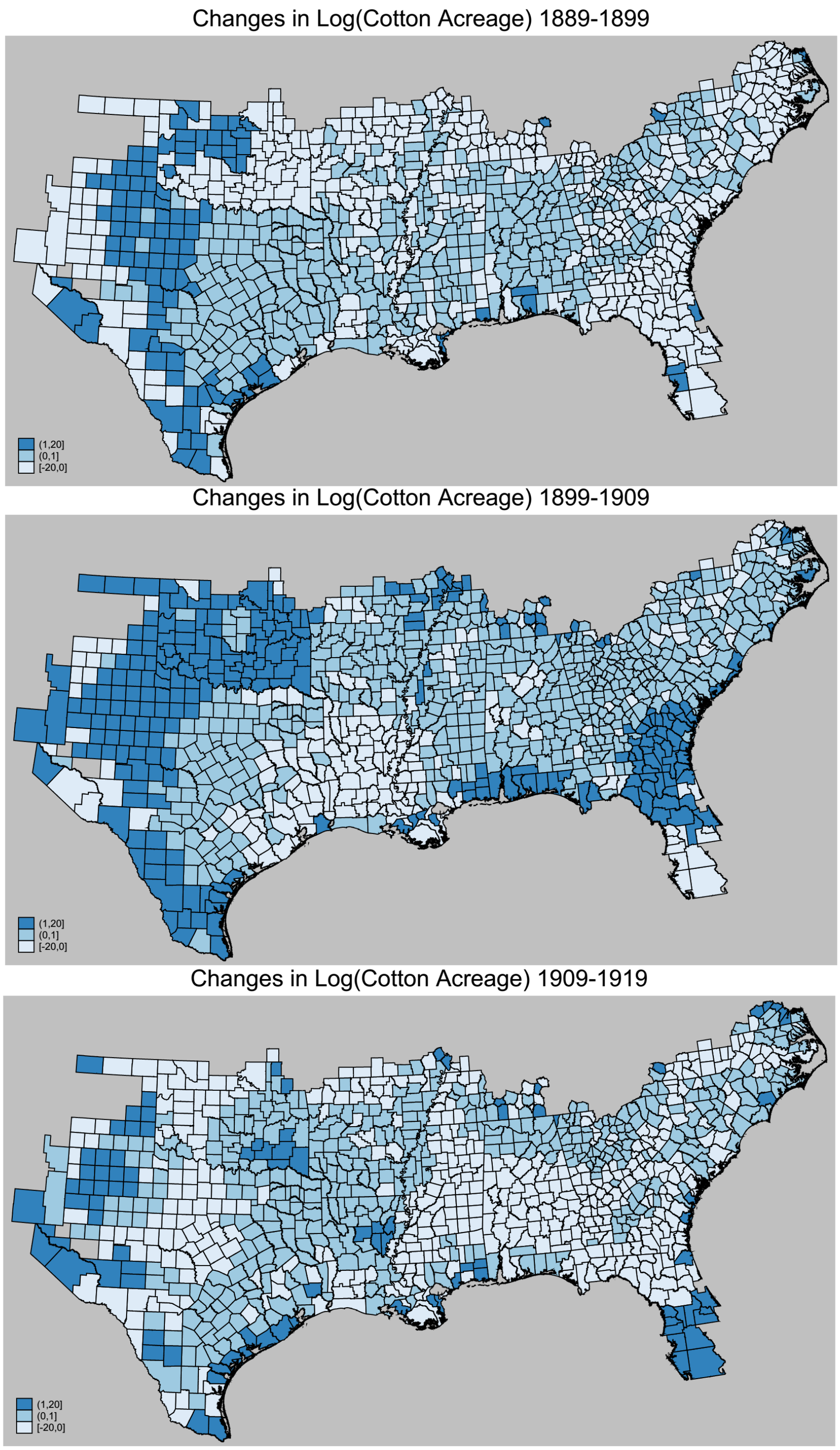
in which w_i is the nominal wage in location i, q_i is the nominal land rent rate in location i, r is the national capital interest rate, L_i is the total amount of agricultural labor in location i, H_i is the total amount of farmland in location i.

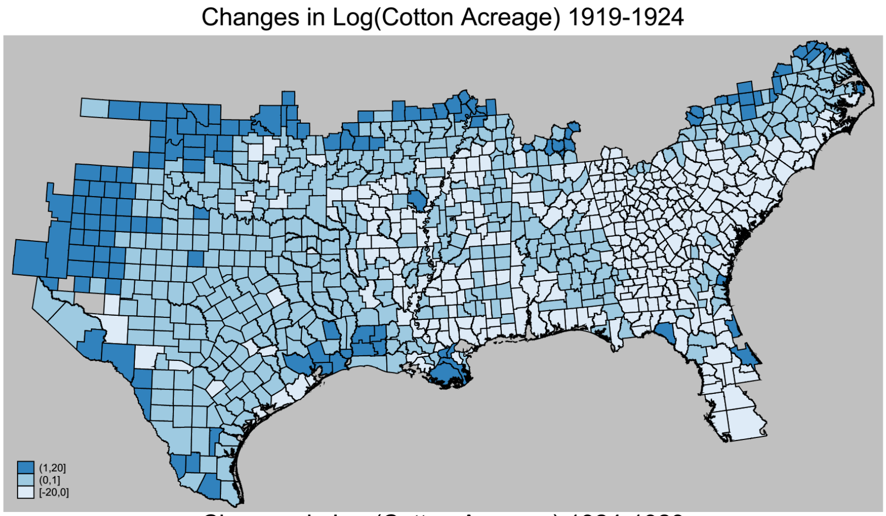
Each location draws an idiosyncratic productivity z_i^j for each location-sector pair (i, j). z_i^j is independently drawn from a Frechet distribution $F_i^j(z) = \exp(z^{-\theta})$. The Frechet distribution contains a shape parameter θ.

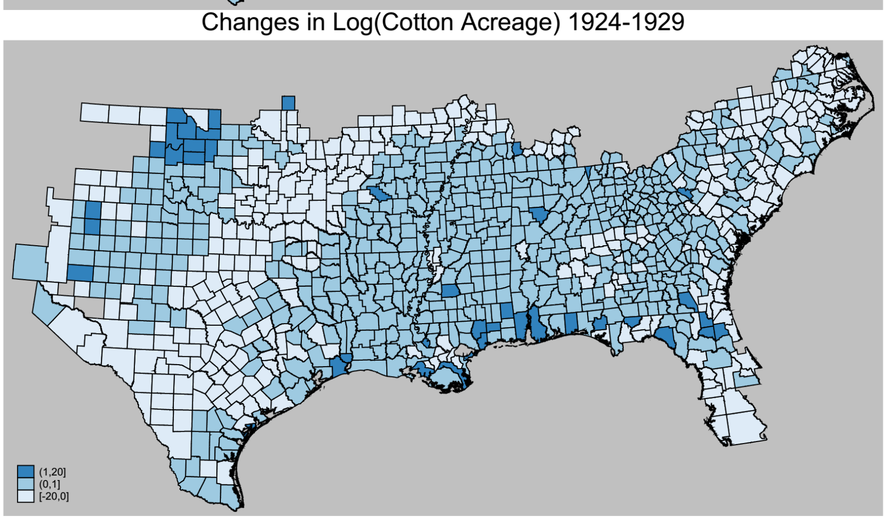
The productivity in sector j and location i is also affected by endowments of location i, which generates a location-sector specific productivity level T_i^j.

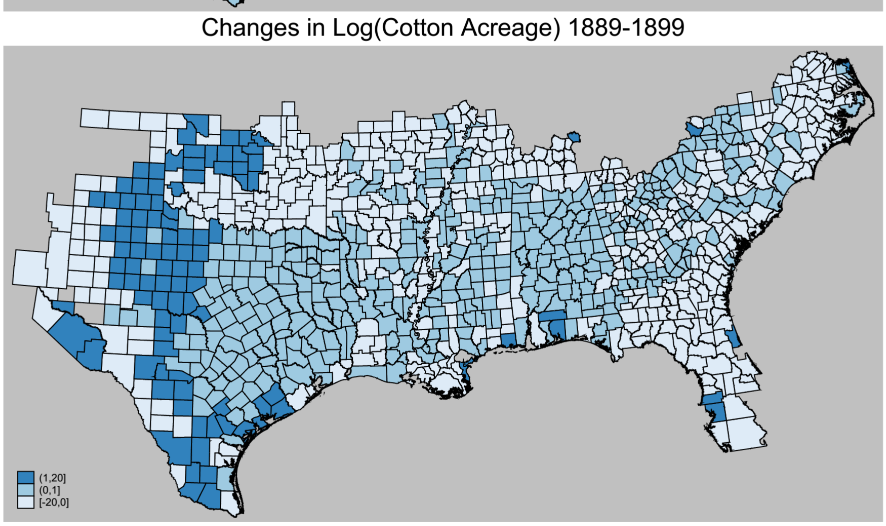
The production function is also Cobb-Douglass. The marginal cost of producing goods in sub-sector j and location i is expressed as:

Figure 2.2: Redistribution of Cotton Production 1889–1929









Notes: The darkest blue indicates a large increase in the cotton acreage. The second darkest blue indicates a small increase in the cotton acreage. The lightest blue indicates a decrease in the cotton acreage.

$$MC_i^j = \frac{q_i^{\beta_i^j} w_i^{\gamma_i^j} r^{1-\beta_i^j-\gamma_i^j}}{z_i^j T_i^j}. \tag{2.3}$$

Goods are tradable with iceberg cost. One unit of any good in any sub-sector j shipped from location i to location n requires producing $\kappa_{ni} > 1$ units goods in location i.

In the world of complete competition, the good price in sub-sector j and location n is given by:

$$p_n^j(z^j) = \min_i \left\{ \frac{\kappa_{ni} q_i^{\beta_i^j} w_i^{\gamma_i^j} r^{1-\beta_i^j-\gamma_i^j}}{z_i^j T_i^j} \right\} \tag{2.4}$$

With the Frechet distribution of z_i^j, the price of goods consumed in sub-sector j location n can be expressed as:

$$P_n^j = \delta_1 [\sum_i (q_i^{\beta_i^j} w_i^{\gamma_i^j} r^{1-\beta_i^j-\gamma_i^j} \kappa_{ni})^{-\theta} T_i^j]^{-\frac{1}{\theta}} \tag{2.5}$$

The share of location n's total expenditure on the goods of sector j purchased from location i is given by:

$$\pi_{ni}^j = \frac{X_{ni}^j}{X_n^j} = \frac{(q_i^{\beta_i^j} w_i^{\gamma_i^j} r^{1-\beta_i^j-\gamma_i^j} \kappa_{ni})^{-\theta} T_i^j}{\sum_m (q_m^{\beta_m^j} w_m^{\gamma_m^j} r^{1-\beta_m^j-\gamma_m^j} \kappa_{nm})^{-\theta} T_m^j} \tag{2.6}$$

Rewriting equation (6), we can get:

$$X_{ni}^j = \frac{(q_i^{\beta_i^j} w_i^{\gamma_i^j} r^{1-\beta_i^j-\gamma_i^j} \kappa_{ni})^{-\theta} T_i^j}{\sum_m (q_m^{\beta_m^j} w_m^{\gamma_m^j} r^{1-\beta_m^j-\gamma_m^j} \kappa_{nm})^{-\theta} T_m^j} X_n^j \tag{2.7}$$

The total value of goods produced by sub-sector j in location i is expressed as:

$$Y_i^j = \sum_n X_{ni}^j = \alpha^j \sum_n \frac{(q_i^{\beta_i^j} w_i^{\gamma_i^j} r^{1-\beta_i^j-\gamma_i^j} \kappa_{ni})^{-\theta} T_i^j}{\sum_m (q_m^{\beta_m^j} w_m^{\gamma_m^j} r^{1-\beta_m^j-\gamma_m^j} \kappa_{nm})^{-\theta} T_m^j} Y_n \equiv FMA_i^j \tag{2.8}$$

Following (DH16), I define the sector-location specific producer market access FMA_i^j for sector j in location i. It includes the factor price in location i $q_i^{\beta_i^j} w_i^{\gamma_i^j} r^{1-\beta_i^j-\gamma_i^j}$, location-sector specific productivity level T_i^j, the trade cost between the producer i and market n, κ_n^j, the competition from all other producers m, $\sum_m (q_m^{\beta_m^j} w_m^{\gamma_m^j} r^{1-\beta_m^j-\gamma_m^j} \kappa_{nm})^{-\theta} T_m^j$, consumption share of sub-sector j, α^j, and the size of market n, Y_n.

IMPLICATION 1: Holding other variables constant, increasing the productivity of sector j in location i, T_i^j, will cause producer market access of sector j in location i, FMA_i^j, to increase.

IMPLICATION 2: Holding other variables constant, increasing the productivity of sector j in location i's competitor m T_m^j will cause producer market access of sector j in location i, FMA_i^j, to decrease.

IMPLICATION 3: Holding other variables constant and allowing T_i^j and T_m^j to change simultaneously, increasing the relative advantage of sector j in location i's competitor m, $\frac{T_m^j}{T_i^j}$, will cause producer market access of sector j in location i, FMA_i^j, to decrease.

Given the acreage of crop j in location i, $H_i^j = \frac{\beta_i^j Y_i^j}{q_i}$, the ratio of crop j acre to the total farm acre can be expressed as:

$$\frac{H_i^j}{H_i} = \frac{\beta_i^j Y_i^j}{\sum_j \beta_i^j Y_i^j} = \frac{\beta_i^j FMA_i^j}{\sum_j \beta_i^j FMA_i^j}. \tag{2.9}$$

Transforming both sides to logarithmic form:

$$\ln(H_i^j) = \ln(H_i) + \ln(\frac{\beta_i^j FMA_i^j}{\sum_j \beta_i^j FMA_i^j}) \tag{2.10}$$

IMPLICATION 4: Holding other variables constant, increasing sector j's producer market access in location i, FMA_i^j, will cause the land use in sector j, H_i^j, to increase.

Combining **IMPLICATION 1, 2, 3, 4**, we can get one testable implication:

IMPLICATION 5: Holding other variables constant, negative productivity shock in sector j location i will decrease T_i^j and cause the land used in sector j, H_i^j, to decrease;

while the negative productivity shock in sector j in competing location m will decrease T_m^j, decrease competition faced by producers in sector j location i, and cause the land use in sector j location i, H_i^j, to increase. The negative impact of negative productivity shocks in sector j on land use in sector j location i is attenuated by negative productivity shocks in competing locations.

2.4 Data and Econometric Model

2.4.1 From Theory to Empirical Strategy

IMPLICATION 5 above implies the following econometric specification:

$$\ln(Acreage_{it}^{cotton}) = \beta_1 BW_{it} + \beta_2 \overline{BW_t} + \beta_3 BW_{it} \times \overline{BW_t} + \eta W_{it} + \theta_i + \gamma T + \delta g(P_{t-1}^{cotton}) + \alpha X_{it} + \eta f(x_{it}, y_{it}) + \delta_{st} + \varepsilon_{it}, \tag{2.11}$$

where $\ln(Acreage_{it}^{cotton})$ is the logarithm form of cotton acreage in county i at the census year t. BW_{it} is an indicator variable that is 1 if county i was infested by the boll weevil at the census year t, and 0 otherwise. $\overline{BW_t}$ is the percentage land infested by the boll weevil in the Cotton Belt in the census year t. **IMPLICATION 5** implies that (1) the boll weevil, a negative productivity shock in cotton productivity, would negatively affect the land use in cotton in location i ($\beta_1 < 0$); (2) the expansion of the boll weevil would reduce the cotton yield in the competing locations and positively affect the land use in cotton in location i ($\beta_2 > 0$); (3) the negative effects of the boll weevil on the cotton acreage in location i would reduce as the boll weevil expanded in the Cotton Belt ($\beta_3 > 0$). Following (LOR09) and (ABH17), I control county fixed effects θ_i. Unlike (LOR09) and (ABH17) which control year fixed effects, I control the time trend T (year and its quadratic term), since the key variable of interest, $\overline{BW_t}$, only varies by year but not varies by locations. Similar to previous literature, I also control weather variables, which include the average temperature and the precipitation in each month (from December in the year before the census year to November in the census year).[2]

Previous literature focuses on the impact of the boll weevil on cotton yield. They omit several potentially important controls. To reduce the problem of endogeneity, I also control several variables which might affect the land use in cotton. (1). I control the cotton price in the year before the census year and its quadratic term $g(P_{t-1}^{cotton})$ since it is believed to be a driven force of cotton production expansion in this sample period. (2). I control the interaction terms of initial production condition with time trend X_{it} to account for the time trend of soil conservation, irrigation, and mechanization. These initial production conditions include average topography roughness, average farm size, the farm share of all land ((JD54)), and the difference in land suitability in cotton between rainfall and irrigation. Besides, following (LOR09), I control the interaction of cotton share in total farm acreage in 1889 and time trend to account for the growth of new cotton producers. (3). I control whether there was any railway in the census year. (Gie04) believes that it was the construction of the railway in western Cotton Belt which stimulated cotton production. (4). I further control the interaction between latitude, longitude, their quadratic and cubic terms with time trend $f(x_{it}, y_{it})$, and the interaction between state fixed effects with time trend δ_{st}.

2.4.2 Potential Problems in Previous Literature

There are two concerns that previous literature may biasedly estimate the effects of the boll weevil on the cotton acreage. The first concern is that they do not control many variables which affected the cotton acreage and also correlated with the presence of the boll weevil. The second concern is that they do not include the competition effects $\overline{BW_t}$ which also correlates with BW_{it}. Without including the competition effects and its interaction term with the presence of the boll weevil, we cannot tell the true treatment effects from the regression results.

First, regarding the cotton acreage, the arrival of the boll weevil is not exogenous.

[2] (LOR09) and (ABH17) only control winter temperature and summer precipitation. (Dai29) indicates that weather conditions in Spring and Autumn also affect cotton production.

(LOR09) and (ABH17) argue that the arrival of the boll weevil is exogenous since "farmers and local authorities could do little to present the boll weevil from entering their territory." To deal with the possibility of the reverse causality that "locally favorable production conditions for cotton are likely to have favored the boll weevil's spread", (LOR09) instrument the presence of the boll weevil using each county's distance in longitude east and west from Brownsville, Texas (the entry point of the boll weevil) as well as the latitude of each county interacted with the year dummies. Their results of the impact of the boll weevil on the cotton bales and the cotton yields are consistent in the IV and the OLS. However, the impact of the boll weevil on the cotton acreage loses statistical significance in the IV strategy. It shows that regarding the cotton acreage, the arrival of the boll weevil might be endogenous. It is true that locally favorable production conditions for the cotton acreage were likely to had favored the boll weevil's spread.

However, their IV strategy could not completely solve the problem of endogeneity. One concern is that their instrument variables are correlated with some omitted variables. The geographical location of one county, especially its distance to Brownsville, Texas (the entry point of the boll weevil), is correlated with some production conditions which might affect the expansion of the cotton production. In fact, counties which are closer to the entry of the boll weevil enjoyed more production advantages. (JD54) point out that compared to the eastern states of the Cotton Belt, western states had many advantages: "topography in the western cotton areas is smoother, fields are larger, and the average cotton acreage per farm is greater." All these advantages were good for soil conservation and mechanization. Western states also got more advantages from the development of irrigation system. Besides, Delta cotton states (Arkansas, Mississippi, and Louisiana), which are also closer to the entry point of the boll weevil than eastern cotton states, had more fertile soil, flatter topography, and larger farm size than eastern cotton areas. These advantages for mechanization, soil conservation, and irrigation contributed more to the expansion of the cotton production as time passed, so they could not be controlled by county fixed effects. Omitting these factors from regression will lead to biases.

Second, if there exist competition effects, omitting the competition effects $\overline{BW_t}$ will also lead to biases. The reason is that the correlation between BW_{it} and $\overline{BW_t}$ is obvious. The presence of the boll weevil in each county depended on its geographical location and the percentage of the boll weevil infestation in the whole Cotton Belt.

Besides, without including $\overline{BW_t}$ in the econometric model, we cannot explain the dynamic rise and fall of the cotton acreage in the different regions of the Cotton Belt. We cannot fully explain that why the cotton acreage increased in the eastern cotton states when the boll weevil was still in the western states and why the cotton acreage recovered in western states while the boll weevil expanded to the eastern Cotton Belt.

2.4.3 Data

In this section, I discuss the data and methods used to access the impact of the expected cotton yield loss caused by the boll weevil on the agricultural land use. I get the county-level land use data (acreage in cotton and acreage in other farm products) from the United States Censuses of Agriculture for the years 1889, 1899, 1909, 1919, 1924, and 1929. Besides, I get the population data from the United States Censuses for the years 1889, 1899, 1909, 1919, and 1929.[3]

The geographic area of the Cotton Belt and the infection of the boll weevil in the county i at the census year t are based on the USDA boll weevil map ((HC23)).[4] I exclude counties with less than 10 acres of cotton plantation in all census years from 1889 to 1929. There are 979 counties in my analysis. (ABH17) indicates that many counties changed boundaries over the sample period and exclude those counties with boundary change from their analysis. Instead, I do not exclude these counties from my analysis. I adjust data from 1889, 1899, 1909, 1924, and 1929 to reflect 1919 boundaries. In fact, only very few counties have big adjustments in this sample period.

The weather data is from The United States Historical Climatology Network (USHCN).

---

[3]These data are digitized by The Inter-University Consortium for Political and Social Research (ICPSR).

[4](LOR09) and (ABH17) also use the same data.

The county-level terrain roughness is calculated as the average Topographic Position Index ((Jen06)) within a county. It is based on elevation raster data which is from OpenTopography.

The railway data is from (Ata16). One problem in this dataset is that the information of the railway construction after 1911 is not available. I have to use the railway in 1911 as a proxy for the existence of railway after 1911. It will lead to some measurement errors. However, the measurement errors are not very serious since more than 84% railways which existed in 1929 had been operated in 1911.

When I discuss the effects of the boll weevil on the black population, I also control the racial violence which is measured as the number of the black lynching victims in the last ten years of the census year. I got this data from http://www.thiscruelwar.com/the-long-list/. The anonymous author combines several databases together (including the databases from The National Association for the Advancement of Colored People (NAACP), Beck and Tolnay Inventory ((BT04)), Michael J. Pfeifer's work ((Pfe13))), corrects inaccurate information in these databases, and adds more information from newspapers. In This Cruel War database, data for CO, KS, MO, NM, OK, TX and WV is only available from 1889 to 1923. To add the missing information, I also combine This Cruel War database with more databases (including (Fra15), (Gre02), (Woo13), (Cha73)). My database includes almost all known lynching cases in the Cotton Belt from 1889 to 1929. I use the county-level number of black victims from 1889 to 1899 to capture the racial violence in the census years 1889 and 1899. Regarding the census years in 1909, 1919, and 1929, I use the county-level number of black lynching victims in the last ten years to capture the level of the racial violence.

2.5 Results

2.5.1 Biases Caused by Omitting Other Factors Affected Cotton Acreage

Before I estimate Equation 2.11, I re-estimate the impact of the presence of the boll weevil on the cotton acreage without including $\overline{BW_t}$. (LOR09) indicates the boll weevil's negative impact on the cotton acreage is smaller than its impact on the cotton yield. (ABH17) also indicates the boll weevil negatively affected cotton acreage. One problem is that previous literature only controls fixed effects and weather, but do not control other factors which might affect the cotton acreage as time goes by. In fact, this problem leads to biased results. Table 2.1 reports the comparisons between without and with controls of other variables which might affect cotton production as time goes by. In column (1) of Table 2.1, I do not control variables other than county fixed effects, time trend, cotton price trend, and weather. The impact of the boll weevil on the cotton acreage is significantly negative. It is consistent with the discussion in (LOR09) and (ABH17). However, once I control other variables which might affect cotton acreage as time goes by, the boll weevil's impact on the cotton acreage becomes significantly positive (column (2) of Table 2.1). Columns (3) and (4) of Table 2.1 indicate the boll weevil's negative effects are robust on the cotton yield no matter without or with other controls. Columns (5) and (6) of Table 2.1 indicate that with other controls, the negative effects of the boll weevil on the cotton bales becomes insignificant.

Does it mean the boll weevil's impact on the cotton acreage is positive? In fact, **IMPLICATION 5** implies that the effects of a productivity shock on the land use are not so simple. The marginal effects of the negative productivity shock vary by the productivity shock in other locations. In other words, the effects of the boll weevil on the location i also depend on the boll weevil infestation in the whole Cotton Belt.

2.5.2 Baseline: Cotton Acreage

Table 2.2 reports the results of the baseline analysis. In column (1), I do not control factors which might affect land use in cotton other than county fixed effects, time trend, weather, and last year cotton price. In column (2), I control more variables which were believed by historians to affect land use in cotton (see details in subsection 2.4.1). Results are robust and consistent with **IMPLICATION 5**. First, with the presence of the boll weevil and no boll weevil in other competing locations, the cotton acreage decreased substantially by 61% on average (column (2)). Second, regarding counties which were not infested by the

Table 2.1: The Boll Weevil and Cotton Production 1889–1929

	Log (Cotton Acreage)		Log (Cotton Yield)		Log (Cotton Bales)	
	(1)	(2)	(3)	(4)	(5)	(6)
Boll Weevil	-0.145*	0.236***	-0.315***	-0.284***	-0.424***	-0.0642
	(0.0787)	(0.0711)	(0.0196)	(0.0197)	(0.0727)	(0.0649)
Year	4.823***	-84.95**	-2.557***	36.11***	1.694	-24.48
	(1.623)	(34.40)	(0.387)	(8.257)	(1.551)	(32.20)
$Year^2$	-0.00126***	0.0234***	0.00068***	-0.0102***	-0.000432	0.00762
	(0.000427)	(0.00796)	(0.000102)	(0.00201)	(0.000408)	(0.00730)
Last Year Cotton Price	0.225**	0.287***	-0.151***	-0.145***	0.0618	0.118
	(0.110)	(0.0879)	(0.0231)	(0.0200)	(0.103)	(0.0825)
$(\text{Last Year Cotton Price})^2$	-0.00568**	-0.00734***	0.00372***	0.00360***	-0.00165	-0.00315
	(0.00268)	(0.00215)	(0.000560)	(0.000485)	(0.00251)	(0.00201)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes	No	Yes
N	5874	5874	5517	5517	5874	5874
R^2	0.159	0.623	0.093	0.349	0.150	0.599
adj. R^2	0.154	0.618	0.088	0.339	0.146	0.594

Notes: All specifications include county fixed effects and controls for monthly average temperature and precipitation in the whole crop year (from last year December to this year November). There is no other controls in columns (1) (3) (5). Columns (2) (4) (6) include other controls which would affect the land use in cotton. See Section 2.4.1 for details. Standard errors are clustered by county. Robust standard errors are in parentheses.

* p ¡ 0.10, ** p ¡ 0.05, *** p ¡ 0.01

boll weevil, as the boll weevil infested 1% more lands in the Cotton Belt, the cotton acreage would increase 3.3% on average. Third, the highly significant positive coefficient on the interaction term $BW_{it} \times \overline{BW_t}$ shows that regarding counties which were infested by the boll weevil, the negative effects of the presence of the boll weevil on the land use in cotton were attenuated by the boll weevil's expansion in the Cotton Belt. As the boll weevil infested 1% more lands in the Cotton Belt, the cotton acreage would recover 1.6% on average.

(LOR09) indicates that agricultural production responded in anticipation of the boll weevil's arrival. One explanation is just the competition effects in this paper. It is a response to enlarged local labor pools swollen by cotton hands moving to unaffected regions to escape the boll weevil. However, there might be other explanations. (LOR09) points out "It might be economically rational to seek to depreciate rapidly cotton-specific assets (equipment and soils) before the insect's attack lowered their productivity." To separate the anticipation effects from the competition effects, I include $BW_{i,t+1}$, $BW_{i,t+2}$, and $BW_{i,t+3}$ into the regression. $BW_{i,t+d}$ $(d = 1, 2, 3)$ refers to the dummy variable which is equal to one if the census year is d years before the boll weevil. I report the results in column (3) of 2.2. My results are in line with the findings in (LOR09). I find the evidence of significant anticipation effects. However, controlling for the anticipation effects does not change my finding in column (2).

Another concern is that cotton production might adapt to the existence of the boll weevil as time goes by. People may have opportunities to find out partial control methods. The adaption effects might mix with the competition effects. As time goes by and the boll weevil infests more lands, the cotton acreage will recover. To separate the adaption effects from the competition effects, I control the number of years after the arrival of the boll weevil and its quadratic term in column (4) of Table 2.2. In fact, as time goes by, the negative impact of the boll weevil on the cotton acreage became stronger and stronger. However, results still support my theoretical implications. I control both anticipation effects and adaption effects in column (5), my results are still robust.

Table 2.2: Effects of The Boll Weevil and Reduced Competition on the Land Use in Cotton 1889–1929

Dependent Variable: Log (Cotton Acreage)	(1)	(2)	(3)	(4)	(5)
Boll Weevil	-0.351**	-0.606***	-0.599***	-1.199***	-1.209***
	(0.144)	(0.145)	(0.146)	(0.177)	(0.182)
Percentage of Infested Land in Cotton Belt	0.0378***	0.0334***	0.0324***	0.0203**	0.0206**
	(0.0113)	(0.00925)	(0.00931)	(0.00929)	(0.00930)
Boll Weevil × Percentage of Infested Land in Cotton Belt	0.00407	0.0162***	0.0175***	0.0289***	0.0283***
	(0.00326)	(0.00287)	(0.00311)	(0.00339)	(0.00346)
One Year Before the Boll Weevil			0.314***		-0.125
			(0.0860)		(0.117)
Two Years Before the Boll Weevil			0.228**		-0.148
			(0.114)		(0.130)
Three Years Before the Boll Weevil			0.0874		-0.0890
			(0.175)		(0.179)
Years After the Arrival of the Boll Weevil				-0.149***	-0.153***
				(0.0213)	(0.0236)
(Years After the Arrival of the Boll Weevil)^2				-0.000108	-0.0000497
				(0.000732)	(0.000735)
County FE	Yes	Yes	Yes	Yes	Yes
Time Trend	Yes	Yes	Yes	Yes	Yes
Last Year Cotton Price	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes
Other Controls	No	Yes	Yes	Yes	Yes
Anticipation Effects	No	No	Yes	No	Yes
Adaption Effects	No	No	No	Yes	Yes
N	5874	5874	5874	5874	5874
R^2	0.160	0.627	0.627	0.639	0.639
adj. R^2	0.156	0.622	0.622	0.634	0.634

Notes: All specifications include county fixed effects and controls for monthly average temperature and precipitation in the whole crop year (from last year December to this year November). There is no other controls in column (1). Columns (2)-(5) include other controls which would affect the land use in cotton. See Section 2.4.1 for details. Column (3) includes the anticipation effects. Column (4) includes the adaption effects. Column (5) includes both anticipation effects and adaption effects. Standard errors are clustered by county. Robust standard errors are in parentheses.

* p ¡ 0.10, ** p ¡ 0.05, *** p ¡ 0.01

2.6 From Cotton to The Southern Economy

2.6.1 Land Use in Agricultural Production and Land Value

How did the boll weevil affect other crops? Table 2.3 shows that the boll weevil had similar effects on corn as it had on cotton. After controlling other factors and taking the competition effects into account, I find that contrary to (LOR09) and (ABH17), the corn acreage also reduced when the boll weevil struck but grew back (together with cotton) when more areas of the Cotton Belt were infested (column (1) of Table 2.3). I do not observe a significant effect of the boll weevil expansion on the corn acreage in the unaffected counties. The effects on the acreage of all farm products other than cotton are not significant (column (2) of Table 2.3). Total farm acreage has the similar trends to corn (column (3) of Table 2.3). My findings are more in line with the arguments of some historians that corn and other crops grew with cotton. (JD54) indicates that:

> The usual plan was to plant as much cotton as the available labor force could thin and pick. The remaining acreage would then be planted to crops that could be fitted around the cotton enterprise with available tools and labor. Corn was a common alternative, although yields in the South were typically very low compared with yields in the Corn Belt.

(JD54) also indicates that many factors, such as the lack of capital and experience, small farms size, prevented farmers from shifting to other crops.

Is there any evidence for diversification? Column (4) of Table 2.3 shows there exists evidence for diversification. It shows that when few lands in the Cotton Belt were affected by the boll weevil, the boll weevil had negative effects on the share of cotton acreage in all farmland. However, as the boll weevil plague expanded, the cotton share also gradually recovered.

To sum it up, my findings are much more complicated than the simple land-abandonment story and diversification story. When a county was infested by the boll weevil, but most lands in the Cotton Belt were still safe, southern farmers abandoned cultivation and also diversified the land use. They reduced the land use in cotton as well as corn. The reduction in cotton acreage was larger than other crops. We observed both diversification and land abandonment. As more and more lands were infested by the boll weevil, both cotton and corn came back, but more cotton came back than corn.

The boll weevil's impacts on the agricultural land value are also similar to the ones on the cotton acreage. When few lands in the Cotton Belt were infested by the boll weevil, the boll weevil negatively affected land value. As more and more lands in the Cotton Belt were infested by the boll weevil, land value would also recover.

Table 2.3: Effects of The Boll Weevil and Reduced Competition on Corn Acreage, Other Farm Products, Total Farm Acreage, and Land Value 1889–1929

	(1) Log (Corn Acreage)	(2) Log (Acreage in Other Farm Products)	(3) Log (Total Farm Acreage)	(4) Cotton Share	(5) Log (Farmland Value)
Boll Weevil	-0.471***	-0.150	-0.199*	-0.0454***	-0.273**
	(0.130)	(0.115)	(0.114)	(0.00585)	(0.135)
Percentage of Infested Land in Cotton Belt	0.00464	0.00760	0.00835	0.000332	0.0148
	(0.00798)	(0.00871)	(0.00866)	(0.000483)	(0.0101)
Boll Weevil × Percentage of Infested Land in Cotton Belt	0.00684***	0.00357*	0.00489***	0.00111***	0.00682***
	(0.00217)	(0.00186)	(0.00187)	(0.0000890)	(0.00224)
County FE	Yes	Yes	Yes	Yes	Yes
Time Trend	Yes	Yes	Yes	Yes	Yes
Cotton Price Controls	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes
Anticipation Effects	Yes	Yes	Yes	Yes	Yes
Adaption Effects	Yes	Yes	Yes	Yes	Yes
N	5874	5874	5874	5733	5874
R^2	0.675	0.767	0.772	0.349	0.813
adj. R^2	0.670	0.764	0.769	0.340	0.810

Notes: Standard errors are clustered by county. Robust standard errors are in parentheses.

* p ¡ 0.10, ** p ¡ 0.05, *** p ¡ 0.01

2.6.2 Population in the Cotton Belt

Since cotton is a labor-intensive crop, we would expect population swang together with cotton. This expectation is proved by my analysis (Table 2.4). I find that when few lands in the Cotton Belt were infested by the boll weevil, the boll weevil decreased total population by 23% and rural population by 25%. Regarding counties which were not affected by the boll weevil, a 1% expansion of the boll weevil increased population by 3.9% and rural population by 3.8%. Regarding counties which were affected by the boll weevil, population and rural population recovered with the 1% expansion of the boll weevil by 0.40% and 0.55%. The effects on the share of the rural population of all population are not significant (column (3) of Table 2.4). The effects on the black population are similar to the total population. When few lands in the Cotton Belt were affected by the boll weevil, the boll weevil has negative effects on the black population. The boll weevil decreased the black population by about 20% (column (4) of Table 2.4). One concern is that the presence of the racial violence which might coincide with the presence of the boll weevil, we cannot conclude that the boll weevil alone pushed millions of black people from the Cotton Belt ((Gie04)). To separate the effects of the boll weevil from the effects of the racial violence, I control the number of the black lynching victims in the last ten years from the census year. It does not change the effects of the boll weevil on the black population (column (5) of Table 2.4). The boll weevil did not significantly change the share of the black population (columns (6) and (7) of Table 2.4).

2.6.3 Tenancy System in the Cotton Belt

As the most important crop in the Cotton Belt, cotton was linked with many historical institutions: the plantation system and slavery before the Civil War and the tenancy system after the Civil War. The Civil War ended slavery, but landless former slaves and poor whites were still landless and became tenants. (Gie04) indicates that a region's turn to cotton meant a concurrent turn to tenancy. A tenancy ladder evolved with the tenancy system. (JEA35) elaborates the three main classes of tenants:

(a) renters who hire land for a fixed rental to be paid either in cash or its

Table 2.4: Effects of The Boll Weevil and The Reduced Competition on Population 1889–1929

	(1) Log (Population)	(2) Log (Rural Population)	(3) Rural Share	(4) Log (Black Population)	(5) Log (Black Population)	(6) Black Share	(7) Black Share
Boll Weevil	-0.227***	-0.253***	-0.00296	-0.203**	-0.207**	-0.00580	-0.00636
	(0.0878)	(0.0910)	(0.00945)	(0.101)	(0.101)	(0.00413)	(0.00415)
Percentage of Infested Land in Cotton Belt	0.0388**	0.0378**	-0.00113	0.00936	0.00949	-0.0000665	-0.0000455
	(0.0152)	(0.0151)	(0.00110)	(0.0138)	(0.0138)	(0.000481)	(0.000478)
Boll Weevil × Percentage of Infested Land in Cotton Belt	0.00397***	0.00546***	0.0000550	0.00264	0.00268	0.0000101	0.0000166
	(0.00140)	(0.00164)	(0.000175)	(0.00201)	(0.00201)	(0.0000721)	(0.0000723)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cotton Price Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Anticipation Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adaption Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Last 10 Years Black Lynching Victims	No	No	No	No	Yes	No	Yes
N	4895	4895	4775	4895	4895	4775	4775
R^2	0.739	0.724	0.310	0.566	0.567	0.356	0.359
adj. R^2	0.734	0.719	0.297	0.559	0.559	0.344	0.347

Notes: Standard errors are clustered by county. Robust standard errors are in parentheses.

* p ¡ 0.10, ** p ¡ 0.05, *** p ¡ 0.01

equivalent in crop values; (b) share tenants, who furnish their own farm equipment and work animals and obtain use of land by agreeing to pay a fixed percent of the cash crop which they raise; (c) sharecroppers who have to have furnished to them not only the land but also farm tools and animals, fertilizer, and often even the food they consume, and who in return pay a larger percent of the crop.

(Con65) points out only when tenants accumulated enough capital, they could move up the tenancy ladder. If a sharecropper could accumulate enough equipment and money, he could become a share-tenant. If a share-tenant owned everything for production but the land, he could become a cash tenant. (JEA35) believed that cash tenants are quite different from share-tenants and sharecroppers. They are very independent and do not belong to the system of subservient tenancy. Given this reason, I divide all tenants into cash-tenants and other tenants (share-tenants and sharecroppers) in the following discussion and see how the boll weevil affect the southern tenancy system.

Table 2.5 examines how the boll weevil affected the tenancy system. Unlike (ABH17), after controlling several variables which might affect the cotton acreage, anticipation effects, and adaption effects, I do not find evidence that the boll weevil increased the average farm size in the Cotton Belt (column (1) of the Table 2.5). In line with (ABH17), I find that when few lands were infested by the boll weevil in the Cotton Belt, the boll weevil decreased the number of farms by 29%, the number of owned farms by 16%, and the number of tenant farms by 37%. However, as the boll weevil expanded 1% in the Cotton Belt, it recovered the number of total farms by 0.58%, the number of owned farms by 0.35%, and the number of tenant farms by 0.64%. The boll weevil expansion does not have a significant effect on the number of farms in the unaffected counties (columns (2) (3) (4) of Table 2.5).

(ABH17) finds that the boll weevil decreased the number of fixed-rent tenants but did not affect the number of other tenants. The explanation is based on (Hig73) which argues that share-rent contracts are more common than fixed-rent contracts when there is a higher agricultural risk as landlords, by having more wealth and better access to capital markets, are inclined to take more risks than tenants. As a result, the higher agricultural risk with the arrival of the boll weevil leads to a shift from fixed-rent contracts to share contracts.

However, contrary to their findings, I find that the boll weevil did not significantly affect the number of fixed-rent tenant farms, while it decreased the number of other tenants by 36% when the boll weevil infested few lands in the Cotton Belts. As the boll weevil expanded 1% more in the Cotton Belt, the number of other tenant farms recovered by 0.69%. Regarding counties which were temporarily not infested by the boll weevil, 1% of the expansion of the boll weevil in the Cotton Belt increased the number of other tenant farms by 1.6%. It is consistent with the story in (Gie04) that when the boll weevil came, tenants moved from infested lands to temporarily unaffected lands. My finding shows that most of these moved tenants are share-tenants and sharecroppers.

Columns (7) (8) (9) of Table 2.5 show that when the boll weevil infested few lands, the share of owned farms in all farms increased by 4.8%, the share of fixed-rent farms decreased by 2.5%, the share of other tenants decreased by 2.4%. As the boll weevil expanded 1% in the Cotton Belt, the share of owned farms decreased by 0.09% and the share of other tenant farms recovered by 0.072%. Regarding uninfested counties, the 1% expansion of the boll weevil in the Cotton Belt did not significantly affect the share of owned farms, but decreased the share of fixed-rent farms by 0.24%, and increased the share of other tenant farms by 0.33%. These findings are consistent with the arguments in (Gie04) that landowners did not move from place to place, but tenants did.

How to explain the contrary story to the findings in (ABH17)? The tenancy ladder might explain the finding in this paper. As (Con65) and (JEA35) indicate, share-tenants are at the bottom of the tenancy system. They are much less wealthy than fixed-rent tenants. As a result, they are more vulnerable to negative productivity shocks and have higher incentives to move from place to place and look for better economic opportunities and freedom.

2.7 Conclusion

This paper re-examines how the boll weevil reshaped the Cotton Belt under a framework of regional competition. I find that initial infestation of the boll weevil decreased cotton

Table 2.5: Effects of The Boll Weevil and Reduced Competition on Tenancy System 1889–1929

	(1) Log (Farm Size)	(2) Log (Number of Farms)	(3) Log (Number of Owned Farms)	(4) Log (Number of Tenant Farms)	(5) Log (Number of Fixed-rent Tenant Farms)	(6) Log (Number of Other Tenants)	(7) Owned Farms Share	(8) Fixed-rent Farms Share	(9) Other Tenant Farms Share
Boll Weevil	0.0798	-0.290***	-0.155**	-0.372***	-0.133	-0.359***	0.0481***	-0.0247***	-0.0235**
	(0.0682)	(0.0788)	(0.0670)	(0.0965)	(0.103)	(0.108)	(0.0111)	(0.00806)	(0.0119)
Percentage of Infested Land in Cotton Belt	0.00138	0.00658	0.00419	0.00781	-0.0165***	0.0159***	-0.000875	-0.00239***	0.00326***
	(0.0047)	(0.00532)	(0.00486)	(0.00523)	(0.00553)	(0.00540)	(0.000809)	(0.000392)	(0.000062)
Boll Weevil × Percentage of Infested Land in Cotton Belt	-0.00084	0.00582***	0.00349***	0.00639***	0.00126	0.00688***	-0.00090***	0.000179	0.000722***
	(0.0012)	(0.00124)	(0.00108)	(0.00153)	(0.00174)	(0.00170)	(0.000196)	(0.000134)	(0.000182)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cotton Price Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Anticipation Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adaption Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	5874	5874	5874	5874	5874	5874	5732	5732	5732
R^2	0.702	0.737	0.724	0.741	0.584	0.726	0.572	0.504	0.647
adj. R^2	0.698	0.733	0.720	0.737	0.577	0.722	0.565	0.497	0.642

Notes: Standard errors are clustered by county. Robust standard errors are in parentheses.

* p ¡ 0.10, ** p ¡ 0.05, *** p ¡ 0.01

acreage, but the effects were attenuated as the boll weevil infested more lands in the Cotton Belt. Besides, temporarily uninfested counties expanded their cotton production as the boll weevil infested other locations.

The movement of many economic outcomes which were closely associated with the cotton acreage, like corn acreage, total farm acreage, population, rural population, agricultural land value, number of farms, mirrored that of the cotton acreage. Contrary to previous literature ((LOR09) and (ABH17)) I find that the diversification story and land-abandonment story coexisted with the arrival of the boll weevil. When the boll weevil infested few lands in the Cotton Belt, the arrival of the boll weevil decreased the cotton acreage, corn acreage, and total farm acreage. It supports the land-abandonment story. Meanwhile, the share of cotton in all farmlands also decreased. It supports the story of diversification. However, as the boll weevil expanded to more lands, all losses in the cotton acreage, corn acreage, all farm acreage, and the share of cotton in all farmlands recovered.

The arrival of the boll weevil also affected the tenancy system when the boll weevil infested few lands in the Cotton Belt. Both numbers of owned farms and tenant farms decreased with the arrival of the boll weevil when the boll weevil infested few lands in the Cotton Belt, but then recovered as the boll weevil expanded in the Cotton Belt. One interesting finding is that controlling other variables which might affect the cotton acreage, the changes in the number of tenant farms should be attributed to the changes in the number of share tenant farms. This finding is contrary to (ABH17) which find that the decrease in fixed-rent tenant farms drove the decrease in tenant farms. In fact, the presence of the boll weevil did not affect the number of fixed-rent tenant farms but affected the number of share tenant farms. The share of owned farms and the share of fixed-rent tenant farms decreased with the arrival of the boll weevil but recovered as the boll weevil expanded in the Cotton Belt. The class ladder of the tenancy system might explain these findings. As the bottom class in the tenancy system, share tenants (include sharecroppers) were the most vulnerable groups with the arrival of the boll weevil. They had the highest incentive to migrate within the Cotton Belt to look for better economic opportunities.

This paper also quantitatively supports the arguments of some economic historians.

(Gie04) points out that the boll weevil rarely changed the whole Cotton Belt. The South grew more cotton than before the presence of the boll weevil. Landowners still enjoyed great advantages over the large majority tenantry. Diversification and land abandonment were limited. For share tenants, the arrival of the boll weevil simply meant internal migration from farm to farm. (Hig76) concludes that ”the boll weevil infestation was neither a necessary nor a sufficient condition underlying the Great Migration.” This paper proves his conclusion. The boll weevil contributed only to the internal migration in the Cotton Belt.

Table 2.6: Summary Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Boll Weevil Presence	5,874	0.512	0.500	0	1
Percentage of Infested Land in Cotton Belt	5874	52.038	40.746	0	100
Log (Cotton Acreage)	5,874	8.909	3.006	0	12.76
Log (Cotton Yield)	5,517	-1.108	0.395	-3.611	0.922
Log (Cotton Bales)	5,874	7.873	2.827	0	11.99
Cotton Share	5,733	0.128	0.117	0	1
Log (Corn Acreage)	5,874	9.450	2.091	0	12.48
Log (Acreage in Other Farm Products)	5,874	11.86	2.008	0	14.68
Log (Total Farm Acreage)	5,874	12.00	2.025	0	14.69
Log (Farmland Value)	5,874	14.80	2.578	0	18.37
Log (Population)	4,895	9.377	1.797	0	13.04
Log (Rural Population)	4,895	9.220	1.790	0	11.73
Log (Black Population)	4,895	7.434	2.738	0	12.03
Rural Share	4,775	0.897	0.176	0	1
Black Share	4,775	0.310	0.245	0	0.942
Log (Farm Size)	5,874	4.781	1.270	0	13.02
Log (Number of Farms)	5,874	7.231	1.515	0	9.676
Log (Number of Owned Farms)	5,874	6.478	1.354	0	8.800
Log (Number of Tenant Farms)	5,874	6.398	1.733	0	9.615
Log (Number of Fixed-rent Tenant Farms)	5,874	4.606	1.696	0	8.802
Log (Number of Other Tenant Farms)	5,874	6.052	1.804	0	9.503
Owned Farm Share	5,732	0.512	0.210	0	1
Fixed-rent Tenant Farm Share	5,709	0.251	0.226	0	1
Other Tenant Farm Share	5,732	0.369	0.194	0	0.927
Land Roughness	5,874	0.472	0.0475	0.298	0.698
Log (Land Suitability with Irrigation) – Log (Land Suitability with Rain Fall)	5,874	0.316	0.905	-0.0005 2	6.805
Cotton Share in 1889	5,874	0.0907	0.0922	0	1
Log (Farm Size in 1889)	5,874	4.696	1.661	0	11.16
Last Year Cotton Price	5,874	16.67	9.262	6.980	28.88
Average Temperature in Last Year December	5,874	35.48	10.82	4.400	67.90
Average Temperature in January	5,874	30.42	12.75	-11.30	67
Average Temperature in February	5,874	31.66	12.58	-6.500	68.50
Average Temperature in March	5,874	41.78	11.33	3.400	71.20
Average Temperature in April	5,874	52.83	8.925	26.60	77
Average Temperature in May	5,874	61.04	7.945	38.30	83.10
Average Temperature in June	5,874	70.98	6.854	48.60	87.20
Average Temperature in July	5,874	74.40	5.621	54.70	91.40
Average Temperature in August	5,874	73.88	6.700	52.30	91
Average Temperature in September	5,874	65.84	7.172	43.20	86.40
Average Temperature in October	5,874	55.88	8.800	27.90	79.30

Average Temperature in November	5,874	44.03	9.935	11.80	73
Average Precipitation in Last Year December	5,874	2.667	1.928	0	15.37
Average Precipitation in January	5,874	2.797	2.306	0	27.30
Average Precipitation in February	5,874	2.942	2.264	0	18.58
Average Precipitation in March	5,874	3.331	2.280	0.01000	17.32
Average Precipitation in April	5,874	3.175	1.642	0	8.460
Average Precipitation in May	5,874	4.080	2.038	0	12.54
Average Precipitation in June	5,874	3.968	1.948	0	12.99
Average Precipitation in July	5,874	3.615	2.000	0	14.05
Average Precipitation in August	5,874	2.879	1.687	0	11.16
Average Precipitation in September	5,874	3.022	2.018	0	14.91
Average Precipitation in October	5,874	2.836	2.136	0	15.59
Average Precipitation in November	5,874	2.552	2.168	0	28

Bibliography

[ABH17] Philipp Ager, Markus Brueckner, and Benedikt Herz. "The boll weevil plague and its effect on the southern agricultural sector, 1889–1929." *Explorations in Economic History*, **65**:94 – 105, 2017.

[Ata16] Jeremy Atack. "Historical Transportation of Navigable Rivers, Canals, and Railroads in the United States." *ICPSR36353-v1*, 2016.

[Bak15] Richard B. Baker. "From the Field to the Classroom: The Boll Weevil's Impact on Education in Rural Georgia." *The Journal of Economic History*, **75**(4):1128–1160, 2015.

[BT04] E.M. Beck and Stewart E Tolnay. "Confirmed Inventory of Southern Lynch Victims, 1882–1930." 2004.

[Cha73] David L. Chapman. "Lynching in Texas." *Texas Tech University Master degree thesis*, 1973.

[Con65] David Eugene Conrad. *The forgotten farmers: the story of sharecroppers in the New Deal*. University of Illinois Press, 1965.

[CPR14] Lorenzo Caliendo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel G. Sarte. "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy." International Finance Discussion Papers 1119, Board of Governors of the Federal Reserve System (U.S.), August 2014.

[Dai29] Daingerfield. "WEATHER AND COTTON YIELD IN TEXAS, 1899–1929, INCLUSIVE." *Monthly Weather Review*, **57**(11):451–453, 1929.

[DH16] Dave Donaldson and Richard Hornbeck. "Railroads and American Economic Growth: A "Market Access" Approach *." *The Quarterly Journal of Economics*, **131**(2):799–858, 2016.

[EK02] Jonathan Eaton and Samuel Kortum. "Technology, Geography, and Trade." *Econometrica*, **70**(5):1741–1779, September 2002.

[Fra15] H.C. Frazier. *Lynchings in Kansas, 1850s-1932*. McFarland, Incorporated, Publishers, 2015.

[Gie04] James Conrad Giesen. "The South's greatest enemy? [electronic resource] : the cotton boll weevil and its lost revolution, 1892-1930 /." 04 2004.

[Gre02] David Mark Greaves. "View From Here: Never Forget -(The Lynching List)." *http://www.ourtimepress.com/view-from-here-never-forget-the-lynchings-list/*, 2002.

[HC23] W.D. Hunter and B.R. Coad. *The boll-weevil problem*. Farmers bulletin. U.S. Dept. of Agriculture, 1923.

[Hig73] Robert Higgs. "Race, Tenure, and Resource Allocation in Southern Agriculture, 1910." *The Journal of Economic History*, **33**(1):149–169, 1973.

[Hig76] R. Higgs. "The Boll Weevil, the Cotton Economy, and Black Migration 1910-1930." *Agricultural History*, **50**(2):335–350, 1976.

[JD54] Lawrence A. Jones and David Durand. *The Cotton Belt*, pp. 90–110. Princeton University Press/NBER, 1954.

[JEA35] Charles Spurgeon Johnson, Edwin R. Embree, and W.W. Alexander. *The collapse of cotton tenancy: Summary of Field studies and statistical surveys, 1933-35*. Books for Libraries Press, 1935.

[Jen06] Jenness. "Topographic position index." *http://www.jennessent.com/arcview/tpi.htm*, 2006.

[LOR09] Fabian Lange, Alan L. Olmstead, and Paul W. Rhode. "The Impact of the Boll Weevil, 1892–1932." *The Journal of Economic History*, **69**(3):685–718, 2009.

[Osb85] Kent Osband. "The Boll Weevil Versus "King Cotton"." *The Journal of Economic History*, **45**(3):627–643, 1985.

[Pfe13] M.J. Pfeifer. *Lynching Beyond Dixie: American Mob Violence Outside the South*. Bienestar y proteccion social. University of Illinois Press, 2013.

[RS01] Roger L. Ransom and Richard Sutch. *One Kind of Freedom: The Economic Consequences of Emancipation*. Cambridge University Press, 2001.

[Str55] J.H. Street. "Cotton mechanization and economic development." *American Economic Review*, **45**(4):566–583, 1955.

[Woo13] Evan Woodson. "Strange Fruit on the Southern Plains: Racial Violence, Lynching, and African Americans in Oklahoma, 1830-1930." *Oklahoma State University Master degree thesis*, 2013.

[Wri86] C. Wright. *Old South, new South: Revolutions in the southern economy since the Civil War*. LSU Press, 1986.

CHAPTER 3

The Power of Propaganda: The Long-term Impact of Early Exposure to Propaganda on People's Gender Related Attitudes, Evidence from China

3.1 Introduction

Among developing countries, China has one of the lowest levels of gender inequality. In 2015, China ranked 37th on the United Nations Development Programme's Gender Inequality Index (GII) among 159 countries for which the index was calculated. One important manifestation of gender equality in China is the high female labor force participation rate: Women's labor force participation rate was 63.6% compared to 77.9% for men in 2015. These numbers are especially striking given the long tradition of Confucianism in China, which had restricted women's role to the private sphere until the establishment of People's Republic of China in 1949. After 1949, the Chinese Communist Party undertook several waves of propaganda campaigns promoting gender equality. This paper shows the causal impact of gender-equality promoting propaganda on individuals' attitudes towards women's participation in the workforce.

Government's influence on political ideology formation is usually taken as given, but no prior literature has proven the existence of this long run effect. The primary contribution of this study is to empirically establish the long-term impact of gender-equality promoting propaganda on individual's attitudes towards women in the workforce. An advantage of examining this question using evidence from China is that not only is China one of the world's largest economies, it is also a country ruled by one party with strict media censorship.

Information from all media outlets are thus reflections of the central government's viewpoint. The media censorship eliminates effects from other factors and enables us to obtain clean causal effects.

We are able to gauge variations in the political climate in the period between 1952 and 2008 by using the official newspaper of the central government, *People's Daily*, since it has been under the direct control of the Chinese Communist Party (CCP)'s top leadership. Numerous top-down political campaigns have been communicated through well-chosen stock-phrases. We can observe the CCP's ideological emphasis in any time period through the frequency variation of these political phrases. In addition, to show that the variation in individual's attitudes is caused by propaganda rather than a mere time trend, we exploit provincial variation in propaganda intensities. We use the provincial coverage of mass media (radio in the early period and television in more recent years) as a proxy for propaganda intensity. We obtain individual attitudes and demographic information from the China General Social Survey (CGSS) (2010-2013), a multi-year cross-sectional household level survey and a Chinese counterpart of the General Social Survey (GSS) in the US.

There may be concerns that propaganda and/or media coverage are not exogenous. To overcome the potential endogeneity concern regarding propaganda, we use three sets of exogenous events to generate an instrumental variable for intensity peaks of gender-equality propaganda. The first set of events is National Women's Conference that was held every four years from 1949 to 1957. Then it was interrupted by a series of political movements; it did not recommence until 1978. Since then, it has been held every five years. The second set of events is the UN World Conference on Women held in Beijing 1995. It had its five-year anniversary in 2000 and ten-year anniversary in 2005. We also take advantage of the impact of Jiang Qing (Mao Zedong's wife) coming into prominence and gaining full media and propaganda control after a political incident in 1971, and her subsequent removal after Mao's death in 1976. During this time period, Jiang heavily promoted gender-equality in order to achieve personal political ambition. These events are highly correlated with the gender-equality propaganda intensity peaks.

The main determinant of variation in media coverage across provinces and years is the level of economic development. There is no evidence showing that media coverage is correlated with other issues. It is therefore reasonable to deduce that conditional on the level of economic development, provincial media coverage is exogenous to people's attitudes towards gender equality.

Furthermore, we perform a placebo test using survey data from Taiwan. The majority of people in mainland China and Taiwan are ethnic Chinese. They share similar language and traditional culture, and face similar shocks in East Asia. However, the CCP has never governed Taiwan so its propaganda should not affect Taiwanese' attitudes. We find no spurious correlation between gender-equality promoting propaganda in mainland China and Taiwanese' attitudes on women's role in the workforce. We also take advantage of the one similar question in both surveys by using difference-in-difference strategy. We are able to confirm that for women from mainland China, propaganda has a positive effect on their attitudes towards gender equality.

In China, the gender-equality promoting propaganda has generally focused on encouraging women to pursue a career. We find that women with more intense exposure to propaganda promoting gender equality before age 26, and men with more intense exposure before age 18 tend to endorse women's participation in the workforce. The effect of early exposure to propaganda on attitude persists in the long run (after respondents are at least 30 years old), and this effect is larger on women than men.

It is worth noting that gender-related propaganda is not entirely progressive: while propaganda encourages women's participation in the workforce it does not emphasize men's role in the household. We find empirical evidence of the ”superwoman complex” in individuals' attitudes: Women are expected to strive for a career and do the bulk of the housework. And this attitude is consistent with observed household chore sharing patterns. According to a 2014 report, Chinese women, on average, complete 190 minutes of housework every day; by contrast, Chinese men spend 49 minutes on housework.[1] This further evidence suggests that propaganda is able to transmit a more nuanced message, rather than a singularly progressive

[1] http://www.womenofchina.cn/womenofchina/html1/survey/1411/2106-1.htm

one.

In addition, we find that more intense exposure to gender-equality promoting propaganda between age 18 to 25 reduces both women's and men's preference for sons. This is perhaps no coincidence since the average age of the first marriage and first-time mothers in China fall within this range.

The policy implication of our finding is nontrivial and multi-fold: government is capable of changing individuals' preferences through propaganda, and individuals' preferences will determine their economic behaviors. In this specific instance, government makes people endorse women's role in the workplace through propaganda, as a result, more women may participate in the labor force. This boost in productivity may further contribute to the China's economic growth miracle in the recent decades. Moreover, an ideology change would lead to further institutional transition. One potential implication is that an autocratic regime may be capable of gradual, endogenous institutional change ((AR06)). Furthermore, propaganda is a channel through which government and the elite class sustain their political power.

This paper is organized as follows: Section 2 discusses related literature. Section 3 introduces the background of the Chinese propaganda system and gender related propaganda. Section 4 describes the data source in detail. Section 5 discusses empirical strategies, including baseline specification, IV strategies, placebo tests, and difference-in-differences strategies. Section 6 reports regression results, and Section 7 concludes and delineates future work.

3.2 Related Literature

First, this study contributes to the large literature of media effects. (BC16) discussed the effect of foreign media exposure. In particular, they measure the effect of Western German television on consumer behavior of the former residents in East Germany. Several studies explore the impact of particular forms of media on attitudes and behaviors in various countries: change in gender attitudes and behaviors as a consequence of introduction of cable television in India ((JO09)), and the impact of television and radio on social capital in Indonesian villages ((Olk09)). Even more specifically, (LCD12) have shown the effects of soap operas on Brazilian fertility from 1970s to the early 1990s. In contrast, we reveal the effect of the domestic media; in particular, the propagated ideology that the government considers to be fitting for the public. Additionally, the propaganda we focus on is not restricted to one singular media form (e.g. radio, or television, or television shows), rather, propaganda messages transmitted through all media channels that are censored by the central government.

Second, our study contributes to the research on the effect of the communist regime on individuals' preferences. Notably, using West Germany as a control, (AF07) showed that individual's preferences were shaped by the political regime under which they lived. The authors used the broadly defined "Communist" regime as the treatment, but do not parse out the channel through which the autocratic regime affected individual preferences. Instead, our study narrow down the focus to the role of propaganda, a vital tool exploited by autocratic regimes. (BFM16) undertook an experiment finding that females in Beijing growing up during the planned economy system are more competitive than males, and they are also more competitive than their female counterparts who grew up under the market economic system. However, the authors could not distinguish the effect of the economic system from cohort effects. (CCY17) showed that in China, a high school political science text book reform lead students to report political attitudes that were more in accord with the updated version of the textbook when students were surveyed later in college. The authors focused on the short-run effects of government's influence on specific individuals' attitudes (college students), but that people's attitudes are likely to change after college graduation. In contrast, we reveal the long-run effects of early exposure to propaganda (before age 26) on individuals who are at least 30 years of age and whose preferences are relatively stable and are more likely to persist into older age.

3.3 Background

3.3.1 Chinese Propaganda System

As documented in many political science studies ((Sha07); (Bra08)), the Chinese propaganda system is highly structured, with a senior leader (a member of the roughly seven[2] People Politburo Standing Committee) overseeing all propaganda and thought work, both internal and foreign, permeating all aspects and levels of governance:

"The propaganda system in China consists of four connected parts: the network of propaganda cadres and offices installed in Party committees and branches at all levels of organizations in both the State bureaucracy, as well as Chinese and foreign-run private enterprises with CCP cells; the political department system of the People's Liberation Army, through which the CCP controls the military in China; the State-run culture, education, sport, science, technology, health, and media sectors; and all mass organizations such as the Journalist Association, the Internet Association..., as well as the government-operated non-governmental organization..." ((Bra08))

In particular, the Central Propaganda Department is the administrative body and the most important organization in the propaganda system. It guides the overall ideological development by issuing written and oral instructions on various propaganda topics to senior bureaucrats and leaders in the media and culture sectors. And it has a leadership role over the nation-wide system of provincial, local level of propaganda departments.

3.3.2 Gender Equality Related Propaganda in China

Mao Zedong has frequently quoted Karl Marx, "the proletariat must emancipate not only itself but all mankind." The aim of the Communist Party of China in Mao's era was to emancipate all mankind, especially those who were oppressed in the "Old Society". In the 12th century, the most famous neo-Confucian scholar, Zhu Xi, indicated that the "Three

[2]This number can change at the Chinese Communist Party National Congress, held every 5 years in Beijing.

Bonds"—"ruler's authority over subject, father over son, and husband over wife" are the foundation of the society. A woman needs to obey three figures throughout her life: her father, as a daughter; her husband, as a wife; and her sons in widowhood. Until the establishment of People's Republic of China in 1949, the custom of foot-binding by the upper and middle class women prevented them from having a public presence and confined them to a life of domesticity. The Communist Party of China aimed to emancipate all socially vulnerable groups, including the oppressed women.

Aside from the consideration of the communist ideology, encouraging women's participation in the labor force was also an act out of concern for economic development. Labor participation was much needed for the reconstruction of the post-war economy, that had been destroyed by the long-lasting Sino-Japanese War and the civil War between the Nationalist Party of China and the Communist Party of China, and the demand for labor remained high during the Great Leap Forward. In order to fully mobilize labor force participation, gender-related propaganda in Mao's era mainly emphasized that women's ability is equal to that of men, and they are capable of making equal contributions to social and economic construction. Propaganda campaign took many forms: for example, straight-forward slogans such as "women hold up half the sky" appeared frequently in posters and news articles. Another more elaborate example is a well-known number from the regional opera "Mulan" written in 1951, "Who Says Girls Can't Compete with Guys?":

"What Brother Liu said doesn't make any sense. Who says that girls just enjoy the leisure time...There were many heroines making contributions. How could you say girls can't compete with guys?" (Chen and Wang, 1951)

While women were encouraged to join the labor force, propaganda regarding the division of labor between men and women were not entirely progressive. On the one hand, there were a few articles criticizing the traditional gender roles and encouraging men to share some housework. On the other hand, the mainstream gender-equality propaganda rarely mentioned men's responsibility in a household. The editorial of *People's Daily* on the International Women's Day in 1962 featured a female model worker who had "overcome the difficulties of having many children and heavy housework, and participated in revolution and production." These propaganda campaigns indicated that Women were expected to strive for a career and do the bulk of the housework. Originating in Mao's era, propaganda that applauds "superwomanhood" has continued to this day. When discussing the meaning of feminism in a BBC interview in 2010, the chief editor of *China Women's News*, the official paper of the All-China Women's Federation in China, an NGO under the leadership of the CCP, indicated that she spent most of her time participating in the feminist movement in public, but still managed to enjoy doing housework at home.[3]

3.4 Data

There are three main groups of data in this paper: top-down propaganda, provincial variation in propaganda implementation, and people's demographic information and attitudes towards gender equality.

3.4.1 Top-down Propaganda

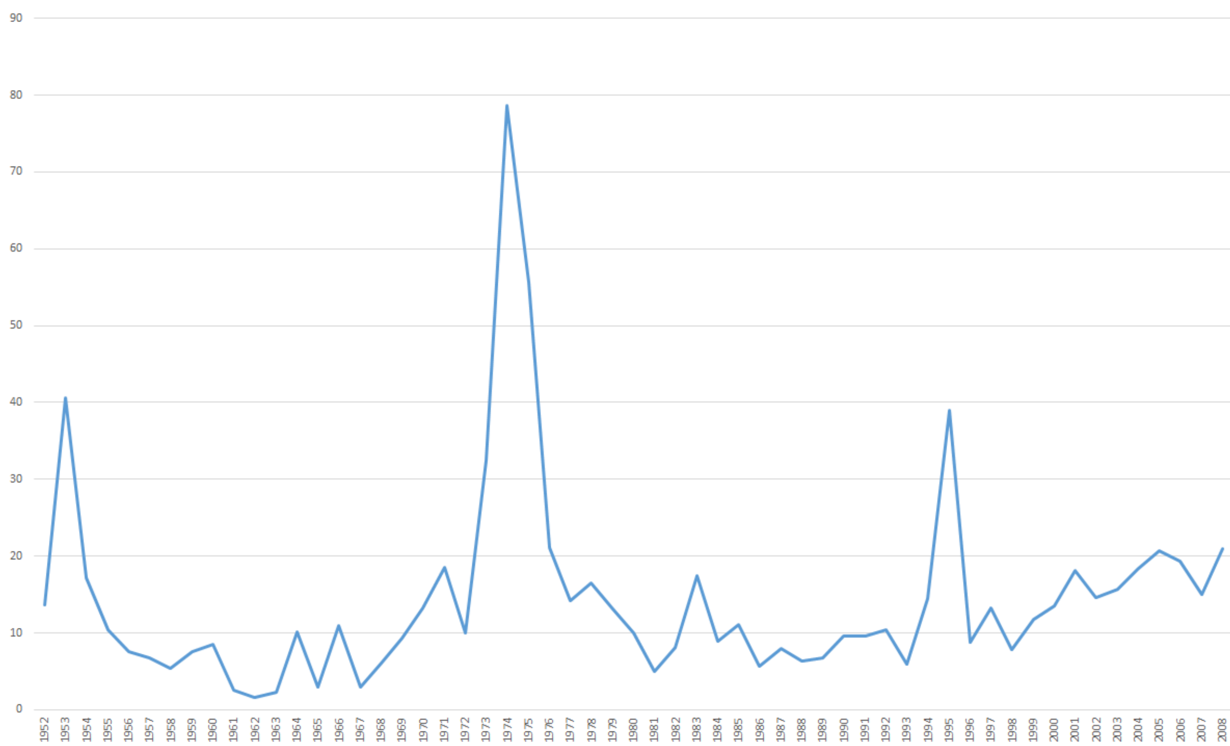
As a first step, we use the official newspaper of the central government, *People's Daily*, to gauge the political climate starting in 1952 until 2008: all *People's Daily* material has been under the direct control of the Chinese Communist Party (CCP)'s top leadership[4], and it provides direct information on the policies and viewpoints of the central government on a daily basis. Throughout the past decades, the CCP has started numerous top-down campaigns and movements, and each movement is communicated through well-chosen stock phrases. Through the changes in the appearance frequencies of these political phrases, we can observe the CCP's ideological emphasis in any given time period. In this study, we counted the annual number of articles that contain the phrases "Gender Equality" and "Half Sky"

[3] http://www.bbc.com/zhongwen/simp/indepth/2010/03/100308_ana_chinese_women.shtml

[4] People's Daily is under the direct control of the Central Propaganda Department, and the Central Propaganda Department decides the senior appointments of People's Daily. The managing director of People's Daily is also a government official whose administrative level is equivalent to province governor or department governor.

and divided that by the total number for articles that year. We report this ratio in Figure 3.1. The mainstream media in China is rarely politically incorrect on topic of gender equality, and when this topic was discussed, the mainstream media would always take the progressive stance. Therefore, when this phrase is mentioned by *People's Daily*, the article can only be supportive to the value of gender equality. As a result, the more articles mentioning gender equality indicates the more emphasis the CCP placed on propagating gender equality. From the graph, we can see that there was significant year to year variation in the frequency of these articles.

Figure 3.1: Number of Articles Mentioning "Gender Equality" or "Half Sky" per 10,000 Articles 1952-2008



Data Sources: https://www.oriprobe.com/peoplesdaily.shtml

3.4.2 Provincial Variation in Propaganda Implementation

Second, in order to show that the variation in individual's attitudes is caused by propaganda rather than a mere result of a time trend, we use the coverage of mass media to exploit provincial variation in propaganda intensities. From 1952 to 2008, the form of mass media has expanded from radio to television. We collect the provincial level data on radio and television signal coverage,[5] and use the maximum signal coverage of radio and television at the provincial level to proxy the provincial variation of media intensity.

3.4.3 Attitudes and Demographic Information

The third main dataset we use is the China General Social Survey (CGSS), a multi-year cross-sectional survey of China's urban and rural households. The survey contains several questions related to people's attitudes towards gender equality. It also contains individual level demographic information that we used as control variables in our regression model. Modeled after the General Social Survey (GSS) in the U.S., this survey is designed to gather longitudinal data on social trends and the changing relationship between social structure and quality of life in China. This annual or biannual survey started in 2003 by researchers at Renmin University and Hong Kong University of Science and Technology. Some survey modules were conducted in every wave, while others were conducted in selected waves for feasibility reasons. Our study uses the 2010, 2012 and 2013 waves since the module containing questions on gender-related attitudes was conducted in these three waves. There are approximately 30,000 observations total in these three cross-sections. Among the 30,000 observations, 7477 individuals were born between 1952 and 1983, answered the gender attitude related questions, and always lived within one province[6] (see more details in Data Appendix).

We use five questions from the relevant module. These questions asked the respondents how much they agreed with a certain statement on a scale of 1-5, with 5 being "completely agree":

[5]We obtained this data from the National Bureau of Statistics of the People's Republic of China.

[6]We exclude observations from individuals who used to migrate cross provinces, since I cannot identify the true media intensity they received. With this exclusion, the number of observations falls from 9109 to 7477.

1. Within a family, women should spend most of their time at home, and men should spend most of their time on career.

2. Male's ability is innately higher than that of the female.

3. It is more important for women to marry well than having a career.

4. When the economy is bad, women should be fired first.

5. Couple should share housework equally.

The above five statements can be divided into two groups: Statements 2, 3, 4 put emphasis on women's careers, while statements 1 and 5 put emphasis on women's responsibility to family and housework. Since related propaganda focused on the idea that women should have a career and take care of their family, we expect that propaganda has negative effects on people's agreement to statements 2, 3, 4, while related propaganda will have less or reversed effects on statements 1 and 5.

Another question that we can discuss in CGSS 2010-2013 is "Without One-Child Policy, how many boys and girls will you give birth to?" Based on the question, we generate one variable "OnlyBoy". If the number of boys is greater than zero and the number of girls is equal to zero, then $OnlyBoy = 1$. It shows a strong preference on a boy.

In the following empirical strategies, we discuss these questions separately.

3.5 Empirical Strategy

3.5.1 Baseline Strategy

Our baseline specification is the following:

$$
\begin{aligned}
Attitudes_{ipct} = \delta_p + \eta_t + \phi_c + \pi_{pc} + \sum_k \beta_k (PeoplesDaily \times MediaCoverage_p)_{ik} \\
+ \sum_k \gamma_k ProvincialReality_{ipk} + X_i \theta + \varepsilon_{ipct}, \qquad (3.1)
\end{aligned}
$$

where $Attitudes_{ipct}$ indicates the response to each question of individual i who belongs to birth cohort c, was interviewed in year t, and resided in province p from birth to the year of interview. k donates our interested age range: preschool 0-5, elementary school 6-12, middle school 13-14, and young adult 18-25. $PeoplesDaily$ is a vector of the number of articles with the keywords "Gender Equality" per 10,000 articles in years during age range k. $MediaCoverage_p$ is the average percentage of population that was covered by radio or television signal in province p during age range k. $ProvincialRealiy_{pk}$ is a vector of provincial observed reality during age range k, include average GDP per capita, average percentage of Internet users, average gender gap of year of schooling among people between 25-29, average percentage of female deputy in National Congress. X_i is a vector of demographic variables[7].

We also include provincial fixed effects δ_p, cohort fixed effects π_c[8], province-cohort fixed effects π_{pc} and interview year dummy η_t in all the specifications.

We further discuss the effects gap between women and men by adding the interaction of the male dummy with $(PeoplesDaily \times MediaCoverage_p)_{ik}$:

$$
\begin{aligned}
Attitudes_{ipct} = {} & \delta_p + \eta_t + \phi_c + \pi_{pc} \\
& + \sum_k \beta_k (PeoplesDaily \times MediaCoverage_p)_{ik} \\
& + \sum_k \tau_k (PeoplesDaily \times MediaCoverage_p)_{ik} \times Male_i \\
& + \sum_k \gamma_k ProvincialReality_{ipk} + X_i\theta + \varepsilon_{ipct} \qquad (3.2)
\end{aligned}
$$

Additionally, for the five statements, we undertake ordered probit and logit regressions instead of the simple OLS strategy, since the answers are ranked by integers. For whether people only want boys without the One-Child Policy, we undertake probit and logit regressions, since the dependent variable is a dummy.

[7]Demographic variables include gender dummy, minority dummy, rural household registration dummy, employment status variables, education variables, communist party member dummy, father's communist party member dummy, mother's communist party member dummy, mother's educations variables, mother's employment status when respondent was 14 years old, employer variables, house rent dummy, log household income per capita, children's gender.

[8]According to birth year, we divide all people who were born between 1952 and 1983 into 7 cohorts: 1952-1955, 1956-1960, 1961-1965, 1966-1970, 1971-1975, 1976-1980, 1981-1983.

3.5.2 Problem of Endogeneity: IV Strategy

One crucial identification concern is whether $(PeoplesDaily \times MediaCoverage_p)_{ik}$ is exogenous? In detail, are $MediaCoverage_p$ and $PeoplesDaily$ exogenous? There is no reverse causality in either variable since survey respondents could not affect either top-down propaganda or provincial media coverage. The endogeneity problem mainly comes from possible omitted variables.

For $MediaCoverage_p$, there are several factors that determine the coverage of mass media. The main factor is economic development. The correlation between $\ln(GDPpercapita_p)$ and $MediaCoverage_p$ is as high as 0.8395. Economic development also positively affected people's attitudes towards gender equality. From the 1940s until now, the key function of media in China is acting as the mouthpiece of the party, and the main aim is to propagate socialism with Chinese characteristics. As Mao Zedong said "Women can be truly emancipated only if class is emancipated", gender equality is always affiliated to socialism. It is just a small part of socialism with Chinese characteristics, thus there is no evidence to show that government increased media coverage to promote gender equality. In general, conditional on economic development, provincial media coverage is exogenous to people's attitudes towards gender equality.

The top-down propaganda $PeoplesDaily$ may not be exogenous, but some exogenous events increase or decrease the propaganda intensity toward gender equality. It helps us to generate an instrument variable to deal with the problem of endogeneity. The first event is the National Women's Conference. It was held every four years from 1949 to 1957, interrupted by a series of political movements and re-started in 1978 and has been held every five years since. The red line in Figure 3.2 shows the timing of all the National Women's Conferences. The conferences held in 1953, 1978, and 1983 are correlated with the propaganda peaks.

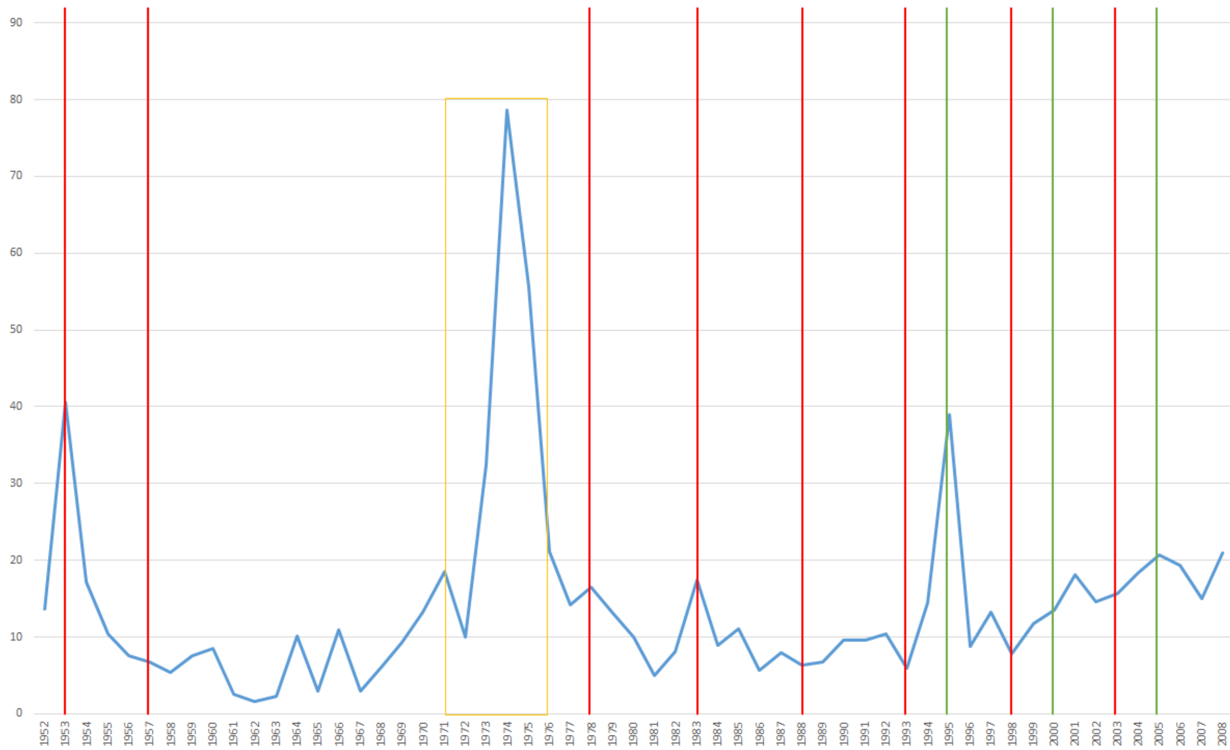
The second event is after the UN World Conference on Women held in Beijing in 1995. There were a series of commemorative activities to propagate gender equality in the five year anniversary in 2000 and ten year anniversary in 2005. The timing of the conference and the following anniversaries are exogenous, since the timing is determined by the United Nations and China received the bid for hosting the conference in 1990. The green line in Figure 3.2 shows the timing of the UN World Conference on Women held in Beijing and the following every five year anniversaries. These events are correlated with propaganda peaks.

Third, there was a propaganda peak during the last five years of the Cultural Revolution (1971-1976), especially in 1974 and 1975[9]. This was when "Gang of Four" led by Mao Zedong's last wife Jiang Qing fully controlled the power organs of the Communist Party of China. They controlled many of China's political institutions, including the media and propaganda. On the one hand, as a woman, Jiang Qing wanted to achieve a higher political position, thus promoting the propaganda about gender equality. On the other hand, as "Gang of Four", the political power struggles among different political factions decreased relatively, thus class struggle was not the only focus of the propaganda and it could focus on more issues other than class struggle. The beginning and end of her influence was determined by other political emergencies. In 1971, Mao's designated successor Lin Biao attempted to flee following a botched coup against Mao and died in air crash. It became Jiang Qing's opportunity. However, in 1976, Mao Zedong died and Jiang Qing lost power soon after. Therefore, we believe the propaganda peak during 1971 and 1976 are exogenous. The yellow rectangle in Figure 3.2 shows the timing.

Another concern is that $(PeoplesDaily \times MediaCoverage_p)_{ik}$ might be correlated with provincial reality about gender equality, and the reality may have affected attitudes formation. To deal with this problem, we control several variables that reflect provincial reality about gender reality, including the average gender gap of year of schooling among people between 25-29 in age range k and the average percentage of female deputies to National Congress in age range k.

[9](BFM16) argues that the Cultural Revolution promoted the propaganda of gender equality. However, Figure 3.2 shows that in the first five years of the Cultural Revolution, the propaganda of gender equality hit bottom since at the beginning of the Cultural Revolution, propaganda focused on "Class Struggle" and ignored all other issues. The peak of the propaganda of gender equality was in the last five years of the Cultural Revolution, following when Mao Zedong's wife Jiang Qing fully took power of propaganda and media. It is hard to believe the Cultural Revolution itself promoted gender equality.

Figure 3.2: Number of Articles Mentioning "Gender Equality" or "Half Sky" per 10,000 Articles and Related Exogenous Events 1952-2008



Notes: Red lines indicate the year of National Women's Conference. Green lines refer to UN World Conference on Women and its following 5-year and 10-year anniversary. The yellow rectangle refers to the timing that Mao Zedong's wife Jiang Qing fully took power of propaganda and media.

We also control for average percentage of Internet users in age range k[10] because the development of Internet attenuates the influence of both traditional mass media and top-down propaganda. Although the Chinese government built what is known as "the Great Firewall" to block residents from accessing certain contents on the Internet, Chinese people had more freedom on Internet before 2008 than today.[11] People could get more information than what the government intended to deliver to them.

3.5.3 Problem of Endogeneity: Taiwan as A Placebo Test

To further eliminate the potential endogeneity of top-down propaganda, we use Taiwan as a placebo to discuss whether there exists any spurious correlation between the top-down propaganda in mainland China times county level mass media coverage in Taiwan and Taiwanese' attitudes towards gender equality. Taiwan and mainland China have been ruled by different authorities since 1949. Although the government of China (PRC) claims the sovereignty of Taiwan, it never has the effective jurisdiction, and the contacts between the two sides were forbidden until 1987, thus the propaganda in mainland China should not affect Taiwanese attitudes towards gender equality. Besides, the majority of people in both sides are ethnic Chinese and share a similar culture. The similarity makes them comparable with each other. If there exists significant spurious effects, there might be omitted variables that correlate people's attitudes in Greater China. In addition, same as CGSS, the 2012 wave of the Taiwan Social Change Survey (TSCS) is also a part of the International Social Survey Programme (ISSP). All surveys in this programme included similar demographic information and the data availability provides us the possibility of comparison. The gender attitudes related questions are not exactly the same in the TSCS and CGSS. We use seven questions from the related module in TSCS 2012. As with CGSS, these questions asked the respondents how much they agreed with a certain statement on a scale of 1-5, with 5 being

[10] The data is available after 1997. We can assume that the percentage of Internet users prior to 1997 was virtually zero even in the most economically advanced cities since the Internet was not common before 1997.

[11] Google was blocked in 2010 and Wikipedia was blocked in 2004.

"completely agree":

1. A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.

2. A pre-school child is likely to suffer if his or her mother works.

3. All in all, family life suffers when the woman has a full-time job.

4. A job is all right, but what most women really want is a home and children.

5. Both the man and woman should contribute to the household income.

6. A man's job is to earn money; a woman's job is to look after the home and family.

7. Politics is man's thing; a woman is best not to participate.

We do the same analysis for the 2012 TSCS. We hope there is no spurious correlation between $(PeoplesDaily \times MediaCoverage_p)_{ik}$ and Taiwanese' attitudes towards gender equality.

3.5.4 Problem of Endogeneity: Difference-in-Differences

Statement 6 in TSCS 2012, "A man's job is to earn money; a woman's job is to look after the home and family", is very similar to the statement 1 in CGSS, "Within a family, women should spend most of their time at home, and men should spend most of their time on career". We append data in CGSS with data in TSCS 2012 together and perform the following difference-in-differences analysis:

$$\begin{aligned}
Attitudes_{ipct} = {} & \delta_p + \eta_t + \phi_c + \pi_{pc} \\
& + \sum_k \beta_k (PeoplesDaily \times MediaCoverage_p)_{ik} \\
& + \sum_k \lambda_k (PeoplesDaily \times MediaCoverage_p)_{ik} \times MainlandChina_i \\
& + \sum_k \gamma_k ProvincialReality_{ipk} + \alpha MainlandChina_i + X_i\theta + \varepsilon_{ipct}, \qquad (3.3)
\end{aligned}$$

in which Taiwanese are a control group. If propaganda affects people's attitudes towards gender equality, the coefficients λ_k should be significantly negative.

We also discuss the effects gap between women and men by the following specification:

$$\begin{aligned}
Attitudes_{ipct} = {} & \delta_p + \eta_t + \phi_c + \pi_{pc} \\
& + \sum_k \beta_k (PeoplesDaily \times MediaCoverage_p)_{ik} \\
& + \sum_k \lambda_k (PeoplesDaily \times MediaCoverage_p)_{ik} \times MainlandChina_i \\
& + \sum_k \tau_k (PeoplesDaily \times MediaCoverage_p)_{ik} \times Male_i \\
& + \sum_k \kappa_k (PeoplesDaily \times MediaCoverage_p)_{ik} \times Male_i \times MainlandChina_i \\
& + \mu MainlandChina_i \times Male_i \\
& + \sum_k \gamma_k ProvincialReality_{ipk} + \alpha MainlandChina_i + X_i\theta + \varepsilon_{ipct} \qquad (3.4)
\end{aligned}$$

3.6 Results and Discussions

In this section, we show the results of the above strategies and discuss the effects of propaganda on each question separately.

3.6.1 Baseline and IV Strategy

3.6.1.1 Within a family, women should spend most of their time at home, and men should spend most of their time on career

Table 3.1 shows the results of propaganda effects on people's attitudes towards the statement "Within a family, women should spend most of their time at home, and men should spend most of their time on career". In all columns, propaganda is negatively correlated with people's agreement to the statement in all age ranges, but the effects are not significant in all columns of the age range 0-5, 13-17, 18-25, and the significance in the age range 6-12 is only significant in the weighted regression.

Table 3.1: Propaganda Effects on Division of Labor between Men and Women

Within a family, women should spend most of their time at home, and men should spend most of their time on career	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.00450	-0.0158	-0.00818	-0.0302	-0.00460	-0.0106
	(0.0135)	(0.0134)	(0.0224)	(0.0227)	(0.0206)	(0.0221)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0223	-0.0359**	-0.0403	-0.0643**	-0.0215	-0.0281*
	(0.0140)	(0.0166)	(0.0260)	(0.0311)	(0.0198)	(0.0164)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0149	-0.0191	-0.0288	-0.0327	-0.0109	-0.0127
	(0.0152)	(0.0174)	(0.0266)	(0.0299)	(0.0174)	(0.0147)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0263	-0.0408	-0.0469	-0.0688	-0.0241	-0.0305
	(0.0226)	(0.0314)	(0.0396)	(0.0532)	(0.0366)	(0.0328)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputies in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.2: Propaganda Effects on Division of Labor between Men and Women: Difference by Gender

Within a family, women should spend most of their time at home, and men should spend most of their time on career	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda $\times$Media Coverage $0-5$	-0.00839 (0.0160)	-0.0219 (0.0157)	-0.0142 (0.0273)	-0.0400 (0.0270)	-0.00929 (0.0219)	-0.0159 (0.0237)
Gender Equality in Propaganda $\times$Media Coverage $6-12$	-0.0264** (0.0133)	-0.0403** (0.0157)	-0.0463* (0.0244)	-0.0711** (0.0299)	-0.0244 (0.0204)	-0.0292* (0.0175)
Gender Equality in Propaganda $\times$Media Coverage $13-17$	-0.0176 (0.0158)	-0.0216 (0.0173)	-0.0312 (0.0279)	-0.0348 (0.0299)	-0.0135 (0.0178)	-0.0148 (0.0144)
Gender Equality in Propaganda $\times$Media Coverage $18-25$	-0.0319 (0.0202)	-0.0449 (0.0302)	-0.0552 (0.0347)	-0.0753 (0.0505)	-0.0284 (0.0354)	-0.0320 (0.0318)
Male $\times$Gender Equality in Propaganda $\times$Media Coverage $0-5$	0.00525 (0.00531)	0.00953 (0.00642)	0.00868 (0.0106)	0.0160 (0.0124)	0.00706 (0.00545)	0.0102 (0.00689)
Male $\times$Gender Equality in Propaganda $\times$Media Coverage $6-12$	0.00732 (0.00605)	0.0101 (0.00679)	0.0116 (0.0107)	0.0168 (0.0126)	0.00770 (0.00612)	0.00883 (0.00715)
Male $\times$Gender Equality in Propaganda $\times$Media Coverage $13-17$	0.00484 (0.00373)	0.00585** (0.00273)	0.00442 (0.00681)	0.00571 (0.00528)	0.00557* (0.00288)	0.00699*** (0.00235)
Male $\times$Gender Equality in Propaganda $\times$Media Coverage $18-25$	0.0101 (0.00788)	0.00967 (0.00866)	0.0149 (0.0145)	0.0151 (0.0161)	0.0101 (0.00719)	0.00993 (0.00772)
Male	-0.130 (0.128)	-0.196 (0.132)	-0.142 (0.218)	-0.273 (0.244)	-0.144 (0.124)	-0.194 (0.130)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.2 shows the difference of the propaganda effects on the gender index. For this statement, the difference between men and women's attitudes are not robustly significant. Propaganda's effects are significantly negative on women's agreement towards the statement in the age range 6-12, except for unweighted IV strategies. If there was 1 more articles related to gender equality per 10,000 articles effectively delivered to people in the age range 6-12, it would reduce women's agreement towards the statement by 0.0292 (out of 4) (column 4).

3.6.1.2 Male's ability is innately higher than that of the female

Table 3.3 indicates the negative effects of propaganda on people's agreement on the statement "Male's ability is innately higher than that of the female". The effects are not always robustly significant in all age ranges. In the age range 0-5, the effects are significant in all ordered probit regressions and the weighted ordered logit regression, but not significant in the unweighted ordered probit regression or in any IV regressions. In the age range 6-12, the effects are significant in all columns except for the unweighted logit regression. If there was 1 more article related to gender equality per 10,000 articles effectively delivered to people in the age range 6-12, it would reduce people's agreement towards the statement by 0.0496 (out of 4) (column 4). In the age range 13-17, the negative effects are robustly significant in all columns. If there was 1 more articles related to gender equality per 10,000 articles effectively delivered to people in the age range 13-17, it would reduce people's agreement towards the statement by 0.0487 (out of 4) (column 4). In the age range 18-25, the effects are only significant in the weighted IV regression.

Table 3.4 shows the difference of the propaganda effects on people's agreement to the statement between women and men. Results are significant across most columns of our interested variables. In general, holding other variables constant, without propaganda, men had more progressive attitudes toward women's ability, but propaganda had larger effects on women than men. On average, the agreement of men was 0.551 lower than women (out of 4) (column 4). Related propaganda significantly promoted women's attitudes towards acknowledging women's ability in all age ranges. If there was 1 more article related to

Table 3.3: Propaganda Effects on Attitudes towards Women's Ability

Male's ability is innately higher than that of the female	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0135*	-0.0241***	-0.0184	-0.0377***	-0.0123	-0.0170
	(0.00785)	(0.00691)	(0.0125)	(0.0117)	(0.0134)	(0.0109)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0291**	-0.0368**	-0.0429	-0.0574**	-0.0470**	-0.0496**
	(0.0144)	(0.0143)	(0.0266)	(0.0275)	(0.0224)	(0.0218)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0299**	-0.0343***	-0.0453*	-0.0548**	-0.0468**	-0.0487***
	(0.0134)	(0.0130)	(0.0250)	(0.0249)	(0.0202)	(0.0184)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0147	-0.0306	-0.0142	-0.0454	-0.0525	-0.0545*
	(0.0179)	(0.0187)	(0.0350)	(0.0353)	(0.0344)	(0.0292)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

gender equality per 10,000 articles effectively delivered to people in the age range 0-5, it would reduce women's agreement to the statement by 0.0276 (out of 4); in the age range 6-12, it would reduce women's agreement to the statement by 0.0580 (out of 4); in the age range 13-17, it would reduce women's agreement to the statement by 0.0540 (out of 4); in the age range 18-25, it would reduce women's agreement to the statement by 0.0684 (out of 4) (column 4). Propaganda effects on men were significantly lower than the ones on women. The difference was 0.0151 in the age range 0-5, 0.0205 in the age range 6-12, 0.0104 in the age range 13-17, 0.0258 in the age range 18-28 (column 4).

3.6.1.3 It is more important for women to marry well than having a career

Table 3.5 indicates the negative effects of propaganda on people's agreement on the statement "It is more important for women to marry well than having a career". The effects are only significant in the age range 0-5. If there was 1 more articles related to gender equality per 10,000 articles effectively delivered to people in the age range 0-5, it would reduce people's agreement towards the statement by 0.0366 (out of 4) (column 4).

Table 3.6 shows the difference of the propaganda effects on people's agreement towards the statement between women and men. In general, without propaganda, men have more progressive attitudes than women, but the difference is not robustly significant. Related propaganda significantly promoted women's attitudes towards the importance of career in the age range 0-5. If there was 1 more article related to gender equality per 10,000 articles effectively delivered to people in the age range 0-5, it would reduce women's agreement to the statement by 0.0444 (out of 4). The negative propaganda effects on women's agreement towards this statement are not significant in other age ranges. Propaganda effects on men are significantly lower than the ones on women. The difference is 0.0130 and significant in the age range 0-5 (column 4). The difference is not significant in other age ranges.

Table 3.4: Propaganda Effects on Attitudes towards Women's Ability: Difference by Gender

Male's ability is innately higher than that of the female	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0240**	-0.0361***	-0.0366**	-0.0576***	-0.0219	-0.0276**
	(0.00999)	(0.00859)	(0.0164)	(0.0139)	(0.0138)	(0.0115)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0410***	-0.0456***	-0.0644**	-0.0744***	-0.0578***	-0.0580***
	(0.0144)	(0.0138)	(0.0257)	(0.0263)	(0.0222)	(0.0211)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0381***	-0.0402***	-0.0608**	-0.0664***	-0.0545***	-0.0540***
	(0.0131)	(0.0133)	(0.0243)	(0.0255)	(0.0193)	(0.0184)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0288*	-0.0425**	-0.0396	-0.0670**	-0.0661*	-0.0684**
	(0.0167)	(0.0177)	(0.0330)	(0.0334)	(0.0354)	(0.0280)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	0.0142***	0.0176***	0.0238***	0.0288***	0.0123***	0.0151***
	(0.00404)	(0.00364)	(0.00689)	(0.00557)	(0.00420)	(0.00356)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.0209***	0.0183***	0.0367***	0.0328***	0.0231***	0.0205***
	(0.00630)	(0.00574)	(0.0107)	(0.00983)	(0.00741)	(0.00730)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	0.0143***	0.0119**	0.0262***	0.0217**	0.0144***	0.0104**
	(0.00458)	(0.00489)	(0.00814)	(0.00866)	(0.00467)	(0.00530)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	0.0247***	0.0242***	0.0417***	0.0416***	0.0255***	0.0258***
	(0.00858)	(0.00732)	(0.0161)	(0.0131)	(0.00885)	(0.00781)
Male	-0.593***	-0.555***	-1.023***	-0.960***	-0.600***	-0.551***
	(0.180)	(0.141)	(0.327)	(0.252)	(0.195)	(0.155)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.5: Propaganda Effects on Attitudes towards the Importance of Marriage and Career to Women

It is more important for women to marry well than having a career	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda $\times$Media Coverage $0-5$	-0.0302***	-0.0364***	-0.0603***	-0.0714***	-0.0423***	-0.0366**
	(0.0109)	(0.0128)	(0.0181)	(0.0203)	(0.0139)	(0.0180)
Gender Equality in Propaganda $\times$Media Coverage $6-12$	-0.0183	-0.0242	-0.0362	-0.0472	-0.0325	-0.0266
	(0.0193)	(0.0218)	(0.0325)	(0.0362)	(0.0227)	(0.0265)
Gender Equality in Propaganda $\times$Media Coverage $13-17$	-0.0105	-0.0136	-0.0202	-0.0257	-0.00811	0.000271
	(0.0170)	(0.0186)	(0.0311)	(0.0335)	(0.0206)	(0.0235)
Gender Equality in Propaganda $\times$Media Coverage $18-25$	-0.0160	-0.0280	-0.0254	-0.0485	-0.0343	-0.0322
	(0.0244)	(0.0247)	(0.0459)	(0.0440)	(0.0312)	(0.0320)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.6: Propaganda Effects on Attitudes towards the Importance of Marriage and Career to Women: Difference by Gender

It is more important for women to marry well than having a career	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0378***	-0.0441***	-0.0740***	-0.0847***	-0.0510***	-0.0444**
	(0.0122)	(0.0133)	(0.0199)	(0.0209)	(0.0147)	(0.0189)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0208	-0.0250	-0.0380	-0.0468	-0.0350	-0.0273
	(0.0187)	(0.0214)	(0.0310)	(0.0352)	(0.0221)	(0.0260)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0127	-0.0136	-0.0233	-0.0244	-0.0101	0.000743
	(0.0162)	(0.0181)	(0.0299)	(0.0327)	(0.0193)	(0.0229)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0190	-0.0297	-0.0289	-0.0511	-0.0370	-0.0352
	(0.0233)	(0.0231)	(0.0435)	(0.0402)	(0.0301)	(0.0304)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	0.0117***	0.0126***	0.0216***	0.0221***	0.0132***	0.0130***
	(0.00403)	(0.00435)	(0.00624)	(0.00708)	(0.00337)	(0.00354)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.00473	0.00334	0.00568	0.00375	0.00501	0.00310
	(0.00520)	(0.00553)	(0.00960)	(0.0104)	(0.00600)	(0.00651)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	0.00425	0.000879	0.00714	-0.0000834	0.00317	-0.000512
	(0.00358)	(0.00464)	(0.00640)	(0.00871)	(0.00371)	(0.00457)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	0.00534	0.00470	0.00703	0.00785	0.00442	0.00462
	(0.00820)	(0.00899)	(0.0144)	(0.0158)	(0.00785)	(0.00867)
Male	-0.293**	-0.235*	-0.459*	-0.366	-0.286*	-0.220
	(0.145)	(0.137)	(0.258)	(0.248)	(0.153)	(0.150)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputies in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

3.6.1.4 When the economy is bad, women should be fired first

Table 3.7 indicates the negative effects of propaganda on people's agreement on the statement "When the economy is bad, women should be fired first". In the age range 0-5, the significance is not robust in probit and logit regressions, but robust in IV strategies. If there was 1 more article related to gender equality per 10,000 articles effectively delivered to people in the age range 0-5, it would reduce people's agreement towards the statement by 0.0318 (out of 4). The results are robustly significant in the age range 6-12, it would reduce people's agreement towards the statement by 0.0534 (out of 4) (column 4). In other age ranges, the effects are not robustly significant.

Table 3.7: Propaganda Effects on Attitudes towards Whether Fire Women First in Bad Economy

When the economy is bad, women should be fired first	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda ×Media Coverage 0 − 5	-0.0181* (0.0103)	-0.0154 (0.0123)	-0.0286 (0.0181)	-0.0255 (0.0217)	-0.0312*** (0.00723)	-0.0318*** (0.00766)
Gender Equality in Propaganda ×Media Coverage 6 − 12	-0.0336** (0.0170)	-0.0284* (0.0165)	-0.0543* (0.0293)	-0.0468* (0.0270)	-0.0501*** (0.0165)	-0.0534*** (0.0131)
Gender Equality in Propaganda ×Media Coverage 13 − 17	-0.00633 (0.0180)	-0.00900 (0.0166)	-0.0150 (0.0307)	-0.0212 (0.0280)	-0.00948 (0.0167)	-0.0149 (0.0135)
Gender Equality in Propaganda ×Media Coverage 18 − 25	-0.00570 (0.0275)	-0.00750 (0.0273)	-0.00927 (0.0454)	-0.0127 (0.0451)	-0.0363 (0.0231)	-0.0408* (0.0210)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.8 shows the difference of the propaganda effects on people's agreement towards the statement between women and men. In general, without propaganda, the difference between men and women is not significant. Related propaganda significantly reduces women's agreement towards the statement in all columns in the age range 0-5 and 6-12. The effects are also significant in the IV strategies of the age range 18-25. The effects are not significant in the age range 13-17. If there was 1 more article related to gender equality per 10,000 articles effectively delivered to people in the age range 0-5, it would reduce women's agreement to the statement by 0.0417 (out of 4); in the age range 6-12, it would reduce women's agreement to the statement by 0.0549 (out of 4); in the age range 18-25, it would reduce women's agreement to the statement by 0.0490 (out of 4) (column 4). In all age ranges, propaganda effects on men are not significantly different from the ones on women.

3.6.1.5 Couple should share housework equally

Different from other statements that propaganda has positive effects on people's attitudes towards gender equality, propaganda has reversed effects on the statement that "Couple should share housework equally". Table 3.9 indicates the positive effects of propaganda on people's agreement on the statement. In the age range 0-5, the significance is not significant in weighted IV strategies but is significant in all other columns. If there was 1 more article related to gender equality per 10,000 articles effectively delivered to people in the age range 0-5, it would reduce people's agreement towards the statement by 0.0117 (out of 4). The results are robustly significant in the age range 6-12 and 13-17. It would reduce people's agreement towards the statement by 0.0379 (out of 4) in the age range 6-12, and 0.0376 in the age range 13-17 (column 4). In the age range 18-25, the significance is not robust in IV strategies.

Table 3.10 shows the difference of the propaganda effects on people's agreement towards the statement between women and men. Propaganda has negative effects on women's agreement to this statement. In the age range 0-5, the significance is not significant in weighted IV strategies but significant in all other columns. If there was 1 more article related to gender equality per 10,000 articles effectively delivered to people in the age range 0-5, it

Table 3.8: Propaganda Effects on Attitudes towards Whether Fire Women First in Bad Economy: Difference by Gender

When the economy is bad, women should be fired first	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0267** (0.0113)	-0.0268** (0.0127)	-0.0443** (0.0198)	-0.0451** (0.0224)	-0.0389*** (0.00813)	-0.0417*** (0.00807)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0357** (0.0160)	-0.0299* (0.0159)	-0.0585** (0.0274)	-0.0505** (0.0251)	-0.0527*** (0.0160)	-0.0549*** (0.0123)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0112 (0.0184)	-0.0128 (0.0171)	-0.0242 (0.0312)	-0.0291 (0.0286)	-0.0145 (0.0173)	-0.0181 (0.0150)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0131 (0.0266)	-0.0124 (0.0284)	-0.0211 (0.0431)	-0.0216 (0.0446)	-0.0457** (0.0214)	-0.0490*** (0.0185)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	0.0124*** (0.00289)	0.0178*** (0.00294)	0.0229*** (0.00477)	0.0312*** (0.00491)	0.0103*** (0.00342)	0.0150*** (0.00374)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.00304 (0.00668)	0.00421 (0.00690)	0.00537 (0.0113)	0.00706 (0.0124)	0.00299 (0.00552)	0.00415 (0.00624)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	0.00825 (0.00535)	0.00791 (0.00597)	0.0153* (0.00879)	0.0149 (0.0102)	0.00649 (0.00584)	0.00512 (0.00686)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	0.0126 (0.00815)	0.0106 (0.00987)	0.0189 (0.0135)	0.0175 (0.0164)	0.0143* (0.00733)	0.0144 (0.00899)
Male	-0.0914 (0.144)	-0.0990 (0.132)	-0.129 (0.239)	-0.151 (0.225)	-0.123 (0.114)	-0.144 (0.105)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputies in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.9: Propaganda Effects on Attitudes towards Housework

Couple should share housework equally	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda	-0.0205***	-0.0187***	-0.0361***	-0.0311***	-0.0138*	-0.0117
$\times$Media Coverage $0-5$	(0.00686)	(0.00490)	(0.0136)	(0.0108)	(0.00775)	(0.0104)
Gender Equality in Propaganda	-0.0492***	-0.0503***	-0.0770***	-0.0813***	-0.0336*	-0.0379**
$\times$Media Coverage $6-12$	(0.0147)	(0.0167)	(0.0247)	(0.0271)	(0.0191)	(0.0182)
Gender Equality in Propaganda	-0.0390***	-0.0432***	-0.0581***	-0.0660**	-0.0298**	-0.0376***
$\times$Media Coverage $13-17$	(0.0128)	(0.0163)	(0.0219)	(0.0263)	(0.0133)	(0.0127)
Gender Equality in Propaganda	-0.0525**	-0.0549**	-0.0860**	-0.0908*	-0.0224	-0.0231
$\times$Media Coverage $18-25$	(0.0216)	(0.0272)	(0.0364)	(0.0467)	(0.0236)	(0.0227)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic information, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputies in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

would reduce women's agreement towards the statement by 0.0141 (out of 4). The results are robustly significant in the age range 6-12 and 13-17. It would reduce women's agreement towards the statement by 0.0337 (out of 4) in the age range 6-12, and 0.0388 (out of 4) in the age range 13-17 (column 4). In the age range 18-25, the significance is not robust in IV strategies. In general, the reversed effects are larger on men. The difference is robustly significant in the age range 0-5 and 13-17. The difference is 0.00753 (out of 4) in the age range 0-5, and 0.0138 (out of 4) in the age range 13-17.

3.6.1.6 Five Questions Summary

Figure 3.3 shows a comparison of propaganda effects on all people's attitudes towards gender equality measured by the above five questions and in all age ranges. In the age range 0-5, propaganda increased people's attitudes towards gender equality in the statements "It is more important for women to marry well than having a career" and "When the economy is bad, women should be fired first". In the age range 6-12, propaganda increased people's gender equality attitudes in the statements "Male's ability is innately higher than that of the female" and "When the economy is bad, women should be fired first". In the age range 13-17, propaganda increased people's attitudes towards gender equality in the statement "Male's ability is innately higher than that of the female". In the age range 18-25, propaganda effects are not robustly significant for any of the statements. For the statement "Couples should share housework equally", propaganda has reversed effects on people's attitudes towards gender equality in the age range 6-12 and 13-17.

Figure 3.4 shows a comparison of propaganda effects on all women's attitudes towards gender equality measured by the above five questions and in all age ranges. In the age range 0-5, propaganda increased women's attitudes towards gender equality in the statements "It is more important for women to marry well than having a career" and "When the economy is bad, women should be fired first". In the age range 6-12, propaganda increased women's gender equality attitudes in the statements "Male's ability is innately higher than that of the female" and "When the economy is bad, women should be fired first". In the age range 13-17,

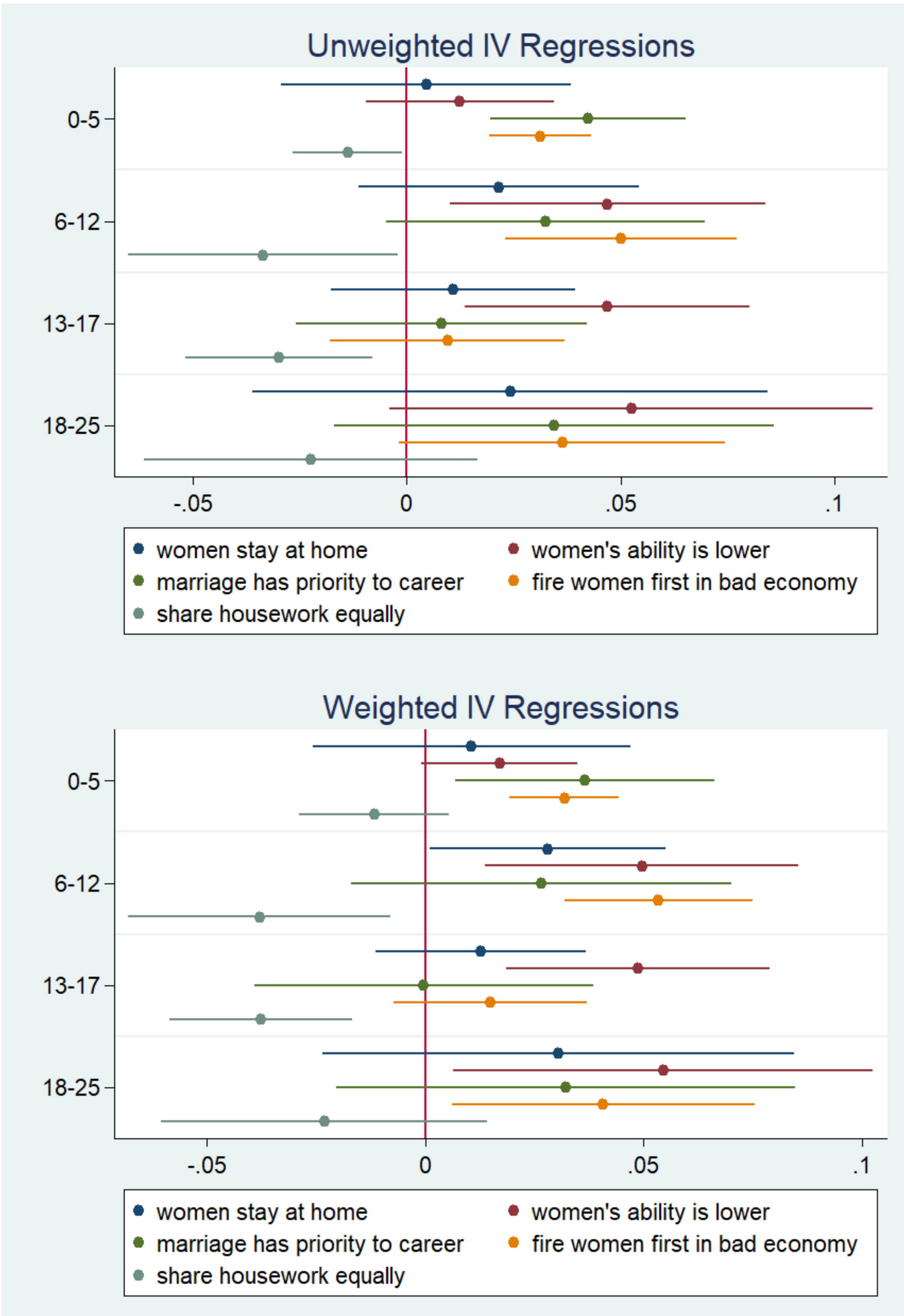
Table 3.10: Propaganda Effects on Attitudes towards Housework: Difference by Gender

Couple should share housework equally	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0217***	-0.0211***	-0.0384**	-0.0354***	-0.0144*	-0.0141
	(0.00819)	(0.00580)	(0.0162)	(0.0123)	(0.00836)	(0.0110)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0429***	-0.0454***	-0.0673***	-0.0736***	-0.0278	-0.0337*
	(0.0145)	(0.0170)	(0.0239)	(0.0275)	(0.0185)	(0.0183)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0383***	-0.0439**	-0.0577***	-0.0679**	-0.0279**	-0.0388***
	(0.0130)	(0.0177)	(0.0218)	(0.0283)	(0.0116)	(0.0127)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0441**	-0.0448	-0.0728**	-0.0742	-0.0154	-0.0163
	(0.0216)	(0.0277)	(0.0365)	(0.0478)	(0.0229)	(0.0218)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	0.00330	0.00572	0.00488	0.00876	0.00328	0.00585
	(0.00375)	(0.00371)	(0.00675)	(0.00640)	(0.00335)	(0.00381)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0110***	-0.00798**	-0.0185***	-0.0142**	-0.00881***	-0.00753**
	(0.00335)	(0.00369)	(0.00520)	(0.00565)	(0.00267)	(0.00378)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	-0.000825	0.00228	-0.000582	0.00430	-0.00113	0.00306
	(0.00921)	(0.00889)	(0.0167)	(0.0157)	(0.00956)	(0.00934)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0149***	-0.0174***	-0.0252***	-0.0304***	-0.0115*	-0.0138**
	(0.00512)	(0.00546)	(0.00957)	(0.00941)	(0.00646)	(0.00568)
Male	-0.0111	-0.0616	-0.0259	-0.0909	-0.0382	-0.0820
	(0.102)	(0.125)	(0.174)	(0.204)	(0.0967)	(0.113)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7477	7477	7477	7477	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Figure 3.3: Propaganda Effects on People's Attitudes to Specific Questions, by Age Range



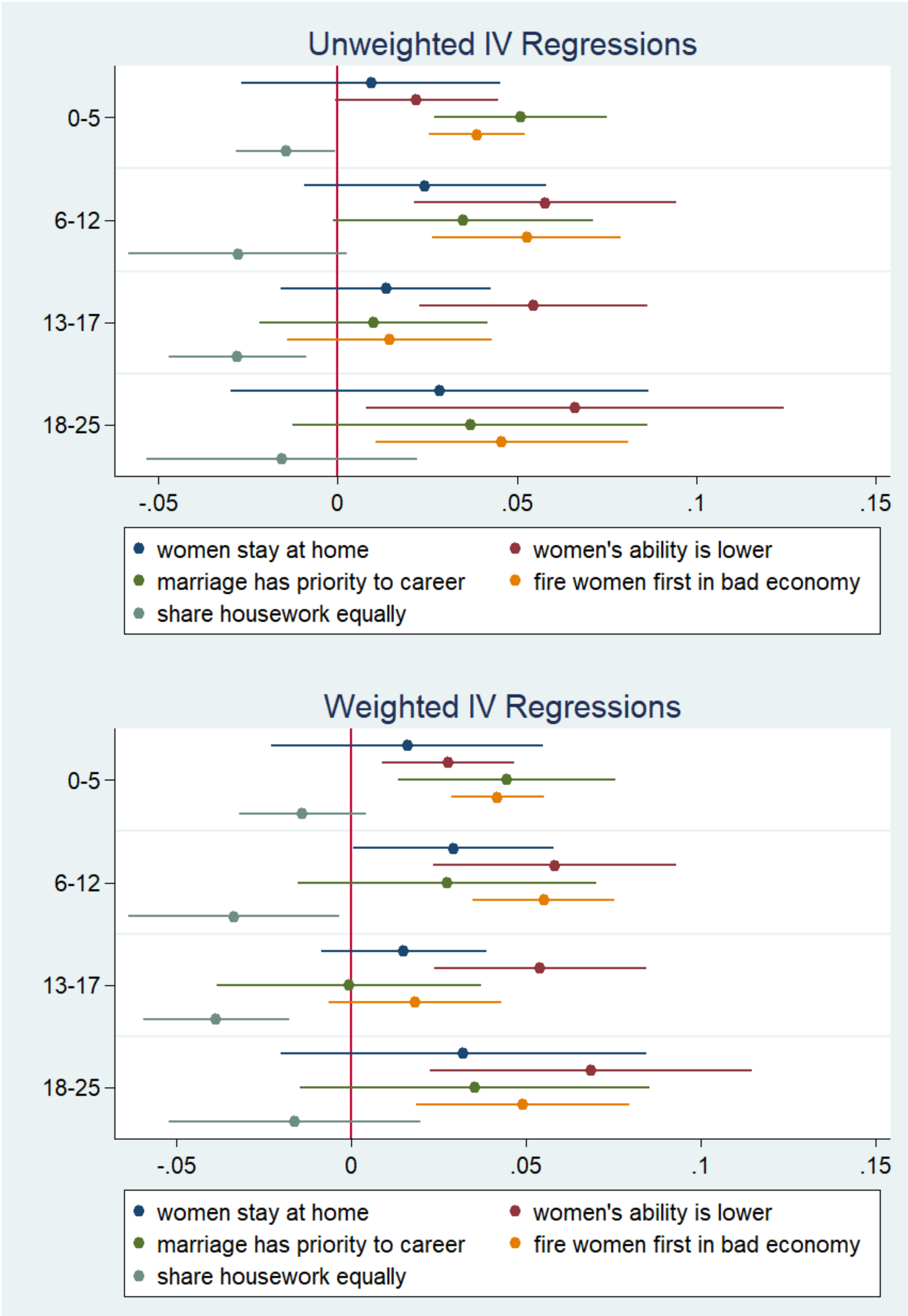
Notes: Effects larger than 0 indicates propaganda has positive effects on people's attitudes towards gender equality. Points indicate average effects. Lines indicate 90% confidence intervals.

propaganda increased women's attitudes towards gender equality in the statement "Male's ability is innately higher than that of the female". In the age range 18-25, propaganda increased women's gender equality attitudes in the statements "Male's ability is innately higher than that of the female" and "When the economy is bad, women should be fired first". For the statement "Couples should share housework equally", propaganda has significantly reversed effects on women's attitudes towards gender equality in the age range 13-17.

Figure 3.5 shows a comparison of propaganda effects on all men's attitudes towards gender equality measured by the above five questions and in all age ranges. In the age range 0-5, propaganda increased men's attitudes towards gender equality in the statements "It is more important for women to marry well than having a career" and "When the economy is bad, women should be fired first". In the age range 6-12, propaganda increased men's gender equality attitudes in the statement "When the economy is bad, women should be fired first". In the age range 13-17, propaganda increased men's attitudes towards gender equality in the statement "Male's ability is innately higher than that of the female". In the age range 18-25, propaganda had no significant effects on men's gender equality attitudes in any of the statements. For the statement "Couples should share housework equally", propaganda has significantly reversed effects on men's attitudes towards gender equality in the age range 13-17.

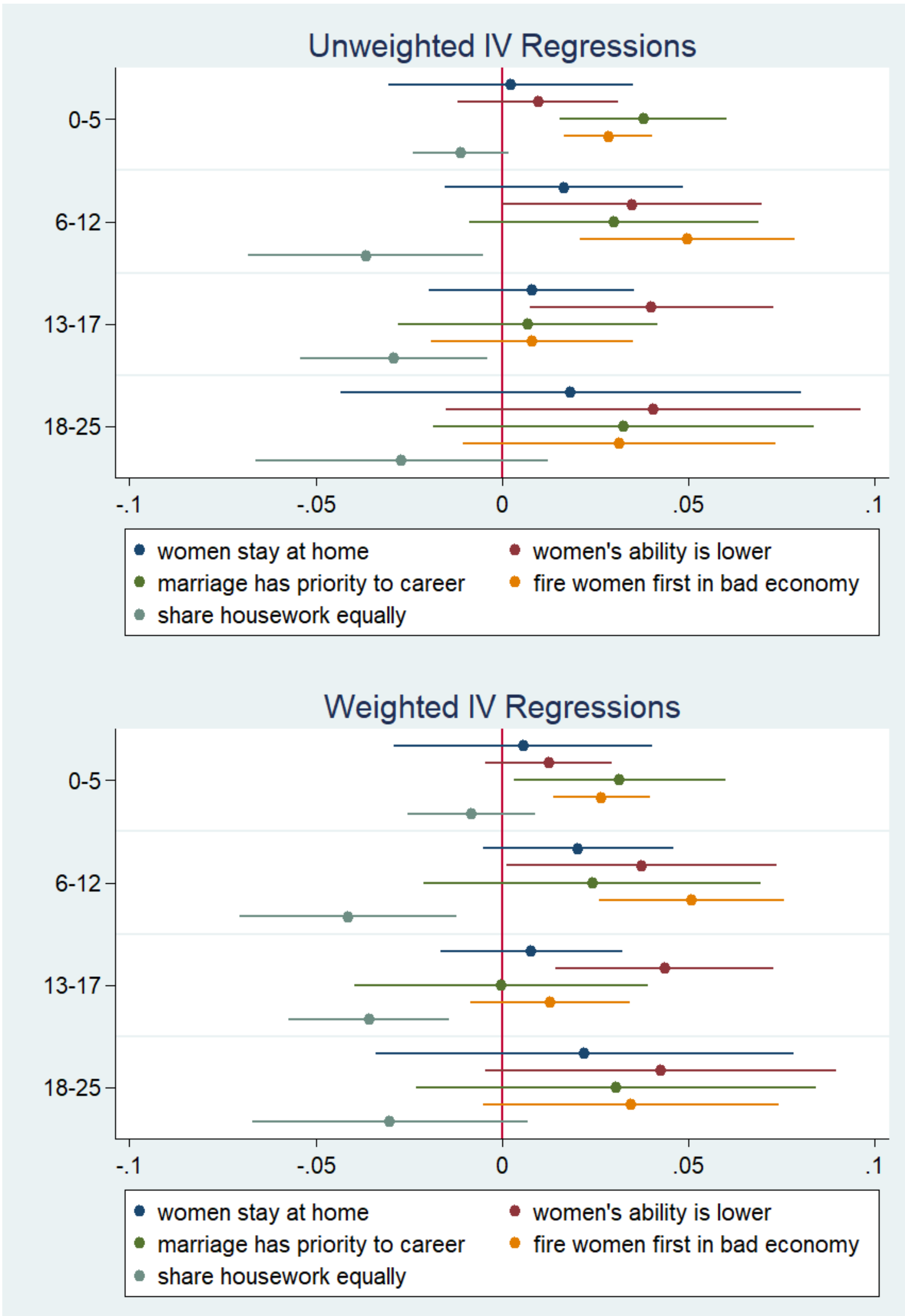
In summary, the results are consistent with our specification. For statements focusing on acknowledging women's ability and encouraging women to pursue their career, "Male's ability is innately higher than that of the female", "It is more important for women to marry well than having a career", and "When the economy is bad, women should be fired first", propaganda has more significant effects. Women are affected in all age ranges. "Male's ability is innately higher than that of the female" is affected in all age ranges. It is consistent with the propaganda that women can do anything that men can do, especially the slogan "Women hold up half the sky" and the song "Who Says Girls Can't Compete with Guys?". "It is more important for women to marry well than having a career" is affected in the younger ages of 0-5 and 6-12. "When the economy is bad, women should be fired first" is significantly affected in all age ranges except for 13-17.

Figure 3.4: Propaganda Effects on Women's Attitudes to Specific Questions, by Age Range



Notes: Effects larger than 0 indicates propaganda has positive effects on people's attitudes towards gender equality. Points indicate average effects. Lines indicate 90% confidence intervals.

Figure 3.5: Propaganda Effects on Men's Attitudes to Specific Questions, by Age Range



Notes: Effects larger than 0 indicates propaganda has positive effects on people's attitudes towards gender equality. Points indicate average effects. Lines indicate 90% confidence intervals.

The effects on men are smaller than women, and are significantly affected before adulthood. "Male's ability is innately higher than that of the female" is affected in the age ranges 6-12 and 13-17. "It is more important for women to marry well than having a career" is affected in the age range 0-5. "When the economy is bad, women should be fired first" is affected in the age ranges 0-5 and 6-12.

For the statement "Within a family, women should spend most of their time at home, and men should spend most of their time on career", the effects are not significant in any of the age ranges for both men and women. For the statement "Couple should share housework equally", the reversed effects are significant in the age range 6-12 and 13-17 for both women and men but has larger reversed effects on men. It is consistent with the propaganda that women are encouraged to do well in both family and career, and as a result, they experience the double burden.

3.6.1.7 Without One-Child Policy, Only Want Boys

Table 3.11 shows propaganda effects on people's boy preference. For women and men as a whole, propaganda has significantly negative effects in the age range 18-25, consistent with the marriage and childbearing ages. Table 3.12 shows there is no significant difference between men and women.

3.6.2 Placebo Test

In this section, we discuss whether there exists any spurious effects of the propaganda in mainland China on Taiwanese' attitudes towards gender equality. I use the baseline specification to discuss the correlation. In addition to all provincial realities used in our baseline analysis, we also control for democracy in all age ranges, since in a democratic society, women have the right to vote and it will affect people's attitudes towards gender equality. Taiwan began its process toward democracy in 1987. In the following subsections, we discuss the spurious effects on people's attitudes towards seven statements. Instead of province fixed effects, we control for county fixed effects.

Table 3.11: Propaganda Effects on Boy Preference

Without "One Child" Policy, I only want boys.	Probit (1)	Probit (2)	Logit (3)	Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.0197	0.0293**	0.0317	0.0523*	-0.00159	0.000325
	(0.0138)	(0.0135)	(0.0293)	(0.0271)	(0.00330)	(0.00275)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0286	-0.0249	-0.0562	-0.0493	-0.0113**	-0.00893*
	(0.0264)	(0.0328)	(0.0525)	(0.0641)	(0.00571)	(0.00493)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0135	-0.0244	-0.0256	-0.0500	-0.00527	-0.00424
	(0.0182)	(0.0258)	(0.0356)	(0.0515)	(0.00398)	(0.00455)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0939***	-0.0870**	-0.172***	-0.162**	-0.0217***	-0.0165**
	(0.0304)	(0.0378)	(0.0578)	(0.0716)	(0.00680)	(0.00726)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7247	7247	7247	7247	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputies in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.12: Propaganda Effects on Boy Preference: Difference by Gender

Without "One Child" Policy, I only want boys.	Probit (1)	Probit (2)	Logit (3)	Logit (4)	IV (5)	IV (6)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.0243* (0.0135)	0.0316** (0.0138)	0.0394 (0.0279)	0.0552** (0.0267)	-0.000750 (0.00327)	0.000679 (0.00283)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0310 (0.0260)	-0.0261 (0.0312)	-0.0590 (0.0526)	-0.0487 (0.0621)	-0.0115* (0.00599)	-0.00903* (0.00495)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0200 (0.0170)	-0.0327 (0.0234)	-0.0364 (0.0333)	-0.0637 (0.0474)	-0.00603 (0.00382)	-0.00535 (0.00434)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0951*** (0.0326)	-0.0856** (0.0404)	-0.174*** (0.0614)	-0.159** (0.0766)	-0.0218*** (0.00708)	-0.0161** (0.00758)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	-0.00825 (0.00937)	-0.00500 (0.00934)	-0.0150 (0.0180)	-0.00761 (0.0172)	-0.00142 (0.00189)	-0.000624 (0.00179)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.00365 (0.00814)	-0.000203 (0.00761)	0.00431 (0.0166)	-0.00425 (0.0154)	0.000190 (0.00137)	-0.000146 (0.00116)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	0.0117 (0.00840)	0.0150* (0.00883)	0.0210 (0.0156)	0.0264 (0.0167)	0.00124 (0.00112)	0.00176 (0.00107)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	0.00170 (0.0119)	-0.00247 (0.0118)	0.00423 (0.0218)	-0.00382 (0.0211)	0.0000944 (0.00192)	-0.000551 (0.00164)
Male	-0.0445 (0.153)	-0.00129 (0.138)	-0.0741 (0.305)	0.00637 (0.288)	0.00196 (0.0241)	0.00315 (0.0202)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes
N	7247	7247	7247	7247	7477	7477

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, are controlled. Robust standard errors clustered by province are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

In addition to the IV strategy we undertake in the baseline analysis, we also drop one instrument variable, UN World Conference on Women in Beijing and its anniversaries, since these events could also affect Taiwanese' attitudes towards gender equality.

Table 3.13 to Table 3.26 show the results. For most cases, the mainland propaganda effects do not exist in Taiwan. For the 1st statement "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work", in the age range 0-5, there exists a significantly positive effect on women's agreement to the statement in all IV strategies, but the effects are not significant in ordered probit and ordered logit strategies (Table 3.14). For the 6th statement "A man's job is to earn money; a woman's job is to look after the home and family", in the age range 0-5 and 18-25, men's agreement is significantly lower than the one of women in ordered probit and logit strategies, but the effects are not significant in all IV strategies. For the 7th statement "Politics is man's thing; a woman is best not to participate", in the age range 13-17, men's agreement is significantly higher than women's in all strategies, but it is the incorrect sign.

In summary, there is no enough evidence to show there exists a spurious correlation between propaganda and people's attitudes towards gender equality.

3.6.3 Difference-in-Differences

In this subsection, we undertake a difference-in-differences analysis on one similar statement in both CGSS (2010, 2012, 2013) and TSCS (2012). In CGSS, the statement is "Within a family, women should spend most of their time at home, and men should spend most of their time on career". In TSCS, the statement is "A man's job is to earn money; a woman's job is to look after the home and family". The only similar statement might not be a good one for us, since in the baseline CGSS analysis, we do not find that propaganda has robustly significant effects on all people, women, and men's agreement to the statement (Figure 3.3, 3.4, 3.5). However, with the difference-in-differences analysis, we find that the baseline strategies may underestimate the propaganda effects on mainland Chinese women but overestimate the effects on mainland Chinese men.

Table 3.13: Spurious Propaganda Effects on Statement 1

A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda	0.0330	0.0325	0.0741	0.0708	0.0705*	0.0784*	0.0366	0.0402
×Media Coverage $0-5$	(0.0461)	(0.0465)	(0.0820)	(0.0835)	(0.0425)	(0.0439)	(0.0337)	(0.0319)
Gender Equality in Propaganda	0.0737	0.0826	0.174	0.191	-0.0360	-0.0299	0.00427	0.0125
×Media Coverage $6-12$	(0.111)	(0.110)	(0.202)	(0.203)	(0.0671)	(0.0697)	(0.0706)	(0.0717)
Gender Equality in Propaganda	0.0532	0.0541	0.120	0.123	-0.0256	-0.0197	-0.0251	-0.0242
×Media Coverage $13-17$	(0.0758)	(0.0749)	(0.139)	(0.139)	(0.0740)	(0.0788)	(0.0761)	(0.0803)
Gender Equality in Propaganda	0.00940	0.00430	0.0535	0.0431	-0.0190	-0.00111	-0.0873	-0.0870
×Media Coverage $18-25$	(0.0855)	(0.0845)	(0.152)	(0.150)	(0.142)	(0.149)	(0.133)	(0.139)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.14: Spurious Propaganda Effects on Statement 1: Difference by Gender

A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.0278 (0.0460)	0.0282 (0.0459)	0.0654 (0.0817)	0.0637 (0.0823)	0.0951** (0.0402)	0.104** (0.0409)	0.0588* (0.0310)	0.0624** (0.0292)
Gender Equality in Propaganda ×Media Coverage $6-12$	0.0750 (0.110)	0.0846 (0.109)	0.172 (0.202)	0.190 (0.201)	-0.0650 (0.0671)	-0.0619 (0.0699)	-0.0288 (0.0799)	-0.0298 (0.0797)
Gender Equality in Propaganda ×Media Coverage $13-17$	0.0537 (0.0734)	0.0551 (0.0732)	0.117 (0.135)	0.122 (0.136)	-0.0634 (0.0721)	-0.0610 (0.0744)	-0.0801 (0.0782)	-0.0909 (0.0779)
Gender Equality in Propaganda ×Media Coverage $18-25$	0.0142 (0.0857)	0.00859 (0.0856)	0.0542 (0.155)	0.0448 (0.154)	-0.0662 (0.130)	-0.0544 (0.132)	-0.173 (0.126)	-0.193 (0.124)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	0.00572 (0.0126)	0.00413 (0.0130)	0.0123 (0.0241)	0.00932 (0.0247)	-0.0082 (0.0103)	-0.0095 (0.0104)	-0.0052 (0.0101)	-0.0062 (0.0103)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.00107 (0.0131)	-0.0005 (0.0132)	0.0100 (0.0239)	0.00647 (0.0246)	-0.0080 (0.0121)	-0.0087 (0.0117)	-0.0073 (0.0121)	-0.0081 (0.0118)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	0.00246 (0.0145)	0.00059 (0.0148)	0.00958 (0.0275)	0.00632 (0.0281)	-0.0063 (0.0119)	-0.0071 (0.0122)	-0.0062 (0.0122)	-0.0073 (0.0126)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0078 (0.0240)	-0.0084 (0.0233)	-0.0003 (0.0454)	-0.0022 (0.0445)	-0.0203 (0.0188)	-0.0207 (0.0174)	-0.0190 (0.0193)	-0.0198 (0.0182)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.15: Spurious Propaganda Effects on Statement 2

A pre-school child is likely to suffer if his or her mother works.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.0148 (0.0317)	0.0114 (0.0335)	0.0124 (0.0637)	0.00318 (0.0673)	-0.0452 (0.0651)	-0.0472 (0.0663)	-0.0122 (0.0534)	-0.0176 (0.0530)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0348 (0.0927)	-0.0251 (0.0936)	-0.111 (0.164)	-0.0884 (0.169)	0.0222 (0.0619)	0.0236 (0.0685)	-0.0466 (0.0751)	-0.0451 (0.0825)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0578 (0.0836)	-0.0423 (0.0843)	-0.135 (0.144)	-0.102 (0.147)	-0.0706 (0.0769)	-0.0648 (0.0815)	-0.0908 (0.0893)	-0.0847 (0.0935)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0454 (0.111)	-0.0101 (0.118)	-0.160 (0.197)	-0.0884 (0.211)	-0.207 (0.160)	-0.184 (0.163)	-0.158 (0.176)	-0.138 (0.177)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.16: Spurious Propaganda Effects on Statement 2: Difference by Gender

A pre-school child is likely to suffer if his or her mother works.	Ordered Probit	Ordered Probit	Ordered Logit	Ordered Logit	IV	IV	IV (no world)	IV (no world)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.0125	0.00933	0.00896	-0.0003	-0.0512	-0.0555	-0.0251	-0.0323
	(0.0261)	(0.0281)	(0.0526)	(0.0561)	(0.0577)	(0.0608)	(0.0443)	(0.0445)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0328	-0.0251	-0.111	-0.0918	0.0281	0.0314	-0.0020	0.00021
	(0.0863)	(0.0874)	(0.153)	(0.158)	(0.0621)	(0.0663)	(0.0831)	(0.0904)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0559	-0.0417	-0.136	-0.105	-0.0748	-0.0670	-0.0347	-0.0270
	(0.0792)	(0.0806)	(0.137)	(0.141)	(0.0777)	(0.0829)	(0.0943)	(0.101)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0646	-0.0315	-0.198	-0.130	-0.241	-0.219	-0.0910	-0.0712
	(0.107)	(0.115)	(0.193)	(0.208)	(0.163)	(0.169)	(0.180)	(0.185)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	0.00916	0.00887	0.0177	0.0178	0.0122	0.0119	0.00871	0.00831
	(0.0224)	(0.0222)	(0.0385)	(0.0383)	(0.0187)	(0.0193)	(0.0187)	(0.0192)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.0123	0.0144	0.0241	0.0283	0.0108	0.0122	0.0100	0.0113
	(0.0170)	(0.0168)	(0.0319)	(0.0313)	(0.0147)	(0.0150)	(0.0150)	(0.0153)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0055	-0.0053	-0.0077	-0.0072	-0.0031	-0.0040	-0.0034	-0.0043
	(0.0166)	(0.0165)	(0.0304)	(0.0302)	(0.0136)	(0.0143)	(0.0143)	(0.0149)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	0.0243	0.0262	0.0429	0.0464	0.0210	0.0223	0.0195	0.0207
	(0.0226)	(0.0225)	(0.0422)	(0.0418)	(0.0212)	(0.0217)	(0.0217)	(0.0221)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.17: Spurious Propaganda Effects on Statement 3

All in all, family life suffers when the woman has a full-time job.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0188 (0.0404)	-0.0143 (0.0432)			-0.00417 (0.0506)	-0.00187 (0.0498)	-0.00828 (0.0312)	-0.00975 (0.0323)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0547 (0.0949)	-0.0311 (0.103)			-0.0278 (0.0766)	-0.00224 (0.0833)	-0.0552 (0.0830)	-0.0256 (0.0893)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0577 (0.0810)	-0.0415 (0.0830)			-0.0695 (0.0857)	-0.0387 (0.0891)	-0.0911 (0.0962)	-0.0607 (0.0970)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0326 (0.124)	-0.0179 (0.128)			-0.126 (0.144)	-0.0741 (0.155)	-0.154 (0.159)	-0.110 (0.158)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Ordered logit regressions are not applicable in this case. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.18: Spurious Propaganda Effects on Statement 3: Difference by Gender

All in all, family life suffers when the woman has a full-time job.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0130 (0.0386)	-0.0081 (0.0406)			-0.0190 (0.0417)	-0.0174 (0.0416)	-0.0143 (0.0271)	-0.0187 (0.0283)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0615 (0.0906)	-0.0384 (0.0993)			-0.0143 (0.0766)	0.0109 (0.0844)	-0.0215 (0.0734)	0.0153 (0.0856)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0562 (0.0757)	-0.0407 (0.0781)			-0.0451 (0.0864)	-0.0135 (0.0914)	-0.0324 (0.0760)	0.00917 (0.0875)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0453 (0.122)	-0.0334 (0.127)			-0.109 (0.152)	-0.0571 (0.165)	-0.0663 (0.138)	-0.0071 (0.156)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0065 (0.0131)	-0.0062 (0.0133)			0.00435 (0.0074)	0.00354 (0.0075)	0.00340 (0.0077)	0.00297 (0.0079)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.0110 (0.0162)	0.0122 (0.0168)			0.0160 (0.0104)	0.0163 (0.0106)	0.0158 (0.0102)	0.0163 (0.0104)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0062 (0.0126)	-0.0066 (0.0131)			0.00071 (0.0064)	-0.0004 (0.0071)	0.00057 (0.0065)	-0.0003 (0.0072)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	0.0138 (0.0261)	0.0163 (0.0257)			0.0239 (0.0177)	0.0253 (0.0173)	0.0234 (0.0177)	0.0252 (0.0172)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.19: Spurious Propaganda Effects on Statement 4

A job is all right, but what most women really want is a home and children.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda	0.00442	0.0125	0.0127	0.0312	0.0339	0.0377	-0.00100	0.00188
×Media Coverage $0-5$	(0.0323)	(0.0325)	(0.0610)	(0.0623)	(0.0459)	(0.0470)	(0.0446)	(0.0470)
Gender Equality in Propaganda	0.0971	0.107	0.165	0.188	-0.0107	0.00596	0.0310	0.0456
×Media Coverage $6-12$	(0.0983)	(0.0940)	(0.188)	(0.181)	(0.0902)	(0.0870)	(0.0870)	(0.0850)
Gender Equality in Propaganda	0.0553	0.0565	0.0853	0.0887	0.0635	0.0839	0.0641	0.0796
×Media Coverage $13-17$	(0.0804)	(0.0781)	(0.153)	(0.150)	(0.105)	(0.108)	(0.105)	(0.107)
Gender Equality in Propaganda	-0.0282	-0.0295	-0.0784	-0.0796	0.166	0.202	0.0956	0.122
×Media Coverage $18-25$	(0.121)	(0.121)	(0.220)	(0.218)	(0.202)	(0.213)	(0.199)	(0.203)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.20: Spurious Propaganda Effects on Statement 4: Difference by Gender

A job is all right, but what most women really want is a home and children.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.00871 (0.0286)	0.0186 (0.0298)	0.0174 (0.0545)	0.0391 (0.0578)	0.0293 (0.0408)	0.0347 (0.0440)	-0.0137 (0.0361)	-0.0065 (0.0393)
Gender Equality in Propaganda ×Media Coverage $6-12$	0.0879 (0.102)	0.100 (0.0976)	0.157 (0.195)	0.184 (0.188)	-0.0078 (0.0980)	0.0106 (0.0933)	0.00701 (0.101)	0.0224 (0.0983)
Gender Equality in Propaganda ×Media Coverage $13-17$	0.0535 (0.0828)	0.0555 (0.0804)	0.0866 (0.158)	0.0920 (0.154)	0.0774 (0.104)	0.0967 (0.107)	0.0444 (0.113)	0.0590 (0.117)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0230 (0.124)	-0.0222 (0.125)	-0.0669 (0.226)	-0.0633 (0.226)	0.196 (0.196)	0.228 (0.209)	0.0652 (0.204)	0.0913 (0.214)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0103 (0.0119)	-0.0138 (0.0118)	-0.0116 (0.0251)	-0.0182 (0.0241)	-0.0061 (0.0074)	-0.0091 (0.0069)	-0.0039 (0.0076)	-0.0068 (0.0070)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.00056 (0.0125)	-0.0059 (0.0153)	0.00589 (0.0218)	-0.0067 (0.0272)	0.00054 (0.0075)	-0.0041 (0.0087)	0.00037 (0.0080)	-0.0042 (0.0093)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0001 (0.0159)	-0.0032 (0.0162)	0.00133 (0.0312)	-0.0050 (0.0317)	0.00179 (0.0108)	-0.0005 (0.0107)	0.00098 (0.0108)	-0.0015 (0.0108)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0132 (0.0170)	-0.0195 (0.0207)	-0.0168 (0.0280)	-0.0288 (0.0352)	-0.0100 (0.0098)	-0.0142 (0.0112)	-0.0100 (0.0103)	-0.0143 (0.0119)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.21: Spurious Propaganda Effects on Statement 5

Both the man and woman should contribute to the household income.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.0288	0.0334	0.0499	0.0577	0.00325	0.00686	-0.00098	-0.00154
	(0.0285)	(0.0308)	(0.0548)	(0.0584)	(0.0357)	(0.0342)	(0.0269)	(0.0263)
Gender Equality in Propaganda ×Media Coverage $6-12$	0.0587	0.0625	0.112	0.122	0.0488	0.0467	0.0197	0.0194
	(0.0879)	(0.0889)	(0.167)	(0.168)	(0.0645)	(0.0637)	(0.0618)	(0.0629)
Gender Equality in Propaganda ×Media Coverage $13-17$	0.0494	0.0471	0.0972	0.0943	0.0183	0.0203	-0.00445	-0.00474
	(0.0696)	(0.0684)	(0.131)	(0.126)	(0.0684)	(0.0690)	(0.0744)	(0.0744)
Gender Equality in Propaganda ×Media Coverage $18-25$	0.0357	0.0342	0.0664	0.0630	-0.0376	-0.0233	-0.0663	-0.0631
	(0.147)	(0.153)	(0.287)	(0.296)	(0.150)	(0.151)	(0.144)	(0.143)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.22: Spurious Propaganda Effects on Statement 5: Difference by Gender

Both the man and woman should contribute to the household income.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.0231 (0.0239)	0.0279 (0.0261)	0.0345 (0.0453)	0.0423 (0.0483)	0.00180 (0.0301)	0.00699 (0.0290)	-0.0090 (0.0264)	-0.0070 (0.0253)
Gender Equality in Propaganda ×Media Coverage $6-12$	0.0615 (0.0821)	0.0674 (0.0826)	0.114 (0.154)	0.129 (0.154)	0.0511 (0.0598)	0.0518 (0.0578)	0.0297 (0.0464)	0.0385 (0.0465)
Gender Equality in Propaganda ×Media Coverage $13-17$	0.0459 (0.0652)	0.0464 (0.0642)	0.0853 (0.122)	0.0893 (0.117)	0.0270 (0.0528)	0.0298 (0.0513)	0.00949 (0.0550)	0.0193 (0.0568)
Gender Equality in Propaganda ×Media Coverage $18-25$	0.0536 (0.136)	0.0540 (0.142)	0.0900 (0.265)	0.0926 (0.273)	0.00233 (0.126)	0.0133 (0.127)	-0.0273 (0.102)	-0.0102 (0.105)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	0.00656 (0.0164)	0.00508 (0.0163)	0.0202 (0.0305)	0.0169 (0.0303)	0.00169 (0.0130)	-0.0002 (0.0130)	0.00094 (0.0134)	-0.0006 (0.0133)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0060 (0.0163)	-0.0088 (0.0169)	-0.0022 (0.0333)	-0.0091 (0.0348)	-0.0072 (0.0114)	-0.0093 (0.0120)	-0.0082 (0.0113)	-0.0100 (0.0118)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	0.0180 (0.0127)	0.0132 (0.0130)	0.0398 (0.0252)	0.0298 (0.0259)	0.00742 (0.0104)	0.00373 (0.0111)	0.00631 (0.0103)	0.00284 (0.0109)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0080 (0.0276)	-0.0138 (0.0282)	-0.0054 (0.0541)	-0.0183 (0.0559)	-0.0084 (0.0205)	-0.0127 (0.0212)	-0.0098 (0.0203)	-0.0137 (0.0210)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.23: Spurious Propaganda Effects on Statement 6

A man's job is to earn money; a woman's job is to look after the home and family.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda $\times$Media Coverage $0-5$	0.0117	0.0189	0.0160	0.0290	0.0635	0.0663	0.0340	0.0391
	(0.0476)	(0.0493)	(0.0911)	(0.0939)	(0.0636)	(0.0702)	(0.0523)	(0.0564)
Gender Equality in Propaganda $\times$Media Coverage $6-12$	-0.00332	0.00302	-0.0126	-0.00175	-0.0996	-0.0738	-0.0579	-0.0374
	(0.112)	(0.114)	(0.210)	(0.214)	(0.117)	(0.120)	(0.121)	(0.127)
Gender Equality in Propaganda $\times$Media Coverage $13-17$	-0.0180	-0.0102	-0.0399	-0.0266	-0.119	-0.0951	-0.115	-0.0942
	(0.0871)	(0.0873)	(0.160)	(0.161)	(0.126)	(0.134)	(0.126)	(0.134)
Gender Equality in Propaganda $\times$Media Coverage $18-25$	-0.0269	-0.0180	-0.0240	-0.00692	-0.141	-0.110	-0.197	-0.168
	(0.122)	(0.121)	(0.223)	(0.223)	(0.252)	(0.262)	(0.223)	(0.228)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.24: Spurious Propaganda Effects on Statement 6: Difference by Gender

A man's job is to earn money; a woman's job is to look after the home and family.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.0214	0.0296	0.0306	0.0443	0.0595	0.0631	0.0232	0.0277
	(0.0485)	(0.0500)	(0.0953)	(0.0978)	(0.0646)	(0.0715)	(0.0508)	(0.0548)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0192	-0.0122	-0.0334	-0.0198	-0.0972	-0.0709	-0.0248	0.00135
	(0.117)	(0.118)	(0.223)	(0.225)	(0.121)	(0.124)	(0.120)	(0.125)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0169	-0.0096	-0.0299	-0.0168	-0.104	-0.0791	-0.0433	-0.0162
	(0.0896)	(0.0897)	(0.169)	(0.168)	(0.122)	(0.130)	(0.125)	(0.133)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0256	-0.0190	-0.0017	0.0111	-0.133	-0.104	-0.0880	-0.0515
	(0.130)	(0.128)	(0.243)	(0.239)	(0.231)	(0.239)	(0.216)	(0.222)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	-0.023**	-0.025**	-0.041**	-0.043**	-0.0159	-0.0179	-0.0134	-0.0155
	(0.0118)	(0.0121)	(0.0192)	(0.0197)	(0.0116)	(0.0119)	(0.0113)	(0.0116)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.00022	-0.0032	-0.0006	-0.0056	0.00919	0.00630	0.0108	0.00796
	(0.0089)	(0.0095)	(0.016)	(0.0179)	(0.0083)	(0.0091)	(0.0087)	(0.0095)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0149	-0.0172	-0.0249	-0.0279	-0.0085	-0.0106	-0.0072	-0.0094
	(0.0124)	(0.0122)	(0.0219)	(0.0219)	(0.0123)	(0.0122)	(0.0132)	(0.0131)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0285	-0.0312*	-0.0578*	-0.0611*	-0.0128	-0.0147	-0.0105	-0.0125
	(0.0179)	(0.0183)	(0.0311)	(0.0327)	(0.0171)	(0.0172)	(0.0168)	(0.0172)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.25: Spurious Propaganda Effects on Statement 7

Politics is man's thing; a woman is best not to participate.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0121	-0.0107	0.0160	0.0290	-0.00249	0.00106	-0.0222	-0.0220
	(0.0452)	(0.0435)	(0.0911)	(0.0939)	(0.0266)	(0.0276)	(0.0261)	(0.0271)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0602	-0.0353	-0.0126	-0.00175	-0.0559	-0.0393	-0.0812	-0.0661
	(0.0766)	(0.0777)	(0.210)	(0.214)	(0.0773)	(0.0781)	(0.0685)	(0.0700)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0535	-0.0333	-0.0399	-0.0266	-0.0583	-0.0405	-0.0906	-0.0777
	(0.0682)	(0.0685)	(0.160)	(0.161)	(0.0735)	(0.0783)	(0.0717)	(0.0748)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0395	0.00112	-0.0240	-0.00692	-0.0732	-0.0348	-0.142	-0.117
	(0.0981)	(0.101)	(0.223)	(0.223)	(0.119)	(0.132)	(0.108)	(0.115)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.26: Spurious Propaganda Effects on Statement 7: Difference by Gender

Politics is man's thing; a woman is best not to participate.	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit (3)	Ordered Logit (4)	IV (5)	IV (6)	IV (no world) (7)	IV (no world) (8)
Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0166	-0.0146			-0.0062	-0.0011	-0.0324	-0.0338
	(0.0456)	(0.0431)			(0.0250)	(0.0257)	(0.0243)	(0.0251)
Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0646	-0.0383			-0.0576	-0.0415	-0.0657	-0.0480
	(0.0812)	(0.0828)			(0.0760)	(0.0766)	(0.0695)	(0.0716)
Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0611	-0.0392			-0.0526	-0.0359	-0.0651	-0.0494
	(0.0703)	(0.0709)			(0.0678)	(0.0714)	(0.0694)	(0.0724)
Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0375	0.00405			-0.0520	-0.0157	-0.0951	-0.0683
	(0.103)	(0.105)			(0.108)	(0.119)	(0.105)	(0.110)
Male ×Gender Equality in Propaganda ×Media Coverage $0-5$	0.0102	0.0100			0.00696	0.00722	0.00688	0.00740
	(0.0090)	(0.0102)			(0.0048)	(0.0054)	(0.0048)	(0.0055)
Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.00856	0.00921			0.00754	0.00824	0.00676	0.00751
	(0.0163)	(0.0156)			(0.0068)	(0.0063)	(0.0066)	(0.0061)
Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	0.0204**	0.0205*			0.0112**	0.0118**	0.0100**	0.0105*
	(0.0099)	(0.0106)			(0.0052)	(0.0057)	(0.0051)	(0.0055)
Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	0.0163	0.0208			0.0138	0.0171*	0.0127	0.0158*
	(0.0178)	(0.0178)			(0.0086)	(0.0088)	(0.0086)	(0.0087)
Weighted Sample Regression	No	Yes	No	Yes	No	Yes	No	Yes
N	942	942	942	942	942	942	942	942

Notes: County fixed effects, cohort fixed effects, province-cohort fixed effects, demographic information, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputies in National Congress, and democracy, are controlled. Ordered logit regressions are not applicable in this case. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.27 shows the results of the difference-in-differences analysis. Consistent with the baseline, propaganda has no effects on people's agreement to the statement. Table 3.28 shows the results of the analysis with male interaction. There exists a robust positive correlation between mainland China's propaganda at the age range 0-5 and Taiwanese women's agreement to the statement. However, the coefficient of the interaction term of $(PeoplesDaily \times MediaCoverage_p)_{ik}$ and $MainlandChina_i$ is significantly negative in the age range 0-5. This indicates propaganda promoted the attitudes to gender equality of women in mainland China and the baseline strategies are likely to underestimate the effects. However, there is a different story for men in mainland China. Compared with men in Taiwan without propaganda, mainland Chinese men tend to disagree with the statement, but under the influence of propaganda in the age range 0-5, 13-17, and 18-25, men in mainland China tend to agree with the statement. In baseline CGSS analysis among all statements, we only find that propaganda has robustly negative effects on mainland Chinese men's agreement to the statement "When the economy is bad, women should be fired first" in the age range 0-5 and 6-12 (Figure 3.5). Now we find that propaganda may have counter-acting effects on men's attitudes to gender equality.

Table 3.27: Difference-in-differences Analysis

Within a family, women should spend most of their time at home, and men should spend most of their time on career	Ordered Probit (1)	IV (2)	IV (no world conference) (3)
Gender Equality in Propaganda ×Media Coverage $0-5$	0.0177	0.0251	0.0198
	(0.0369)	(0.0452)	(0.0549)
Gender Equality in Propaganda ×Media Coverage $6-12$	0.0665	-0.0267	0.0161
	(0.0935)	(0.104)	(0.116)
Gender Equality in Propaganda ×Media Coverage $13-17$	0.0343	-0.0781	-0.0124
	(0.0751)	(0.115)	(0.124)
Gender Equality in Propaganda ×Media Coverage $18-25$	0.0167	-0.156	-0.0452
	(0.108)	(0.207)	(0.210)
MainlandChina ×Gender Equality in Propaganda ×Media Coverage $0-5$	-0.0230	-0.0332	-0.0280
	(0.0385)	(0.0472)	(0.0565)
MainlandChina ×Gender Equality in Propaganda ×Media Coverage $6-12$	-0.0750	0.0150	-0.0324
	(0.0952)	(0.106)	(0.117)
MainlandChina ×Gender Equality in Propaganda ×Media Coverage $13-17$	-0.0392	0.0721	0.00260
	(0.0775)	(0.115)	(0.125)
MainlandChina ×Gender Equality in Propaganda ×Media Coverage $18-25$	-0.0380	0.129	0.00751
	(0.110)	(0.211)	(0.215)
MainlandChina	23.55	-27.77	11.95
	(68.21)	(110.5)	(105.4)
N	8419	8419	8419

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges are controlled. Regressions are not weighted. Logit model is not applicable in this case. Robust standard errors clustered by county are reported in parentheses.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3.28: Difference-in-differences Analysis with Interactions

Within a family, women should spend most of their time at home, and men should spend most of their time on career; A man's job is to earn money; a woman's job is to look after the home and family.	Ordered Probit (1)	IV (2)	IV (no world conference) (3)
Gender Equality in Propaganda ×Media Coverage 0 − 5	0.0267***	0.0308***	0.0244*
	(0.00999)	(0.0102)	(0.0136)
Gender Equality in Propaganda ×Media Coverage 6 − 12	0.0571*	-0.0309**	0.0105
	(0.0332)	(0.0128)	(0.0170)
Gender Equality in Propaganda ×Media Coverage 13 − 17	0.0401	-0.0629***	0.00336
	(0.0270)	(0.0188)	(0.0329)
Gender Equality in Propaganda ×Media Coverage 18 − 25	0.0255	-0.135***	-0.0235
	(0.0245)	(0.0313)	(0.0592)
MainlandChina×Gender Equality in Propaganda ×Media Coverage 0 − 5	-0.0360**	-0.0435**	-0.0381*
	(0.0179)	(0.0172)	(0.0199)
MainlandChina×Gender Equality in Propaganda ×Media Coverage 6 − 12	-0.0705**	0.0157	-0.0307
	(0.0358)	(0.0204)	(0.0252)
MainlandChina×Gender Equality in Propaganda ×Media Coverage 13 − 17	-0.0482	0.0543**	-0.0164
	(0.0322)	(0.0237)	(0.0381)
MainlandChina×Gender Equality in Propaganda ×Media Coverage 18 − 25	-0.0526*	0.106**	-0.0171
	(0.0302)	(0.0462)	(0.0748)
Male×Gender Equality in Propaganda ×Media Coverage 0 − 5	-0.0193***	-0.0163***	-0.0152***
	(0.00128)	(0.000800)	(0.00103)
Male×Gender Equality in Propaganda ×Media Coverage 6 − 12	0.000237	0.00647***	0.00891***
	(0.00165)	(0.00113)	(0.00140)
Male×Gender Equality in Propaganda ×Media Coverage 13 − 17	-0.0105***	-0.00833***	-0.00519***
	(0.00170)	(0.00141)	(0.00170)
Male×Gender Equality in Propaganda ×Media Coverage 18 − 25	-0.0157***	-0.0126***	-0.00979***
	(0.00430)	(0.00238)	(0.00296)
MainlandChina×Male ×Gender Equality in Propaganda ×Media Coverage 0 − 5	0.0255***	0.0241***	0.0239***
	(0.00556)	(0.00539)	(0.00530)

Difference-in-differences Analysis with Interactions, continue

MainlandChina×Male ×Gender Equality in Propaganda ×Media Coverage $6-12$	0.00826 (0.00675)	0.00219 (0.00661)	-0.00153 (0.00697)
MainlandChina×Male ×Gender Equality in Propaganda ×Media Coverage $13-17$	0.0160*** (0.00437)	0.0142*** (0.00343)	0.0110*** (0.00305)
MainlandChina×Male ×Gender Equality in Propaganda ×Media Coverage $18-25$	0.0260*** (0.00864)	0.0225*** (0.00673)	0.0161** (0.00684)
MainlandChina	8.759 (48.54)	-39.58 (52.21)	-2.081 (63.45)
Male	0.942*** (0.135)	0.760*** (0.0743)	0.625*** (0.0947)
MainlandChina×Male	-1.115*** (0.202)	-0.938*** (0.148)	-0.763*** (0.162)

Notes: Province fixed effects, cohort fixed effects, province-cohort fixed effects, demographic informations, provincial reality at all age ranges, including GDP per capita, percentage of Internet user, gender gap of year of schooling among people between 25-29, percentage of female deputy in National Congress, and democracy, are controlled. Regressions are not weighted. Logit model is not applicable in this case. Robust standard errors clustered by county are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

3.7 Conclusions

This study reveals that gender-equality promoting propaganda has long-run effects on people's attitudes towards women's participation in the workforce in China. We observe that propaganda has larger effects on women than men. Women are affected by propaganda exposure before age 26, while men were affected by exposure before adulthood. It is shown that gender-equality promoting propaganda is remarkably effective in convincing people to endorse women's participation in the workforce, despite the stronghold of traditional gender roles in China. We also note that the propaganda message is not entirely progressive: while it liberates women in the public sphere, women are still expected to be heavily burdened by housework. This further evidence corroborates the persuasiveness of propaganda; public opinion shifts not because of the progressiveness of the message, but it changes exactly as propaganda intends.

Consistent results from different strategies confirm the long-term causal effect of propaganda on people's attitudes. The media censorship in China helps us to eliminate confounding factors, but this special context does not necessarily preclude the generalization of the results. Most developing countries still have very high gender inequality and low female labor force participation rates, even though economic growth and gender equality are positively correlated with each other. Gender inequality, especially the disapproval of women's role in the workforce would reduce female labor force participation and hinder economic development. It would be more efficient for governments in developing countries with severe gender inequality to change people's attitudes first.

We also find that more intense exposure to gender-equality promoting propaganda between the ages 18 to 25 reduce both women's and men's preference for sons in the absence of the one-child policy. As the one-child policy began to be phased out in 2015, this finding suggests that more intense gender-equality propaganda could help to alleviate China's heavily skewed sex ratio.

The message of gender-related propaganda is also changing in China: compared to his predecessors, Xi Jingping, the president of China, has repeatedly emphasized women's familial responsibility on different occasions. In his speech on the 2015 International Women's day, he stressed that "We should pay attention to the unique role women plays in social life and family life." Given the findings of this study, we expect this updated propaganda message to further aggravate women's double burden in China.

In future research, we plan to examine whether changes in gender equality attitudes leads to an increase in female labor force participation. We also plan to use the same methodology to explore whether local government officials who grew up in an era in which the central government endorsed the free market system, such as the period immediately after the Open-Door Policy in 1978, were more open to business-friendly policies when they were in office, thus leading to better economic growth, partially explaining the recent Chinese economic growth miracle.

Data Appendix

Table 3.29: Summary Statistics

N=7477	Mean	National Mean	Women Mean	Men Mean	Std.Dev.	Range
Age	43.978		43.491	44.426	8.498	30-61
Male	0.520	0.520			0.500	0-1
Married	0.930		0.946	0.915	0.255	0-1
Minority	0.107	0.0849	0.110	0.104	0.309	0-1
Household Registration (*Hukou*)						
Rural	0.629	0.634	0.634	0.625	0.483	0-1
Education						
No School	0.194	0.0464	0.255	0.137	0.395	0-1
Traditional Primary School	0.00187		0.00251	0.00129	0.432	0-1
Modern Primary School	0.202	0.237	0.202	0.203	0.402	0-1
Middle School	0.303	0.481	0.276	0.327	0.459	0-1
Academic High School	0.120	0.151	0.0996	0.139	0.325	0-1
Vocational High School	0.0713		0.0622	0.0797	0.257	0-1
Associate Degree	0.0615	0.0524	0.0583	0.0645	0.240	0-1
Bachelor Degree	0.0437	0.0298	0.0402	0.0470	0.245	0-1
Master or Doctoral Degree	0.00241	0.00330	0.00335	0.00154	0.0490	0-1
Employment						
Employed by Government	0.0156		0.0103	0.0206	0.124	0-1
Employed by State Owned Organization	0.0999		0.0800	0.118	0.300	0-1
Employed by Private Organization	0.226		0.186	0.263	0.418	0-1
Retired	0.0401		0.0594	0.0224	0.196	0-1
Unemployed	0.139		0.209	0.0743	0.346	0-1
Student	0.000401		0.000279	0.00514	0.0200	0-1
Politics Status						
Member of CPC	0.0880		0.0494	0.124	0.283	0-1
Financial Situation						
Household Income Per Capita (RMB)	14400.09		13872	13886.79	21664.45	0-435000.1
Living in Rented House/Apartment	0.0768		0.0764	0.0771	0.266	0-1
Living in House/Apartment Owned by Relatives	0.00455		0.00502	0.00411	0.0673	0-1
Mother's Education						
No School	0.614		0.618	0.610	0.487	0-1

Traditional Primary School	0.0263	0.0259	0.0267	0.160	0-1
Modern Primary School	0.234	0.231	0.237	0.423	0-1
Middle School	0.0841	0.0831	0.0851	0.278	0-1
Academic High School	0.0203	0.0190	0.0216	0.141	0-1
Vocational High School	0.0143	0.0159	0.0129	0.119	0-1
Associate Degree	0.00428	0.00363	0.00488	0.0653	0-1
Bachelor Degree	0.00267	0.00335	0.00206	0.0517	0-1
Master or Doctoral Degree	0	0	0	0	0-1
Politics Status of Parents					
Father is CCP Member	0.160	0.158	0.161	0.366	0-1
Mother is CCP Member	0.0242	0.0243	0.0242	0.154	0-1
Mother's Employment Status When the Person Was at 14					
Unemployed	0.121	0.118	0.124	0.327	0-1
Attitudes to Gender Equality					
Within a family, women should spend most of their time at home, and men should spend most of their time on career	2.563	2.554	2.571	1.154	0-4
Male's ability is higher than that of the female	2.012	2.026	1.999	1.221	0-4
It is more important for women to marry well than having a career	2.103	2.178	2.034	1.182	0-4
When the economy is bad, women should be fired first	1.166	1.101	1.226	1.007	0-4
Couples should share housework equally	2.808	2.905	2.718	1.035	0-4
If there's no One-Child Policy, would you only want boys	0.0885	0.0862	0.0907	0.284	0-1
Gender of Children					
Bother daughters and sons	0.321	0.342	0.303	0.467	0-1
Only have daughters	0.255	0.257	0.254	0.436	0-1
Only have sons	0.370	0.368	0.372	0.483	0-1
Survey Year					
2010	0.348	0.384	0.315	0.476	0-1
2012	0.370	0.369	0.370	0.483	0-1
2013	0.282	0.247	0.315	0.450	0-1

Bibliography

[AF07] Alberto Alesina and Nicola Fuchs-Schundeln. "Goodbye Lenin (or Not?): The Effect of Communism on People's Preferences." *American Economic Review*, **97**(4):1507–1528, September 2007.

[AR06] Daron Acemoglu and James A. Robinson. "De Facto Political Power and Institutional Persistence." *American Economic Review*, **96**(2):325–330, May 2006.

[BC16] Leonardo Bursztyn and Davide Cantoni. "A Tear in the Iron Curtain: The Impact of Western Television on Consumption Behavior." *The Review of Economics and Statistics*, **98**(1):25–41, 2016.

[BFM16] Alison L. Booth, Elliott Fan, Xin Meng, and Dandan Zhang. "Gender Differences in Willingness to Compete: The Role of Culture and Institutions." IZA Discussion Papers 10364, Institute for the Study of Labor (IZA), November 2016.

[Bra08] Anne-Marie Brady. *Marketing Dictatorship: Propaganda and Thought Work in Contemporary China*. Rowman and Littlefield, 2008.

[CCY17] Davide Cantoni, Yuyu Chen, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang. "Curriculum and Ideology." *Journal of Political Economy*, **125**(2):338–392, 2017.

[JO09] Robert Jensen and Emily Oster. "The Power of TV: Cable Television and Women's Status in India*." *The Quarterly Journal of Economics*, **124**(3):1057–1094, 2009.

[LCD12] Eliana La Ferrara, Alberto Chong, and Suzanne Duryea. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics*, **4**(4):1–31, July 2012.

[Olk09] Benjamin Olken. "Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages." *American Economic Journal: Applied Economics*, **1**(4):1–33, 2009.

[Sha07] David Shambaugh. "China's Propaganda System: Institutions, Processes and Efficacy." *The China Journal*, **57**:25–58, 2007.