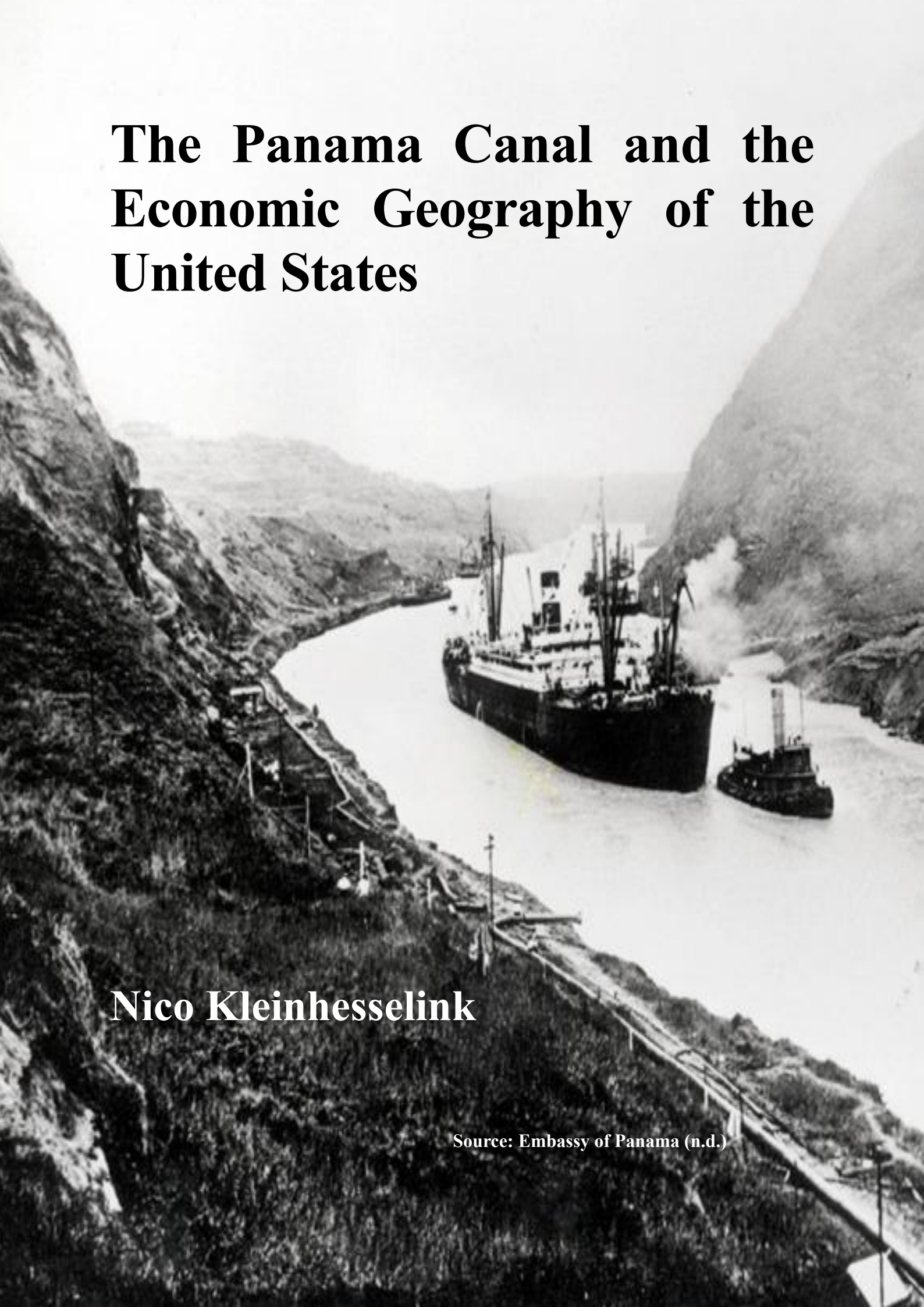


# The Panama Canal and the Economic Geography of the United States

**Nico Kleinhesselink**

Source: Embassy of Panama (n.d.)



# The Panama Canal and the Economic Geography of the United States

Student: Nico Kleinhesselink (1562142)

Supervisor: dr. Daniel Gallardo Albarrán

Second reader: dr. Pim de Zwart

Course code: RHI80436

Date of completion: 17-02-2026

Wageningen University,

Wageningen

## **Abstract**

Trade infrastructures are important for economic development by facilitating trade; however, their effects vary across regions and may therefore reshape the economic geography of countries. This paper studies whether changes in market potential induced by the Panama Canal stimulated economic growth in 2,424 U.S. counties over the first two decades after its full opening in 1920. I consider three economic outcomes: manufacturing output per capita, population, and agricultural land value. The overall results indicate that canal-driven gains in market potential are not associated with higher manufacturing output per capita, but they are associated with increases in population and land values. However, the existence of pre-trends across all outcomes limits a causal interpretation of these estimates. I further find heterogeneity across pre-canal characteristics, but only for selected outcomes: population growth responds more strongly in high-human-capital counties, while land values increase less in counties with agricultural labor market tightness. Additionally, I find strong spatial heterogeneity, with the effects on manufacturing and agriculture often moving in opposite directions, and gains in market potential even coinciding with decreases across all economic outcomes in the West North Central region. These findings deliver important insights for policymakers by documenting the complex role of transportation infrastructure in shaping the economic geography of nations.

## **Acknowledgements**

This thesis has been a big learning process for me. From the start, I experienced that things rarely go as planned, and I learned to accept that. There are many things I would have done differently, but in the end, I'm proud of the work that I put in. I also got a taste of the great amount of work that goes into doing research and publishing a paper, and my appreciation for research has grown.

I would like to thank my supervisor, Dr Daniel Gallardo Albarrán. I sincerely appreciate that you have always read my output with great attention and provided incisive feedback. I also appreciate that you took the time to respond to my questions via email. You have challenged me throughout the entire process, and I know that your supervision greatly improved the quality of this thesis.

I also want to thank my family, girlfriend, and friends for being there throughout the process. You have all been supportive and helpful in different ways. I also want to thank my study mates. Having lunch, sharing our thesis experiences, and sitting together on the fourth floor of the library always gave me energy to keep working.

# Table of Contents

- 1. Introduction .....1
- 2. Historical Background: The United States and The Panama Canal .....3
  - 2.1. The Panama Canal.....3
    - 2.1.1. The Building of the Canal.....3
    - 2.1.2. How the Canal Shortened Shipping Routes .....4
  - 2.2. The American Economy .....7
    - 2.2.1. Broad Patterns of Growth.....7
    - 2.2.2. Regional Growth, the Panama Canal, and Market Potential .....10
- 3. Theoretical Framework.....12
  - 3.1. Overall Effect of Market Potential on Economic Growth .....12
  - 3.2. Moderating Effects.....13
    - 3.2.1. The Moderating Effect of Human Capital .....13
    - 3.2.2. The Moderating Effect of Agricultural Labor Market Tightness .....14
- 4. Data .....15
  - 4.1. Market Potential .....15
  - 4.2. Economic Outcomes .....16
    - 4.2.1. Manufacturing Output Per Capita .....16
    - 4.2.2. Population Growth .....17
    - 4.2.3. Agricultural Land Value .....18
  - 4.3. Moderators .....18
    - 4.3.1. Human Capital .....18
    - 4.3.2. Agricultural Labor Market Tightness .....19
  - 4.4. Descriptive Statistics .....20
- 5. Method.....22
  - 5.1. Replication Maurer and Rauch (2023): Empirical Strategy.....22
  - 5.2. Overall Analysis: Empirical Strategy .....22

5.3.	<b>Moderating Variables: Empirical Strategy</b>	24
5.4.	<b>Spatial Heterogeneity Across Census Divisions: Empirical Strategy</b>	25
6.	<b>Results</b>	26
6.1.	<b>Replication of Maurer and Rauch (2023): Empirical Results</b>	26
6.2.	<b>Overall Analysis: Empirical Results</b>	26
6.3.	<b>Moderating Variables: Empirical Results</b>	30
6.4.	<b>Spatial Heterogeneity Across Census Divisions: Empirical Results</b>	32
7.	<b>Robustness Checks</b>	33
8.	<b>Conclusion</b>	34
	<b>References</b>	36
	<b>Use of AI</b>	42
	<b>Appendix A. World War I, the Roaring Twenties, and the Great Depression</b>	42
	<b>Appendix B. Additional Figures</b>	43
	<b>Appendix C. Construction of the Market Potential Measure</b>	44
	<b>Appendix D. Replication Maurer and Rauch (2023): Details</b>	46
	<b>Appendix E. Alternative Manufacturing Outcomes</b>	47
	<b>Appendix F. Alternative Clustering</b>	48
	<b>Appendix G. Multicollinearity Tests</b>	51
	<b>Appendix H. Long Difference Specification</b>	52
	<b>Appendix I. Alternative Functional Forms</b>	54
	<b>Appendix J. Additional time trends</b>	55
	<b>Appendix K. Additional Covariates</b>	56

# 1. Introduction

Global trade has increased dramatically over the past two centuries, with the total volume of imports and exports in 2024 more than 2,000 times that in 1800 (Ortiz-Ospina et al., 2018). Economic geography theory predicts that reductions in transportation costs increase trade flows and can stimulate regional economic growth (Krugman, 1991). Thus, economic development is higher in locations near coastlines or navigable rivers, due to lower transportation costs (Gallup et al., 1999).

The Panama Canal has been one of the most important global trade infrastructures for reducing transportation costs. It was especially important for the U.S., given its unique position with major ports on both the Atlantic and Pacific Oceans.<sup>1</sup> The opening of the Panama Canal reduced the journey of a conventional ship travelling between New York and San Francisco by about 33 days.<sup>2</sup> At a time when international trade occurred predominantly by sea and intercoastal U.S. trade relied on shipping or railroads,<sup>3</sup> the opening of the Panama Canal thus constituted a major transportation shock that dramatically altered market potential, with highly heterogeneous effects across U.S. counties.

Numerous studies show that transportation infrastructure stimulated economic growth in regions that became better connected, with the majority focusing on railroads (Atack et al., 2010; Duranton & Turner, 2012; Hornung, 2015; Donaldson & Hornbeck, 2016; Jedwab & Moradi, 2016; Berger & Enflo, 2017; Donaldson, 2018; Hornbeck & Rotemberg, 2019; Chan, 2021), and roads and highways (Chandra & Thompson, 2000; Faber, 2014; Baum-Snow et al., 2020; Herzog, 2021; Jedwab & Storeygard, 2022). Yet, only a handful of studies analyze whether the opening of the Panama Canal reshaped the spatial distribution of regional economic activity (Dajud, 2017; Belmar, 2023; Maurer & Rauch, 2023; Galiani et al., 2025), and only

---

<sup>1</sup> Brancaccio et al. (2020) estimate that closing the Canal today would reduce Northeast American trade by up to 28%.

<sup>2</sup> Shipping distance between New York and San Francisco fell by approximately 7,873 nautical miles, and conventional steamships traveled at roughly 10 knots (Johnson, 1912). The implied time saving is therefore about  $\frac{7,873}{10} = 787.3$  hours, or  $787.3/24 \approx 33$  days, abstracting from port time and other operational delays.

<sup>3</sup> Historical evidence indicates that, despite rapid growth in truck ownership - from 700 vehicles in 1904 to over 1.1 million by 1920 - trucks largely operated as short-distance complement to railroads rather than substitutes for long-distance freight, mainly due to the poor condition of intercity roads (Herbst & Wu, 1973). Similarly, early commercial aviation was oriented toward passenger and mail services; air freight consisted primarily of low-volume, high-value goods and represented a negligible share of total intercity freight flows (Schmitt & Gollnick, 2015).

one of these studies focuses on the U.S. (Maurer & Rauch, 2023). Consequently, the evidence on the effect of the Panama Canal on regional economic growth is sparse.

This knowledge gap is exacerbated by studies that provide an incomplete picture by focusing on individual outcomes, despite evidence that transportation infrastructure can have heterogeneous sectoral effects (Chandra and Thompson, 2000). Indeed, Belmar (2023) documents structural transformation in Colombian municipalities, with employment shifting from agriculture to manufacturing. In contrast, Galiani et al. (2025) and Maurer and Rauch (2023) report positive associations for both manufacturing and agriculture. However, these studies focus primarily on individual outcomes - manufacturing in the former and population in the latter - leaving a gap in our understanding of how the Canal reshaped manufacturing and agricultural activity across the U.S.

This paper aims to fill this gap by estimating the effects of plausibly exogenous variation in market potential induced by the Panama Canal on manufacturing output per capita, population growth, and agricultural land values across U.S. counties. Additionally, I analyze whether this association depends on pre-existing levels of human capital and agricultural labor market tightness, and whether it varies across the U.S. census divisions. To do so, I construct a dataset using Maurer and Rauch's (2023) measure of market potential change, which is calculated for 2,424 counties as the difference between a county's market potential before and after the full opening of the Canal in 1920. I combine this with U.S. Census data for the period 1900-1940, digitized by Haines and ICPSR (2010), for the outcome and moderating variables, allowing me to examine growth in manufacturing output per capita, population, and agricultural land values. For this, I employ a DiD model with no untreated group and a continuous treatment, defined as the change in market potential induced by the opening of the Canal.

I find no significant association between canal-induced changes in market potential and manufacturing output per capita, but a positive and significant association with population growth and agricultural land values. However, pre-trends limit a causal interpretation of these associations. Additionally, I find selected heterogeneity across pre-canal characteristics: human capital increases the effect on population growth, and agricultural labor tightness decreases the effect on agricultural land values. Lastly, I find that the effect of market potential varies across divisions, with manufacturing and agriculture often moving in opposite directions across regions, and even coinciding with declines across all economic outcomes in the West North Central region.

This paper builds on Maurer and Rauch (2023), as I use their measure of market potential change and examine the same U.S. counties. This study adds to their paper in four important

ways. First, while they focus on population growth, I extend the analysis to manufacturing output per capita and agricultural land values, providing a broader assessment of growth and showing that pre-trends limit a causal interpretation of both their findings and mine. Second, whereas they find a positive association between market potential and manufacturing wages, I find no corresponding effect for manufacturing output per capita, suggesting that wage responses did not necessarily translate into production growth. Third, I analyze heterogeneity in pre-canal county characteristics, identifying how initial human capital and agricultural labor-market tightness influence the canal-induced effect on selected outcomes. Fourth, I document substantial spatial heterogeneity across census divisions, indicating that overall estimates may conceal pronounced regional differences in responses.

## **2. Historical Background: The United States and The Panama Canal**

This section provides the historical and economic context needed to understand how the Panama Canal could have influenced U.S. economic growth in the early twentieth century. I start by documenting the construction of the Panama Canal and how it altered U.S. shipping routes in Section 2.1. I then examine key features of the U.S. economy and outline the channels through which the canal could have stimulated economic growth in Section 2.2.

### **2.1. The Panama Canal**

#### **2.1.1. The Building of the Canal**

The first attempt to build a Panama Canal began in 1881 by the French Panama Canal Company under the leadership of Ferdinand de Lesseps, whose earlier achievement at Suez fostered fatal overconfidence. De Lesseps underestimated the difficulties that a combination of yellow fever, malaria, and tropical climate presented, resulting in the death of around 20,000 workers, mainly from the West Indies. In addition, the determination to construct the canal at sea level was an error in judgment. The company eventually went bankrupt in 1889 after spending \$287 million, leaving the canal unfinished and the French initiative discredited (McCullough, 1977, pp. 203-235).

The Americans had long been interested in building the Canal to facilitate commerce and to strengthen the efficiency and striking power of the American Navy (Johnson, 1912). Accordingly, they took over the project on January 4, 1902, buying all the Panama Assets

of the French company for \$40 million (Maurer & Yu, 2008). Panamanian leaders were also eager to complete the Canal, which would bring their country business and prestige. However, Panama was still part of Colombia at the time, which deliberately stalled negotiations with the United States, hoping to delay an agreement until the French concession expired in 1904 and thereby capture part of the \$40 million valuation of French assets (Maurer & Yu, 2008). Recognizing the opportunity presented by Panama's interests, President Roosevelt engaged with the Panamanians. With encouragement and military support from the U.S., Panama launched a coup in 1903 (McCullough, 1977, p. 376). Within days, Panama became an independent nation and quickly signed a treaty granting the U.S. rights to construct the canal.

Just a decade after the French left, the Americans were ready to start digging. The construction would come at an unprecedented financial and human cost.<sup>4</sup> However, the Americans eventually succeeded. On January 7th, 1914, the first ship finally transited the Canal (McCullough, 1977, p.607), inaugurating a new maritime route that drastically reduced the effective distance between the U.S. coasts and international ports.

### **2.1.2. How the Canal Shortened Shipping Routes**

Before the opening of the Canal, all trade requiring movement between the Atlantic and Pacific oceans relied on four interoceanic routes. These routes imposed significant constraints not only on domestic trade between the U.S. coasts but also on international trade involving the U.S. East and South Coasts and Asia, Oceania, and the west coast of South America, as well as trade between the U.S. West Coast and both Europe and the east coast of South America.

First, the Cape Horn Route required a lengthy and hazardous passage around the southern tip of South America. This route was slow, dangerous, and expensive, and was generally unsuitable for time-sensitive or high-value cargoes (Hichborn & Thiesen, 2010). For example, a voyage from New York to San Francisco using the Cape Horn Route covered approximately 13,135 nautical miles and required about 55 days for a 10-knot freight steamer<sup>5</sup> (Johnson, 1912).

---

<sup>4</sup> The Americans had spent \$352 million since 1904, four times what the Suez Canal had cost, without even considering the sum spent by the French. Taken together, the French and American expenditures came to about \$639 million. In addition, if deaths incurred during the French era are included, the total price in human life may have been as high as 25,500 lives per mile of the Canal (McCullough, 1977, p. 610).

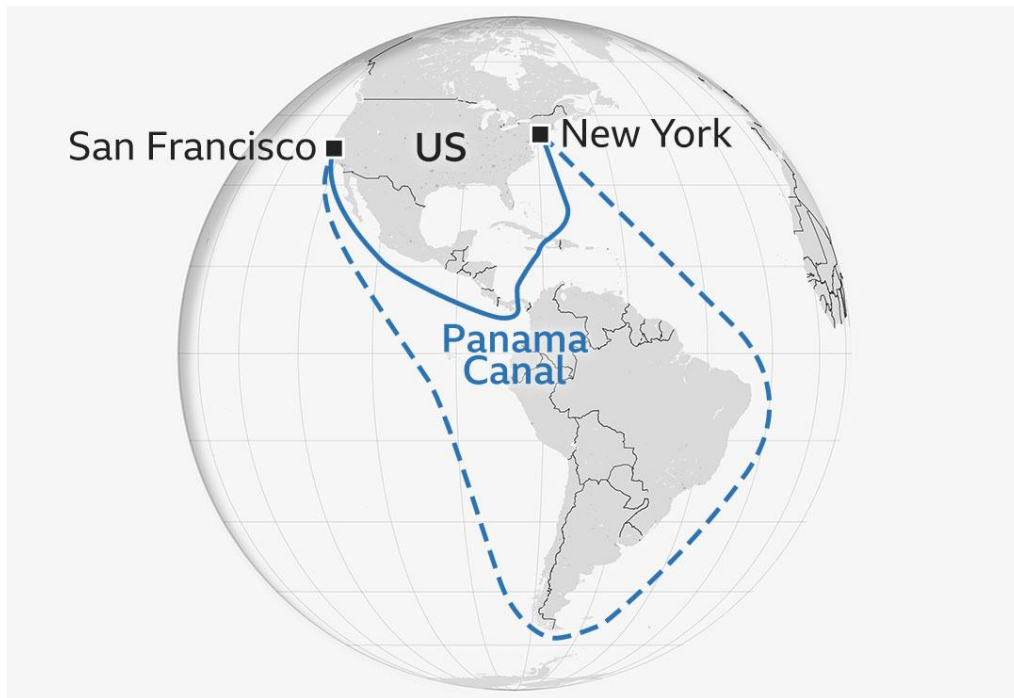
<sup>5</sup> For clarity, all travel time calculations in this section correspond to a 10-knot freight steamship, a representative shipping speed in the early-twentieth century. The implied travel time is calculated by dividing the approximate Cape Horn distance between New York and San Francisco ( $\approx 13,135$  nautical miles) by a speed of 10 knots (nautical miles per hour), following Johnson (1912), yielding roughly 1,314 hours, or approximately 55 days.

Second, the Transcontinental Railroads, completed in 1869 (Churella, 2018), established a continuous land connection across the United States and became the predominant means of freight transportation in the nineteenth century (Donaldson & Hornbeck, 2016). Nevertheless, the cost of transporting bulk freight by rail was substantially higher than by sea; estimates suggest that per-mile rail freight rates were at least an order of magnitude greater than those of ocean transport in the early twentieth century (Maurer & Rauch, 2023).

Third, the Panama Railroad, inaugurated in 1855, was a 76-kilometre railway that facilitated interoceanic transit near the future site of the Panama Canal. Despite its financial success and status as the world's busiest railroad by freight volume during the nineteenth century, it was fundamentally inefficient relative to maritime routes (Maurer & Yu, 2010). The necessity of unloading cargo, rail transport, and subsequent reloading rendered it unsuitable for bulky or heavy goods, as the resultant "transshipment costs" constituted a significant implicit tariff on trade.

Fourth, the Tehuantepec Railroad, managed by the Mexican National Railway, offered a route from New York to San Francisco that was 1,016 nautical miles shorter than the Panama Canal route. However, similar to the Panama Railroad, the economic viability of this alternative was undermined by significant transshipment costs at Tehuantepec ultimately making it less competitive than the Panama Canal (Johnson, 1912).

The opening of the Panama Canal offered a far more efficient alternative, significantly reducing the cost of transporting bulk goods between oceans. The East and West Coasts of the United States became economically closer (see Figure 1). For instance, the shipping distance between New York and San Francisco decreased by 7,873 miles, resulting in a reduction of transit time by approximately 33 days (Johnson, 1912).



**Figure 1. The Reduction in Shipping Distance due to Panama Canal.**

*Source:* Murphy, Horton, and Rivault (2025, February 14)

The dramatic reductions in distance and transit time induced by the Panama Canal translated into pronounced canal usage by U.S. vessels. (see Table 1). Although traffic remained low prior to 1920 due to landslides and World War I, which delayed the Canal’s full opening until July 1920 (Maurer & Yu, 2008), traffic expanded rapidly thereafter. From 1920 onward, traffic surged<sup>6</sup> and remained well above pre-1920 levels throughout most of the 1930s, justifying the designation of 1920 as the baseline year after which the Canal started having the most significant effect on U.S. trade. These sustained traffic levels show the Canal’s importance for U.S. trade during this period. Between 1915 and 1940, U.S. vessels accounted for nearly half of all Canal transits (almost 50,000) and transported approximately half of the total cargo tonnage (nearly 250,000 tons).

---

<sup>6</sup> In the ten years after 1920, the average annual cargo tonnage was nearly four times that recorded during the five years prior. Between 1915 and 1919, the Panama Canal handled an annual average of approximately 5,894 thousand tons of cargo (29,472 thousand tons divided by five years). In contrast, during the period 1921–1930, average annual cargo volumes increased to approximately 23,704 thousand tons (237,041 thousand tons divided by ten years). This corresponds to a fourfold rise in average annual cargo throughput between the two periods (23,704 divided by 5,894 equals 4.02).

**Table 1. Panama Canal Commerce: 1915-1940**

Year	Number of vessels	Tons of cargo (thousands)	U.S. vessels	U.S. cargo
1915	1,058	4,888	459	2,188
1916	724	3,093	213	849
1917	1,738	7,055	404	1,476
1918	1,989	7,526	567	2,008
1919	1,948	6,910	784	2,759
1920	2,393	9,372	1,129	4,547
1921	2,791	11,596	1,210	5,163
1922	2,665	10,883	1,095	4,951
1923	3,908	19,566	1,994	11,055
1924	5,158	26,993	2,947	16,654
1925	4,592	23,957	2,326	13,080
1926	5,087	26,030	2,432	13,711
1927	5,293	27,734	2,685	15,242
1928	6,253	29,616	2,753	14,249
1929	6,289	30,648	2,700	14,076
1930	6,027	30,018	2,885	14,499
1931	5,370	25,065	2,417	11,805
1932	4,362	19,799	1,917	8,835
1933	4,162	18,161	1,686	7,988
1934	5,234	24,704	2,269	11,578
1935	5,180	25,310	2,143	10,826
1936	5,382	26,506	1,936	10,701
1937	5,387	28,108	1,670	9,844
1938	5,524	27,386	1,780	9,893
1939	5,903	27,867	1,788	9,909
1940	5,370	27,299	1,997	12,385
Total	109,787	526,090	46,186 (42,1%)	240,271 (45,7%)

*Note:* The fiscal year ends on June 30 of the respective calendar year. Bracketed values indicate the share of U.S. vessels and cargo of the total.

*Sources:* Department of Commerce (1930, p. 442) for U.S.-specific data for 1915–1919; Department of Commerce (1935, p. 400) for U.S.-specific data for 1919–1934; Department of Commerce (1938, p. 418) for U.S.-specific data for 1935–1937; Department of Commerce (1939, p. 442) for the U.S.-specific 1938 figure; Department of Commerce (1941, p.501) for U.S.-specific data for 1940; and Department of Commerce (1940, p. 466) for aggregated figures.

## 2.2. The American Economy

### 2.2.1. Broad Patterns of Growth

The first four decades of the twentieth century were among the most formative and turbulent periods in American economic history. During this time, production and trade were reconfigured by three major macroeconomic shocks: World War I, the Roaring Twenties, and the Great Depression (see Appendix A for a discussion). Despite facing several economic downturns, this period experienced robust overall growth, largely fueled by technological change. Table 2 shows that TFP - the residual after accounting for physical and human capital

- accounted for approximately 60 percent of labor productivity growth between 1899 and 1941, compared to 14 percent and 25 percent for physical and human capital, respectively, and its contribution increased over time.

**Table 2. Sources of Growth in the Private Domestic Economy, 1899-1941  
(percent per annum)**

	1899-1929	1929-1941	1899-1941
1. Labor productivity (2) + (3) + (4)	2.04	2.48	2.16
Contributions from:			
2. Physical capital	0.38	0.15	0.31
3. Human capital	0.59	0.46	0.55
4. Total factor productivity	1.07	1.86	1.29

*Notes:* All variables are expressed in per hours term. Growth rates in natural logarithms.  
*Source:* Author’s elaboration based on Bakker et al. (2019).

TFP growth was especially pronounced in manufacturing. In particular, the sub-inventions stemming from electric light and power, and the internal combustion engine, such as electric hand and machine tools, fundamentally increased manufacturing productivity (Gordon, 2017). These productivity gains were reinforced by the diffusion of mass-production methods, pioneered by Henry Ford’s introduction of the assembly line in 1913, which revolutionized production away from artisan-based, steam-powered workstations toward standardized production enabled by electric machinery (Gordon, 2017). To fully capture these productivity gains, the newly organized manufacturing sector had to reach larger markets to benefit from economies of scale. By lowering transport costs and integrating previously segmented markets, the Panama Canal may have supported this scale expansion, stimulating economic growth in manufacturing.

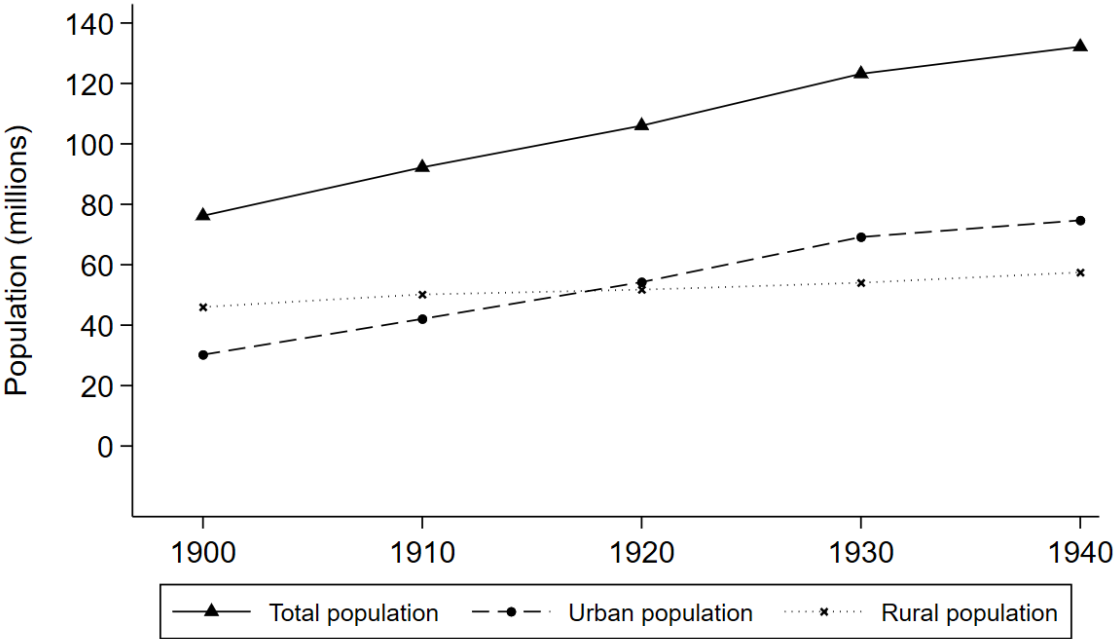
Consistent with this trend, the manufacturing sector gained economic importance over the agricultural sector (see Figure B1). Yet agriculture also saw substantial increases in productivity and yields from widespread mechanization, especially through the adoption and improvement of tractors.<sup>7</sup> These improvements in productivity and yield provided farmers with an

---

<sup>7</sup> In the first two decades of the twentieth century, the number of tractors in the U.S. was marginal, with the vast majority of farmers using horses and mules. In 1920, there were still only about 100,000 tractors across the U.S. Over the following two decades, this number increased fifteenfold, reaching around 1.5 million tractors in

opportunity to commercialize and increase their exports. Thus, the Panama Canal may have increased agricultural economic growth by lowering transportation costs and increasing the profitability of exportable crops.

Throughout the period 1900-1940, the U.S. population shifted from rural to urban areas (see Figure 2). Although the rural population continued to increase, it did so at a very slow pace and almost stagnated after 1920. Importantly, during the two decades after the opening of the Panama Canal, foreign migration was low (see Figure B2), implying that observed regional population adjustments primarily reflect internal migration. Therefore, the steady expansion of the urban population throughout the period suggests that urban areas became increasingly more attractive in the early twentieth century. The Panama Canal may have accelerated this trend by unevenly improving economic opportunities in certain areas, thereby stimulating population reallocation toward these locations and fostering the expansion or emergence of urban hubs.



**Figure 2. Total, Urban, and Rural Population, 1900-1940.**

*Note:* Urban places are those with 2,500 or more people.

*Source:* U.S. Census Bureau (1995).

1940 (see Figure 1 in Steckel & White, 2012). During the latter period, the quality of tractors also improved drastically, allowing tractors to deliver mobile power at a far lower cost.

### **2.2.2. Regional Growth, the Panama Canal, and Market Potential**

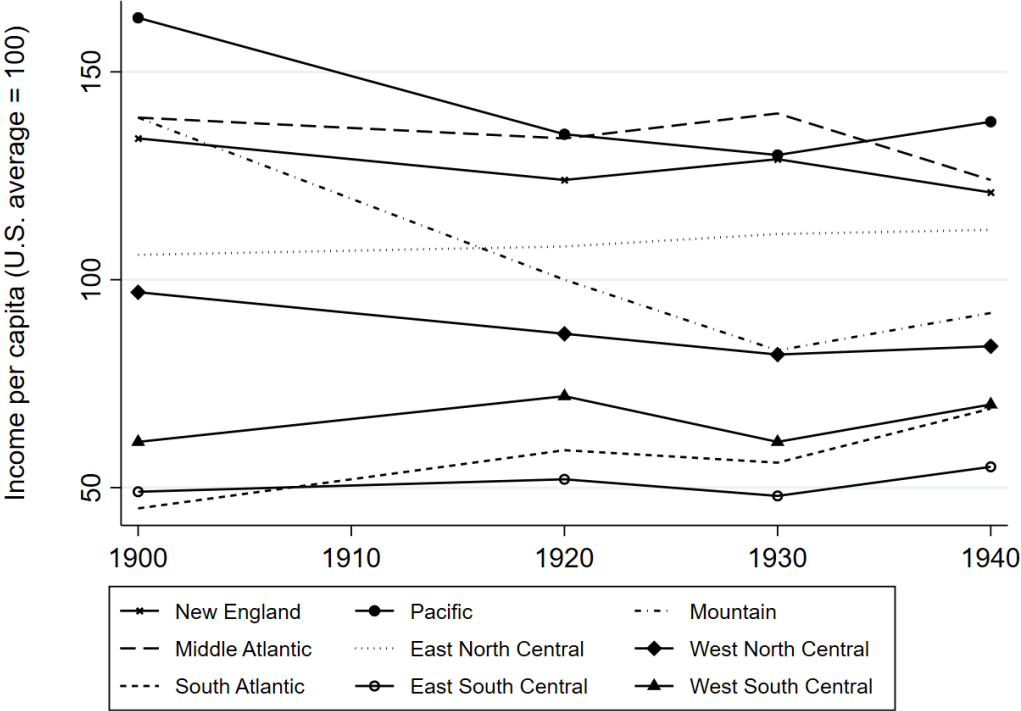
Before the opening of the Panama Canal in 1920, there was already pronounced economic heterogeneity between counties. There was strong industrial divergence, with the formation of a manufacturing belt, composed mostly of New England, Middle Atlantic, and East North Central counties, and the continued specialization of agriculture in the South. This also led to significant divergence in income per capita, especially between the North and the South, with the latter lagging behind (Kim & Margo, 2004). These industrialized counties were generally the most urbanized, and the most populated regions of the United States (Heim, 2000, p.93). Manufacturing tended to cluster in counties with high market potential to capture scale effects and to benefit from forward linkage effects (Klein & Crafts, 2012). Hence, industrialized counties were typically characterized by superior connectivity to domestic and international markets.

Notably, however, manufacturing was not the sole pathway to regional prosperity during this period: the Pacific region was among the most wealthy between 1900 and 1940 despite limited manufacturing specialization. The Pacific Northwest relied heavily on exports of crude and processed primary products, in particular forestry, agriculture, and animal products (Pfister, 1963), while California specialized primarily in distributive and service industries (Heim, 2000, p.154). This demonstrates that high income levels were not exclusively associated with manufacturing specialization, suggesting that the Panama Canal could also have stimulated economic growth through agriculture and other sectors.

The Panama Canal substantially altered the effective distance between many U.S. counties - especially those near the coasts - and international markets, thereby increasing counties' market potential. The Canal's effect on market potential depended on factors such as proximity to a U.S. port, the proximity of that port to other international or U.S. ports, and rail connections (see Appendix C for details on the market potential measure). As a result, the plausibly exogenous market potential shock induced by the opening of the Canal was heterogeneous across U.S. counties. Accordingly, we would expect counties experiencing larger gains in market potential to exhibit greater economic growth than those experiencing smaller gains.

Nevertheless, this prediction is not reflected in aggregate regional trends. There is no obvious structural break in regional economic growth coinciding with the opening of the Panama Canal in 1920. Figure 3, which displays regional personal income per capita as a percentage of the national average over the period 1900–1940, shows no clear post-1920 upward shift in relative income levels for any region. Instead, most of the regional economic

differences existing in 1900, prevailed through 1940. Of all the regions with above average income in 1900 - Pacific, Middle Atlantic, Mountain, New England, and East North Central - only Mountain was not able to retain above the national average throughout the entire period 1900-1940. Conversely, all the below average regions remained below the national average throughout the entire period.



**Figure 3. Regional Income per Capita Relative to the U.S. Average, 1900–1940**

*Note:* Regional income per capita is expressed as an index relative to the U.S. national average, which is normalized to 100 in each year. Regions correspond to the nine U.S. Census Bureau divisions.

*Source:* Author’s elaboration based on Easterlin (1961).

Taken together, these patterns indicate strong persistence in relative regional income levels over the first four decades of the twentieth century, rather than a clear structural break around 1920. However, as Figure 3 aggregates outcomes across nine broad regions, these averages may conceal substantial within-region heterogeneity. Consequently, the absence of a visible regional break does not rule out economically meaningful effects of the Panama Canal; rather, it underscores the need for a county-level analysis that exploits within-region variation in changes to market potential.

### **3. Theoretical Framework**

This section outlines the mechanisms through which market potential may affect economic growth. I first discuss the channels underlying the overall effect in Section 3.1, and then examine how these effects are moderated by human capital and agricultural labor market tightness in Section 3.2.

#### **3.1. Overall Effect of Market Potential on Economic Growth**

The role of market potential in shaping regional economic development originates with Harris (1954), who argued that the potential demand faced by producers in a given location can be represented by the distance-weighted economic mass (e.g., GDP) of surrounding locations. This notion of market potential is central to many New Economic Geography (NEG) models (Fujita et al., 1999; Combes et al., 2008), which build upon Krugman's (1991) core-periphery model. The framework in this thesis adopts the core mechanisms of Krugman's model while incorporating additional empirical and theoretical considerations.

The primary intuition is that higher market potential - reflecting cheaper access to consumer and input markets - increases the effective demand faced by local producers and reduces delivered input costs. Consequently, economic activity expands in counties with higher market potential as firms locate or relocate there to economize on shipping costs for both outputs and inputs. In turn, rising labor demand increases local employment and wages, thereby attracting in-migration. This migration further reinforces agglomeration because new residents also increase local demand, enlarging the consumer market and strengthening the incentives for firms to concentrate in these counties (Krugman, 1991).

This mechanism is expected to be particularly strong in manufacturing, given substantial increasing returns to scale in that sector (Krugman, 1991). Improvements in market potential raise effective demand and support larger production scales, allowing manufacturing firms to spread fixed costs over greater output and thereby reduce average costs. By contrast, the mechanism may be weaker in agriculture, where economies of scale are less pronounced, and production is more tightly constrained by local soil and climate suitability (Mohana et al., 2009).

Even though the pull of better access to consumer and input markets is widely documented (Redding & Venables, 2004; Hanson, 2005; Jacks & Novy, 2018), several countervailing mechanisms can dampen agglomeration in counties with higher market potential. For example, as firms enter and expand in these counties, product-market competition intensifies, limiting

profitability (Krugman, 1991). Moreover, agglomeration can increase the cost of living, most notably through higher housing costs, decreasing further agglomeration (Hanson, 2005).

In addition, economic activity could leak out of counties as their market potential improves. Because market access is bidirectional, improved connectivity also increases exposure to competition from more productive core locations. In such cases, peripheral counties may lose market share and experience slower industrial output growth as production re-concentrates in the core (Faber, 2014).

Lastly, manufacturing and agriculture may be inversely linked if they compete for scarce factors, such as land and labor (Matsuyama, 1992). In this case, an expansion of manufacturing in a county may coincide with a contraction of agriculture, or vice versa, as factors reallocate toward the sector with higher relative returns.

Overall, these mechanisms imply that exogenous improvements in market potential should, on average, shift economic activity toward counties that experience larger increases in market potential. This leads to the following hypothesis:

***Hypothesis 1:** Increases in market potential induced by the opening of the Panama Canal stimulate regional economic activity by increasing effective demand and reducing delivered input costs.*

## **3.2. Moderating Effects**

Although the overall effect of market potential is expected to be positive, U.S. counties differ, and the extent to which they respond to market potential might depend on pre-existing characteristics. Therefore, it is important to consider variables that might moderate the canal's effect. The literature provides particular motivation for analyzing human capital (Cohen & Levinthal, 1990) and agricultural labor market tightness (Picard & Zeng, 2005) as potential moderating variables. The mechanisms through which they might operate are discussed below.

### **3.2.1. The Moderating Effect of Human Capital**

Human capital refers to the collection of skills and knowledge that people acquire through education, training, and experience that improve their productive capacities (Schultz, 1961). Human capital increases an economy's absorptive capacity - its ability to recognize valuable external opportunities, assimilate information, and apply it commercially (Cohen & Levinthal, 1990). Accordingly, when market potential rises due to better access to external consumer and

input markets, counties with more human capital are likely better able to scale production, adopt new technologies, and reorganize activity toward higher-value uses.

This mechanism is expected to be particularly strong in manufacturing, as industrial production increasingly relied on more complex machinery and processes in the 1920s and 1930s (Gordon, 2017), raising the relative demand for skilled labor (Goldin, 2000, p.612). Therefore, pre-existing human capital levels probably influenced whether firms could expand in response to better market access.

Human capital may also strengthen agricultural responses to market potential by facilitating technology adoption and commercialization (Huffman, 2020). However, the moderating effect of human capital on agricultural growth may be weaker if educated and skilled workers disproportionately sort into non-agricultural employment, given substantially higher nonfarm wages (Caselli & Coleman, 2001).

Taken together, increases in market potential create opportunities for expansion, but the extent to which counties can realize prolonged growth likely depends on their pre-existing productive capabilities. Human capital enhances counties' abilities to adapt to emerging opportunities and technologies, making growth responses to market potential more pronounced where it is higher. This leads to the following hypothesis:

***Hypothesis 2:** The growth effect of Panama Canal-induced increases in market potential is larger in counties with higher pre-existing human capital because it enables firms to scale production, adopt new technologies, and reorganize economic activity in response to improved market potential.*

### **3.2.2. The Moderating Effect of Agricultural Labor Market Tightness**

Increases in market potential are expected to contribute to local agglomeration of economic activity, but the magnitude of the response likely depends on whether counties can access additional labor at relatively stable wages. Both manufacturing expansion and commercial agriculture relied heavily on unskilled labor. In practice, manufacturing frequently drew workers from the agricultural sector because farm workers entering manufacturing were typically paid less than incumbent industrial workers with similar observable characteristics (Schultz, 1961). However, local labor supply is not perfectly elastic in the short run, meaning that migration can only partially offset labor needs, and local labor-market conditions can constrain where firms expand (Picard & Zeng, 2005).

We would therefore expect the market-potential shock to generate a smaller growth response in counties where agriculture already absorbs substantial unskilled labor. When local agricultural labor demand is already high, there is plausibly less “surplus” labor available to reallocate. As a result, additional expansion is more likely to bid up wages than increase employment and output. Consequently, pre-existing agricultural labor market tightness likely attenuates the positive effect of market potential on economic growth. This leads to the following hypothesis:

*Hypothesis 3: The growth effect of Panama Canal-induced increases in market potential is smaller in counties with higher pre-existing agricultural labor market tightness because limited availability of unskilled labor constrains firms’ ability to expand production in response to improved market potential.*

## 4. Data

The substantial importance of the Panama Canal to U.S. trade, together with the heterogeneous changes in market potential it induced, motivates an empirical test of the hypotheses outlined in the previous section. To do so, I construct a dataset covering 2,424 U.S. counties - nearly all continental counties<sup>8</sup> - with decadal observations from 1900 to 1940. I exclude post-1940 observations because the measure of market potential becomes less reliable over time for at least two reasons: (i) U.S. transport infrastructure likely adapted endogenously to the Panama Canal, and (ii) expanding global trade increased the Canal’s value over time. This section describes the data sources and their limitations, examines data patterns, and explains the empirical approach used to evaluate the hypotheses.

### 4.1. Market Potential

Market potential is often used as a proxy for trade flows when actual trade flows are not observed (Harris, 1954; Hanson, 2005; Maurer & Rauch, 2023). This study uses data on county-specific changes in market potential induced by the opening of the Panama Canal, calculated by Maurer and Rauch (2023) for 2,424 U.S. counties. This section briefly explains how they

---

<sup>8</sup> The states and territories included in the 1900 census were divided into 2,867 counties or equivalent subdivisions (U.S. Census Bureau, 1901). Accordingly, my sample covers approximately 85% (2,424 of 2,867) of all counties enumerated in the 1900 census.

constructed this measure and outlines some concerns (see Appendix C for a detailed explanation).

Similar to Harris (1954), Maurer and Rauch’s market potential measure weights reachable populations by income per capita and by effective distance. They calculate both domestic and international market potential and sum them to obtain the total market potential. To compare market potential before and after the Panama Canal’s opening, they construct effective distances under both scenarios using least-cost paths between U.S. counties and between U.S. counties and international ports, assuming overseas shipping is ten times cheaper than inland transport.

Dividing the total market potential after the opening by the total market potential before, then gives the change in market potential due to the opening of the Panama Canal:

$$\Delta MP_c^{tot} = \frac{MP_c^{tot} | Canal}{MP_c^{tot} | No Canal} \quad (1)$$

This ratio provides the percentage change in market potential induced by the Canal and offers the advantage of easy-to-interpret summary statistics. In addition, it does not change with the arbitrary unit of distance measurement.

One limitation of this measure is the somewhat ambiguous parameter for the relative cost of shipping, which determines whether a county uses a canal route, and, consequently, the least-cost paths. Furthermore, the assumption that GDP per capita is constant across U.S. counties is unrealistic, and tariffs are not considered. Lastly, the canal-induced change in market potential is constant over time, whereas many developments between 1920 and 1940 could have influenced it. For instance, a shift in the relative cost of shipping to rail transport, or new or improved infrastructure.

While a comprehensive sensitivity analysis is beyond the scope of this thesis, Maurer and Rauch (2023) demonstrate that their main results are robust to alternative specifications of market potential that attempt to incorporate these concerns.

## 4.2. Economic Outcomes

### 4.2.1. Manufacturing Output Per Capita

The first outcome variable is the total value of manufacturing output per capita, calculated as a county’s total annual manufacturing output in dollars divided by that county’s population. I

obtained the yearly value of total manufacturing output from Haines and ICPSR (2010), who digitized the 1900, 1920, 1930, and 1940 censuses of manufacturing; 1910 was missing, and then divided it by the total county population in each year.

From 1899 onward, the reliability of the Census of Manufactures improved drastically. The participation rate of large firms improved substantially, securing nearly universal participation among significant establishments. The Census also drew a sharper distinction between small shops and factory manufacturing, improving consistency and comparability over time (U.S. Census Bureau, 1902a, 1933). Data quality also improved in 1900 with greater reliance on technically trained “special agents,” thereby drastically improving the quality of the complicated data collection process. Even so, measurement issues remained: suspicious reports from smaller establishments were often imputed using industry ratios; some errors persisted where regular enumerators collected schedules; and, beginning in 1921, the reporting threshold rose from \$500 to \$5,000, which may mechanically depress measured growth in counties with many small establishments. Finally, county-level manufacturing output reflects gross output rather than value added; the latter would be preferable, but, to the best of my knowledge, is unavailable at the county level for this period.

#### **4.2.2. Population Growth**

The second outcome variable is the county population growth from the 1900-1940 U.S. Censuses sourced from Haines and ICPSR (2010). Population growth is widely used as a proxy for economic growth in economic geography and structural transformation studies (Dajud, 2017; Maurer & Rauch, 2023; Belmar, 2023; Galiani et al., 2025). Population can be a valuable proxy for economic activity, as historically opportunities to earn a living mainly determined the areas where people settled (Easterlin, 2000, p. 526).

Nevertheless, despite its prevalent use, utilizing population growth as a proxy for local economic growth has limitations. Many non-economic factors can influence population growth, including amenities, fertility rates, and weather conditions. Additionally, there are concerns related to how population data is collected. The population counts were gathered by enumerators - specialized census takers assigned to small districts. By law, they were required to visit every household in their district and record the responses on census schedules (U.S. Census Bureau, 1901, 1942a). However, enumerators often did not receive adequate training, resulting in considerable variations in the quality of their work, with some reports being of poor quality. As a result, the reliability of the total population count may be low in certain counties. Moreover, rural, minority, and migrant populations were systematically undercounted, so

migration involving these groups may be only weakly reflected in measured census population growth.

### **4.2.3. Agricultural Land Value**

The third outcome variable is agricultural land value, measured as the average dollar value of farmland and buildings per acre in the 1900-1940 U.S. Censuses, retrieved from Haines and ICPSR (2010). Land value is also widely used as a proxy for economic growth in economic geography and structural transformation studies (Donaldson & Hornbeck, 2016; Dajud, 2017; Maurer & Rauch, 2023; Galiani et al., 2025). The rationale is that the value of agricultural land, as an immobile asset, reflects the profitability of agricultural practices.

However, agricultural land values pose significant limitations for interpretation and measurement. Changes in values may not reflect agricultural profitability alone; they can also capture increased competition for land from non-agricultural uses associated with urbanization (Kuethe et al., 2011). Measurement is also concerning: from 1900 to 1940, Census enumerators visited farms and recorded the farm operator's self-reported market value of the farm land and buildings under the condition of a willing buyer and seller, rather than a forced sale (U.S. Census Bureau, 1902b, 1942b). Because these values were self-reported rather than based on transactions or appraisals, they capture farmers' perceptions and expectations rather than realized prices. Self-reported values may also be "sticky": Rajan et al. (2015) show Census land values between 1910 and 1930 were far less volatile than contemporaneous transaction data from the Department of Agriculture. Given these limitations, the empirical results should be interpreted with these sources of ambiguity in mind.

## **4.3. Moderators**

### **4.3.1. Human Capital**

To assess the hypothesis that the growth effect of Panama Canal-induced increases in market potential is larger in counties with higher pre-existing human capital because it enables firms to scale production, adopt new technologies, and reorganize economic activity in response to improved market potential, I use data on school attendance for individuals aged 6–20 in 1910<sup>9</sup> from that year's Census, sourced from Haines and ICPSR (2010). I calculate the school attendance share by dividing the number of enrolled 6-to-20-year-olds by the total population

---

<sup>9</sup> This study considers the 1910 rate of human capital, and later the 1900 figure of farm labor expenditures, to avoid the "bad control problem" - a situation where the variable is affected by market potential changes resulting from the Canal's opening (Cinelli et al., 2024).

in that age group. School attendance ratios are a standard proxy of human capital in economic papers (Barro, 1991; Mankiw et al., 1992), as they reflect the flows of education at a point in time, and the accumulation of these flows will be a substantial element in the stocks of human capital available later (Barro & Lee, 1993).

Even so, there are various limitations regarding the measurement of school attendance. First, the measure relies entirely on self-reporting,<sup>10</sup> which can produce inaccuracies when people misinterpret the question or deliberately provide false information. Second, the measure does not distinguish between the amount of time spent attending school, meaning a child who attended once was counted the same as one attending full-time. Lastly, it does not distinguish between different types of education, which limits its suitability as a proxy of human capital, because higher education clearly has a greater effect on one's knowledge and technical know-how. Unfortunately, to the best of my knowledge, type-specific schooling rates are not available for U.S. counties in the early twentieth century.

#### **4.3.2. Agricultural Labor Market Tightness**

To examine the hypothesis that the growth effect of Panama Canal-induced increases in market potential is smaller in counties with higher pre-existing agricultural labor market tightness because limited availability of unskilled labor constrains firms' ability to expand production in response to improved market potential, I use county-level data on total annual expenditures on farm labor for each county in 1900 from the U.S. census, digitized by Haines and ICPSR (2010). I scale this measure by county population to obtain farm labor expenditures per capita, which I interpret as a proxy for pre-canal local agricultural labor market tightness.

Naturally, there are limitations in using farm-labor expenditures as a proxy for pre-canal labor market tightness. First, data-collection concerns are similar to those for the self-reported measures discussed above: enumerators visited farms and recorded total labor outlays, analogous to the procedure used for land values (U.S. Census Bureau, 1902b). Second, relying on 1900 figures may weaken comparability for later decades. This concern is heightened by mechanization in U.S. agriculture between 1900 and 1940, lowering labor requirements on farms. Because tractor adoption varied across counties,<sup>11</sup> the extent to which 1900 farm-labor

---

<sup>10</sup> An enumerator asked every person on the population schedule if they attended school at any time between September 1909 and April 1910. Those answering "yes" and who were in the 6–20 age range were tabulated in the school-attendance reports (U.S. Census Bureau, 1913).

<sup>11</sup> For example, tractor numbers rose by 131% in North Dakota between 1930 and 1940, compared with 191% in Mississippi (U.S. Census Bureau, 1942b).

expenditures proxy for post-1920 labor market tightness also varies. Lastly, farm labor expenditures do not directly measure labor scarcity. Clearly, it would have been preferable to have county-level unemployment data, but, to the best of my knowledge, this is not available for the early twentieth century.

#### **4.4. Descriptive Statistics**

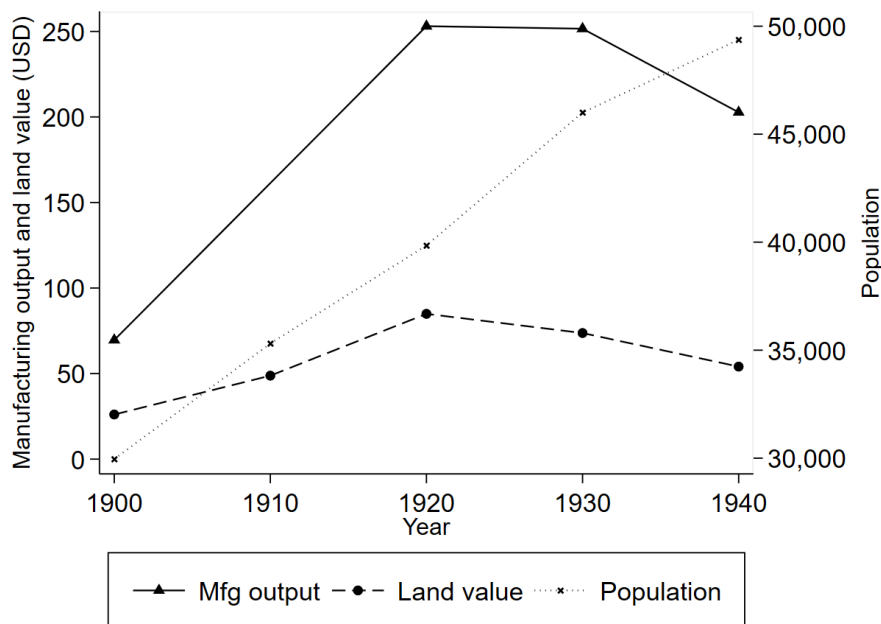
Table 3 displays considerable cross-county heterogeneity in the variables introduced above. The mean change in market potential as ratio in 1930 and 1940 is 1.029, indicating that the Panama Canal increased market potential by 2.9% for the average county. There is pronounced variation, with the majority of counties experiencing minor changes, while a small share face substantial changes ranging up to 31%, providing meaningful identifying variation. The levels of manufacturing activity also differ substantially across counties. The average manufacturing output per capita is \$192, with a large standard deviation (\$306) and values ranging from zero to over \$5,000. Population levels are similarly dispersed, with an average of about 40,097 thousand inhabitants, a standard deviation exceeding 125,000, and county populations ranging from fewer than 20 residents to over 4 million. Land values exhibit extreme variation as well, averaging \$57.60 per acre but spanning from \$1 to nearly \$5,000. Human capital measures also vary considerably: school attendance among those aged 6-20 averages 63%, yet ranges from 24% to 82%. Finally, per-capita farm labor expenditures in 1900 are low on average (\$6.72) but highly variable, suggesting that hired farm labor was concentrated in a limited subset of counties.

**Table 3. Summary statistics for some selected variables**

	Mean	Std. Dev.	Min	Max	N
$\Delta$ MP	1.029	0.027	1.000	1.31	4,850
$\Delta$ lnMP	0.029	0.025	0.000	0.27	4,850
Population	40,097	125,734	17,000	4,063	12,120
Ln population	9.951	0.938	2.833	15.218	12,120
Mfg. output p.c.	191.7	306	0.000	5,232	8,826
Ln mfg. output p.c.	4.341	1.437	-1.403	9.038	8,814
Land value	57.57	139.9	1.000	4,999	12,079
Ln land value	3.543	0.948	0.000	8.517	12,079
Schooling 1910	0.63	0.092	0.240	0.823	12,120
Farm labor 1900	6.26	7.82	0.047	91.94	12,115

*Note:*  $\Delta$ MP is the ratio of total market potential after the Canal's opening in 1920 to the market potential before the opening.  $\Delta$ lnMP is the log difference between the post-Canal and pre-Canal market potential measures. Population is the total population. Mfg. output p.c. is manufacturing output per capita, measured in dollars. Land value is the value of agricultural land and buildings, measured in dollars per acre. Schooling 1910 is the share of the 6–20 year-old population that attended school in 1910. Farm labor 1900 represents per capita expenditures on farm labor in 1900, measured in dollars.

To see how the outcome variables evolved over time, Figure 4 plots decadal values from 1900 to 1940. As illustrated, the average population continued to grow at approximately the same rate after the opening of the Canal. Conversely, manufacturing output per capita and agricultural land values increased from 1900 to 1920, but declined from 1920 to 1940, most likely due to the Great Depression.



**Figure 4. Economic Activity, 1900-1940.**

*Notes:* The left y-axis reports averages across all U.S. counties in the sample of manufacturing output per capita and land values (USD), while the right y-axis reports average county population.

## 5. Method

This section describes the methodology used to assess the hypotheses presented in Section 3. I start by replicating the estimates of Maurer and Rauch in Section 5.1. I proceed by explaining the baseline model and the specification with the moderating variables in Sections 5.2 and 5.3, respectively. Lastly, I introduce an additional model to analyze spatial heterogeneity in Section 5.4.

### 5.1. Replication Maurer and Rauch (2023): Empirical Strategy

I first revisit the findings of Maurer and Rauch (2023), as this study builds on their work. They employ two types of regressions. First, they estimate a long-difference regression on the logs of population, average manufacturing wages, and agricultural land values on changes in market potential, using 1900 as the pre-canal year and 1940 as the post-canal year. Second, their preferred specification is essentially the same as my baseline model, but with clustering at the 5-by-5-degree grid level, and a panel for the period 1880-1940, with 20-year intervals. Therefore, their panel contains only one post-opening year (1940). While they only estimate this specification for the log population, I use their dataset to apply it to average manufacturing wages and agricultural land values, allowing for a direct comparison of the panel coefficients with those obtained by Maurer and Rauch from the long-difference model. The details of the replication are explained in Appendix D.

### 5.2. Overall Analysis: Empirical Strategy

To test the first hypothesis that increases in market potential induced by the opening of the Panama Canal stimulate regional economic activity by increasing effective demand and reducing delivered input costs, I employ the following baseline specification:

$$Y_{c,t} = \beta \Delta \ln MP_{c,t} + \alpha_c + \tau_t + X'_c \cdot \delta_t + \varepsilon_{c,t} \quad (2)$$

Where  $Y_{c,t}$  is the outcome variable in county  $c$ , in year  $t$  (i.e. the log of manufacturing output per capita, the log of population, and the log of agricultural land values). The coefficient of interest ( $\beta$ ) shows the average effect of changes in market potential on economic activity across 1930 and 1940. The log of market potential change due to the Canal's opening in 1920 in county  $c$  is given by  $\Delta \ln MP_{c,t}$ , and equals zero in all years prior to the Canal's opening and takes a county-specific constant value in all post-opening years. Moreover, the specifications

incorporate both county and year fixed effects, denoted by  $\alpha_c$  and  $\tau_t$ , respectively. In contrast to random effects models, fixed effects models do not assume that county-specific characteristics are unrelated to the explanatory variables (Verbeek, 2017, pp. 386-395). This makes fixed effects models more robust in accounting for omitted-variable bias arising from time-invariant, county-specific factors that may be correlated with the regressors, a common feature of panel data analysis. In addition, I include time-fixed effects to capture temporal shocks and global trends that affect all counties simultaneously, such as World War I, the Roaring Twenties, and the Great Depression. Without time-fixed effects, the estimated impact of market potential could be confounded with shocks and trends that affect all regions simultaneously, leading to biased estimates. Moreover, the set of controls, which is interacted with time-fixed effects ( $\delta_t$ ), is constant across all specifications and includes longitude, latitude, the county's 1880 population, and a coastal dummy variable. Longitude and latitude are introduced to control for broad geographical trends, ensuring that geographic factors, such as climate, do not confound the estimated effects of changes in market potential. The 1880 population is included because it might influence subsequent growth, for instance, through agglomeration effects (Krugman, 1991). Lastly, a coastal dummy is introduced to control for general population movement to the coast.

Because equation (2) is equivalent to a difference-in-differences design with multiple periods, conventional standard errors tend to be biased downward and therefore lead to over-rejection of the null hypothesis. This problem arises because the treatment variable often varies only at the group level and the outcome variables are typically serially correlated. To address this issue, I follow the recommendation of Bertrand et al. (2004) and cluster standard errors at the county level, which allows for arbitrary serial correlation in the error terms within each county over time. In Appendix F, I also show that the results are robust to clustering at the 5-by-5-degree level following Maurer and Rauch (2023).

According to Callaway et al. (2024), our coefficient of interest ( $\beta$ ) can only be interpreted as causal if we strengthen the parallel trends assumption, as our design resembles a difference-in-differences approach with continuous treatment and no untreated group. The parallel trends assumption requires that, in the absence of treatment, the average outcome trajectories of counties receiving any dose  $d$  would have evolved in the same way as the trajectories observed for untreated counties. Since there is no untreated group in this analysis, it is essential to strengthen this assumption. This is achieved through the strong parallel trends assumption, which asserts that there is no selection bias into treatment doses. In other words, it indicates that the observed outcomes in each dose group reflect what would have occurred in all other

groups had they received that same dose. Although the strong parallel trends assumption is fundamentally untestable, Section 6.2 examines whether it might hold in the context of this study.

One situation in which parallel trends would be violated is if the log change in market potential caused by the canal were endogenous (Cunningham, 2021). Therefore, the underlying assumption - that the Panama Canal was indeed an exogenous shock - must hold. This implies that its construction and subsequent impact on market potential were not determined by, nor systematically correlated with, pre-existing local economic, geographic, or institutional conditions that could independently influence the outcome of interest. In econometric terms, this implies that the variation in market potential induced by the Canal is independent of the error term in the regression model, allowing it to be treated as a plausibly exogenous source of variation for identifying causal effects (Verbeek, 2017, p. 145). One phenomenon that might violate this assumption is simultaneity, the possibility that the dependent variable also influences one or more independent variables (Verbeek, 2017, p. 148). Although simultaneity cannot be ruled out, it is unlikely to hold in this context, as market potential is constructed without post-opening data, and counties could not influence the treatment intensity because the canal's location was fixed. Another concern is that the canal might be correlated with other economic, geographical, or institutional drivers of economic activity outside the model. This possibility is investigated in Appendix K.

### 5.3. Moderating Variables: Empirical Strategy

To test the second hypothesis that the growth effect of Panama Canal-induced increases in market potential is larger in counties with higher pre-existing human capital because it enables firms to scale production, adopt new technologies, and reorganize economic activity in response to improved market, I employ the following specification:

$$Y_{c,t} = \beta_1 \Delta \ln MP_{c,t} + \beta_2 (\Delta \ln MP_{c,t} \cdot HighSchooling_{c,1910}) + \alpha_c + \tau_t + X'_c \cdot \delta_t + \varepsilon_{c,t} \quad (3)$$

Here,  $HighSchooling_{c,1910}$  is a dummy variable equal to 1 if the share of 6-to-20-year-olds enrolled in school in county  $c$  in 1910 is above the median, and 0 otherwise. This dummy is interacted with the change in market potential induced by the canal. The coefficient of the interaction term ( $\beta_2$ ) reflects whether the effect of the change in market potential on economic activity is positively moderated by human capital. In addition,  $HighSchooling_{c,1910}$  is

interacted with year dummies to control for the possibility that counties would have had different trajectories based on their 1910 schooling levels, even in the absence of the canal. The rest of the model is identical to equation (2).

Additionally, to test the third hypothesis that the growth effect of Panama Canal–induced increases in market potential is smaller in counties with higher pre-existing agricultural labor market tightness because limited availability of unskilled labor constrains firms’ ability to expand production in response to improved market potential, I estimate the following model:

$$Y_{c,t} = \beta_1 \Delta \ln MP_{c,t} + \beta_2 (\Delta \ln MP_{c,t} \cdot HighLabor_{c,1900}) + \alpha_c + \tau_t + X'_c \cdot \delta_t + \varepsilon_{c,t} \quad (4)$$

The equation is identical to equation (3), except that the interaction term now includes a dummy variable that is 1 if the percentage of farm labor expenditures per capita in 1900 is above the median, and 0 otherwise. The coefficient of the interaction term ( $\beta_2$ ) displays whether the impact of changes in market potential on economic activity is negatively moderated by agricultural labor market tightness.  $HighLabor_{c,1900}$  is also interacted with year dummies to account for the possibility that counties with different levels of labor market tightness would have experienced distinct trajectories, even in the absence of the canal.

## 5.4. Spatial Heterogeneity Across Census Divisions: Empirical Strategy

To test the possibility that the overall results mask spatial differences, I adjust equation (2) to allow the marginal effect of changes in market potential on the economic outcomes to vary across the nine US census divisions. The specification reads as follows:

$$Y_{c,t} = \beta_{ENC} \Delta \ln MP_{c,t} + \sum_{d \neq ENC} \theta_d (1\{Div_c = d\} \times \Delta \ln MP_{c,t}) + \alpha_c + \tau_t + X'_c \cdot \delta_t + \varepsilon_{c,t} \quad (5)$$

Here, East North Central - a large industrialized division around the 1920s - is the omitted division. Thus,  $\beta_{ENC}$  is the marginal effect of the log change in market potential ( $\Delta \ln MP$ ) in East North Central and  $\theta_d$  captures the difference in the effect of  $\Delta \ln MP$  for division  $d$  relative to East North Central. The rest of the specification is identical to equation (2).

## 6. Results

This section presents the empirical results. I first report the replication of Maurer and Rauch (2023) in Section 6.1. I then present the results of the main analysis, including the event study estimates, in Section 6.2. Finally, I report the results of the analyses of the moderating variables and spatial heterogeneity in Sections 6.3 and 6.4, respectively.

### 6.1. Replication of Maurer and Rauch (2023): Empirical Results

The replication results are presented in Table 4. Columns 1-3 present estimates from their long-difference specification, showing that changes in market potential have a positive and significant effect across all outcomes. Columns 4-6 present the panel model results. While Maurer and Rauch report only population, my replication shows that the effects on average manufacturing wages and agricultural land values are also positive and significant, though substantially smaller in magnitude than in the long-difference estimates.

**Table 4. Replication of Maurer and Rauch (2023)**

VARIABLES	(1) Dif. wages	(2) Dif. pop	(3) Dif. land	(4) Ln wages	(5) Ln pop	(6) Ln land
$\Delta \ln MP$	1.962*** (0.483)	3.733*** (1.117)	5.120*** (1.294)	0.764* (0.387)	2.303** (1.058)	4.196*** (1.128)
Observations	1,928	2,424	2,397	6,659	12,125	11,958
R-squared	0.060	0.156	0.381	0.835	0.898	0.745
Specification	Long difference	Long difference	Long difference	TWFE	TWFE	TWFE
County FE	No	No	No	Yes	Yes	Yes
Year FE	No	No	No	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	47	47	47	47	47	47

*Note:* All coefficients are elasticities. In the long-difference specifications, the dependent variables are the changes between 1900 and 1940 in the logs of average manufacturing wages, population, and the average agricultural land value per acre. In the panel specifications, the dependent variables are the logs of manufacturing wages, population, and the average agricultural land value per acre.  $\Delta \ln MP$  denotes the change in market potential induced by the canal. The set of control variables includes longitude, latitude, log 1880 population, and a coastal county indicator. Standard errors are clustered at the 5-by-5-degree grid level.\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

### 6.2. Overall Analysis: Empirical Results

Table 5 reports the results for the first hypothesis that increases in market potential induced by the opening of the Panama Canal stimulate regional economic activity by increasing effective demand and reducing delivered input costs (H1). Contrary to this hypothesis, the estimates provide no evidence that canal-induced increases in market potential raised manufacturing

output per capita. Although the coefficient of interest is positive and statistically significant in the specification that includes only county fixed effects (column 1), it loses statistical significance once year fixed effects are introduced (column 2), and remains insignificant after adding the full set of control variables (column 3). This finding contrasts with Maurer and Rauch (2023), Belmar (2023), and Galiani et al. (2025), who report positive effects on alternative manufacturing outcomes in the U.S., Colombia, and Canada, respectively. The results also differ from the positive association between market potential and three alternative manufacturing outcomes reported in Appendix E. The absence of a robust effect on manufacturing output, which directly captures production in arguably the most important export sector,<sup>12</sup> suggests that average estimates may conceal spatial heterogeneity. This is examined in Section 6.4.

**Table 5. Impact of the Change in Market Potential on Economic Activity**

VARIABLES	(1) Ln mfg	(2) Ln mfg	(3) Ln mfg	(4) Ln pop	(5) Ln pop	(6) Ln pop	(7) Ln land	(8) Ln land	(9) Ln land
$\Delta \ln MP$	6.829*** (0.443)	0.359 (0.500)	0.608 (0.597)	3.977*** (0.203)	2.467*** (0.343)	2.046*** (0.360)	7.939*** (0.311)	5.344*** (0.427)	4.760*** (0.378)
Observations	8,764	8,764	8,764	12,120	12,120	12,120	12,079	12,079	12,079
R-squared	0.783	0.875	0.877	0.946	0.951	0.956	0.712	0.925	0.945
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes	No	No	Yes
Clusters	2361	2361	2361	2424	2424	2424	2422	2422	2422

*Note:* All coefficients are elasticities. The dependent variables are: the log of manufacturing output per capita, the log of population, and the log of the average agricultural land value per acre.  $\Delta \ln MP$  refers to the change in market potential induced by the canal. The set of control variables consists of longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors, clustered at the county level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Conversely, the population responds strongly and positively to increases in market potential, supporting H1. The specification that only includes county fixed effects (column 4) shows that the coefficient of interest has the expected positive sign. Including year fixed effects (column 5) and the control variables (column 6) does not change the sign, but it reduces their size considerably. In my preferred specification, a 1% increase in market potential is associated with a 2.046% increase in population on average in the post-opening years. This corresponds to roughly 20 per cent of the observed population growth, indicating a substantial reallocation of

<sup>12</sup> In the early twentieth century, manufacturing formed the core of most cities' export base, particularly in the Northeast and North Central states (Heim, 2000, p. 154).

people to regions benefiting from improved market potential created by the canal.<sup>13</sup>The elasticity is very similar to the estimate by Maurer and Rauch (2023) and consistent with the association found in Canadian counties by Galiani et al. (2025).

Agricultural land values also exhibit a strong positive response to changes in market potential due to the canal, consistent with H1. Column 7 shows that the specification with only county fixed effects yields a large, positive, and statistically significant coefficient. Adding year fixed effects slightly reduces the magnitude (column 8), and including the set of controls further reduces the effect, though it remains significant (column 9). The estimated elasticity of 4.76 implies that a 1% increase in market potential raised agricultural land values by approximately 4.76% in the post-opening years. This accounts for roughly 16.7% of observed growth,<sup>14</sup> indicating that the canal substantially influenced agricultural growth. The estimate is consistent with findings for U.S. counties reported by Donaldson and Hornbeck (2016) and Maurer and Rauch (2023), as well as with the estimate of Galiani et al. (2025) for Canadian counties.

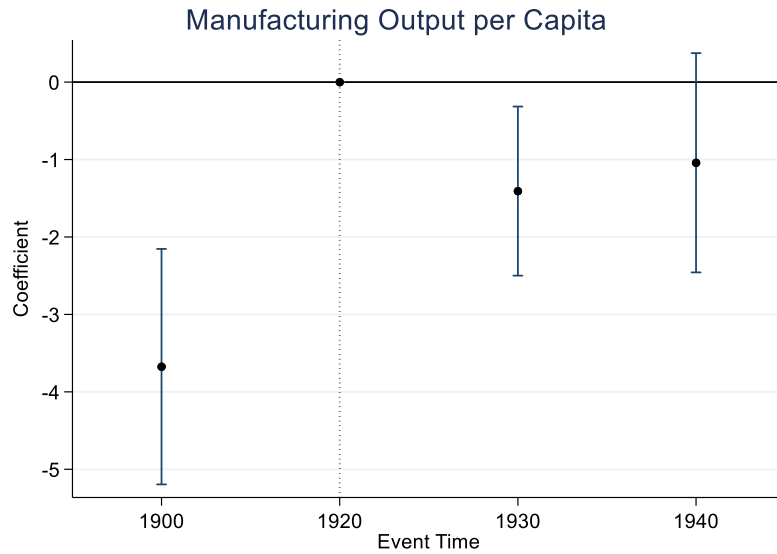
To assess the causal interpretation of the previous findings, I examine whether the outcome variables respond to future changes in market potential before the canal was opened. I interact the log market potential change with dummies for 1900, 1910, 1930, and 1940, with 1920 being the omitted reference category. If market potential “predicts” population before the canal opened, that would suggest that pre-existing trends or omitted variables are driving the results.

For manufacturing output per capita, the main analysis showed no significant effect of changes in market potential. Figure 5 reinforces this conclusion by highlighting two critical issues. First, the level of manufacturing output per capita was systematically lower in counties experiencing the greatest changes in market potential before 1920, undermining the plausibility of the parallel trends assumption required for causal interpretation (Cunningham, 2021). Second, even if we could interpret the post-opening coefficients causally, Figure 5 shows a negative relationship in 1930 and 1940 compared to 1920, directly contradicting H1.

---

<sup>13</sup> Computed as the product of the mean change in market potential (0.0287) and the estimated elasticity (2.046), implying an average 5.9% population increase. Since total population grew by roughly 29.4% over the period, changes in market potential account for about 20% of overall population growth.

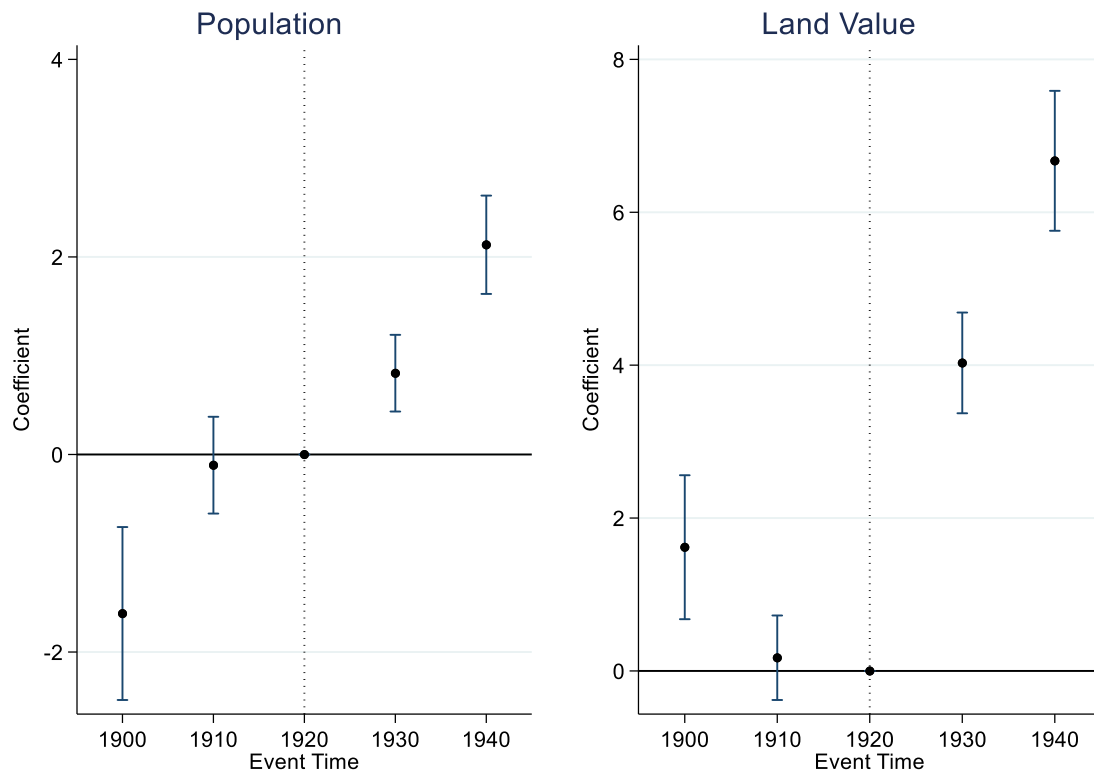
<sup>14</sup> Computed as the product of the mean change in market potential (0.0287) and the estimated elasticity (4.76), implying an average 13.66% land value increase. Since total population grew by roughly 82% over the period, changes in market potential account for about 16.7% of overall growth in land values.



**Figure 5. Event study plot of Market Potential Change - Manufacturing.**

*Notes:* The figure shows the effects of the Panama Canal's change in market potential on manufacturing output per capita. The vertical bars are the 95% confidence intervals. Standard errors are clustered at the county level. Data for the year 1910 is missing.

Similarly, the presence of pre-trends for both population and agricultural land values in Figure 6 raises concerns about a causal interpretation of their estimated effects. Although the left panel shows a clear increase in population following the canal's opening, counties that would later experience larger gains in market potential already exhibited systematically different population growth patterns as early as 1900, when the canal shock is estimated to have had a substantial negative effect. A comparable pattern emerges for agricultural land values: counties with greater anticipated gains in market potential had higher land values in 1900, experienced declines through 1920, and only then saw increases after the canal opened. Taken together, these patterns complicate a causal interpretation of the estimates presented in Table 5.



**Figure 6. Event study plot of Market Potential Change.**

*Notes:* The figure shows the effects of the Panama Canal's change in market potential on population growth (left) and agricultural land values (right) over time. The vertical bars are the 95% confidence intervals. Standard errors are clustered at the county level.

### 6.3. Moderating Variables: Empirical Results

Table 6 presents the results of the analysis of the moderating variables. The findings offer mixed support for the second hypothesis that the growth effect of Panama Canal-induced increases in market potential is larger in counties with higher pre-existing human capital, as this enables firms to scale production, adopt new technologies, and reorganize economic activity in response to improved market potential (H2). Consistent with this hypothesis, column 3 shows that the interaction coefficient for population growth is 4.06 (significant at the 1% level), indicating that the market potential effect was 4.06 percentage points higher in counties with above-median human capital in 1910 than in below-median counties. In contrast, there is no significant interaction effect for manufacturing output per capita and agricultural land values (see columns 1 and 5). It is especially notable that human capital did not significantly influence the response

of manufacturing output per capita, as manufacturing heavily relied on human capital.<sup>15</sup> One explanation is that high-skilled workers migrated rapidly towards counties experiencing gains in market potential, and therefore pre-canal levels of human capital were not constraining. Another possibility is that manufacturing output per capita was mainly limited by capital availability rather than local human capital endowments. This is particularly plausible during the 1930s as the Great Depression reduced the credit availability, limiting firm activity (Bernanke, 1983). Alternatively, the share of 6-20-year-olds attending school in 1910 may fail to capture the relevant features of human capital required for industrial expansion (see Section 4.3.1 for a discussion of data limitations).

**Table 6.**

**Moderating Variables**

VARIABLES	(1) Ln mfg	(2) Ln mfg	(3) Ln pop	(4) Ln pop	(5) Ln land	(6) Ln land
$\Delta \ln MP$	1.013 (1.008)	1.632 (1.221)	-0.927 (0.618)	2.624*** (0.714)	3.983*** (0.701)	6.033*** (0.781)
$\Delta \ln MP \times \text{HighSchooling}$	-0.493 (1.150)		4.059*** (0.746)		0.795 (0.805)	
$\Delta \ln MP \times \text{HighLabor}$		-1.142 (1.311)		-0.720 (0.775)		-1.897** (0.830)
Observations	8,764	8,764	12,120	12,120	12,079	12,079
R-squared	0.877	0.877	0.958	0.956	0.946	0.946
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	2361	2361	2424	2424	2422	2422

*Note:* All coefficients are elasticities. The dependent variables are the log of manufacturing output per capita, the log of population, and the log of average agricultural land values per acre.  $\Delta \ln MP$  measures the average effect of increases in market potential from the opening of the Panama Canal in 1920 for counties with below-median school attendance (ages 6–20) in 1910 or below-median farm labor expenditures per acre in 1900.  $\Delta \ln MP \times \text{HighSchooling}$  captures the differential effect for counties with above-median 1910 school attendance.  $\Delta \ln MP \times \text{HighLabor}$  captures the differential effect for counties with above-median 1900 farm labor expenditures per acre. Control variables include longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors are clustered at the county level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

The results are also mixed for the third hypothesis that the growth effect of Panama Canal–induced increases in market potential is smaller in counties with higher pre-existing agricultural labor market tightness because limited availability of unskilled labor constrains firms’ ability to expand production in response to improved market potential (H3). Column 6 shows that the interaction coefficient is only significant for agricultural land values (-1.897), indicating that

<sup>15</sup> Industrial production increasingly relied on more complex machinery and processes in the 1920s and 1930s (Gordon, 2017), raising the relative demand for skilled labor (Goldin, 2000, p.612).

the market potential effect was roughly 1.9 percentage points smaller in counties with above median agricultural labor market tightness, compared to below median. For manufacturing output per capita and population growth, the interaction is insignificant, indicating that low-skilled agricultural labor scarcity did not bind a county's ability to increase manufacturing activity or attract population in response to increased market potential. A plausible explanation is that low-skilled workers rapidly migrated towards expanding counties, thereby pre-empting labor availability is not a binding constraint. Another possibility is that, consistent with Acemoglu and Guerrieri (2008), mechanization or capital deepening in the manufacturing sector decreased or muted this sector's demand for unskilled agricultural labor. Furthermore, agricultural labor expenditures per capita in 1900 may not accurately reflect actual unskilled labor supply during 1920-1940 (see Section 4.3.2 for a discussion of data limitations).

## **6.4. Spatial Heterogeneity Across Census Divisions: Empirical Results**

Table 7 presents the results of the spatial analysis across the nine U.S. census divisions, showing strong spatial heterogeneity in the impact of changes in market potential. The findings suggest that increases in manufacturing output per capita in response to market potential are not widespread; instead, they are primarily located in the industrial core of the East North Central, while agricultural values decreased in this region. This pattern is consistent with Matsuyama (1992), who predicts that increases in market potential shift economic activity toward the sector of comparative advantage - manufacturing or agriculture - at the expense of the other due to competition for land and labor. By contrast, this prediction does not hold in the other two industrial cores: in the Middle Atlantic and New England, the effect on manufacturing output is small or negative, while the effect on land values is positive.<sup>16</sup> A similar trade-off is evident in most other divisions: agricultural land values often respond positively to gains in market potential, while manufacturing output per capita responds little or even negatively. The market potential effect on population growth is also very dispersed across divisions, with population growth not systematically aligning with changes in manufacturing output per capita or agricultural land values. Most notably, the estimates for the West North Central are negative across all economic outcomes, suggesting that increases in market potential may induce an

---

<sup>16</sup> The elasticity is calculated as the sum of the baseline estimate and the estimate for that specific division relative to the baseline division. Thus, for the Middle Atlantic, the elasticity for manufacturing output per capita equals  $5.893\% - 5.636\% = 0.26\%$ , and for New England it is  $5.893\% - 7.308\% = -1.415\%$ .

outflow of economic activity. A possible explanation, following Faber (2014), is that higher market potential intensifies competition by enabling firms in other regions to more easily access local markets, thereby losing economic activity to other regions.

**Table 7. Spatial Heterogeneity Across Nine U.S. divisions**

VARIABLES	(1) Ln mfg	(2) Ln pop	(3) Ln land
$\Delta \ln MP$	5.893* (3.468)	1.577 (1.551)	-3.542** (1.502)
$\Delta \ln MP \times$ New England	-7.308*** (2.075)	0.615 (1.018)	7.434*** (0.996)
$\Delta \ln MP \times$ Middle Atlantic	-5.636*** (2.006)	2.762*** (1.002)	6.959*** (0.918)
$\Delta \ln MP \times$ West North Central	-23.930*** (5.903)	-8.433*** (2.454)	-4.364** (1.843)
$\Delta \ln MP \times$ South Atlantic	-2.379 (2.253)	-2.574*** (0.997)	5.934*** (0.845)
$\Delta \ln MP \times$ East South Central	-5.643** (2.564)	-2.183** (1.063)	9.395*** (0.933)
$\Delta \ln MP \times$ West South Central	-4.434 (3.577)	-0.504 (1.502)	11.988*** (1.613)
$\Delta \ln MP \times$ Mountain	-5.651 (5.617)	-3.926 (3.130)	3.679 (2.437)
$\Delta \ln MP \times$ Pacific	-6.484 (3.959)	1.175 (1.947)	8.825*** (2.006)
Observations	8,764	12,120	12,079
R-squared	0.878	0.957	0.947
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Clusters	2361	2424	2422

*Note:* All coefficients are elasticities. The dependent variables are the log of manufacturing output per capita, the log of population, and the log of average agricultural land values per acre.  $\Delta \ln MP$  measures the average effect of increases in market potential from the opening of the Panama Canal in 1920 for counties in the East North Central. The other coefficients reflect the effect in counties in the other eight divisions relative to East North Central. Control variables include longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors are clustered at the county level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## 7. Robustness Checks

This section evaluates whether the baseline results are sensitive to potential identification concerns and alternative empirical choices. Appendix F shows that the vast majority of my results are robust to clustering at the 5-by-5-degree grid level, following the approach of Maurer and Rauch (2023). Appendix G shows that there are also no signs of multicollinearity. I further assess robustness to alternative specifications, additional covariates, and different trend

structures. Appendix H shows that the baseline findings are broadly consistent with a long difference specification, although the coefficient of manufacturing output per capita is positively significant in the latter. Similarly, Appendix I shows that alternative functional forms do not provide a better fit to the data than the baseline specification. Lastly, while Appendix J shows that the baseline results are not robust to the addition of state fixed effects, state trends, and county trends, Appendix K shows that the baseline estimates are robust to adding the following covariates interacted with time-fixed effects: soil quality, pre-canal industrial structure, and a coastal dummy.

## 8. Conclusion

This paper studied the hypothesis that increases in market potential induced by the opening of the Panama Canal stimulate regional economic activity by increasing effective demand and reducing delivered input costs. To do so, it exploited the plausibly exogenous variation provided by the opening of the canal to examine its effect on three economic outcomes: manufacturing output per capita, population, and agricultural land values. Using a DiD model with a continuous treatment and no untreated group, the overall analysis finds no significant association between canal-induced changes in market potential and manufacturing output per capita, while the association with population growth and land values was significantly positive. However, the presence of pre-trends across all outcomes undermine a causal interpretation of these findings.

This study also examined two hypotheses concerning heterogeneity in the effect of canal-induced increases in market potential. First, the growth effect is expected to be larger in counties with higher pre-existing human capital, as this facilitates firms' ability to scale production, adopt new technologies, and reorganize economic activity in response to improved market access. Second, the growth effect is expected to be smaller in counties with tighter pre-existing agricultural labor markets because limited availability of unskilled labor constrains production expansion. The results reveal selected heterogeneity: high human capital increased the effect on population growth, while agricultural labor market tightness decreased the effect on agricultural land value. Additionally, by allowing the market potential effect to vary across the nine U.S. Census divisions, I found that the overall results masked pronounced spatial heterogeneity. There was often a trade-off between growth in manufacturing and agriculture, and in the West North Central region, market potential coincided with declines across all economic outcomes.

The results presented in this paper should be interpreted with caution as this analysis is subject to several limitations. First, the results may be sensitive to the construction of the market potential measure and to the assumption that canal-induced changes in market potential remain constant after 1920. Second, most variables are retrieved from historical censuses and are therefore prone to measurement errors. Third, given the presence of pre-trends, the estimates should be interpreted as descriptive associations rather than causal effects. Lastly, the analysis considers manufacturing, agriculture, and population, excluding other potentially important sectors, and provides only a partial assessment of spatial heterogeneity. Future research should consider extending the analysis to other sectors and further address spatial heterogeneity by studying the effects within regional subsamples.

Despite these limitations, the findings presented in this paper have various important implications for the existing literature. The presence of pre-trends across all outcomes suggests that caution is warranted when interpreting estimated canal effects in studies where parallel trends are not established. Moreover, the contrasting results for manufacturing output per capita and average manufacturing wages suggest that relying on wages may reflect labor market conditions rather than actual manufacturing growth. In addition, the selected heterogeneity observed in pre-canal levels of human capital and in agricultural labor market tightness suggests that average estimates in the literature may provide an incomplete picture in some contexts. Finally, pronounced spatial heterogeneity across all outcomes indicates that the overall results reported in the literature might mask regional differences.

This paper highlights the complex role of transportation infrastructure in shaping the economic geography of nations. The findings indicate that trade infrastructure can benefit some regions more than others by increasing their market potential. However, raising a region's market potential may also induce a structural reallocation between manufacturing and agriculture, and can even adversely affect some regions. These results present important guidance for policymakers when designing and evaluating transportation infrastructure investments.

## References

- Acemoglu, D., & Guerrieri, V. (2008). Capital deepening and nonbalanced economic growth. *Journal of political Economy*, 116(3), 467-498.
- Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Atack, J., Bateman, F., Haines, M., & Margo, R. A. (2010). Did railroads induce or follow economic growth?: Urbanization and population growth in the American Midwest, 1850–1860. *Social Science History*, 34(2), 171-197.
- Bakker, G., Crafts, N., & Woltjer, P. (2019). The Sources of growth in a Technologically Progressive Economy: The United States, 1899-1941. *Economic Journal*, 129(622), 2267-2294.
- Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, 106(2), 407-443.
- Barro, R. J., & Lee, J. (1993). International comparisons of educational attainment. *Journal of Monetary Economics*, 32(3), 363-394.
- Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth?. *Journal of Urban Economics*, 115, 103124.
- Belmar, J. (2023). Trade and structural change: Evidence from Colombia and the Panama canal. *working paper*.
- Berger, T., & Enflo, K. (2017). Locomotives of local growth: The short-and long-term impact of railroads in Sweden. *Journal of Urban Economics*, 98, 124-138.
- Bernanke, B. S. (1983). *Non-monetary effects of the financial crisis in the propagation of the Great Depression* (No. w1054). National Bureau of Economic Research.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. (2004). “How Much Should We Trust Differences-in-Differences Estimates?” *Quarterly Journal of Economics* 119 (1): 249–75.
- Brancaccio, G., Kalouptsi, M., & Papageorgiou, T. (2020). Geography, transportation, and endogenous trade costs. *Econometrica*, 88(2), 657-691.
- Caselli, F., & Coleman II, W. J. (2001). The US structural transformation and regional convergence: A reinterpretation. *Journal of political Economy*, 109(3), 584-616.
- Callaway, B., Goodman-Bacon, A., & Sant’Anna, P. H. (2024). Difference-in-differences with a continuous treatment. National Bureau of Economic Research: Technical report.

- Chan, J. (2022). Farming output, concentration, and market access: Evidence from the 19th-century American railroad expansion. *Journal of Development Economics*, 157, 102878.
- Chandra, A., & Thompson, E. (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional science and urban economics*, 30(4), 457-490.
- Churella, A. (2018). Railroads in US History. In *Oxford Research Encyclopedia of American History*.
- Cinelli, C., Forney, A., & Pearl, J. (2024). A crash course in good and bad controls. *Sociological Methods & Research*, 53(3), 1071-1104.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 35(1), 128-152.
- Combes, P. P., Thisse, J. F., & Mayer, T. (2008). Economic geography: The integration of regions and nations.
- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.
- Dajud, C. U. (2017). *Domestic transport costs, Canada, and the Panama Canal* (No. 2017-02). Department of Commerce. (1930) Statistical Abstract of the United States 1930, Government Printing Office, Washington. [Statistical Abstract of the United States: 1930](#)
- Department of Commerce. (1935) Statistical Abstract of the United States 1935, Government Printing Office, Washington. [Statistical Abstract of the United States: 1935](#)
- Department of Commerce. (1938) Statistical Abstract of the United States 1938, Government Printing Office, Washington. [Statistical Abstract of the United States: 1938](#)
- Department of Commerce. (1939) Statistical Abstract of the United States 1939, Government Printing Office, Washington. [Statistical Abstract of the United States: 1939](#)
- Department of Commerce. (1940) Statistical Abstract of the United States 1940, Government Printing Office, Washington. [Statistical Abstract of the United States: 1940](#)
- Department of Commerce. (1941) Statistical Abstract of the United States 1941, Government Printing Office, Washington. [Statistical Abstract of the United States: 1941](#)
- Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A “market access” approach. *The Quarterly Journal of Economics*, 131(2), 799-858.
- Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5), 899-934.
- Durant, G., & Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, 79(4), 1407-1440.

- Easterlin, R. A. (1961). Regional income trends, 1840–1950. In S. Harris (Ed.), *American economic history* (pp. 525–547). McGraw-Hill.
- Easterlin, R. (2000). Twentieth-Century American Population Growth. In S. L. Engerman & R. E. Gallman (Eds.), *The Cambridge Economic History of the United States* (pp. 505–548). chapter, Cambridge: Cambridge University Press.
- Embassy of Panama. (n.d.). *Panama Canal*. Retrieved February 16, 2026, from [Panama Canal — EOP](#)
- Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies*, 81(3 (288)), 1046–1070.
- Fujita, M., Krugman, P., Venables, A. (1999). *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge.
- Galiani, S., Jaramillo, L. F., & Uribe-Castro, M. (2025). Free-Riding Yankees: Canada and the Panama Canal. *Journal of Economic Growth*, 1-44.
- Gallup, J. L., Sachs, J. D., & Mellinger, A. D. (1999). Geography and economic development. *International regional science review*, 22(2), 179-232.
- Goldin, C. (2000). Labor Markets in the Twentieth Century. In S. L. Engerman & R. E. Gallman (Eds.), *The Cambridge Economic History of the United States* (pp. 549–624). chapter, Cambridge: Cambridge University Press.
- Gordon, R. (2017). *The rise and fall of American growth: The US standard of living since the civil war*. Princeton university press.
- Haines, M.R.; Inter-university Consortium for Political and Social Research (ICPSR). (2010) *Historical, Demographic, Economic, and Social Data: The United States, 1790–2002*, Inter-university Consortium for Political and Social Research, Ann Arbor.
- Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of international economics*, 67(1), 1-24.
- Harris, C. D. (1954). The, Market as a Factor in the Localization of Industry in the United States. *Annals of the association of American geographers*, 44(4), 315-348.
- Heim, C. (2000). Structural Changes: Regional and Urban. In S. L. Engerman & R. E. Gallman (Eds.), *The Cambridge Economic History of the United States* (pp. 93–190). chapter, Cambridge: Cambridge University Press.
- Herbst, A. F., & Wu, J. S. (1973). Some Evidence of Subsidization: the US Trucking Industry, 1900–1920. *The Journal of Economic History*, 33(2), 417-433.

- Herzog, I. (2021). National transportation networks, market access, and regional economic growth. *Journal of Urban Economics*, 122, 103316.
- Hichborn, P., & Thiesen, W. H. (2010). *Cruise of the Dashing Wave: Rounding Cape Horn in 1860*. University Press of Florida.
- Hornbeck, R., & Rotemberg, M. (2019). Railroads, Reallocation, and the Rise of American Manufacturing. Working Paper 26594, National Bureau of Economic Research. Series: Working Paper Series.
- Hornung, E. (2015). Railroads and growth in Prussia. *Journal of the European Economic Association*, 13(4), 699-736.
- Huffman, W. E. (2020). Human capital and adoption of innovations: policy implications. *Applied Economic Perspectives and Policy*, 42(1), 92-99.
- Jacks, D. S., & Novy, D. (2018). Market potential and global growth over the long twentieth century. *Journal of International Economics*, 114, 221-237.
- Jedwab, R., & Storeygard, A. (2022). The average and heterogeneous effects of transportation investments: Evidence from Sub-Saharan Africa 1960–2010. *Journal of the European Economic Association*, 20(1), 1-38.
- Jedwab, R., & Moradi, A. (2016). The permanent effects of transportation revolutions in poor countries: evidence from Africa. *Review of economics and statistics*, 98(2), 268-284.
- Johnson, E. R. (1912). The Panama Canal and the Commerce of the United States. *Scientific American*, 107(19), 394–396.
- Kim, S., & Margo, R. A. (2004). Historical perspectives on US economic geography. In *Handbook of regional and urban economics* (Vol. 4, pp. 2981-3019). Elsevier.
- Klein, A., & Crafts, N. (2012). Making sense of the manufacturing belt: determinants of US industrial location, 1880–1920. *Journal of Economic Geography*, 12(4), 775-807.
- Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, 99(3), 483-499.
- Kuethe, T. H., Ifft, J., & Morehart, M. J. (2011). The influence of urban areas on farmland values. *Choices*, 26(2).
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, 107(2), 407-437.
- Matsuyama, K. (1992). Agricultural productivity, comparative advantage, and economic growth. *Journal of economic theory*, 58(2), 317-334.
- Maurer, N., & Yu, C. (2008). What TR took: The economic impact of the Panama Canal, 1903–1937. *The Journal of Economic History*, 68(3), 686-721.

- Maurer, N., & Yu, C. (2010). *The big ditch: how America took, built, ran, and ultimately gave away the Panama Canal*. Princeton University Press.
- Maurer, S., & Rauch, F. (2023). Economic geography aspects of the Panama Canal. *Oxford Economic Papers*, 75(1), 142-162.
- McCullough, D. (1977). *The path between the seas: the creation of the Panama Canal, 1870-1914*. Simon and Schuster.
- Michaels, G., Rauch, F., & Redding, S. J. (2012). Urbanization and structural transformation. *The Quarterly Journal of Economics*, 127(2), 535-586.
- Mohana, P., Mariappan, N. V. E., & Manoharan, N. (2009). Land suitability analysis for the part of Parambikulam Aliyar command area, Udumalpet Taluk using remote sensing and GIS techniques. *International Journal on Design and Manufacturing Technologies*, 3(2), 98-102.
- Murphy, M., Horton, J., & Rivault, E. (2025, February 14). *Mapping China's influence around the Panama Canal*. BBC News.  
[Mapping China's influence around the Panama Canal - BBC News](#)
- Ortiz-Ospina, E., Rohenkohl, B., Samborska, V., Van Teutem, S., Beltekian, D., & Roser, M. (2018). *Trade and globalization*. Our World in Data. <https://ourworldindata.org/trade-and-globalization>
- Pascali, L. (2017). The wind of change: Maritime technology, trade, and economic development. *American Economic Review*, 107(9), 2821-2854.
- Pfister, R. L. (1963). External trade and regional growth: a case study of the Pacific Northwest. *Economic Development and Cultural Change*, 11(2, Part 1), 134-151.
- Picard, P. M., & Zeng, D. Z. (2005). Agricultural sector and industrial agglomeration. *Journal of Development Economics*, 77(1), 75-106.
- Rajan, Raghuram, and Rodney Ramcharan. (2015). The Anatomy of a Credit Crisis: The Boom and Bust in Farm Land Prices in the United States in the 1920s. *American Economic Review* 105 (4): 1439–77.
- Redding, S., & Venables, A. J. (2004). Economic geography and international inequality. *Journal of international Economics*, 62(1), 53-82.
- Rockoff, H. (2004). Until it's over, over there: the US economy in World War I.
- Romer, C. D. (1993). The nation in depression. *Journal of Economic Perspectives*, 7(2), 19-39.
- Schultz, T. W. (1961). Investment in human capital. *The American economic review*, 51(1), 1-17.

- Schmitt, D., & Gollnick, V. (2015). Historical development of air transport. In *Air transport system* (pp. 19-38). Vienna: Springer Vienna.
- Steckel, R. H., & White, W. J. (2012). *Engines of growth: farm tractors and twentieth-century US economic welfare* (No. w17879). National Bureau of Economic Research.
- Temin, P. (2000). The Great Depression. In S. L. Engerman & R. E. Gallman (Eds.), *The Cambridge Economic History of the United States* (pp. 301–328). chapter, Cambridge: Cambridge University Press.
- U.S. Census Bureau (1901). 1900 Census: Volume I. Population, Part 1. Retrieved from [1900 Census: Volume I. Population, Part 1](#)
- U.S. Census Bureau (1902a). 1900 Census: Volume VII. Manufactures, Part 1. United States by Industries. Retrieved from [1900 Census: Volume VII. Manufactures, Part 1. Industries](#)
- U.S. Census Bureau (1902b). 1900 Census: Volume V. Agriculture, Part 1, Farms, Live Stock, and Animal Products. Retrieved from [1900 Census: Volume V. Agriculture, Part 1](#)
- U.S. Census Bureau (1913). 1910 Census: Abstract of the Thirteenth Census of the United States. <https://www.census.gov/library/publications/1913/dec/abstract.html>
- U.S. Census Bureau (1933). 1930 Census: Manufactures, 1929 Volume 1. General Report. [1930 Census: Manufactures, 1929 Volume 1. Statistics by Subjects](#)
- U.S. Census Bureau (1942a). 1940 Census of Population: Volume 1. Number of Inhabitants. [1940 Census of Population: Volume 1. Number of Inhabitants](#)
- U.S. Census Bureau (1942b). Sixteenth Census of the United States: 1940 — Agriculture: Volume I. Retrieved from [1940 Census — Agriculture: Volume I: 1st & 2nd Series State Reports](#)
- U.S. Census Bureau (1949). *Historical statistics of the United States* , 1789 - 1945. U.S. Government Printing Office. [Historical Statistics of the United States, 1789 - 1945](#)
- U.S. Census Bureau (1960). *Historical statistics of the United States: Colonial times to 1957*. U.S. Government Printing Office. [Historical Statistics of the United States, Colonial Times to 1957](#)
- U.S. Census Bureau (1975). Bicentennial Edition: Historical Statistics of the United States, Colonial Times to 1970. U.S. Government Printing Office. [Historical Statistics of the United States, Colonial Times to 1970](#)
- U.S. Census Bureau. (1995). *Urban and rural population: 1900–1990* (Table 1). Decennial Census, 1990. Retrieved from <https://www2.census.gov/programs-surveys/decennial/tables/1990/1990-urban-pop/urpop0090.txt>
- Verbeek, M. (2017). *A guide to modern econometrics*. John Wiley & Sons.

White, E. N. (1990). The stock market boom and crash of 1929 revisited. *Journal of Economic perspectives*, 4(2), 67-83.

## **Use of AI**

I attended the ‘Responsible Generative AI Use For BSc/MSc Theses’ workshop at the start of my thesis, and have used AI in line with the most recent university guidelines. I used Grammarly to spot and correct grammar and spelling mistakes, and to enhance clarity. I used large language models, mainly ChatGPT, for proofreading, feedback and suggestions, and help with Stata code. In addition, I used its research mode to find data sources and references, alongside elicit, and conventional search engines. I have always checked the content of papers and other AI-generated content myself before referencing or using it in any other way. I have never used AI to generate original content; I used it as an assistant to support my writing and coding. I take full responsibility for the correctness, completeness, and coherence of all my texts

## **Appendix A. World War I, the Roaring Twenties, and the Great Depression**

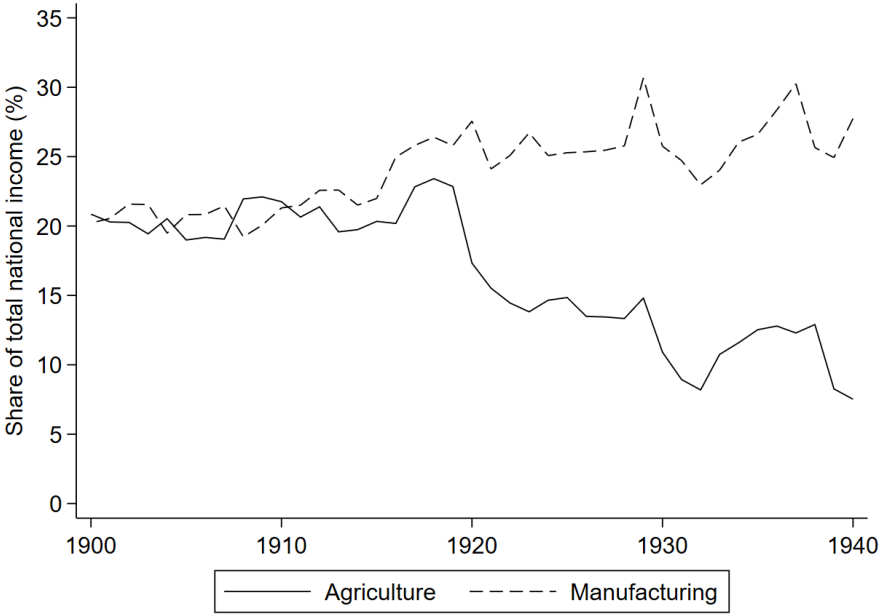
This appendix elaborates on three major economic disturbances to U.S. economic growth and international trade. The first major shock was the outbreak of World War I in 1914, which initially disrupted the American economy, which was emerging from a recession that had begun in 1913. Although financial markets were unsettled in the early months of the war, strong European demand for American goods, primarily munitions and food, triggered a long economic boom. The percentage increase in Real GNP between 1914 and 1918 rose by 26.1 percent (see Table 1 in Rockoff, 2004).

After the war, the American economy experienced a sharp yet brief recession from 1920 to 1921. Soon afterwards, technological advances and structural shifts in industry fueled solid economic growth beginning in 1922, culminating in a stock market boom during the Roaring Twenties. However, these rapid changes complicated the assessment of economic fundamentals - the underlying factors justifying stock prices - ultimately creating conditions ripe for a speculative bubble (White, 1990).

In late 1929, the collapse of the stock market’s speculative bubble triggered broad anxiety among consumers and businesses, sharply reducing durable goods purchases. This shock,

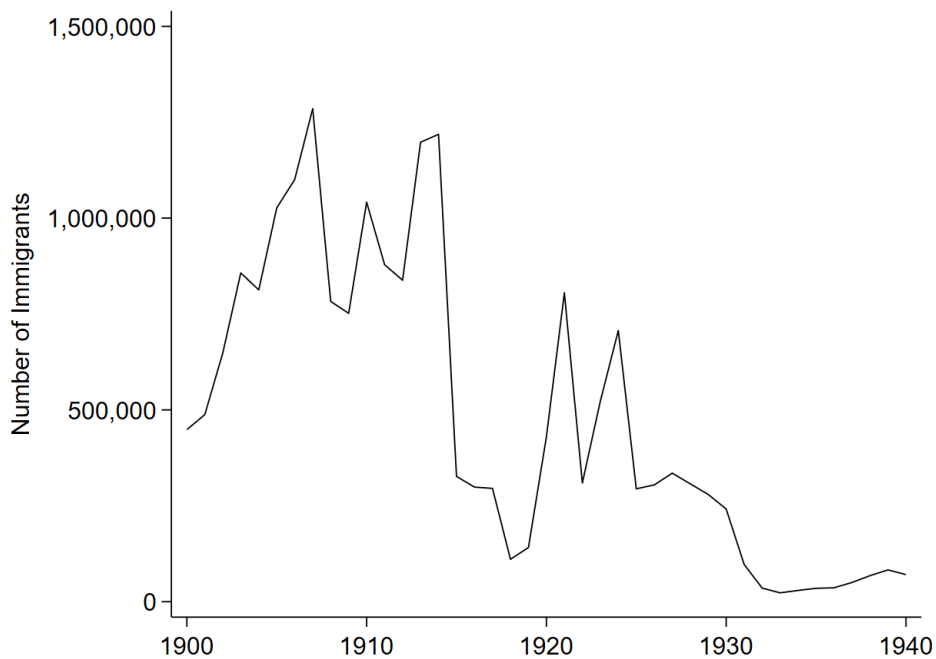
together with banking failures, monetary contraction, and international spillover effects, ushered in the Great Depression (the most severe and prolonged economic crisis in industrial history). This period underwent a rapid decline in the consumption and production of goods and a severe rise in unemployment; by 1932, the unemployment rate had surpassed 24 per cent (Romer, 1993). Businesses and banks shut down, countless people lost their jobs, homes, and savings, and many depended on charity. The direction of change reversed when the Roosevelt administration increased the money supply tremendously after 1933, leading to recovery in the early 1940s (Romer, 1993). Nevertheless, economic conditions remained dire, with unemployment staying above 15 percent until 1940 (Temin, 2000, p. 317).

## Appendix B. Additional Figures



**Figure B1. Manufacturing and Agricultural Shares of Total National Income, 1900-1940.**

*Sources:* U.S. Census Bureau (1949, p.15), for data covering 1900-1938; U.S. Census Bureau (1975, p.293), for data covering 1939-1940.



**Figure B2. Total Immigration, 1900-1940.**

*Source:* U.S. Census Bureau (1960, p.56).

## **Appendix C. Construction of the Market Potential Measure**

This appendix provides a more extensive explanation of Maurer and Rauch’s (2023) market potential measure. Since the Panama Canal eased trade between the US coasts and international ports, the market potential measure includes both effects. Firstly, Maurer and Rauch calculate the change in international market potential induced by the Canal using a dataset on major ports in the 19th century assembled by Pascali (2017). They then use a 20×20 km grid of the world and ArcGIS’s Cost Distance tool to calculate the seaborne least-cost paths and distances between every coastal county in the US and every major international port. They calculate these distances under two scenarios: with the Panama Canal closed and with it open. Subsequently, they calculate the effective distance from each mainland county’s centroid to every international port by summing the distance from the county to the US coast and the distance from the coast to the international port. They estimate the latter using a 1890 cost matrix sourced from Donaldson and Hornbeck (2016), which accounts for canals, railroads, and wagon roads, with multiple cost parameters. Therefore, it gives precise insights into the effective distance between counties during that time.

To combine within-US cost-distances (in US\$) with distances to international harbors (in km), they first convert the domestic distance matrix to kilometers by scaling it to match average great-circle distances between county pairs. The total effective distance between county  $c$  and the international port  $p^{int}$  is then calculated as the shortest distance across all possible routes through domestic ports.

$$d_{cp}^{int} = \min[d_{cp}^{dom} + \alpha d_{p}^{dom,p^{int}}].$$

The first essential parameter is the relative cost of shipping 1 km inland versus 1 km by sea, denoted by  $\alpha$ . If  $\alpha = 1$ , sea and land transport cost the same; if  $\alpha = 1/2$ , sea transport costs half as much. Basing their choice on existing literature, they use  $\alpha = 1/10$  as a plausible yet conservative estimate. Under these assumptions, they calculate the minimum effective distance from each US county to each international port, with and without the Panama Canal. International market potential for county  $c$  is then defined as

$$MP_c^{int} = \sum_{p^{int}} (d_{cp}^{int})^\theta pop_{p^{int}},$$

Where  $d_{cp}^{int}$  is the effective distance between county  $c$  and the international port  $p^{int}$ ,  $pop_{p^{int}}$  denotes the population of the international port  $p^{int}$ , which equals the port's country's population divided by the number of major ports in that country. The fact that incomes per capita, and thus the value as trading partners, differ across countries is captured by weighting country populations by their 1900 GDP per capita relative to the U.S. This comes at the drawback of losing destination countries without GDP data.

The second key parameter, the elasticity of market potential with respect to effective distance, is denoted by  $\theta$ .  $MP_c^{int}$  is measured in terms of population per cost-distance. Regardless of  $\theta$ , this variable increases by 1 when a person with the same GDP per capita as the USA is added at a cost-distance of 1 dollar. Following the standard market potential literature (Harris, 1954), they set  $\theta$  to -1.

The Panama Canal also influenced domestic market potential within the U.S. Therefore, they calculate the domestic market potential analogously to international market potential as

$$MP_c^{dom} = \sum_{c^d} (d_{c^d})^\theta pop_{c^d},$$

where  $c^d$  and  $pop_{c^d}$  represent the potential destination counties in the U.S. and their population, respectively. Importantly, for domestic-market potential, county populations are not weighted by GDP per capita. This data is not available, so the assumption is made that perfect mobility of workers equalizes utility levels across counties in equilibrium.

The total market potential of a county is the sum of domestic and international market potential. This study, however, considers how market potential changed due to the opening of the Canal. Therefore, the main variable of interest is the change in market potential due to the canal:

$$\Delta MP_c^{tot} = \frac{MP_c^{tot} | Canal}{MP_c^{tot} | No Canal}$$

## **Appendix D. Replication Maurer and Rauch (2023): Details**

This appendix explains the replication of Maurer and Rauch's (2023) results. To estimate their long difference model for the logs of population, average manufacturing wages, and agricultural land values, I rebuilt their dataset to estimate the following specification on the change in the outcome variables between 1940 and 1900:

$$\Delta^{1940-1900} y_c = \beta \Delta \ln MP_c + X'_c \cdot \delta + \varepsilon_c$$

Here, the left-hand side represents the change in the outcome variable between 1900 and 1940 in county  $c$ , and  $\Delta \ln MP_c$  represents the change in market potential induced by the opening of the canal in 1920. The coefficient on the change in market potential  $\beta$  indicates whether economic activity in 1940 changed in response to increases in market potential. The set of controls ( $X'_c$ ) are the same as in my baseline model.

Second, their preferred specification is essentially the same as equation (2), but with two differences. First, they cluster at the 5-by-5-degree grid level. Second, for this specification,

they consider a panel for the period 1880–1940 in 20-year intervals. So, their dataset only contains 1940 as post-opening year. In their study, this specification is estimated only for log population. However, I use their dataset to apply the same specification to average manufacturing wages and agricultural land values. This allows for a direct comparison of the panel coefficients with those obtained by Maurer and Rauch from the long-difference model.

## **Appendix E. Alternative Manufacturing Outcomes**

This appendix examines whether canal-induced changes in market potential affected manufacturing outcomes beyond output per capita. Table E1 reports the estimates of employing equation (2) on the logs of the share of the population employed in manufacturing, total manufacturing wages per capita, and manufacturing establishments per capita. I construct these variables using county-level data on manufacturing workers, total manufacturing wages, and total manufacturing establishments from Haines and ICPSR (2010). Table E1 shows that the coefficients of the change in market potential are positive and significant across all three outcomes.

Taken together with the earlier result for manufacturing output per capita, these findings offer a more nuanced picture of how increased market potential shaped the manufacturing sector. Three patterns emerge. First, the positive association between market potential and the share of the population employed in manufacturing - combined with the absence of any detectable response in manufacturing output per capita - indicates that counties experiencing larger gains in market potential induced by the canal absorbed additional manufacturing workers without generating proportional productivity improvements.

Second, column 2 shows that total manufacturing wages per capita – denoted by wages – also increase in regions experiencing higher gains in market potential due to the canal. This suggests that wage gains in counties experiencing increases in market potential due to the canal stemmed from labor-market effects – such as increases in labor demand or greater competition for workers – rather than from growth in manufacturing production.

Third, the number of manufacturing establishments per capita also responds positively to changes in market potential, while manufacturing output per capita does not. This suggests that the average output per establishment decreased in counties experiencing gains in market potential due to the canal.

Overall, the results suggest that gains in market potential from the canal's opening affected labor allocation and the structure of manufacturing activity, but did not translate into detectable gains in output per capita.

**Table E1. Alternative manufacturing outcomes**

VARIABLES	(1) Ln workers	(2) Ln wages	(3) Ln estab
$\Delta \ln MP$	3.208*** (0.626)	3.585*** (0.658)	1.749*** (0.329)
Observations	9,066	8,761	9,242
R-squared	0.849	0.873	0.877
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Clusters	2376	2360	2400

*Note:* All coefficients are elasticities. The dependent variables are: the log of the share of the population employed in manufacturing, the log of total manufacturing wages per capita, and the log of the manufacturing establishments per capita.  $\Delta \ln MP$  refers to the change in market potential induced by the canal. The set of control variables consists of longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors are clustered at the county level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## Appendix F. Alternative Clustering

This appendix addresses the concern that the findings may be sensitive to the choice of clustering approach. It is important to verify that the results are robust across clustering methods because they can substantially affect whether we reject the null hypothesis. Following Maurer and Rauch (2023), who cluster at the 5-by-5-degree grid level to account for geographical correlation, I adopt their approach to assess whether my conclusions still hold. Reassuringly, Tables F2 – F4 show that the results presented in Tables 5-7 are largely robust to these alternative clustering methods. With the exception of Table F4, which shows that, in contrast to clustering at the county level, I find no evidence of spatial variation in the effect on population with 5-by-5-degree grid-level clustering. Table F5 further shows that, consistent with Figures 6 and 7, population growth and manufacturing output per capita display significant pre-trends under 5-by-5-degree clustering, except that the 1900 pre-trend for agricultural land values is statistically insignificant at this wider spatial level.

**Table F2. Baseline analysis with alternative clustering**

VARIABLES	(1) Ln mfg	(2) Ln mfg	(3) Ln mfg	(4) Ln pop	(5) Ln pop	(6) Ln pop	(7) Ln land	(8) Ln land	(9) Ln land
$\Delta \ln MP$	6.829*** (1.276)	0.359 (0.920)	0.608 (1.076)	3.977*** (0.514)	2.467*** (0.742)	2.046*** (0.705)	7.939*** (1.218)	5.344*** (1.373)	4.760*** (0.822)
Observations	8,764	8,764	8,764	12,120	12,120	12,120	12,079	12,079	12,079
R-squared	0.783	0.875	0.877	0.946	0.951	0.956	0.712	0.925	0.945
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes	No	No	Yes
Clusters	47	47	47	47	47	47	47	47	47

*Note:* All coefficients are elasticities. The dependent variables are: the log of manufacturing output per capita, the log of population, and the log of the average agricultural land value per acre.  $\Delta \ln MP$  refers to the log change in market potential induced by the canal. The set of control variables consists of longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors clustered at a 5\*5 degree grid, are shown in parentheses.. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table F3. Heterogeneity analysis with alternative clustering**

VARIABLES	(1) Ln mfg	(2) Ln mfg	(3) Ln pop	(4) Ln pop	(5) Ln land	(6) Ln land
$\Delta \ln MP$	1.013 (1.601)	1.632 (1.727)	-0.927 (1.153)	2.624** (1.208)	3.983*** (1.198)	6.033*** (1.309)
$\Delta \ln MP \times \text{HighSchooling}$	-0.493 (1.674)		4.059*** (1.217)		0.795 (1.066)	
$\Delta \ln MP \times \text{HighLabor}$		-1.142 (1.597)		-0.720 (1.117)		-1.897 (1.197)
Observations	8,764	8,764	12,120	12,120	12,079	12,079
R-squared	0.877	0.877	0.958	0.956	0.946	0.946
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	47	47	47	47	47	47

*Note:* the dependent variables are the log of manufacturing output per capita, the log of population, and the log of average agricultural land value per acre.  $\Delta \ln MP$  measures the average effect of increases in market potential from the opening of the Panama Canal in 1920 for counties with below-median school enrollment (ages 6–20) in 1910 or below-median farm labor expenditures per acre in 1900.  $\Delta \ln MP \times \text{HighSchooling}$  captures the differential effect for counties with above-median 1910 school enrollment.  $\Delta \ln MP \times \text{HighLabor}$  captures the differential effect for counties with above-median 1900 agricultural labor expenditures per acre. Control variables include longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors clustered at a 5\*5 degree grid, are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table F4. Spatial Heterogeneity with Alternative Clustering**

VARIABLES	(1) Ln mfg	(2) Ln pop	(3) Ln land
$\Delta \ln MP$	5.893 (4.679)	1.577 (3.486)	-3.542 (3.650)
$\Delta \ln MP \times \text{New England}$	-7.308** (3.587)	0.615 (2.574)	7.434*** (2.596)
$\Delta \ln MP \times \text{Middle Atlantic}$	-5.636* (3.160)	2.762 (2.488)	6.959** (3.347)
$\Delta \ln MP \times \text{West North Central}$	-23.930*** (5.544)	-8.433 (5.844)	-4.364 (3.335)
$\Delta \ln MP \times \text{South Atlantic}$	-2.379 (3.436)	-2.574 (3.069)	5.934** (2.797)
$\Delta \ln MP \times \text{East South Central}$	-5.643 (4.718)	-2.183 (2.802)	9.395*** (2.223)
$\Delta \ln MP \times \text{West South Central}$	-4.434 (4.779)	-0.504 (3.225)	11.988*** (3.622)
$\Delta \ln MP \times \text{Mountain}$	-5.651 (6.191)	-3.926 (4.020)	3.679 (4.165)
$\Delta \ln MP \times \text{Pacific}$	-6.484 (5.263)	1.175 (3.917)	8.825** (4.230)
Observations	8,764	12,120	12,079
R-squared	0.878	0.957	0.947
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Clusters	47	47	47

*Note:* All coefficients are elasticities. The dependent variables are the log of manufacturing output per capita, the log of population, and the log of average agricultural land values per acre.  $\Delta \ln MP$  measures the average effect of increases in market potential from the opening of the Panama Canal in 1920 for counties in the East North Central. The other coefficients reflect the effect in counties in the other eight divisions relative to East North Central. Control variables include longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors clustered at a 5\*5 degree grid, are shown in parentheses.. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

**Table F5. Event study estimates with alternative clustering**

VARIABLES	(1) Ln mfg	(2) Ln pop	(3) Ln land
$\Delta \ln MP * 1900$	-3.675*** (1.182)	-1.610** (0.755)	1.619 (1.611)
$\Delta \ln MP * 1910$		-0.108 (0.358)	0.173 (1.061)
$\Delta \ln MP * 1930$	-1.407 (0.913)	0.823** (0.400)	4.029*** (1.020)
$\Delta \ln MP * 1940$	-1.042 (1.237)	2.123*** (0.566)	6.673*** (1.337)
Observations	8,764	12,120	12,079
R-squared	0.877	0.956	0.946
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Clusters	47	47	47

*Note:* All coefficients are elasticities. The dependent variables are the log of manufacturing output per capita, the log of population, and the log of average agricultural land value per acre. The coefficients of interest are interactions between  $\Delta \ln MP$  and year dummies, which capture the effect of changes in market potential - driven by the opening of the Panama Canal in 1920—in each specific year. Control variables include longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors are clustered using a 5×5-degree grid. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

## Appendix G. Multicollinearity Tests

This appendix investigates the issue of multicollinearity, which can lead to highly inaccurate estimates and would be problematic if the variable of interest - the change in log market potential - were highly correlated with one of the other independent variables. Multicollinearity refers to the situation in which an approximately linear relationship among independent variables leads to unreliable regression estimates (Verbeek, 2017, p. 45). This can happen when the independent variables are highly correlated, making it difficult for the model to estimate the relationship between each independent variable and the dependent variable, since the independent variables tend to vary together. It is crucial to address this issue by using techniques such as the variance inflation factor (VIF). The VIF reflects the factor by which the variance of the estimator of a single coefficient is inflated compared to the hypothetical situation where the independent variable this coefficient corresponds to is uncorrelated with any other independent variables. This comparison, however, does not guide how the problem should be tackled; therefore, it is not recommended to mechanically exclude variables with VIFs exceeding a

certain threshold (Verbeek, 2017, p. 45). Nevertheless, an inspection of VIFs can still be helpful for gauging whether multicollinearity might be an issue.

Table G6 clearly shows that there is no evidence of multicollinearity, as the VIFs are small (all below 1.30) and thus well below the often-applied threshold of 10, at which the risk of multicollinearity becomes serious. The correlation matrix (Table G7) further supports this, showing that only the log population in 1880 and the longitude are strongly correlated (0.451). The above indicates there is no multicollinearity in the data.

**Table G6. VIF Test**

Variable	VIF	1/VIF
Lnpop1880	1.30	0.77
x	1.29	0.78
y	1.04	0.96
Coastal	1.02	0.98
Mean VIF	1.16	

*Note:* The longitude and latitude are denoted by x and y, respectively. Lnpop1880 refers to the log population in 1880 and coastal indicates whether a county is coastal.

**Table G7. Matrix of correlations**

Variables	(1)	(2)	(3)	(4)	(5)
(1) $\Delta \ln \text{MPC}$	1.000				
(2) x	-0.011	1.000			
(3) y	-0.050	-0.045	1.000		
(4) Lnpop1880	0.081	0.452	0.131	1.000	
(5) Coastal	0.200	0.133	-0.080	0.078	1.000

*Note:*  $\Delta \ln \text{MP}$  refers to the log change in market potential induced by the opening of the canal in 1920. The longitude and latitude are denoted by x and y, respectively. Lnpop1880 refers to the log population in 1880 and coastal indicates whether a county is coastal.

## Appendix H. Long Difference Specification

A common alternative to a TWFE model is a long-difference model. The long-difference specification is also used by Maurer and Rauch (2023) and Galiani et al. (2025), and has the advantage of excluding potentially ambiguous years in which construction already started or the canal was already partly opened. The downside of this approach, however, is that it excludes county- and year-fixed effects, as the dataset is now cross-sectional. As a result, the estimation does not allow for identifying the specific years in which growth occurred. Even so, applying a first difference transformation is a reasonable alternative to a within transformation because it also removes time-invariant effects (Angrist & Pischke, 2009). Rewriting equation (2), and

including a time-invariant dummy, to a first difference specification between 1940 and 1900 gives the following:

$$\Delta^{1940-1900}y_c = \beta_1(\Delta \ln MP_c) + \beta_2(\Delta \ln MP_c \cdot D_c) + D_c \cdot \gamma + X'_c \cdot \delta + \varepsilon_c.$$

The term on the left-hand side now represents the change in the outcome variable between 1940 and 1900 in county  $c$ . While  $\Delta \ln MP_c$  still represents the change in market potential induced by the canal. The coefficient of the change in market potential ( $\beta_1$ ) shows whether economic activity in 1940 changed due to increases in market potential. Moreover,  $D_c$  represents a dummy that is one if either the percentage of 6-21 year olds in school in 1910, or farm labor expenditures per capita in 1900 is above the median, depending on which hypothesis is tested. The coefficient of the interaction term ( $\beta_2$ ) then reflects whether the effect of changes in market potential on economic activity is moderated by human capital or agricultural labor expenditures. The rest of the specification is identical to equation (2).

The results are presented in Table H8. Columns 1, 4, and 7 indicate that, similar to the main specification, the effect of market potential is positive across all economic outcomes. The impact on population growth is considerably larger than in the main specification, whereas the effect on agricultural land values only mildly increased. More notably, Column 2 demonstrates that, unlike in the main specification, the effect of changes in market potential on manufacturing output per capita is both positive and statistically significant. Overall, the above supports the results of the main analysis, which show that the canal led to population growth and higher agricultural land values, while it refutes the finding that there was no significant effect on manufacturing output per capita.

**Table H8. Long run specification**

VARIABLES	(1) Dif. Mfg	(2) Dif. Mfg	(3) Dif. Mfg	(4) Dif. Pop	(5) Dif. Pop	(6) Dif. Pop	(7) Dif. Land	(8) Dif. Land	(9) Dif. Land
$\Delta \ln MP$	1.986** (0.947)	2.660 (1.672)	3.167 (2.076)	3.631*** (0.557)	-1.165 (0.869)	5.355*** (0.973)	5.120*** (0.606)	5.072*** (1.239)	8.363*** (1.469)
$\Delta \ln MP \times$ HighSchooling		-0.885 (1.884)			6.217*** (1.090)			-0.495 (1.326)	
$\Delta \ln MP \times$ HighLabor			-1.217 (2.203)			-2.223** (1.096)			-4.397*** (1.517)
Observations	1,928	1,928	1,928	2,425	2,425	2,425	2,397	2,397	2,397
R-squared	0.026	0.027	0.035	0.183	0.234	0.185	0.378	0.387	0.389
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	1928	1928	1928	2424	2424	2424	2397	2397	2397

*Note:* All coefficients are elasticities. The dependent variables are: the differences between 1940 and 1900 in population, manufacturing output per capita, and agricultural land values.  $\Delta \ln MP$  measures the average effect of increases in market potential from the opening of the Panama Canal in 1920 for counties with below-median school enrollment (ages 6–20) in 1910 or below-median farm labor expenditures per acre in 1900.  $\Delta \ln MP \times$  HighSchooling captures the differential effect for counties with above-median 1910 school enrollment.  $\Delta \ln MP \times$  HighLabor captures the differential effect for counties with above-median 1900 farm labor expenditures per acre. The set of control variables consists of longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors, clustered at the county level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table H8 further shows that the observed interaction effects are also very similar to those presented in Table 6. Columns 2, 5, and 8 show that in line with Table 6, human capital only moderates the effect of changes in market potential on population growth; this effect is slightly stronger than in the main specification. Lastly, columns 3, 6, and 9 show that in line with Table 6, there is an adverse moderating effect for agricultural land values. However, in contrast to Table 6, it also mutes the effect on population growth in the long-difference specification.

It should be noted that caution is warranted when interpreting these results as causal. The estimates reflect only overall growth between 1940 and 1900, so whether this growth occurred before or after the canal's opening remains unclear. Additionally, discussed in Section 6.2, there are pre-trends for each outcome variable

## Appendix I. Alternative Functional Forms

This appendix investigates whether alternative functional forms might better fit the data. Until now, this analysis has only considered a log-log specification, which is standard practice in other papers that use a continuous measure of market potential or access (Donaldson, 2018; Maurer & Rauch, 2023; Galiani et al., 2025). Even so, a log-linear or linear model might better

fit the data, or the relationship between changes in market potential and economic outcomes might be exponential. Table I9 considers these possibilities, showing that the log-linear specification (columns 1-3) yields results similar to the main specification, whereas the linear model (columns 4-6) produces highly insignificant results. In addition, there is little evidence that the relationship between the log change in market potential and the log of the economic outcomes is exponential (columns 7-9).

**Table I9. Alternative functional forms**

VARIABLES	(1) Ln mfg	(2) Ln pop	(3) Ln land	(4) Mfg	(5) Pop	(6) Land	(7) Ln mfg	(8) Ln pop	(9) Ln land
$\Delta \ln MP$							1.951 (1.241)	2.187*** (0.781)	6.419*** (0.768)
$\Delta \ln MPC^2$							-10.805 (7.797)	-1.155 (6.697)	-13.608** (6.763)
$\Delta MP$	0.543 (0.554)	1.917*** (0.344)	4.435*** (0.377)	123.661 (133.775)	254,742.465 (254,779.522)	513.151 (334.750)			
Observations	8,764	12,120	12,079	8,773	12,120	12,079	8,764	12,120	12,079
R-squared	0.877	0.956	0.945	0.816	0.911	0.823	0.877	0.956	0.945
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	2361	2424	2422	2364	2424	2422	2361	2424	2422

*Note:* All coefficients are elasticities. The dependent variables are (the logs of): population, manufacturing output per capita, and agricultural land values.  $\Delta \ln MP$  refers to the log change in market potential induced by the canal.  $\Delta \ln MP^2$  is the square of the log change in market potential.  $\Delta MP$  refers to the ratio of post-opening market potential to pre-opening market potential. The set of control variables consists of longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors, clustered at the county level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## Appendix J. Additional time trends

In Table J10, I further test whether the results are robust to adding state fixed effects (columns 1-3), state-specific linear trends (columns 4-6), and county-level trends (columns 7-9). Under this highly saturated set of controls, all coefficients lose statistical significance, except for the effect on land value when including state fixed effects. Angrist and Pischke (2009) provide a possible explanation. They note that fixed effects and trends act as filters: while they help address omitted-variable bias, they can also remove a considerable part of the identifying variation. In this case, the added state and county trends likely absorb much of the systematic variation used to identify the effects, leaving insufficient residual variation for precise estimation.

**Table J10. State FE, State Trends, and County Trends**

VARIABLES	Including State FE			Including State Trends			Including County Trends		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Ln mfg	Ln pop	Ln land	Ln mfg	Ln pop	Ln land	Ln mfg	Ln pop	Ln land
$\Delta \ln MP$	-1.338 (1.522)	0.284 (1.044)	2.957*** (1.139)	-2.823 (0.000)	0.122 (0.477)	4.412 (0.000)	-2.998 (0.000)	-0.214 (0.281)	5.103 (0.000)
Observations	8,760	12,115	12,079	8,764	12,120	12,079	8,764	12,120	12,079
R-squared	0.888	0.962	0.962	0.885	0.961	0.952	0.957	0.991	0.972
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	2360	2423	2422	2361	2424	2422	2361	2424	2422

*Note:* All coefficients are elasticities. The dependent variables are: the log of manufacturing output per capita, the log of population, and the log of the average agricultural land value per acre.  $\Delta \ln MP$  refers to the change in market potential induced by the canal. The set of control variables consists of longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors, clustered at the county level. columns 1-3 include state fixed effects; columns 4-6 include state-specific linear trends; and columns 7-9 include county-specific linear trends\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## Appendix K. Additional Covariates

This appendix tests whether the baseline results are robust to various potential cofounders, to address the concern that the canal might be correlated with other economic or geographical drivers of economic activity that are not included in the model. First, I control for the time-varying effect of soil quality, measured by potential yields for wheat, maize, and cotton, which I interact with time-fixed effects. The soil quality data are sourced from Maurer and Rauch (2023), who obtained potential yields for wheat, maize, and cotton from the FAO GAEZ database and merged them into counties based on county centroids. Columns 1-3 of Table K11 show that the baseline results are robust to allowing for agricultural productivity to have time-varying effects. Column 1 shows that the effect on manufacturing output per capita changes remarkably little and remains statistically insignificant. Similarly, Columns 2 and 3 show that the effects on population growth and land value only decrease slightly, while remaining statistically significant.

A second worry is that canal-induced changes in market potential correlate with the pre-canal industrial structure, and that the pre-canal industrial structure subsequently influences growth. For instance, if economic activity disproportionately concentrates in manufacturing belts. To address this, I introduce the 1880 share of workers employed in services and manufacturing as additional covariates, which I interact with time fixed effects. The 1880 manufacturing and employment shares are constructed by Michaels et al. (2012), who construct these variables at the county level by aggregating individual-level records from the 1880 U.S.

Population Census. As shown in column 1, the effect on manufacturing output per capita increases slightly but remains statistically insignificant. The effects on population growth and land value remain almost unchanged.

A third concern is that, because market potential change is clustered along the coast, the observed effect of market potential is not due to the canal but to the pre-canal high level of market potential in coastal counties. To test this, I add time-varying effects of the pre-canal log market potential levels. As shown in columns 7-9, all coefficients are very similar to the baseline estimates presented in Table 5.

**Table K11. Additional Covariates**

VARIABLES	Including soil quality			Including pre-canal industrial employment			Including pre-canal market potential		
	(1) Ln mfg	(2) Ln pop	(3) Ln land	(4) Ln mfg	(5) Ln pop	(6) Ln land	(7) Ln mfg	(8) Ln pop	(9) Ln land
$\Delta \ln MP$	0.617 (0.640)	1.295*** (0.373)	4.394*** (0.402)	0.909 (0.606)	2.138*** (0.377)	4.714*** (0.384)	0.558 (0.607)	2.071*** (0.360)	4.890*** (0.366)
Observations	8,764	12,120	12,079	8,764	12,120	12,079	8,764	12,120	12,079
R-squared	0.877	0.959	0.948	0.877	0.959	0.947	0.877	0.956	0.945
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	2361	2424	2422	2361	2424	2422	2361	2424	2422

*Note:* All coefficients are elasticities. The dependent variables are: the log of manufacturing output per capita, the log of population, and the log of the average agricultural land value per acre.  $\Delta \ln MP$  refers to the change in market potential induced by the canal. The set of control variables consists of longitude, latitude, log 1880 population, and an indicator for coastal counties. Standard errors, clustered at the county level. Columns 1-3 additionally include the potential yield for wheat, maize, and cotton; columns 4-6 include the 1880 population share employed in manufacturing and services; and columns 7-9 include the pre-canal level of market potential. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.