Free-Riding Yankees: Canada and the Panama Canal*

Sebastian Galiani[1], Luis F. Jaramillo[2], and Mateo Uribe-Castro[3]

[1]University of Maryland and NBER

[2]University of Maryland

[3]Universidad de los Andes

January 6, 2025

Abstract

We study the impact of the Panama Canal on the development of Canada's manufacturing sector from 1900 to 1939. Using newly digitized county-level data from the Census of Manufactures and a market-access approach, we exploit the plausibly exogenous nature of this historical episode to study how changes in transportation costs influence the process of structural transformation and manufacturing productivity. Our reduced-form estimates show that lowered shipping costs increased manufacturing employment as a share of the population by increasing the number of manufacturing establishments, though not their average size, capital intensity, or skilled labor share. Manufacturing revenues grew 9% more in counties with market access gains at the 75th percentile, compared to counties with 25th percentile gains. Productivity grew by 13% more. These effects persist when we consider general equilibrium effects: the closure of the Canal in 1939 would result in economic losses equivalent to 1.86% of GDP, chiefly as a by-product of the restriction of the country's access to international markets. Altogether, these results suggest that the Canal substantially altered the economic geography of the Western Hemisphere in the first half of the twentieth century.

JEL: F14, F63, N22.

Keywords: Structural Change, Productivity, Market Access, Panama Canal, Canada.

*We are grateful to the editor, Oded Galor, and five anonymous referees for their helpful comments and discussions. We thank John Haltiwanger, Rick Hornbeck, Eunhee Lee, Nuno Limão, Noel Maurer, Chris Meissner, Martin Rotemberg, John Shea, John Wallis and participants at the NBER Summer Institute Seminar - DAE, the Annual Meetings of the EHA, the Annual Meetings of the EEA, Uniandes' Seminario CEDE, RIDGE Historical Development Workshop, the URosario Political Economy Workshop, and the UMD Trade Workshop for their comments and suggestions. Kris Inwood and Ian Keay provided valuable guidance in terms of data sources at an early stage of the paper.

1 Introduction

On March 23, 2021, the Ever Given ran aground and blocked the Suez Canal. Although the blockade was short-lived, the event disrupted supply chains worldwide.[1] Recent research by Brancaccio et al. (2020) indicates that closing the Panama Canal today would reduce global welfare by 3.28%. The episode served as a reminder of the obvious: the world's great man-made waterways have not always been there and play a central role in today's economy. The construction of the Suez and Panama Canals widened the range of economic opportunities everywhere by increasing access to output and input markets. In this paper, we show that reductions in transportation costs led to gains in Canada's manufacturing employment, revenue, and productivity using the opening of the Panama Canal as a natural experiment.

We compute the market access gains obtained by different counties after the Canal began commercial operations in 1920.[2] Counties in Western Canada saw their trade opportunities with Western Europe and the Eastern Seaboard of the United States expand significantly. For the better-connected East, with established connections with those markets, the Canal brought cheaper inputs and new export markets in East Asia for manufactured goods (Table 1). We show that changes in market access resulted in heightened manufacturing activity by 1939. Counties gaining more from the Canal's opening experienced more pronounced sector growth than others who gained less. Both the number of establishments and employment as a share of the population surged. Firms saw increased revenues and use of factor inputs due to reduced transportation costs. Moreover, manufacturing industries in counties that benefited from the opening of the Canal also realized productivity gains.

Canada is well suited to study of the subnational impacts of the Panama Canal's opening for various reasons. First, shipping costs fell because the Canal provided an alternative to the comparatively expensive all-North-American routes. Those routes typically required cumbersome and onerous transshipments and were not, in some cases, available throughout the year due to harsh weather conditions in the winter. Innis (1933) documents shipping rate reductions from Liverpool, England, that ranged from 93% in the case of Vancouver to zero in the case of Regina, located to the north of eastern Montana. Second, since the United States built the Canal without targeting any outcome in Canada,

[1] See The Economist (2021).

[2] The administrative divisions or districts in Canada that are analogous to counties in the United States are referred to by different names in different Canadian provinces. The term "county" is used in this paper in a generic descriptive sense for ease of reference.

the variability in transport costs generated by the episode is plausibly strictly exogenous in a panel data model. This is useful since a key concern when studying the effects of transportation infrastructure on economic activity is whether the potential for economic growth influences the placement of roads, railroads, and canals. The United States built and operated the Canal to achieve strategic objectives that had little to do with its vast yet relatively underpopulated northern neighbor (McCullough, 2001).

Third, the period under study coincides with an important increase in the relevance of manufactured products in the Canadian export basket, providing a setting to study how international trade can drive structural transformation (Table 2).[3] Theories in economic development emphasize input-output linkages (Hirschman, 1958) and economic growth models highlight how sectors with higher productivity growth, which tend to be in manufacturing rather than agriculture, attract workers (Galor and Mountford, 2008). Other models of structural transformation that do not explicitly consider trade, also highlight how supply side mechanisms might generate changes in the economic structure that track the stylized facts about industrialization only when manufacturing's productivity grows faster than agriculture's (Herrendorf et al., 2014). Contemporary observers, such as Mackintosh (1939) and Innis (1933), pointed to increased access to global markets and increased exports of staples, such as lumber and wheat, and related products as the mechanism behind the expansion brought by the Canal.

We show that the Panama Canal impacted Canadian manufacturing through changes in market access. This paper combines newly digitized county-level data from the Canadian Census of Manufactures of 1901, 1911, and 1939; geographic information system (GIS) data for transportation networks in Canada, the United States, and the rest of the world before and after the opening of the Canal; and a market access approach (Redding and Venables, 2004; Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2021). Market access, a concept derived from the international trade literature, measures exposure and proximity to consumers and suppliers by weighting the income-normalized population of destination markets by how costly it is to reach them. We compute each Canadian county's market access in 1910 and 1920 and document considerable variation in market access gains across and within Provinces. Leveraging this variation, we estimate how the reduction in transportation costs affected the county's manufacturing activity.

We observe significant effects of the Canal on various outcomes related to the size of the

[3] We will use *industrialization* and *structural transformation* interchangeably. We refer by structural transformation to the process through which manufacturing becomes more prominent in the economic structure of employment and production, while the role of agriculture decreases (Herrendorf et al., 2014)

manufacturing sector. Shifting a county from the 25th percentile of market access gains in 1920 to the 75th percentile would result in approximately a 6% increase in manufacturing establishments. Population would see a 3.3% increase, while manufacturing employment would rise even more substantially, by 9.1%. Consequently, manufacturing employment as a proportion of the population would also increase by roughly 6%. These gains appear to stem from a broad expansion of economic activity: neither the average establishment size, their share of skilled to unskilled workers, the skill wage premium, nor their capital intensity changed due to the reduction in transportation costs. These results are consistent with a theoretical framework like Eaton and Kortum (2002). Market Access gains generate aggregate output log-linear increases in output but the relative use of capital and labor to output remains constant.

Data from the Manufacturing Census shows that firms' revenues and expenditures in input increased proportionally. Manufacturing revenues grew by 8.9% more in counties with market access gains at the 75th percentile compared to counties with 25th percentile gains. Regarding inputs, labor and materials expenditures grew by 9.7% and 8.3%, respectively. Relative growth in capital expenditures, which are estimated from reported capital stocks using available interest rates data for the period, was 7%, but estimates are less precise. Finally, measured productivity improved by 13.1%. These results are expected since Hornbeck and Rotemberg (2021) show that in an Eaton and Kortum (2002) model, Market Access shocks have log-linear effects on expenditure on inputs, with largest increases in labor and materials.

We decompose productivity gains as coming from either Allocative Efficiency (AE) or from revenue increases obtained while using the same level of inputs given the existing production technology (total factor revenue productivity (TFPR) (Petrin and Levinsohn, 2012). Our evidence suggests that increases in AE– which measures the extent to which inputs are used in locations where they yield their highest productive use due to imperfections in the economy– seem to be behind the changes in measured productivity (Hsieh and Klenow, 2009; Hornbeck and Rotemberg, 2021). Though the theoretical framework we use is agnostic about the origins of such imperfections, these might take, in the context we study, the form of credit market frictions and market power in product markets, issues that have historically plagued the Canadian economy (Innis, 2018).

We do not find that the opening of the Panama Canal had a material impact on those imperfections, but instead that it enabled the expansion of relatively constrained places because, partly at least, of those imperfections.[4] Counties with higher transportation costs and larger manufacturing establishments have lower capital and labor wedges and higher materials wedges. Overall, these results suggest that gains in access to markets drive productivity through better input use, similar to what Hornbeck and Rotemberg (2021) find for American railroads. The decomposition rests on assumptions about the correct measurement of expenditure on inputs and proper estimates of production function elasticities (Haltiwanger, 2016; Haltiwanger et al., 2018; Foster et al., 2016; Collard-Wexler and De Loecker, 2021; Baqaee and Farhi, 2020). Constrained by data availability, we take the results from the decomposition simply as suggestive of the role of transportation cost reduction on productivity dynamics coming from the reallocation of inputs.

We rule out plausible alternative explanations for our results and, thus, attribute the changes in counties' manufacturing activity to the reduction in transportation costs brought by the Panama Canal. First, a concern is that regional convergence might drive our results. The Western provinces of British Columbia and the Territories (Alberta and Saskatchewan) contained 6% of the country's population and 3.4% of the manufacturing employment in 1900. They grew to represent 24.2% of population and 9.4% of employment in 1939. Figure 1 shows that most of the convergence happened before the Canal was built, between 1900 and 1910. After 1910, manufacturing outcomes grew slightly faster in the Western provinces than in Ontario, Quebec, or Manitoba.

More generally, our results could be driven by trends in development outcomes seen before the opening of the Canal that would have given rise to the same outcomes, at least partially, even in the absence of the Canal. However, we show that changes in market access attributable to the Canal cannot predict changes in manufacturing activity between 1900 and 1910. Our estimates for the effect of the Panama Canal-driven reduction in transportation costs on manufacturing outcomes' pre-trends are close to zero in magnitude, with half being even negative, and not statistically significant.

Second, it is plausible that the anticipation of the Canal opening prompted enhancements in local transportation networks. Indeed, areas benefiting from the Canal built a few more kilometers of railroads after 1910. Additionally, there might have been broader improvements in the transportation infrastructure of less connected areas benefiting from the Canal, which we may not observe. We show that both remote and well-connected

[4] As Table A.8 shows, gains in market access fail to predict changes in differences between the value of marginal products and marginal costs, *input wedges*, which measure the extent of imperfections in the economy. This finding also explains why we those wedges as exogenous to market access in the model we expose.

places derived benefits from the Canal. In essence, the change in market access attributed to the Panama Canal is unrelated to the initial level of market access.

Furthermore, we illustrate that the interaction between initial market access and the Canal-related gains fails to account for the outcomes observed in manufacturing activity. Reductions in transportation costs drive the results due to the Canal rather than by the differential behavior of less or more connected areas benefiting from it. Lastly, we confirm that all estimates remain robust even with the inclusion of local railroad construction that occurred *after* the Canal commenced commercial operations.

Third, the Canal helped create opportunities for the lumber and wood products industry, which also benefited from other contemporaneous productivity shocks that relaxed the constraints imposed by railroad networks - like the introduction of trucks and pneumatic tires. We show that the effects of MA gains on manufacturing activity are robust to include a measure of lumber potential interacted with our measure of MA gains. Moreover, the coefficients for the interaction term are small in magnitude and not statistically distinguishable from zero. We interpret these results as suggestive evidence that not only a correlation between lumber potential and MA gains from the Canal does not drive the results, but also that places with high lumber potential do not see differential growth in manufacturing outcomes.

Our estimates contrast outcomes for different counties in terms of the extent of their gain in market access (comparing those that gained relatively *more* market access with those that gained *less* access). A natural, though different, question that follows relates to the *aggregate* impact of the Canal. A naïve exercise could use the input elasticities that follow from our reduced-form exercise to provide an estimation of this effect. However, this approach would be unsatisfactory because our findings speak about *relative* changes, which are not predictive or informative about new input *levels*. These levels are likely to be affected by general equilibrium effects. As Redding (2021) points out, differences-in-differences estimations do not distinguish between the displacement and the creation of economic activity and, therefore, cannot be used for welfare calculations.

A key concern is that the effects we identify are simply the result of a displacement of economic activity toward counties that benefited more from the Canal. Perhaps the Canal triggered shifts in production activities within Canada, but the aggregate benefits derived from the Canal were minimal. To address this concern, we calibrate a benchmark general equilibrium model of economic geography with frictions (Eaton and Kortum, 2002; Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2021) to assess a counterfactual scenario in which the Canal closes permanently in 1939. In calculating the impact of closing the Canal, we allow for the population within North America to be fully mobile. We assume the population adjusts freely, fixing workers' welfare at its pre-closure level. We then use the model's results on counterfactual prices, wages, rents, and population to compute the impact of closing the Canal on manufacturing productivity and property values.

Though not the focus of our analysis, we use the results from our general equilibrium exercise to measure the impact of closing the Canal on agriculture, an important sector of the Canadian economy. We do so to provide a more comprehensive measure of the economy-wide effects of closing the Canal in 1939. We follow Donaldson and Hornbeck (2016) and take the counterfactual prediction on rents from the model to compute the total loss due to decreased land and property values in agriculture.

We find that the closure of the Panama Canal in 1939 would lead to non-negligible population and economic losses for Canada. The total Canadian population would shrink by 2.7%. Our calculations point to economic losses of 1.86% of GDP, reflecting lower productivity in manufacturing (0.27% of GDP) and decreased agricultural land values (1.59% of GDP), which follow from lower land rental rates. Although most counties would have lost population, those closer to the coasts would have experienced the most significant declines. These decreases would be more acute in the western provinces, as British Columbia and Alberta would see their populations fall by 7.8% and 4.2%, respectively. These declines contrast with population gains of 0.07% in Manitoba and 0.2% in Saskatchewan, both central provinces that stood to gain much less – if anything – from the Canal. Our general equilibrium exercise suggests that the Panama Canal facilitated higher aggregate levels of economic activity in Canada.

An important question that follows is the extent to which greater integration with international markets explain the economic expansion caused by the Canal. Given that we do not observe disaggregated county-level export data, we are unable to approach this question using a reduced-form approach. Mackintosh (1939) and Innis (1933) pointed out to greater exports, particularly staples such as lumber and wheat and related products, as the mechanism behind the expansion. As Table 2 shows, the period we study coincides with the increased relevance of manufactured goods in the Canadian export basket. The Canal shock, which lowered trade costs for products across the board, might have differential impacts for manufacturing and induce structural change (Hirschman, 1958; Herrendorf et al., 2014; Galor and Mountford, 2008).

To shed light on the origin of the gains brought about by the Canal, we carry out a second quantitative exercise in which we close the Canal for shipments originating in or destined for countries other than Canada. In other words, we allow the use of the Canal only for domestic trade within Canada. Along the lines of Fajbelgaum and Redding (2014) study of the Argentinian case, the exercise sheds light on the relative importance of domestic and international market access gains in the shock we study.

Using the 1939 factual equilibrium as a base, total population and GDP both fall, by 2.4% and 1.74%, respectively. These losses reflect lower productivity in manufacturing (0.25% of GDP) and land values in agriculture (1.49% of GDP). Given the consequences of completely closing the Canal described in the previous paragraph, our results suggest that the bulk of the gains brought about by the Canal for Canadian counties were derived from their greater exposure to markets outside Canada. Our findings are consistent with the accounts of contemporary observers and the idea that international trade might be a driver of structural transformation, even in the presence of increases in exports of primary products. Though we are unable to disentangle the specific role of different margins of international trade, such as import competition, aggregate data for the interwar period we discuss below suggests that most changes in the composition of Canadian international trade took place at the export margin: exports of manufactured goods gained relevance in the Canadian export basket between 1910 and 1939. In contrast, the composition of the import basket remained relatively unchanged.

Contributions to the Literature

This paper contributes to the literature on the effect of trade on structural transformation and industrialization. Galor and Mountford (2008) argue that in developed countries, trade increased demand for skill-intensive goods, promoting human capital accumulation and higher productivity growth in manufacturing. In two-sector growth models without trade, higher productivity growth in manufacturing than agriculture leads to structural transformation (Herrendorf et al., 2014). For Canada, we show that trade brought about gains in productivity, which may explain part of the shift to manufacturing, through increased use of inputs in places where their marginal revenue was higher than their marginal cost. However, we do not find that lower trade costs shift production towards more skill-intensive employment. Development models like Hirschman (1958) and empirical evidence (e.g Droller and Fiszbein (2021)) emphasize that trade may help the industrial sector through primary exports' linkages. Given historical data constraints, we are unable to precisely pin-down the relevance of these mechanisms in our empirical analysis. For similar reasons, we can not distinguish the importance of margins such as imports relative to exports. Our findings, however, are consistent with the rise to prominence of manufactured products in the Canadian export basket in the interwar period, contemporary remarks that highlight increased exports of wheat and lumber as the key drivers of the economic expansion brought by the Canal (Innis, 1933; Mackintosh, 1939), the relatively high level of input-output linkages of those products, and the market size shock brought by the Canal as a productivity shock that disproportionately favors manufacturing.

This paper is most closely related to recent works on the impact of the Panama Canal on structural transformation. Belmar (2023) and Maurer and Rauch (2020) document the Canal's influence on manufacturing employment growth in Colombia and the United States, respectively. Within both countries, places that benefited more from the Canal increased their manufacturing employment. We document a similar finding for Canada, a developed country (relative to Colombia) specializing in agriculture (relative to the US).

Our work improves upon these papers in two ways. First, we document changes in manufacturing outcomes beyond employment rates. Reductions in transportation costs led to increased manufacturing revenues, input use, and productivity. We report an increase in establishment entry while establishment size, and capital and skill intensities are unaffected by the Canal. Second, local-level effects might simply capture the displacement of economic activity. We use our estimates of transportation costs with and without the Panama Canal to calculate positive general equilibrium effects on manufacturing activity and agriculture. Here we also show that these gains are driven by access to international markets with access to domestic markets being less relevant. Different from Maurer and Rauch (2020), our general equilibrium model introduces market distortions (Hornbeck and Rotemberg, 2021), which our productivity decomposition suggests might be important to determine manufacturing growth.

Finally, Belmar (2023) highlights the role of county-level comparative advantage on the Canal's effect on manufacturing employment in developing economies. In Colombia, the positive impact of transportation cost reductions on manufacturing employment was lower for places with a comparative advantage in coffee production. For the Canadian context, we do not find evidence that places specializing in the lumber industry before the Canal had different manufacturing employment and output growth than others after 1920. Since places that can specialize in staple goods and improve their Market Access may face import competition for their manufactures, it is striking that we do not find a differential effect from the Panama Canal. This might be due to differences in the production function of staple goods. Unlike coffee, which is primarily a final consumption good, lumber or wheat–the Canadian staples, which are used in other industrial processes as intermediate inputs– have more linkages to the rest of the economy, which could explain why we do not see differential effects by lumber intensity (Droller and Fiszbein, 2021).[5]

This paper contributes to a body of work that assesses the impact of transportation costs and infrastructure on economic activity (Asturias et al., 2019; Asher and Novosad, 2020; Banerjee et al., 2020; Donaldson, 2018; Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2021; Cao and Chen, 2022; Fajgelbaum and Redding, 2014; Jacks and Novy, 2018; Krugman, 1991; Limão and Venables, 2001; Liu and Meissner, 2015; Martinez-Galarraga et al., 2015; Missiaia, 2016; Redding and Venables, 2004; Pascali, 2017; Sotelo, 2020; Zárate, 2020). In contrast to the existing literature, the intervention we exploit is located outside the domestic infrastructure network of the country we study. This setting is highly advantageous for several reasons. First, this feature alleviates reverse causality or targeting concerns, which are prevalent in this work. Most studies focus on the gradual development of a transportation network over time, which might be subject to strategic local considerations about the location of investment in infrastructure. Second, our shock affected only transportation costs and did not lower migration barriers. Most studies bundle those two together, leading to difficulties in understanding the drivers behind changes in market access. Third, the Canal affected domestic and international trade costs, enabling us to analyze how they differentially shape economic outcomes. We document a large shock that transformed the economic geography of the Western Hemisphere. Our setting enables us to provide clean, credible evidence of the effects of transportation infrastructure on various endogenous economic outcomes.

A large literature on reallocation documents imperfections – such as regulations or mark-ups – in the economy (Hsieh and Klenow, 2009; Petrin and Levinsohn, 2012; Restuc-

[5]Though Input - Output Tables are only available for Canada in 1961 (StaisticssCanada, 1987) and for Colombia in 1970 (DANE, 1976), figures for those years show that wheat and lumber in Canada had more forward linkages than coffee in Colombia: 48.5% of the Gross Value of Products of "Wheat, flour, meal and others" was used as an input by other sectors. Similarly, 47.5% of the gross value of products of "Lumber and timber" was used as input by other sectors. In contrast, Colombian data for 1970 shows that only 6.3% of the coffee sector's output was used as an input by other sectors. Similarly, backward linkages towards the nonprimary sector were much stronger for Canadian staples relative to the Colombian staple. The coffee sector's Value Added was 48.3% of Gross Output, with 83% of the intermediate input expenditures accruing to the agricultural sector. In contrast, the Forestry sector's Value Added was 52.5% of Gross Output, with only 23.8% of the intermediate input expenditure accruing to agricultural goods (including forestry products themselves).

cia and Rogerson, 2008). These imperfections prevent the equalization of marginal revenue products, dampen productivity, and induce the misallocation of production factors. Recent work assesses the impact of transportation infrastructure on reallocation dynamics in India, the United States, and Mexico (Asturias et al., 2019; Asher and Novosad, 2020; Hornbeck and Rotemberg, 2021; Zárate, 2020). We use the exogenous Canal shock to answer the question as to whether improved transportation technologies can increase productivity by enabling the reallocation of production factors and the role that establishment entry might play in such a process.

We build on previous work on the impact of the Panama Canal on the Western Hemisphere. This literature emphasizes that the Canal was a source of pecuniary externalities, social savings and population changes for the United States (Rockwell, 1971; Maurer and Yu, 2008) and a potential determinant of land and wage values in Canada (Umaña-Dajud, 2017).

Finally, this study contributes to the literature on Canada's economic history. Although contemporary observers were quick to point to the Canal as a significant disruptor of economic life in Canada (Innis, 1933, 2018; Mackintosh, 1939), ours is the first project to systematically assess the impact of the improvement in transportation technology embodied by the Canal on manufacturing outcomes in that country. It also relates to work that emphasizes integration into the global economy as a force that helped to shape Canada's industrialization patterns (Alexander and Keay, 2019; Jaworski and Keay, 2021).[6] Canadian historiography typically overlooked this factor and emphasized scale economies, industrial policy, and domestic market expansion to explain the development of this sector of the economy (Keay, 2007).

The rest of this study is structured as follows: Section 2 provides historical context; Section 3 presents a detailed discussion of the construction of the data; Section 4 describes the empirical analysis that was undertaken; Section 4.2 covers the implementation of robustness checks; Section 5 provides an assessment of the costs of closing the Canal in 1939; and Section 6 concludes.

[6]Innis (1931), MacIntosh (1939), and Lawrence (1957) emphasize access to export markets as the central driver of the transformations triggered by the Panama Canal. We cannot systematically test for and disentangle a role for greater import competition. Historical data suggests, however, that export flows from Vancouver were significantly higher than respective import flows. For example, Innis points out that Vancouver exported 8.1 tons per imported ton from Europe by 1927. Table 1, which includes data for 1927, shows that wheat and lumber made most of the Eastbound, whereas iron and steel, glass and glassware, paper, and sulfur did so for the Westbound traffic.

2 Historical Background: Canada and the Panama Canal

Transportation has been a central factor in shaping Canada's development and its relationship with the rest of the world (Innis, 2018). With the fall in haulage costs that followed the introduction of the steamship and the expansion of the railroads in the nineteenth century, Canada's main staple products transitioned from cod and furs to wheat and lumber. These were primarily produced in the country's eastern and maritime provinces.

However, even after the completion of the transcontinental Canadian Pacific Railway in 1886, some parts of the country's effective access to key domestic and international markets remained limited (Innis, 2018). The vast distances involved, the railroad's market power, and multiple transshipments remained obstacles to cheap haulage. Before the opening of the Panama Canal, a typical producer located west of Winnipeg eager to sell its products in Toronto or New York had four potential shipping options: (i) the Canadian transportation network, which entailed a combined journey by railroad and steamship spanning over two thousand miles over a route that was partly impassable during the winter; (ii) the United States railroad network up to an American port that did not freeze in the winter (as did those on the St. Lawrence River); (iii) the Panama and Tehuantepec railroads; and (iv) Cape Horn, a route that was over 16,000 miles in length linking the cities of Vancouver and New York.

The costliness of these routes hindered economic agents from trading with distant markets. Lawrence (1957) points out that the shipment of lumber via the Canadian transportation network was so cumbersome that, in the case of some products, it was slightly cheaper to ship them from Vancouver to the Atlantic coast of North America via Cape Horn. Even then, these routes were too costly for most products. For Huebner (1915), the minimal capacity of the Isthmian railroads and high transshipment costs voided most of the savings obtained by using shorter routes than the Cape Horn passage. Given these conditions, Canada was set to tangibly benefit from opening a waterway in Central America.

The idea of building a canal in Central America dates back to at least the sixteenth century. Advisors to the King of Castille pointed out how beneficial such a waterway would be for the Spanish crown's profitable trade with Asia. After surmounting what was traditionally regarded as the impossible task of building the Suez Canal, French developers undertook the first serious effort to construct a canal through the Isthmus of Panama. Construction began in 1881, but that attempt eventually failed due to a series of conceptual flaws and challenging conditions on the ground. The *Compagnie Universelle du Canal Interoceánique de Panama* filed for bankruptcy in 1889. The works remained essentially abandoned until the next century when the United States became interested in the project.[7]

In his 1901 State of the Union address, President Theodore Roosevelt asserted that no single material work yet to be undertaken was of such consequence to the American people as an Isthmian canal. One of the strategic considerations behind this statement was the expectation that such a project would eliminate the need for the United States to establish two distinct naval fleets to defend its Pacific and Atlantic shores. As part of an explicit policy to limit the influence of European powers in the Western Hemisphere, this was a powerful reason for renewed interest in constructing a canal. After the Colombian Congress rejected a treaty that would allow the United States to build and manage such a canal in 1903, the Roosevelt Administration supported a revolution that ended in the secession of Panama from Colombia. Panama allowed for the construction of a canal under terms similar to those rejected by Bogotá. At a cost of approximately $10 billion in current terms, the Panama Canal opened for traffic on August 15, 1914, just a few weeks after the Great War erupted in Europe.

Even after it opened, however, the Panama Canal remained underutilized and effectively closed to commercial passage until 1920 (Maurer and Yu, 2010). First, the outbreak of war in Europe depressed maritime shipping markets and increased rates to prohibitive levels that would not recede until after the end of the war. Second, several landslides closed the waterway for several months in 1915, 1916, 1917, and 1920. Third, widespread labor strikes prevented its full operation in 1916 and 1917. As a result, the Canal did not effectively open for commercial traffic until 1920. As per the Panama Canal Act, tolls were set at levels designed only to cover operation and maintenance costs, not to maximize the revenue of the company running the waterway.

Contemporary observers pointed to the Canal as a potential factor behind the expansion of the economies of western North America during the 1920s. Mackintosh (1939) claims that: "It was not until the opening of the Panama Canal that British Columbia experienced the rapid development which comes from increasing access to world markets and a great extension of the hinterland tributary to its metropolitan centre." In a similar vein, in an article entitled 'The Boom in California,' The Economist (1924) asserted that: "Undoubtedly the chief factor in the expansion has been the opening of the Panama Canal, which has given the varied products of Southern California cheapened access to outside markets."

[7] Another company, the *Compagnie Nouvelle du Canal de Panama* was established in 1894 but failed to make material progress towards completion.

Figure 1 introduces perspective to these remarks. First, while the Canal was likely a source of growth, the transformation of the country's western provinces started between 1900 and 1910. Alberta and Saskatchewan began the century with a combined population of just over 150,000 and experienced a rapid economic expansion over the decade. Their growth in population and manufacturing employment was high relative to the Eastern provinces of Ontario and Quebec (Figures 2a and 2b) and to the Western US states (Figures 1c and 1d). British Columbia (178,000 inhabitants in 1900) grew faster than Ontario or Quebec and roughly at the same pace as Manitoba or the US West Coast. Second, the Western provinces' manufacturing activity trends after 1910 seem slightly steeper than the Eastern provinces, but not unequivocally. For instance, in the number of establishments (Figure 1e) and total manufacturing revenue (Figure 1f). The effects on economic activity due to the reduction in transportation costs unleashed by the Canal are not as apparent as contemporary observers asserted.

Figure 2 and Table 2 show that the structure of Canada's export basket changed considerably over the interwar years, which overlaps with our study period. Manufactured products rose to prominence, whereas raw materials decreased their relevance. This development contrasts with the import basket, which remained relatively unchanged over the same period. These numbers suggest that most changes in Canada's international trade structure took place at the export margin, which is consistent with Innis (1933) and Mackintosh (1939) remarks on the relevance of the Canal to increase Canadian exports overall.

We systematically test these contemporary historical observations. Our purpose is to investigate the Panama Canal's contribution to Canada's economic development. We do so while carefully ruling out other contemporaneous and potentially correlated changes that might induce spurious findings. This task requires estimating the local-level reductions in transportation costs that came from the possibility of shipping goods through Panama and showing how they are related to the changes in local economic activity only after 1910 but not before.

3 Data

This study combines newly digitized census data, geographic information system (GIS) data, and key parameter estimates from the recent trade and economic history literature. Our main sample consists of 217 counties in Canada that reported manufacturing activity in 1900, 1910, and 1939. The sample encompasses the universe of counties throughout the study except those in the Yukon and the Northwest Territories. This sample of counties includes 99.6% of the total population in 1911 and 99.9% of the population over subsequent census years (1921, 1931, and 1941). We first measure how the effective opening of the Panama Canal in 1920 changed each county's exposure to other markets in North America and the rest of the world. Then, we estimate how those changes in market access led to changes in economic structure and manufacturing activity. We focus first on employment, establishment size, capital intensity, and skill-premium. We then turn to manufacturing output, input use, and productivity. Finally, we perform a simple productivity decomposition to explore where the changes in productivity might originate.

In this section, we describe how we construct market access and how we measure manufacturing outcomes over time, highlighting how historical data restrictions guide our empirical choices.

3.1 Market Access

To assess the impact of the Panama Canal on manufacturing outcomes, we focus on changes in counties' Market Access. The concept of market access, as developed in the economic geography literature (e.g., Redding and Venables (2004)), captures the effective exposure of agents in any given location to suppliers and consumers elsewhere. This approach is based upon constructing a transportation network that connects every possible origin and destination through the least cost path among all possible paths using different modes of transportation. Intuitively, a county's Market Access equals the total size of potential input and output markets, weighed by the cost of getting to those markets. Each potential market size depends on its population and income level. Focusing on Market Access allows us to highlight that counties benefit from improvements in the transportation network even when those improvements are not close by because those improvements reduce the cost of reaching further out markets.

We compute the gains in Market Access brought about by the Panama Canal as the difference between (log) Market Access in 1920 with the factual transportation network, including the Canal, and its analog, assuming that the Canal is closed. We hold other features of the transportation network constant to focus on gains related to the opportunity to ship goods through the Canal. This approach also allows us to make necessary assumptions and generalizations to estimate transportation costs. This section explains the assumptions required to estimate transportation costs and describes how we estimate

Market Access gains.

We calculate county c's market access as:

$$MA_c = \sum_{d \neq c} \tau_{cd}^{-\theta} L_d Y_d \tag{1}$$

Where τ_{cd} is the iceberg trade cost between county c and destination d, L_d is the destination's population, Y_d is the GDP per capita of the country where d is located relative to Canada's GDP per capita, and θ is the elasticity of trade to trade costs. The trade cost τ takes an iceberg cost form and is computed as:

$$\tau_{cd} = 1 + \frac{t_{cd}}{\bar{P}} \tag{2}$$

Where t_{cd} is the cost of moving one ton from county c to destination d and $\bar{P}$ is the average transportation cost per ton. The estimation of each county's market access requires, therefore, a definition of the possible set of destinations D and estimations of t_{cd}, $\bar{P}$, and θ.

3.1.1 Destinations

The set of destinations D to which we assume each Canadian county has access comprises all counties in Canada, all counties in the United States, and selected countries in the rest of the world. For the last group, we are constrained to use a subset of 63 countries and territories for which population and GDP per capita data are available from the Maddison project from 1910 to 1920. In countries with ocean access, their population is assigned to their most historically relevant port. The population of landlocked countries is assigned to the closest international port, as measured by the distance from its borders. In total, we use 56 ports as destinations to compute transportation costs. Therefore, each Canadian county has 3,069 destinations: 216 of its peers in Canada, 2,797 in the United States, and 56 ports in the rest of the world.[8] Altogether, the destinations in our sample account for approximately 86% of the global population and 93% of the global GDP in 1920.[9]

[8]The transport network data for the US uses 1890 county boundaries. To use population figures for 1910 and 1920, we match counties in 1910 and 1920 to 1890, assuming that the spatial distribution of the population is homogeneous (Eckert et al., 2020). For Canada, we use 1941 boundaries from that year's Census of Population, which were the same as those used for the 1939 Census of Manufactures.

[9]We detail the list of countries we cover and the corresponding port to which they are assigned in the Online Appendix. In practice, the only systematically excluded area is Sub-Saharan Africa. Canadian trade with Africa, some of which is covered in our sample, represented between 0.2 and 0.8% of imports and between 1 and 2% of exports over the 1930s.

To estimate transportation costs between each origin (Canadian counties) and destination (set D), t_{cd}, we need two elements. The first is a transportation network that captures the relevant options open to Canadian producers shipping goods in the first decades of the twentieth century. Those options involved both the United States and Canada's systems of railroads, canals and waterways, wagon routes and ocean liners – and, eventually, the Panama Canal. The second is an assumption about the rates charged for each mode in each country and transshipment costs across different modes whenever possible. As we note below, we build such a network and estimate costs for each mode in each country using rates for wheat – a key Canadian staple with relatively simple transportation features – as inputs for computing the transportation cost estimates (t_{cd}).

3.1.2 Transportation Network

For Canada, our transportation network uses the University of Toronto's GEORIA project shapefiles, which provide georeferenced information on Canadian railroads and stations from the nineteenth century onward, including data on the year each line opened. We built most of the remaining components by using historical sources and then manually georeferencing them: we identified canals, waterways and harbors and drew them based upon information from traffic and other available facilities found in the *Summary of Canal Statistics* (Dominion Bureau of Statistics, 1940) and *Directory of Ports and Harbours of Canada* (Department of Marine and Fisheries, 1922) . For the United States, we rely on the transportation network provided by Donaldson and Hornbeck (2016), based on previous work by Atack et al. (2010). It includes railroads, canals, waterways, and a linear ocean route through Cape Horn.

As Donaldson and Hornbeck (2016) do for the continental United States, each county's geographical centroid is connected to railroad stations and harbors in a 200km radius with straight-line wagon routes. Each county centroid can also reach other county centroids in a 400km radius – both in Canada and the United States – using straight-line wagon routes.[10] This assumption likely underestimates wagon route costs in rugged counties and overestimates them for flatter counties. This assumption might be a concern if we were interested in local-level changes in the transportation network. For instance, if a new railroad station built in a rugged county looks reachable by wagon route when assuming straight lines routes but, in reality, the county's topography makes it prohibitively costly. However,

[10] Straight-line wagon connections passing over the Great Lakes and the St. Lawrence River basin were not included.

we abstract from the more precise modeling of local level transportation costs and follow Donaldson and Hornbeck (2016) since we are interested in the change in Market Access only when the Panama Canal route becomes available, with all other features remaining constant.

To do so, we supplement the domestic networks described above with the Panama Canal as an alternative to transporting goods across the Americas. We use The Panama Canal Records, the official United States government gazette, to identify American and Canadian ports listed as origins or destinations of shipments passing through the Panama Canal from 1914 to 1939. We find 34 ports, 30 in the United States and 4 in Canada. This information helps construct shipping routes between ports in the Pacific (12, 2 of which are in Canada) and the Atlantic (22, 2 of which are in Canada) passing through the Canal.[11]

Finally, we allow for the shipment of goods to destinations outside North America by creating ocean routes between the 34 North American ports and the 56 international ports around the globe that we had previously identified. We first do so by using information on actual distances between ports and key global chokepoints from the United States Navy (1911, 1917, 1920, 1931, 1943) while allowing for either direct routes between ports – whenever possible – or routes passing through global chokepoints *other than* the Panama Canal.[12] We then allow the Canal to be used and incorporate routes that use it as part of the transportation network.

3.1.3 Rates

We construct estimates for rates for each mode of transportation used in the network separately for the United States and Canada. We use wheat as the product of reference to compute our estimates because it is a staple product of central importance for Canada's economy and has relatively simple transportation requirements; moreover, it is a product for which a wealth of historical data exists in both countries. We discuss our sources and compare our estimates to the previous ones below. The specifics of each calculation are discussed in the Online Appendix B. All rates reported in this section are in 1910 Canadian dollars, as the American and Canadian currencies traded at par in 1910.

For Canadian railroads, we use historical data on rates and distances provided by the Canadian Railway Commission (1939) and compute an average rate of 0.514 cents per

[11] See the Online Appendix B.

[12] These chokepoints are the Suez Canal, Cape Horn, Cape of Good Hope, Singapore, the Strait of Gibraltar and Bishop Rock.

ton-mile.[13] For US railroads, we use data on average rates provided by the Interstate Commerce Commission (1913). We retrieve an average rate of 0.626 cents per ton-mile for railways. This estimate is similar to the rate used by Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2021). Notice that the estimated cost of railroad transportation in Canada is 17.9% lower in Canada than in the US. This phenomenon could be due to a combination of stricter enforcement of anti-monopoly regulation, government cash or capital cost subsidies to railroad companies, or even public ownership of railroad lines (in 1939, the Dominion of Canada owned 53% of all railroad mileage) (Royal Commission on Dominion-Provincial Relations, 1939).

For non-oceanic waterway transportation, we use information on rates, distances, storage costs, and insurance premiums provided by the Saskatchewan Grain Commission (1914) and the House of Commons of Canada (1907). We compute an average rate of 0.238 cents per ton-mile. For the US, we retrieve an average rate of 0.260 cents per ton-mile from the Interstate Commerce Commission (1913). These rates include additional storage charges whenever the waterways were frozen, insurance, and fees that were either not applicable (winter storage) or already included in the railroad rates.

We use data from the United States Department of Agriculture (1906) for wagon transportation. Without specific figures for Canada, we compute an average rate of 25.657 cents per ton-mile for border US states, which is a reasonable measure of wagon transportation costs for Canada according to the Saskatchewan Grain Commission. The average rate for all US states is 22.639 cents per ton-mile. We force the payment for any shipment that switches modes of a flat rate of 50 cents per ton, as per information provided by the Saskatchewan Grain Commission.

For maritime routes, we rely on the weekly transportation rates provided by the Georgian Bay Canal Commission (1916) for shipments from Liverpool, United Kingdom, to New York City, Odessa, Karachi, and Buenos Aires. We estimate an average rate of 0.052 cents per ton-mile, including insurance charges, as provided by the Saskatchewan Grain Commission (1914). According to historical records, shippers using routes through the Suez Canal and the Panama Canal paid, respectively, a flat toll fee of $1.48 and $0.95 per ton (The Panama Canal Company, 1971). All rates are in 1910 Canadian dollars.

The rates estimated above are similar for both countries, consistent with accounts by Innis (2018) and the House of Commons (1907) that emphasize the co-determination of transportation rates in both countries. Except for the rates for waterways, all of them are

[13] Standard deviation of rates is 0.04 cents per ton-mile. Rates vary by distance: a trip 100 miles longer is associated with a 1.16 cents per ton-mile cheaper rate.

close to those given in Fogel (1964) for the late nineteenth century in the United States: railroads (0.630 cents per ton-mile), wagons (23.1 cents per ton-mile) and transshipments (50 cents per ton). Although the rate for waterways is similar to Fogel's, our final all-inclusive estimate for non-oceanic waterway transportation (0.26 cents per ton-mile) is lower than his (0.49 cents per ton-mile) because we do not attribute the capital costs of storing wheat inventories during the winter to the transportation mode itself. Doing so might be justified for a social savings exercise, but it is more challenging to justify in this paper. Our estimate of maritime transportation (0.052 cents per ton-mile) is within the boundaries of the literature. It is higher than Maurer and Yu (2008) estimate of 0.036 cents per ton-mile of variable costs and slightly lower than Donaldson's and Hornbeck's implied rate of 0.055 cents per ton-mile for a Cape Horn route.

3.1.4 Computation of Iceberg Trade Costs

To calculate the iceberg trade costs in equation 2, we need estimates of t_{cd} and $\bar{P}$. For t_{cd}, the cost of transporting one ton of cargo from county c to any possible destination d, we find the lowest cost path for each origin-destination pair using the transportation network and freight rates described above. We do so by implementing Dijkstra's algorithm (Dijkstra, 1959), which identifies the shortest path between a given node (origin) and all other nodes in a graph (destinations). In our context, the path length is determined by the economic cost of using the different alternatives available in the transportation network. The algorithm is not instructed to impose any penalty for using or switching between the American and Canadian networks besides the transshipment costs, which apply to both networks.

Following equation 1, these "raw dollar" transportation costs t_{cd} are divided by $\bar{P}$ in order to convert them to iceberg trade costs for use in the computation of each county's market access.

We compute $\bar{P}$ following equation 3. $\bar{P}_c$, computed following equation 4, is county c's average transportation rate and μ_c is county c's share of manufacturing revenues in 1910. We combine those two to weigh transportation costs by the relative economic relevance of each county. In equation 4, we weigh origin-destination transportation costs t_{cd} by $distance_{cd}^{-1}$, the inverse of the straight-line distance between county c and any possible destination d.[14]

[14] The weights used for the computation of the average are the inverse of distance between county c and destination d, normalized by the sum of all weights, so the total weights add up to 1. Exclusively for distance indices, we use the azimuthal equidistant projection of the world centered around the geographic center of

$$\bar{P} = \sum_{c} \bar{P}_c \mu_c \tag{3}$$

$$\bar{P}_c = \sum_{c \neq d} t_{cd} distance_{cd}^{-1} \tag{4}$$

We estimate $\bar{P}$ to be equal to 9.4. Although we follow the literature in using $\bar{P}$ in our baseline estimates, we implement a robustness check in which we use extreme values of $\bar{P}$ to compute market access.

3.1.5 Market Access Calculation

As equations 1 and 2 show, the computation of market access further requires an assumption about trade-to-trade-costs elasticity, θ. We use $\theta = 5$, which is Head's and Mayer's (2014) preferred estimate in the literature. Section 4.2 covers the implementation of several robustness checks, assuming several alternative values of θ ranging from 1 to 7.5. Finally, the computation requires population and GDP per capita statistics for each possible destination. For Canada, we use the 1911 and 1921 population statistics and the corresponding GDP per capita figures. For the United States and other countries, we use the 1910 and 1920 population and GDP per capita figures. We normalize destinations' GDP per capita by simply dividing them by Canada's GDP per capita. The underlying assumption is that access to markets is influenced by transportation costs, relative income, and population. However, there might be other factors that influence Market Access, like transaction costs and policies that change at countries' borders (Liu and Meissner, 2015). Our measure of Market Access gains due to the Panama Canal, to which we turn next, helps us abstract from those issues.

Market Access Gains due to the Panama Canal

We define the change in market access brought about by the opening of the Panama Canal as the difference between actual market access for each Canadian county in 1920 and market access in 1920 in the absence of the Canal:

Canada.

$$\Delta \ln(MA_{1920}) = \ln(MA_{c,1920}|\textit{Canal}) - \ln(MA_{c,1920}|\textit{No Canal}) \tag{5}$$

Where $\ln(MA_{c,1920}|\textit{Canal})$ is the natural logarithm of county c's Market Access in 1920 with the transportation network that includes the Panama Canal. $\ln(MA_{c,1920}|\textit{No Canal})$ is its equivalent without the Panama Canal.[15] Both of our market access measures (Canal, No Canal) use 1920 population figures and infrastructure networks. Note that the only change considered here is the possibility of shipping goods through the Canal. Figure 3 shows that the introduction of the Canal into the transportation network does *not* change the minimum-cost path between Vancouver, British Columbia, and Calgary, Alberta, but *does* change the minimum-cost path between Vancouver and New York City. Our measure indicates that *all* counties observed increased market access in 1920 following the opening of the Canal.

Figure 4 plots a map of Canada for this variable, while Table 3 describes the variation in Market Access gains at the national and provincial levels. These figures indicate that gains in market access due to the Canal: (i) are greater in the western part of the continent. This pattern arises because counties in the West gained access to the relatively wealthy and populated destinations on the eastern seaboard of North America and Western Europe. Counties in the East gained access to the sparsely populated counties along the west coast of North America and the heavily populated but relatively poorer Asian countries. However, Market Access gains (ii) exhibit a substantial degree of variation that goes beyond the comparison between the West and East. Notably, in Ontario, Market Access increased by a factor of six in the most advantaged county compared to the least favored one. This gap is around 3.3 times in Alberta and British Columbia. (iii) Market Access gains are greater in places close to the coast, the Great Lakes, or the St. Lawrence River basin.

Another important fact about our measure of Market Access gains is that it is not correlated with the initial level of Market Access (Table A.1. In other words, when we estimate the effects of the transportation cost reduction brought about by the Panama Canal, we are not simply comparing remote places with places that were initially well-placed in the transportation network. Our measure is idiosyncratic to the Panama Canal shock and does not capture potential baseline Market Access effects.

Finally, Figure 4 reveals a discontinuity between counties in Alberta and Saskatchewan. This follows from the original structure of the transportation network in the latter province,

[15] We follow the market access literature (Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2021; Redding and Venables, 2004) in specifying our covariate of interest in (changes in) natural logarithms.

which favored rail lines that would effectively cut the distance to the eastern part of North America.[16] The by-product of this feature is that the counties located there were, in terms of the transportation network, effectively farther from the Pacific Coast than what might be expected given their location and ended up gaining little Market Access when the Canal started operations.

We consider these facts when presenting empirical evidence of the positive effects of transportation costs reduction on structural transformation and manufacturing outcomes in Section 4. Specifically, we show that our estimates are robust to excluding one province at a time and that even when focusing exclusively on the West Coast, we identify positive effects of the Panama Canal on manufacturing activity.

3.2 Structural Transformation and Manufacturing Outcomes

We digitized information on manufacturing activity from the three Censuses of Manufactures conducted in 1901, 1911, and 1939. These sources include county-level data on the number of establishments, total revenue, capital stock, wage payments, employment, and cost of materials for selected manufacturing establishments. We focus on outcome changes between the 1911 Census (with information for 1910) and the 1939 Census. We also study changes in outcomes between the 1901 Census (with information for 1900) and the 1911 Census to assess our identification assumptions.

Several factors explain our Census choices. First, the 1901 Census was the first Census to comprehensively include the Western Provinces. Second, the 1911 Census was the latest census before the opening of the Canal. Third, the 1939 Census was the last Census before the start of World War II and the most comparable in terms of sector composition after 1932. Despite our choices, the Census of Manufactures considerably changed its coverage and reporting over time.

The 1901 and 1911 Census include some information at the county-industry level. According to the Census, there were 264 manufacturing industries. However, the Census reports aggregate industries with less than three establishments under the "All Other Industries" category. On average, 40% of a county's total manufacturing revenue is classified as "All Other Industries" in 1910 (Table A.3). The 1939 Census only includes aggregate

[16] Contemporaries acknowledged this fact. The Report on the Influence of the Panama Canal on Trade with Western Canada, June, 1930 (C. 3319) concludes that "in most cases Manitoba points are best reached via Eastern Canadian ports; Saskatchewan is border-line territory, but Alberta and British Columbia destinations can most cheaply be reached by all classes of commodities by means of the Panama Canal route."

county-level data.[17] Though not central to our analysis, this data feature prevents us from studying the extent to which reallocation happened across industries. It limits our ability to conduct accurate production function estimations (more on this below).

We adjust the data in three ways to ensure consistency over time. First, we use 1939 county boundaries and adjust data for other years at this level. To do so, we implement the geographic crosswalks proposed by Eckert et al. (2020), which assume economic activity is evenly distributed across space.[18] Second, we account for the fact that earlier censuses surveyed establishments with five or more employees, whereas the 1939 census had full coverage. Following Urquhart and Dales (2007), we use linear expansion factors derived from the (Postal) Census of Manufactures of 1906.[19]. Finally, unlike the 1901 and 1911 censuses, the 1939 census did not include what were described as "hand trades". We remove the hand-trade sectors directly whenever possible in 1901 and 1911. When they were reported in the "All other industries" category, we scale down those values by multiplying the original values in this category by 1 minus the share of those activities in the province's aggregate values.

We complement these sources with information on total county population from the Census of Population and with information on county-level total land value from the Census of Agriculture from 1901, 1911, 1921, and 1941.

3.2.1 Manufacturing Activity

We follow counties over the first half of the 20th century and track their manufacturing activity. We observe the number of manufacturing establishments and their total factory-gate revenue and expenditure in materials, as well as the capital stock, the wage bill, and the number of employees. For labor-related variables, we observe information for two types of workers. First, salaried workers, which include "owners and firm members taking an active part in management or work who draw a living or other allowance out of revenue; and officers, managers, salesmen, etc. with functions of administration." Second, wage workers, defined by the 1901 Census as "the working class, who may be employed either

[17] Moreover, at the province and national level, only the 40 "leading industries" (around 75% of revenue) are reported in detail.

[18] We use the Canadian Century Research Infrastructure boundary files. The procedure we follow is standard in the literature (e.g., Hornbeck (2010); Hornbeck and Rotemberg (2021); Fajgelbaum and Redding (2014)).

[19] We use the information on revenues and inputs provided in the Census of Manufactures of 1906, which is discriminated by establishment size at the province level. In practice, we multiply our 1901 and 1911 figures by a factor that captures how much each variable (revenues, employment, etc.) in 5+ employees establishments in 1906 should expand to equate the variable for *all* establishments in 1906.

in the establishment or out of it."

We focus on three sets of outcomes. The first one is related to the process of structural change. For each county, we track population growth, total employment in manufacturing as a share of total population, average establishment size (by revenue and employment), capital share (capital stock as a share of total revenue), skill share (salaried workers over total employment), and the skill wage-premium (average wage for salaried workers divided by average wage for wage workers). These outcomes are informative about changes in counties' economic structure driven by reductions in transportation costs.

The second set of outcomes is related to manufacturing output and input use. We focus on revenue (value of production), expenditure on materials, total employment in manufacturing, and capital expenditure. We estimate capital expenses using the value of capital stock provided by the different manufacturing censuses and the interest rate paid on mortgages in Ontario (Homer and Sylla, 1996). This implies that the only source of geographic variation in capital expenditures comes from the capital stock and not from the cost of capital, which is a somewhat problematic assumption if different regions have different mixes of capital types and qualities or if there are local-level risk premiums. However, we are constrained by data availability for the period of study.

The third set of outcomes is related to productivity estimates. First, we define total county productivity as the difference between total revenue and total cost of inputs (Basu and Fernald, 2002; Solow, 1957). Following Petrin and Levinsohn (2012) and Hornbeck and Rotemberg (2021), we decompose total productivity into two components: total factor revenue productivity (TFPR) and allocative efficiency (AE). The first component relates to growth in revenue beyond the growth implied by increased input use, which is a classical measure of (revenue) productivity. The second relates to changes in surpluses related to the increased use of inputs in locations where they yield a higher productive use due to imperfections in the economy (such as mark-ups or distortions) that prevent equalizing marginal revenue products. AE increases when relatively distorted locations due to friction increase their input use.

This decomposition rests on strong assumptions about prices, input measurement, and production function estimation. Given the data availability we have described so far, we only attempt the productivity decomposition to provide suggestive evidence about the potential sources of productivity growth after a considerable reduction in transportation costs. The following subsection describes the decomposition and discusses its limitations.

3.2.2 Productivity Decomposition

Our starting point is county aggregate productivity:

$$\ln(Productivity_c) = \xi \left[\ln(P_c Q_c) - \sum_k s_{c,k} \ln(W_{c,k} X_{c,k}) \right] \tag{6}$$

Where P_cQ_c is the gross value of manufacturing establishments (revenue) in county c, $s_{c,k}$ is the revenue share for input k in county c, and $W_{c,k}X_{c,k}$ are expenditures in input k in county c. The set of inputs k consists of capital, labor, and materials (including fuel and power). Though we fully describe productivity, we only observe total revenue and input expenditure. We do not observe prices. Finally, ξ is a scalar that converts the expression from output growth to productivity growth.[20]

Though the productivity measure is directly observed from the data, it relies on the correct measurement of inputs. Leaving aside the issue of capital stocks estimates (Collard-Wexler and De Loecker, 2021), our interest rate assumption generates measurement error on our capital expenditure estimates. On one hand, this assumption is consistent with free capital flows and homogeneous capital goods across counties in Canada. On the other hand, this approach underestimates capital expenditures for counties with higher interest rates than mortgage rates in Ontario, which would lead to overestimating their productivity. The opposite is true for counties with lower interest rates.

It is unclear how these input estimation issues relate to the changes in market access brought about by the Panama Canal. If the interest rate falls to the national level in counties that experience a large reduction in transportation costs while input use increases, we would underestimate productivity growth for places that benefited the most from the Canal. Reassuringly, we showed changes in Market Access due to the Canal are not correlated with initial Market Access. This might be useful if capital costs correlate with market access. Moreover, our main specification includes province fixed effects, which should control for common changes in capital costs by province. Again, unlike the US and to the very best of our knowledge, there are no disaggregated and reliable interest rate estimates by province for Canada during the period we study, so we rely on national-level interest rates.

As in Petrin and Levinsohn (2012), the previous measure of productivity could be further decomposed into growth in TFPR and growth in AE (equations 7 and 8) under some production function assumptions. If we assume a CRS Cobb-Douglas production

[20] The productivity scalar used here is $\xi = 1/[1 - \frac{1}{C}\sum_c \sum_k s_{k,c}]$, where C is the total number of counties.

function, like Hsieh and Klenow (2009), we can estimate output elasticities for each input (α_k) by calculating inputs' cost shares. Hornbeck and Rotemberg (2021) follow this approach, estimating production function elasticities for each industry at the national level and averaging among industries according to each county's industry composition.

$$TFPR_c = \xi \left[\ln(P_c Q_c) - \sum_k \alpha_k \ln(W_{c,k} X_{c,k}) \right] \tag{7}$$

$$AE_c = \xi \left[\sum_k (\alpha_k - s_{c,k}) \ln(W_{c,k} X_{c,k}) \right] \tag{8}$$

Given that we do not observe county-industry level data for all years, we follow a different approach: we estimate production function elasticities using province-level input costs shares, which we then apply to individual counties to estimate input wedges as $\frac{\alpha_k - s_{c,k,t}}{s_{c,k,t}}$.[21] Figure A.1 shows the geographic distribution of estimated input wedges for Canada.

While our production function estimates are relatively coarse (province level, instead of industry level), we show some suggestive patterns in the data consistent with a story in which the estimated wedges come from local level mark-ups or distortions. Estimated wedges for capital and labor are positively related with Market Access in 1910, while wedges for materials are negatively correlated with Market Access (Figure A.2, Table A.6). Moreover, the correlation between average establishment size and capital and labor wedges is negative, while counties with larger establishments tend to have larger materials wedges. Additionally, in counties with a high concentration of manufacturing activity in a few sectors (high sector Herfindahl-Hirschman index in 1910), capital and labor wedges tend to be large, while there is no correlation between sector concentration and wedges for materials (Table A.5).

TFPR and AE refer to different notions of productivity. The former captures whether agents can produce more output for the same level of inputs given the existing production technology. The latter emerges whenever we move from a perfectly competitive environment and allow for imperfections that prevent equalizing marginal products and marginal costs across economic sectors and places. Because we allow for differences between marginal products and costs to exist and for them to vary over sectors and locations, an additional unit of a given input—labor, capital, or materials—yields different increases in production in different places and sectors. The reallocation of inputs between sectors

[21] We also show results for a productivity decomposition that uses national level input cost shares to estimate output elasticities.

and locations can increase productivity and output because those inputs can now be used in places or sectors producing greater economic value. Why? Because we allow for the differences between marginal products and costs to exist and vary over sectors and places. These differences are captured in equation 8 above by the term $(\alpha_k - s_{c,k})$.[22] The approach we use is agnostic about the nature of the differences between marginal revenue products and marginal costs. For example, the wedges we compute are the same from a productivity accounting perspective, whether they originate from products with artificially high prices (e.g., resource rents) or relatively low marginal costs. Resource rents, for example, might have a different economic interpretation in some contexts but are accounted for as other forms of "surplus" in this accounting framework.

4 Empirical Analysis

Main Specification

Equation 9 shows our main specification:

$$\begin{aligned} \Delta_{1939,1910} ln(Y_c) &= \beta \Delta \ln(MA_{1920}) + \gamma_p + \eta G_c + \epsilon_c \\ \Delta_{1939,1910} ln(Y_c) &= ln(Y_{c,1939}) - ln(Y_{c,1910}) \end{aligned} \quad (9)$$

Where Y_c is an outcome of interest for county c and $\Delta \ln(MA_{1920})$ is the causing variable of interest. γ_p are province-level fixed effects. G_c is a vector of county-level controls that includes: (i) a quadratic polynomial on latitude and longitude of county c; (ii) its distance to the coast, the Great Lakes, or the St. Lawrence River; and (iii) (log) population in 1911. The coefficient of interest is β. This empirical specification is derived from the model that we solve for in Section 5 and that predicts a log-linear relation between output, inputs, productivity and market access. We report standard errors clustered at the level of 43

[22] These concepts can be illustrated with a hypothetical example: Toronto is a manufacturing center where producers are highly productive, and markets work relatively well, so the differences between marginal output and marginal costs are small. Vancouver is a city where wages and the return to capital are the same as in Toronto, but where wood is much cheaper because forests surround it. Production there, however, is constrained because it is distant from most major markets on the eastern seaboard of North America. Now, let us assume there are two identical carpenters in both cities. Whenever their carpentry skills improved so that they could use less wood to manufacture a table, Toronto and Vancouver's TFPR would increase. If the Vancouver carpenter started working more hours manufacturing tables while leaving his skills constant, Vancouver's AE would increase. Note that, in a perfectly competitive economy where the difference between marginal products and marginal costs are equalized across sectors and places – the carpenter's increased labor in Vancouver would not increase productivity.

arbitrary grids measuring 300km by 300km to allow for a potential spatial correlation of the residuals.

We estimate equation 9 using ordinary least squares. This specification leverages within-province variation in changes in market proximity, net of geographic and baseline controls. Our coefficient of interest is β, which we can identify under two identification assumptions, given that our specification is equivalent to a 2x2 difference-in-differences design with continuous treatment and no untreated group (Callaway et al., 2024). First, that counties within the same province, after controlling for county variables, would have followed a similar development path had the Panama Canal shock not materialized (parallel trends). Second, that there is no selection into treatment doses. In other words, that the treatment effect of MA gains would have been the same for all dose levels if the treatment dose had been the same (strong parallel trends). We discuss both assumptions in Section 4.3.

4.1 Empirical Results

Panel A in Tables 4 and 5 present the results of the estimation of equation 9. For ease of interpretation, these are reported in terms of the interquartile range of (log) market access due to the Panama Canal in 1920.[23] For instance, moving from the 25th to the 75th percentile in terms of exposure to the Canal shock leads to an increase in population of 3.3%.

The results outlined above indicate that the Panama Canal had economically significant effects on the Canadian economic structure during the study period. They suggest several patterns. First, the Canal altered the economic geography of Canada: production inputs were used relatively more in counties that benefited more from the reduction in transportation costs. Second, increases in input usage occurred across the board and on relatively similar scales. Indeed, the 95% confidence intervals for the coefficients of interest on capital expenditures, the wage bill, and material expenditures overlap. This suggests that the expansion in production brought about by the Canal was not biased toward a specific factor of production.

Given the sizable increases in aggregate productivity, Columns (6) and (7) of Table 5 suggest that transportation infrastructure induced the reallocation of inputs and economic

[23] We use the interquartile range of the residual (within) variation of our causing variable, which is the one we use for identification and equals 0.0027. The corresponding magnitude for the original causing variable is 0.0045.

activity to activities where production levels were inefficiently low due to the presence of distortions and imperfections in the economy. This reallocation increased AE because inputs shifted to activities where the difference between the value of marginal products and their marginal cost was higher, yielding an increase in the surplus generated in the county. We note that these gains in productivity are different from classical gains from trade typically associated with specialization and the division of labor and production – elements captured in our decomposition by TFPR. These findings can be rationalized in a world in which there are constant returns to scale (CRS) in manufacturing and constant elasticity of substitution (CES) demand, so increased scale does not yield improvements in technical productivity (TFP) or its value (TFPR). As discussed before, our productivity decomposition is limited by the available data. However, we believe it still provides valuable insights into the nature of reallocation that occurs with reductions in transportation costs.

4.2 Robustness Checks

This subsection describes some robustness tests related to different samples, specifications, parameter values for estimating Market Access, and estimates of standard errors.

Figure A.4 shows that our results are qualitatively similar when we drop one province at a time. In other words, our results do not depend on any particular province. Table A.9 presents estimates from different specifications of equation 9. Column (1) shows the main specification for comparison. Column (2) removes province fixed effects and geographic controls, including coordinates polynomials. Column (3) removes only province fixed effects, and Column (4) removes geographic controls. Overall, these results highlight that our estimates do not rely on within-province comparisons, though some coefficients are much more precisely estimated when controlling for province fixed effects.

Table A.10 reports a variety of robustness checks that address concerns related to assumptions made while calculating Market Access. Column (1) shows that our results change little when the population figures are fixed at their 1910 level. We interpret this as evidence that the results are driven by the (exogenous) reduction in transportation costs brought about by the Canal rather than by (potentially endogenous) population movements in anticipation of its completion. Columns (2) and (3) present estimates that assume different values for θ, the trade elasticity. Our qualitative results and conclusions remain unchanged when we assume $\theta = 1$ or $\theta = 7.5$.[24] This is because a different trade

[24] Figure A.3 shows that our point estimates and standard errors change very little when we use different values of trade elasticity.

elasticity alters both elements used in the construction of our causing variable in a similar fashion. We then assume different average transportation costs while keeping all other baseline assumptions fixed for the computation of market access. Estimates in Columns (4) and (5) change very little, and the conclusions remain unaltered when we set $\bar{P} = 5$ or $\bar{P} = 20$.

Finally, Table A.11 presents five standard error estimates in addition to the already described estimates. We present standard error estimates from clustering counties at an arbitrary grid level, varying grid size from 150km, 200km, and 300km (our preferred specification). We also show estimates using Conley's (1999, 2008) standard errors with distance cutoffs from 150km to 600km. Most of the inference from the main tables holds when using different standard error estimates.

4.3 Potential Confounders

So far, the specification detailed in this section relies crucially on the assumption that the change in market access is exogenous. Simultaneity can be ruled out in this case because the economic outcomes of Canadian counties did not determine the construction of the Canal. Moreover, Table A.1 shows that MA gains are not correlated with initial MA levels. The MA effect of a Canal opening in an isthmus in Central America disseminated through the existing transportation networks in ways that might be as good as random from the perspective of Canadian counties. This insight is helpful for considering the limitations highlighted by Callaway et al. (2024) for difference-in-differences designs with continuous treatment. While their strong parallel trend assumption is fundamentally untestable, it might hold in the context we study. In other words, one of our identifying assumptions is that the treatment effect of market access gains at any hypothetical level is the same regardless of actual level of treatment.[25]

Still, omitted variables correlated with gains in Market Access could confound the effect of the Panama Canal. One concern is that the Canadian counties that benefited the most from the building of the Panama Canal would have grown faster than other counties, even in the absence of the Canal. To assess this concern, we conduct a pre-trends analysis. We repeat our main specification, using the change in outcomes between 1900 and 1910 as the dependent variable.

Panel B on Tables 4 and 5 show that our causing variable is unable to predict changes in

[25]This assumption is consistent with the theoretical predictions resulting from the model we expose and motivates our empirical exercise. See Equation 14.

the outcomes of interest between 1900 and 1910. Moreover, point estimates for $\Delta \ln(MA_{1920})$ are one order of magnitude smaller than estimates in Panel B and not statistically significant. Taken as a whole, Tables 4 and 5 suggest that the variation that we exploit to identify our parameters of interest is not correlated with unobservable secular county-level trends. We consider this fact as evidence that the parallel trends assumption, more traditional to differences-in-difference models with discrete treatment, is likely to hold.

Another concern relates to potential omitted variables that are not only correlated with Market Access gains due to the Panama Canal but also only start to matter after 1910. If that were the case, we could be confounding the effect of transportation cost reductions due to the Canal with other factors.

The most salient one is perhaps the endogenous response of the transportation network to the introduction of the Panama Canal route. If counties that benefited from the Canal also started to build more railway lines, then it would be hard to disentangle the pure effect of the Canal from the improvement in the railroad network. Table A.2 shows that for the decade after 1910, counties that would benefit from the Canal did lay more kilometers of railroads. In Table 6, Panel A, we control for the construction of railroad lines between 1910 and 1940 and find that our estimates remain unchanged.

Beyond railroads, it might be the case that counties that benefited from the Panama Canal also invested in their local infrastructure to take advantage of the new maritime routes. The challenge with this concern is that, unlike for railroads, we lack a measure of general infrastructure improvements. Table 6, Panels B and C propose two indirect tests. First, we interact our measure of Market Access gains with the initial level of Market Access. Suppose less connected places invested more in their local infrastructure after learning about the Canal's gains. We should see a strong negative coefficient for this interaction in that case. Results from Panel B show that better-connected places did grow faster and became more manufacturing intensive over time. However, the coefficients for our measure of Canal-driven gains remain unchanged. Moreover, the interaction term between Canal-driven gains and initial Market Access is negative but small and not statistically significant in most cases.

Another indirect test comes from our description of the distribution of Market Access gains from Section 3.1. In Panel C, we focus on all counties west of Winnipeg and show that the Market Access effects follow a gradient, especially by latitude. Places further from Vancouver that benefited a lot from the Canal see the most significant improvements in manufacturing outcomes. We should not expect a location gradient if the results were driven by idiosyncratic general improvements in the transportation network and not from a shock related to distance to Vancouver or other ports in the Pacific.

Finally, the opening of the Canal coincided with technological advancements that benefited the lumber sector, such as the introduction of trucks for extracting lumber beyond the constraints of existing railroad lines. Considering the significant role played by the lumber sector in both the rise of the West Coast and the perceived utility of the Panama Canal, it is plausible that gains in manufacturing activity originated from areas with high potential for lumbering. Panel D investigates this hypothesis by interacting our measure of Market Access gains with the 1900-1910 average share of a county's area devoted to forests, a measure of lumber potential. Results indicate that the standalone effect of Market Access gains remains relevant and of similar magnitude to other analyses. Furthermore, counties with high lumber potential were not the primary drivers of the increases in manufacturing activity.[26]

5 Counterfactual Analysis: Closing the Panama Canal in 1939

What were the *aggregate* effects of the Panama Canal on the Canadian economy? What was the role of greater integration with international markets on the expansion brought by the Canal? Our previous analysis does not necessarily enable us to answer these questions since the observed effects may result from the displacement of economic activity across counties and we lack detailed micro-data on county-level exports. Although the mere shifting of inputs could still imply a higher level of productivity and economic activity, we want to establish whether the Canal induced a higher *aggregate* level of economic activity in Canada in a manner that allows for displacement and general equilibrium effects to occur.

To answer the first question, we calibrate an extension of a benchmark economic geography model (Eaton and Kortum, 2002). The extension developed by Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2021) features a market structure in which there are exogenous differences between factor prices and their marginal products, input market frictions or mark-ups. We use the model to compute hypothetic equilibria

[26] We prefer to include a county's area devoted to forest as a measure of pre-existing endowments or lumber potential to measures of baseline lumbering output since output-based measures might be endogenous to market access. We include those measures for completeness in Table A.12, while nothing that our conclusions do not change.

under different scenarios: (i) closing the Panama Canal permanently and (ii) opening the Canal just for Canadian domestic trade. Such equilibria feature new counterfactual quantities (population, production factors, etc.) and prices (goods, factor prices, etc.), which allows us to assess the Canal's relevance for Canada from an aggregate perspective. These new quantities and prices incorporate general equilibrium effects that our reduced-form exercise cannot account for. We use those quantities to compute losses or gains in manufacturing that materialize due to changes in input use in places with different allocative efficiency. Further, we exploit the model's land rental rate prediction to compute agricultural sector losses. We add those two to come up with a measure of the losses that the Canadian economy would face upon the materialization of those two scenarios.

Our two counterfactual exercises allow us to investigate different questions. First, the permanent closure of the Panama Canal in 1939 enables us to provide an answer to the question regarding *aggregate* effects. Second, allowing the use of the Canal *only* for domestic trade within Canada only allows us to study whether most of the gains came from greater exposure to international or domestic markets. In both cases, we assume that the population might freely adjust over time, reflecting long-term equilibria. We then use those new quantities to calculate the changes in productivity levels that result in the new equilibrium. We conclude by providing a calculation that also incorporates effects on the agricultural sector that follow from different land-rental rates. We do so because, as land is assumed to be fixed, land rents capitalize all the gains in market access that do not relate to imperfections in the economy - which are given by changes in manufacturing productivity.

5.1 Model Primitives[27]

Production in each county is undertaken by firms that maximize profits by optimally choosing inputs while taking their price and input frictions as given. The production technology follows a Cobb Douglas production function for variety j that utilizes capital (K), labor (L), land (T), and an intermediate good (M) as inputs. These are paid, respectively, an interest rate r_c, a wage w_c, rent q_c, and price P_c. Producers have CES preferences over the continuum of varieties used as intermediates with a CES of σ. P_c is, therefore, a CES Price Index. Following the standard assumption in Eaton and Kortum (2002) and the results in our reduced form exercise in Section 4, each county has an exogenous technical efficiency level for variety j which is drawn from the Fréchet distribution with CDF $F_c = 1-e^{-A_c z^{(-\theta)}}$.

[27] This subsection follows Hornbeck and Rotemberg (2021) closely.

The marginal cost of production is thus characterized by equation 10:

$$MC_c(j) = \frac{r_c^{\alpha_c^K} w_c^{\alpha_c^L} P_c^{\alpha_c^M}}{z_c(j)} = \frac{\Pi_k (w_c^k)^{\alpha_c^k}}{z_c(j)} \tag{10}$$

Where K inputs are capital, labor, land, and intermediate inputs. By assumption, producers face frictions in each one of the input markets in which they participate. These are taken as given and exogenous. They prevent firms from using inputs up to the point where price equals marginal cost. In this context, $1 + \psi_c^k$ in equation 11 captures the factor k specific input friction that embodies firms' inability or unwillingness to expand production beyond a given level. Note that these are fixed and exogenous to the model. We make this assumption because our covariate of interest cannot predict changes in these wedges (See Table A.8).

$$p_c(j) = \frac{\Pi_k ((1 + \psi_c^k) w_c^k)^{\alpha_c^k}}{z_c(j)} > MC_c(j) \tag{11}$$

This implies that, although nominal wages can differ across places, they simply reflect higher nominal prices. When we conduct the counterfactual exercise, we fix worker's utility at this initial level and allow for changes in the total population in our area of study after the Canal shock. Thus, Canadian and American populations move freely and without cost, are not fixed in the aggregate, and might draw or expulse individuals (international migration) from abroad.

Workers supply labor inelastically, receive a wage w_c, and have CES preferences over the j good varieties like firms do. The indirect utility is $V = w_c/P_c$, with both w_c and P_c being endogenously determined within the model. To focus on a hypothetical long-run equilibrium, we assume that workers are perfectly mobile across counties, so any difference in indirect utility can be arbitraged out. Note that this is the case for both Canada and the United States, which we see as areas where internal migration might be plausible over the long run. [28] This leads us to set $V = \bar{U}$. This implies that, although nominal wages can differ across places, they simply reflect higher nominal prices. When we conduct the counterfactual exercise, we fix worker's utility at this initial level and allow for changes in the total population in our area of study *after* the Canal shock.

We assume that capital is perfectly mobile so that interest rates are equalized across

[28] The United States was an important source of migration to Canada over the first half of the twentieth century. By 1921, over 4% of the Canadian population and 19% of Canada's foreign-born population were born in the United States. Furthermore, between 1931 and 1940, the United States was the top country of origin of immigrants to Canada (StatisticsCanada, 2016).

counties ($r_o = r$). Further, we assume that Canada faces a perfectly elastic supply of capital, with the interest rate being exogenous and set abroad. We also assume that the land supply is completely inelastic and set at exogenous levels. The remuneration of land –rents– is, however, endogenous to the model. Trade in final or intermediate goods and between counties c and d can occur while incurring an iceberg trade cost τ_{cd}, defined in the same fashion as equation 2. The price in county c of a good produced in county d is $p_{cd}(j) = \tau_{cd} p_{dd}(j)$, where $p_{dd}(j)$ is the price of good-variety j in county d and $\tau_{cd} > 1$ for all $c \neq d$. Goods markets clear in general equilibrium, so demand and supply are equal. Production in each county equals the sum of exports to all possible destinations plus within-county sales.

5.2 Analytical Results

Solving the model yields a set of equations that are useful in understanding how changes in transportation costs might affect economic activity. The first important result, in line with Eaton and Kortum (2002), is given in equation 12. Also known as a "gravity equation", it states that exports from county c to county d (E_{cd}) are positively related to the origin's technical efficiency (A_c), the destination's income (Y_d) and its price level (P_d) (Chaney, 2018). On the other hand, it is inversely related to transportation costs (τ_{cd}), input prices (W_c^k), and distortions in the county of origin ($1 + \psi_c^k$). This rationalizes an important stylized fact in the international trade literature: countries with larger economic mass (income) and closer to each other (with lower trade costs) tend to trade more among themselves.

$$E_{cd} = \kappa_1 A_c \left(\Pi_k \left((1 + \psi_c^k) W_c^k \right)^{\alpha_c^k} \right)^{-\theta} \tau_{cd}^{-\theta} Y_d P_d^{\theta} \tag{12}$$

$$\kappa_1 = \left(-\frac{\theta}{1 - \sigma} \right) ln \left(\Gamma \left(\frac{\theta + 1 - \sigma}{\theta} \right) \right)$$

A second result is that market access- a concept we exploit in our reduced-form work- is an inverse transformation of the CES price index. Moreover, and as Equation 13 shows, a county's market access (MA_c) is inversely related to transportation costs (τ_{cd}) and is greater whenever the locations to which a given county has access have a higher income or greater market access themselves. A third result is that, in general equilibrium, changes in a county's market access summarize how changes in transportation costs affect each county's economic activity through changes in both goods and factor markets.[29] This relation is described in equation 13 below.[30]

$$P_c^{-\theta} = MA_c = \kappa_2 \sum_d \tau_{cd}^{-\theta} Y_d MA_d^{\frac{-(1+\theta)}{\theta}} \tag{13}$$

$$lnY_c = \kappa_1 + \kappa_{2c} + \left(\frac{\alpha_c^M + \alpha_c^L + 1}{1+\theta_c^T}\right) ln(MA_c) \tag{14}$$

A fourth result is that the model predicts how changes in market access will affect productivity and input prices. Given the assumption that technical efficiency is exogenous, changes in county-level productivity should come from changes in AE. Equation 15 shows that this is indeed the case, with changes in county-level productivity being driven by changes in equilibrium input quantities. Equations 16 and 17 show that increases in market access yield log-linear responses in equilibrium input prices. Note that there is no capital remuneration equation as it is assumed to be exogenous to the model.

$$\frac{d \ln PR_c}{d \ln MA_c} = \frac{P_c Q_C}{Pr_c} \sum_k \left(\alpha_c^k - s_c^k\right) \frac{d \ln X_c^k}{d \ln MA_C} \tag{15}$$

$$\frac{d \ln q_c}{d \ln MA_c} = \frac{\alpha_c^M + \alpha_c^L + 1}{1 + \theta \alpha_c^T} \tag{16}$$

$$\frac{d \ln w_c}{d \ln MA_C} = \frac{d \ln P_c}{d \ln MA_C} = -\frac{1}{\theta} \tag{17}$$

5.3 Data and Calibration

We allow for trade to take place between counties in the United States and Canada. To solve for prices, we use population data from the Canadian Census of Population of 1941 and the United States Population Census of 1940. We adjust these figures to obtain the size of the population in 1939 under the assumption that the population in each county grew at a constant rate from 1931/1930 to 1941/1940. We use our data on transportation costs, with and without the Panama Canal, to compute iceberg trade costs. We use data from the Canadian Census of Manufactures of 1939 to compute measures of the distortions and output elasticities. For this counterfactual exercise, we assume output elasticities and

[29] For a detailed discussion of consumer market access and firm market access, see Redding and Venables (2004) and Hornbeck and Rotemberg (2021).

[30] Where $\kappa_2 = \bar{U}\rho^{\frac{1+\theta}{\theta}}$ and $\kappa_{2c} = \frac{\kappa_{1c} + ln\psi_c - \theta\alpha_c^T \ln \frac{\alpha_c^T}{T_c}}{1+\theta\alpha_c^T}$

distortions in each input that result from output-weighted measures of the county-level information that we retrieve from the Census of Manufactures.[31] As in our empirical work, we follow the literature assuming $\theta = 5$.

We follow the procedure outlined in Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2021) to compute the effects of our two counterfactual exercises. This entails: (i) solving for prices using the iceberg trade costs assuming that the Canal is open – that is, factual trade costs. Then: (ii) we use each county's factual population to solve for an "amenity" that captures each county's fixed endowments of land and productivity, which are not observed in the data. Next: (iii) we use the "amenity" and counterfactual trade costs to compute the counterfactual distribution of the population under the assumption that the Canal closes. Then: (iv) we solve for prices again using the counterfactual distribution of population. Next: (v) we use equation 15 to compute changes in output. Then: (vi) use our assumption that revenue shares are constant to retrieve changes in input bills. Finally: (vii) use equations 16 and 17 to determine what fraction of the change in input bills relates to changes in real input demand rather than input prices. In our counterfactual scenario, we then measure how much equilibrium quantities of labor, capital, and intermediate goods inputs change.

To compute the impact on manufacturing productivity of such changes as a fraction of the manufacturing sector's GDP, we use equation 18:

$$\begin{aligned} ChangeProductivity =& \frac{1}{GDP}\sum_c dProductivity_c \\ ChangeProductivity =& \frac{1}{GDP}\sum_c Productivity_c dln(Productivity_c) \end{aligned} \tag{18}$$

Given that we assume technical efficiency to be exogenous, changes in productivity only come from changes in AE[32].

[31] We set $\alpha_c^T = 0.1748$ (Caselli and Coleman II, 2006), $\alpha_c^L = 0.255$, $\alpha_c^K = 0.0768$, and $\alpha_c^M = 0.669$. These come directly from the data.

[32] Changes in AE might be written as $d\ln(AE_c) = \frac{P_cQ_c}{Productivity_c}[\sum_k(\alpha_c^k - s_c^k)d\ln(X_c^k)]$. See equation (20) in Hornbeck and Rotemberg (2021).

$$
\begin{aligned}
ChangeProductivity =& \frac{1}{GDP}\sum_c Productivity_c dln(AE_c) \\
=& \frac{1}{GDP}\sum_c P_c Q_c \sum_k \left(\alpha_c^k - s_c^k\right) d\ln(X_c^k) \\
=& \sum_c D_c \sum_k \left(\alpha_c^k - s_c^k\right) d\ln(X_c^k)
\end{aligned} \tag{18}
$$

Where $D_c = \frac{P_c Q_c}{GDP}$ are Domar weights and $d\ln(X_c^k)$ is the change in input k quantity as predicted by the model. Here, the K inputs are labor, capital, and materials. Land is not included, as it is assumed to be in fixed supply and, hence, always used at efficient levels, which implies $d\ln(X_c^T) = 0$. We take $(\alpha_c^k - s_c^k)$ directly from Section 4 of this study.

5.4 Results

We find that closing the Panama Canal permanently for all traffic in 1939 would substantially change Canada's economic geography. First, its total population would decrease by 2.7% as Canada loses population to the United States and overseas. This result follows from the assumption of fixed real wages, which implies that migration is costless, leading to frictionless population adjustments once the losses in market access materialize. Second, as shown in Figure 5, population losses concentrate in places closer to the coasts and would be particularly severe in counties located in the western provinces. British Columbia's and Alberta's population would fall, respectively, by 7.8% and 4.2%. Counties closer to the Atlantic Ocean or the St. Lawrence River basin would see decreases too, but smaller ones than those in counties closer to the Pacific Ocean. This result suggests that the Panama Canal shock was probably stronger for the Western Hemisphere Pacific basin than for the Atlantic basin. Finally, some counties in the prairie provinces, such as Manitoba and Saskatchewan, would observe marginal changes in population. The populations of these two provinces would increase by 0.07% and 0.2%, respectively.

Our new counterfactual equilibrium features factor prices and demand, enabling us to compute the economic losses of closing the Canal. In our setting, land is the fixed factor and is always used at efficient levels. Its rents, therefore, capitalize on all market access gains in the absence of imperfections in the economy. However, as we discuss above, the presence of imperfections in the economy opens the way to gains from reallocating inputs across places and sectors. We move to discuss those losses below.

We calculate substantial impacts in terms of changes in productivity as defined in equation 18. Here, we use changes in real input use from our counterfactual equilibrium for labor, capital, and materials, and the wedges between marginal costs and products observed in the data. Note that we do not include land because its supply is fixed and inelastic, which implies that it is always used at the efficient level. The losses coming from equation 18 add up to 0.27% of Canadian GDP in 1939. These losses, as shown in equation 10, reflect both changes in aggregate inputs and differences in distortions in different counties of Canada. Our results here indicate that, although there could be relevant displacements of production within Canada, the Canal paved the way for increases in the country's aggregate level of economic activity.

As we discuss above and as emphasized in Donaldson and Hornbeck (2016), land values capitalize gains stemming from increases in market access due to its fixed nature. Though always used at the efficient level, a reduction in market access lowers land rents and yields economic losses. Those are different from the ones we find in the previous paragraph. We find that land rents fall 2.6% in our counterfactual scenario. If land values are the present discounted value of rents, we find that closing the Canal in 1939 would yield a permanent economic loss equivalent to 1.59% of Canadian GDP. Thus, the total general equilibrium losses caused by closing the Panama Canal come to 0.27% + 1.59% = 1.86% of Canadian GDP in 1939.

To contrast the model's predictions, we estimate the county level effect of Market Access gains on land values. We do so by digitizing data on total land values and total area in farms from 4 different Canadian Census of Agriculture: 1901, 1911, 1921, and 1941. In an event study framework, we regress log average land values on our measure for MA gains due to the Panama Canal interacted with year dummies, leaving 1921 as the reference year. Table 7 shows the results from these exercises. Column (1) only includes year and county fixed effects. Column 2 includes 1911 population interacted with year dummies, our geographic controls (coordinates and distance to the coast) interacted with year dummies, and province by year fixed effects. Regardless of the specification, the message of Table 8 is consistent with the model's implications: moving one county from the 25th percentile of MA gains to the 75th percentile would increase land rents by 3.3%.

Table 8 compares the results obtained from our counterfactual exercise with those found in the international trade and economic geography literature, which suggest that closing the Panama Canal would be much less consequential for Canada than removing the railroad network for the United States in 1890 (Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2021) or for Argentina in 1914 (Fajgelbaum and Redding, 2014). Our quantitative results, however, are similar in magnitude to those from Hornbeck and Rotemberg's counterfactual exercise in which the United States railroad network is held fixed as it was in 1880 (as opposed to enjoying a decade, from 1880 to 1890, of (frenzied) railroad construction). Closing the Canal in 1939, for instance, would be much more consequential for Canada than freezing the railroad network in 1920 (with the Panama Canal open). Compared to the 1939 factual equilibrium, doing so would yield population losses of just 0.34% and economic losses of only 0.2% of GDP.

We argue that the small, although certainly non-negligible, impacts of the Canal on Canada relate to: (i) the relatively greater importance of existing railroads for the transportation network than the Canal; and (ii) the fact that the economies of the primary beneficiaries of the Canal – the western provinces of Canada – were still relatively small in 1939. Although we see substantial population losses in British Columbia and Alberta, these two provinces accounted for only 14% of the total population and 8.4% of manufacturing employment at the time. This reduces the aggregate relevance of the Canal shock for the Canadian economy.

To address the question of whether greater integration with international markets explain the economic expansion caused by the Canal, we perform a second quantitative exercise in which we close the Panama Canal permanently but only for traffic originating from or destined for locations outside Canada. That is, we allow the use of the Panama Canal only for domestic trade within Canada. We motivate this exercise by theories that emphasize the relevance of integration into the global economy for structural change (Hirschman, 1958; Galor and Mountford, 2008) and our inability to provide detailed microeconometric analysis on this margin due to data constraints.

Compared to the initial 1939 equilibrium that features a fully open Canal, we find that the counterfactual population falls by 2.4% and GDP by 1.74%. The losses come from lower productivity (0.25% of GDP) and lower land values (1.49% of GDP) due to reduced land rents. These represent roughly 93% of the losses we find in our first counterfactual exercise, which suggests that most of the gains brought about by the Panama Canal stemmed from cheaper access to international markets. Gains originating from domestic trade are much smaller. These findings are consistent with remarks by contemporary observers emphasizing increased exports and integration to the global economy as drivers of the expansion brought about by the Canal (Innis, 1933; Mackintosh, 1939). Similarly, they are consistent with theories that emphasize expansion in exports of primary products with greater linkages, such as lumber and wheat, as drivers of economic expansion in the manufacturing sector.

6 Conclusion

We find that the opening of the Panama Canal had substantial impacts on Canadian economic life. First, by leveraging comparatively cheap maritime transportation, the Canal paved the way for gains in market access for all Canadian counties. Thus, the Panama Canal changed Canada's economic geography. Second, those gains drove the growth of the manufacturing sectors of the counties that benefited more from the Canal than others. Productivity increased, thanks mainly to improvements in the allocative efficiency of the use made of production factors. The Canal allowed greater use of inputs in places where they yielded a higher productive use – i.e., marginally productive counties. Third, these effects were not purely local, persisting in general equilibrium. The calibration of a benchmark economic geography model indicates that closing the Canal permanently in 1939 would have led to losses of 1.86% of Canadian GDP in 1939. Most of those losses would be attributable to decreased access to overseas markets. Consequently, it turns out that there was, after all, a free ride.

References

Alexander, P. D. and Keay, I. (2019). Responding to the First Era of Globalization: Canadian Trade Policy, 1870–1913. *The Journal of Economic History*, 79(3):826–861. Publisher: Cambridge University Press.

Asher, S. and Novosad, P. (2020). Rural Roads and Local Economic Development. *American Economic Review*, 110(3):797–823.

Asturias, J., García-Santana, M., and Ramos, R. (2019). Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India. *Journal of the European Economic Association*, 17(6):1881–1940.

Atack, J., Bateman, F., Haines, M., and 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. Publisher: Cambridge University Press.

Banerjee, A., Duflo, E., and Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, 145:102442.

Baqaee, D. R. and Farhi, E. (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, 135(1):105–163.

Basu, S. and Fernald, J. G. (2002). Aggregate productivity and aggregate technology. *European Economic Review*, 46(6):963–991.

Belmar, J. (2023). Trade and structural change: Evidence from colombia and the panama canal. *working paper.*

Bolt, J. and van Zanden, J. L. (2020). Maddison project database, version 2020. *URl: https://www. rug. nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020 (visited on 04/23/2021)(cit. on p. 25)*.

Brancaccio, G., Kalouptsidi, M., and Papageorgiou, T. (2020). Geography, Transportation, and Endogenous Trade Costs. *Econometrica*, 88(2):657–691. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA15455.

Callaway, B., Goodman-Bacon, A., and Sant'Anna, P. H. (2024). Difference-in-differences with a continuous treatment. Technical report, National Bureau of Economic Research.

Cao, Y. and Chen, S. (2022). Rebel on the Canal: Disrupted Trade Access and Social Conflict in China, 1650–1911. *American Economic Review*, 112(5):1555–1590.

Carlos, A. M. and Lewis, F. (1992). The Profitability of Early Canadian Railroads: Evidence from the Grand Trunk and Great Western Railway Companies. In *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel*, pages 401–426. University of Chicago Press.

Caselli, F. and Coleman II, W. J. (2006). The World Technology Frontier. *American Economic Review*, 96(3):499–522.

Chaney, T. (2018). The Gravity Equation in International Trade: An Explanation. *Journal of Political Economy*, 126(1).

Collard-Wexler, A. and De Loecker, J. (2021). Productivity and capital measurement error. *NBER Working Paper*.

DANE (1976). *III Censo Industrial*.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108(4-5):899–934.

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

Droller, F. and Fiszbein, M. (2021). Staple products, linkages, and development: Evidence from argentina. *The Journal of Economic History*, 81(3):723–762.

Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, 70(5):1741–1779. Publisher: [Wiley, Econometric Society].

Eckert, F., Gvirtz, A., Liang, J., and Peters, M. (2020). A method to construct geographical crosswalks with an application to us counties since 1790. Technical report, National Bureau of Economic Research.

Fajgelbaum, P. and Redding, S. J. (2014). Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914. Working Paper 20217, National Bureau of Economic Research. Series: Working Paper Series.

Fogel, R. (1964). *Railroads and American Economic Growth: Essays in Economic History*. Johns Hopkins Press, Baltimore.

Foster, L., Grim, C., Haltiwanger, J., and Wolf, Z. (2016). Firm-level dispersion in productivity: is the devil in the details? *American Economic Review*, 106(5):95–98. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Galor, O. and Mountford, A. (2008). Trading population for productivity: Theory and evidence. *The Review of Economic Studies*, 75(4):1143–1179.

Haltiwanger, J. (2016). Firm dynamics and productivity: tfpq, tfpr, and demand-side factors. *Economía*, 17(1):3–26. Publisher: Brookings Institution Press.

Haltiwanger, J., Kulick, R., and Syverson, C. (2018). Misallocation Measures: The Distortion That Ate the Residual.

Head, K. and Mayer, T. (2014). Chapter 3 - Gravity Equations: Workhorse,Toolkit, and Cookbook. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*, volume 4 of *Handbook of International Economics*, pages 131–195. Elsevier.

Herrendorf, B., Rogerson, R., and Ákos Valentinyi (2014). Chapter 6 - growth and structural transformation. In Aghion, P. and Durlauf, S. N., editors, *Handbook of Economic Growth*, volume 2 of *Handbook of Economic Growth*, pages 855–941. Elsevier.

Hirschman, A. (1958). *The Strategy of Economic Development*. Yale University Press.

Homer, S. and Sylla, R. E. (1996). *A History of Interest Rates*. Rutgers University Press.

Hornbeck, R. (2010). Barbed wire: Property rights and agricultural development. *The Quarterly Journal of Economics*, 125(2):767–810. Publisher: MIT Press.

Hornbeck, R. and Rotemberg, M. (2021). Railroads, Reallocation, and the Rise of American Manufacturing. Working Paper 26594, National Bureau of Economic Research. Series: Working Paper Series.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Huebner, G. G. (1915). Economic Aspects of the Panama Canal. *The American Economic Review*, 5(4):816–829. Publisher: American Economic Association.

Innis, H. (2018). *Essays in Canadian Economic History*. University of Toronto Press. Publication Title: Essays in Canadian Economic History.

Innis, H. A. (1933). *Canada and the Panama Canal*.

Jacks, D. S. and Novy, D. (2018). Market Potential and Global Growth over the Long Twentieth Century. *Journal of International Economics*, 114:221–237.

Jaworski, T. and Keay, I. (2021). Globalization and the Spread of Industrialization in Canada, 1871–1891.

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

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3):483–499. Publisher: The University of Chicago Press.

Lawrence, J. C. (1957). *Markets and capital: a history of the lumber industry of British Columbia (1778-1952)*. PhD thesis, University of British Columbia.

Limão, N. and Venables, A. J. (2001). Infrastructure, Geographical Disadvantage, Transport Costs, and Trade. *The World Bank Economic Review*, 15(3):451–479.

Liu, D. and Meissner, C. M. (2015). Market potential and the rise of US productivity leadership. *Journal of International Economics*, 96(1):72–87.

Mackintosh, W. A. (1939). *The economic background of dominion-provincial relations : a study prepared for the Royal Commission on Dominion-Provincial Relations*. J.O. Patenaude, Ottawa.

Martinez-Galarraga, J., Tirado, D. A., and González-Val, R. (2015). Market potential and regional economic growth in Spain (1860–1930). *European Review of Economic History*, 19(4):335–358.

Maurer, N. and Yu, C. (2008). What T. R. Took: The Economic Impact of the Panama Canal, 1903–1937. *The Journal of Economic History*, 68(3):686–721.

Maurer, N. and Yu, C. (2010). The big ditch. In *The Big Ditch*. Princeton University Press.

Maurer, S. and Rauch, F. (2020). Economic Geography Aspects of the Panama Canal.

McCullough, D. (2001). *The Path Between the Seas: The Creation of the Panama Canal, 1870-1914*. Simon and Schuster. Google-Books-ID: uVaeIXMxJpQC.

Missiaia, A. (2016). Where do we go from here? Market access and regional development in Italy (1871–1911). *European Review of Economic History*, 20(2):215–241.

of Marine and Fisheries, D. (1922). *Ports and Harbours Directory of Canada*.

of Statistics, D. B. (1937). *Summary of Canal Statistics*.

Pascali, L. (2017). The Wind of Change: Maritime Technology, Trade, and Economic Development. *American Economic Review*, 107(9):2821–2854.

Petrin, A. and Levinsohn, J. (2012). Measuring aggregate productivity growth using plant-level data. *The RAND Journal of Economics*, 43(4):705–725. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12005.

Redding, S. and Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, 62(1):53–82.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720. Publisher: Elsevier.

Rockwell, A. E. (1971). The Lumber Trade and the Panama Canal, 1921-1940. *The Economic History Review*, 24(3):445–462. Publisher: [Economic History Society, Wiley].

Solow, R. M. (1957). Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics*, 39(3):312–320. Publisher: The MIT Press.

Sotelo, S. (2020). Domestic Trade Frictions and Agriculture. *Journal of Political Economy*, 128(7):2690–2738. Publisher: The University of Chicago Press.

StaisticssCanada (1987). *The Input-Output Structure of the Canadian Economy: 1961 - 1981*.

StatisticsCanada (2016). 150 years of immigration in canada.

Umaña-Dajud, C. (2017). Domestic transport costs, Canada, and the Panama Canal.

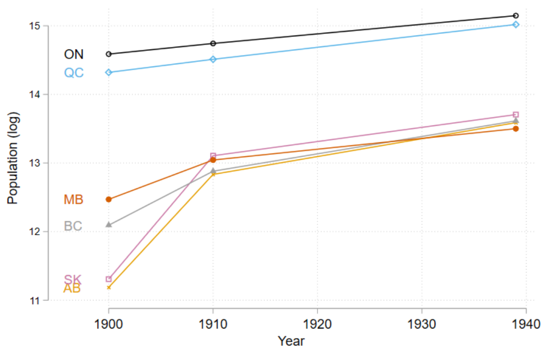
Urquhart, M. C. and Dales, J. H. (2007). 2. New Estimates of Gross National Product, Canada, 1870-1926: Some Implications for Canadian Development. In *2. New Estimates of Gross National Product, Canada, 1870-1926: Some Implications for Canadian Development*, pages 9–94. University of Chicago Press.

Zárate, R. D. (2020). Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City.

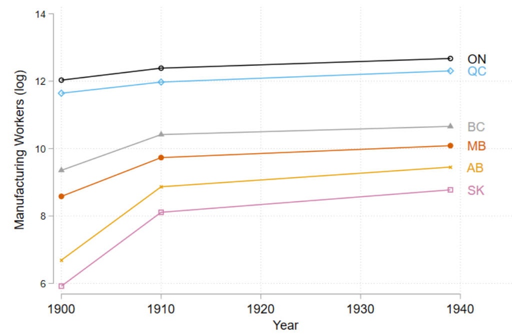
Figures and Tables

Figure 1: Population and Manufacturing Activity in Canada and the US West Coast

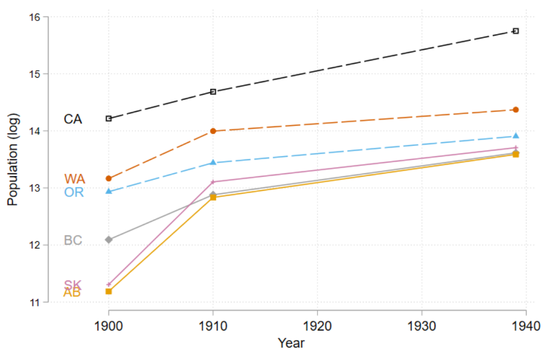
(a) Population by Province



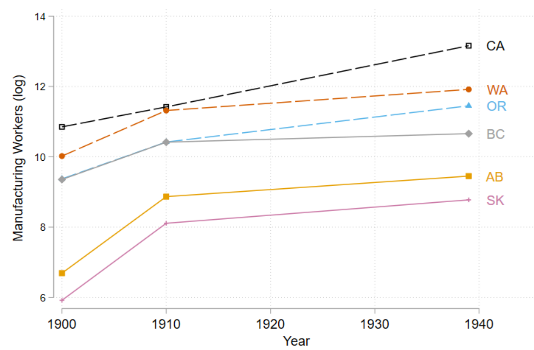
(b) Manuf. Employment by Province



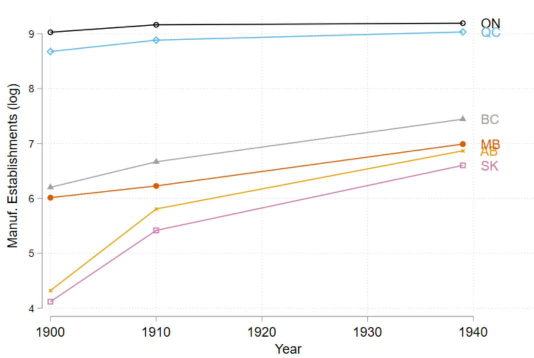
(c) Population, Canada and US West Coast



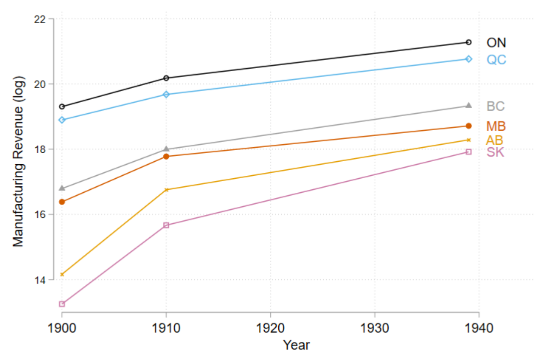
(d) Manuf. Employment, Canada and US West Coast



(e) Manuf. Establishments by Province



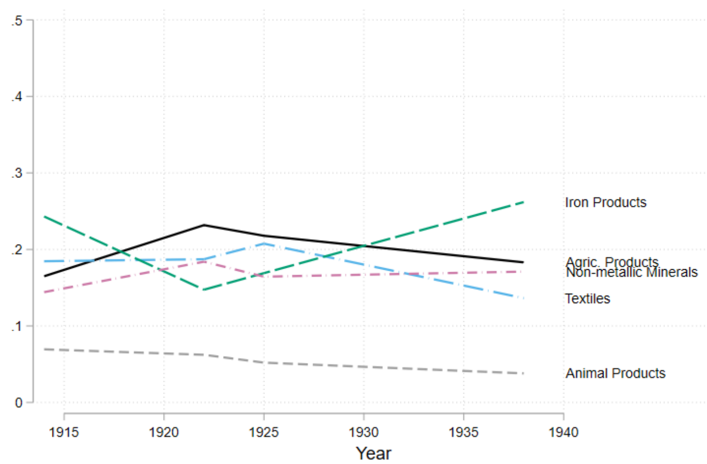
(f) Manuf. Revenue by Province



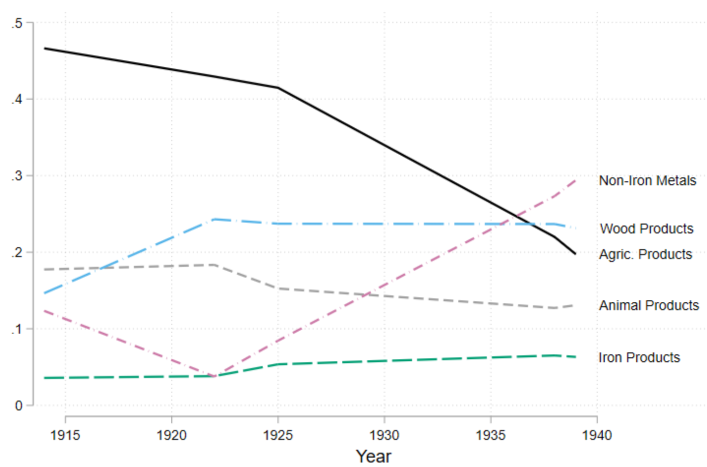
Note: Figures show population, manufacturing employment, establishments, and revenue for selected Canadian provinces (Ontario (ON), Quebec (QC), Saskatchewan (SK), Alberta (AB), Manitoba (MB), and British Columbia (BC)). Selected provinces include 90.1% of total population in 1939. Figures 1c and 1d include selected Canadian provinces (BC, AB, SK) and selected US states (California (CA), Oregon (OR), and Washington (WA)). Data from Canadian Population Census (1901, 1911, 1941), Canadian Manufacturing Census (1901, 1911, 1939) and United States Decennial Census (1900, 1910, 1940).

Figure 2: Industry Composition of Imports and Exports

(a) Share of Imports



(b) Share of Exports



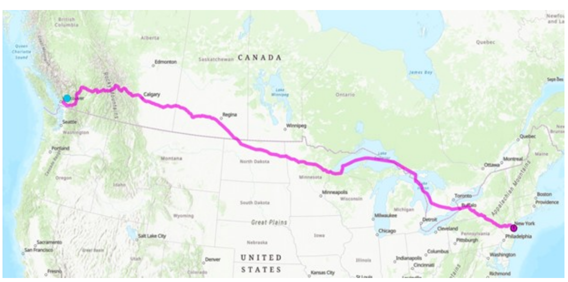
Note: Figures show the industry shares of total imports (a) and of total exports (b). Each graph shows industry shares for the top 5 industries in 1914. Top-5 industries by import share account for 73% of imports in (a). Industries shown on Figure (b) account for 92% of total exports . Data from the Condensed Preliminary Report on the Trade of Canada (1914, 1922, 1924, 1939).

Figure 3: Minimum Cost Path Illustrations With and Without the Panama Canal

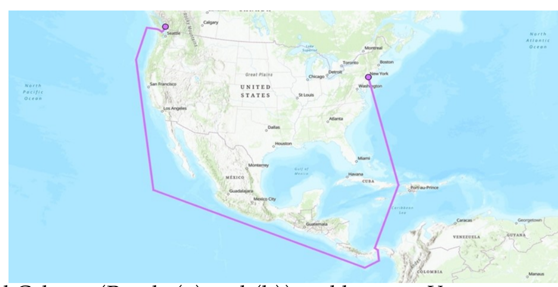
(a) Vancouver, British Columbia, to Calgary, Alberta - Without Canal

(b) Vancouver, British Columbia, to Calgary, Alberta - With Canal

(c) Vancouver, British Columbia, to New York City, NY - Without Canal

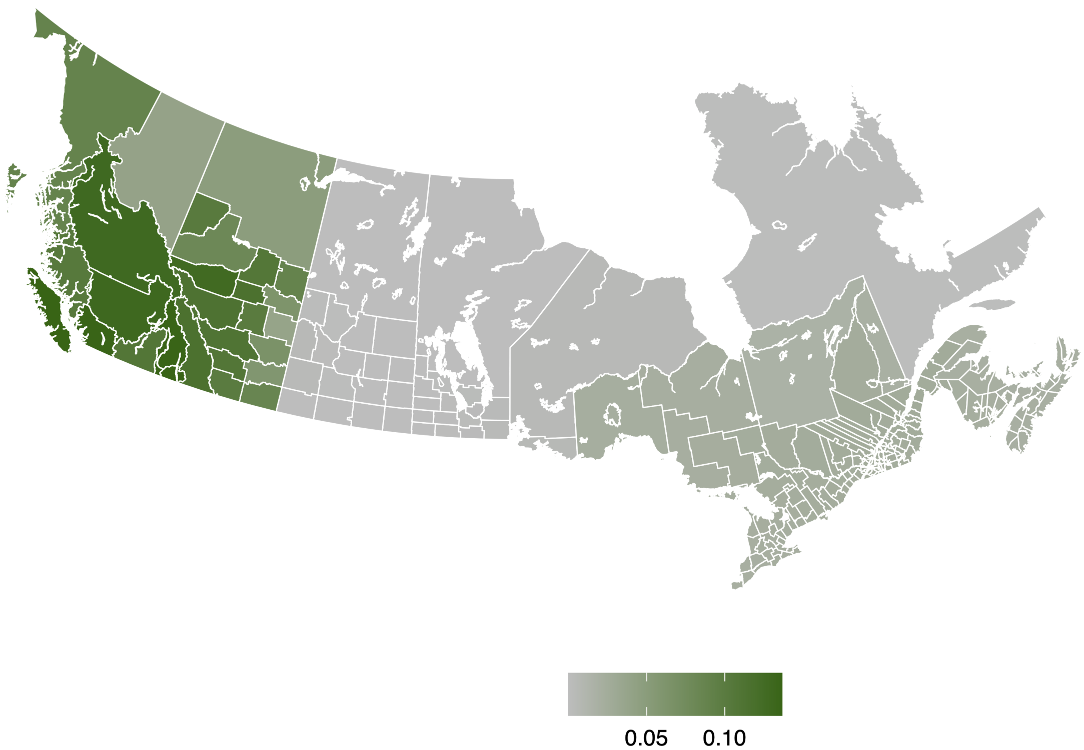


(d) Vancouver, British Columbia, to New York City, NY - With Canal



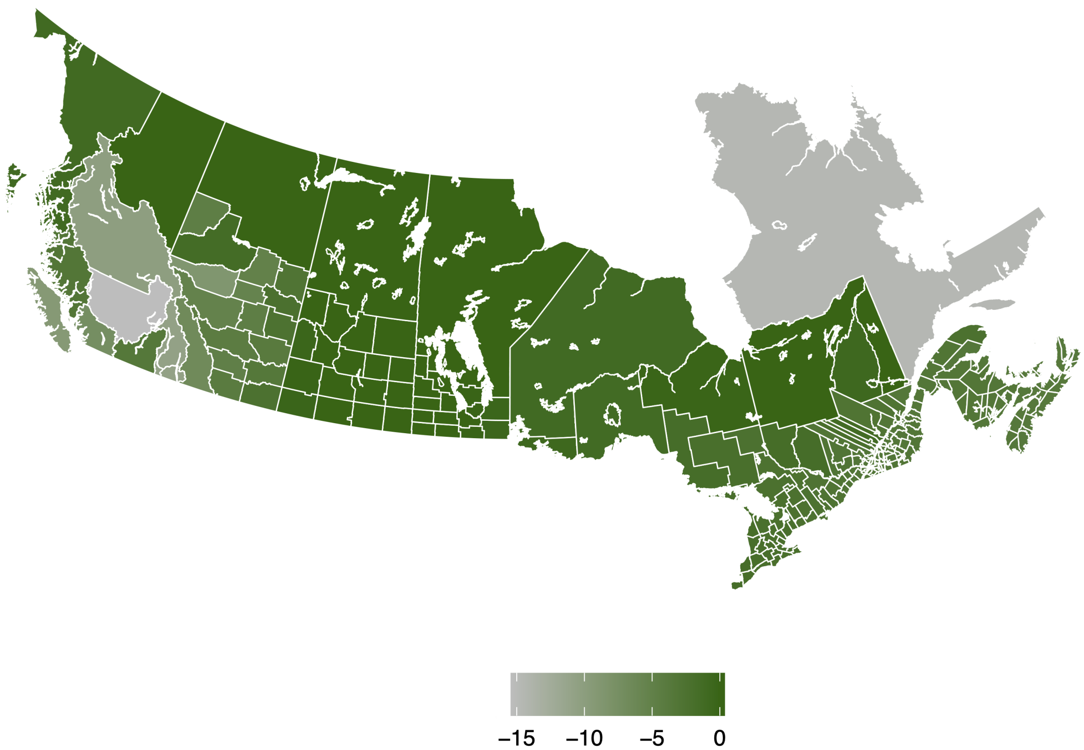
Note: The figure shows the minimum cost path between Vancouver and Calgary (Panels (a) and (b)) and between Vancouver and New York City (Panels (c) and (d)). When the Panama Canal is open, the path changes only between Vancouver and New York City.

Figure 4: Changes in Log Market Access Due to the Opening of the Panama Canal in 1920



Note: The map shows the difference in log market access in 1920 between a scenario where the Panama Canal is open and another where it is closed for all Canadian counties.

Figure 5: Percentage-Point Changes in Population in a Counterfactual Scenario Where the Panama Canal Closes in 1939



The figure shows the change in total population between the actual 1939 data and a counterfactual scenario where the Panama Canal is closed. Details about the counterfactual scenario and the calculations are given in Section 5.

Table 1: Trade Flows Through the Panama Canal in 1927 (in Tons.)

	Eastbound Trade			
Route	Lumber	Wheat	Other	Total
West Coast of Canada to US East Coast	472,637	1,452	7,600	481,689
West Coast of Canada to Canada East Coast	100,178		9,105	109,283
West Coast of Canada to Europe	107,490	1,876,105	191,214	2,174,809
Total	680,305	1,877,557	207,919	2,765,781

	Westbound Trade	
Route	Main Products (Tons in parenthesis)	Total
East Coast of the US to Canada West Coast	Iron and Steel (12,345), Sulphur (19,412)	44,420
East Coast of Canada to Canada West Coast		
Europe to Canada West Coast	Glass and glassware (36,030), Iron and Steel (35,721), Liquors (12,184)	148,171
East Coast of Canada to Australasia	Automobiles (18,568), Iron and Steel (17,847), Paper (31,691)	108,622

Note: Data from the Panama Canal Records.

Table 2: Composition of Imports and Exports by Type of Goods

Period	% of Imports			% of Exports		
	Raw Materials	Partly Manufactured Goods	Fully Manufactured Goods	Raw Materials	Partly Manufactured Goods	Fully Manufactured Goods
Average 1910-1914	23.00	9.65	67.35	57.20	13.10	29.70
Average 1920-1929	26.32	9.84	63.84	44.80	15.84	39.36
Average 1930-1939	27.98	8.96	63.07	33.13	27.29	39.60

Note: Data from the Condensed Preliminary Report on the Trade of Canada (1924-25, 1938-39).

Table 3: Changes in Log Market Access Induced by the Panama Canal

Province	Counties	Average	Std. Dev.	Min	Max
Alberta	17	0.089	0.026	0.039	0.129
British Columbia	10	0.112	0.03	0.04	0.137
Manitoba	16	0.001	0.001	0.000	0.003
New Brunswick	15	0.021	0.001	0.019	0.025
Nova Scotia	18	0.020	0.000	0.019	0.02
Ontario	54	0.022	0.003	0.004	0.024
Quebec	66	0.024	0.003	0.000	0.026
Prince Edward	3	0.019	0.000	0.019	0.02
Saskatchewan	18	0.001	0.001	0.000	0.004
Canada	217	0.028	0.029	0.000	0.137

Note: The table gives descriptive statistics for the difference in market access in 1920 when the Canal is open and closed for all Canadian provinces.

Table 4: Structural Transformation

Outcomes: y_c	(1) Pop.	(2) Manuf. Empl. Pop. Share	(3) N Estab.	(4) Avg. Estab. size Labor	(5) Avg. Estab. size Revenue	(6) K Share	(7) Skill Ratio	(8) Ratio Salary to Wage
				Panel A: Δy_c 1939-1910				
$\Delta Ln(MA_{1920})$	0.033***	0.058**	0.066***	0.025	0.022	-0.019	-0.001	0.012
	(0.008)	(0.026)	(0.015)	(0.020)	(0.029)	(0.013)	(0.003)	(0.010)
Observations	217	217	217	217	217	217	209	208
r2	0.471	0.372	0.552	0.419	0.228	0.222	0.332	0.114
				Panel B: Pre-Trends Δy_c 1910-1900				
$\Delta Ln(MA_{1920})$	-0.007	0.002	-0.007	0.002	0.007	-0.003	0.000	-0.003
	(0.016)	(0.025)	(0.026)	(0.009)	(0.011)	(0.008)	(0.000)	(0.006)
Observations	217	217	217	217	217	217	217	217
r2	0.678	0.196	0.522	0.336	0.224	0.261	0.201	0.313

Note: Coefficients are standardized for moving from the 25th to the 75th percentile in terms of gains in market access in 1920. Outcomes are, in order: Population, manufacturing employment as share of population, number of establishments, average establishment size (measured as employment per establishment and revenue per establishment), capital expenditure as share of revenue, salaried workers as share of employment, and the ratio between average salary and average hourly wage. Dependent variables are in log-changes between 1910 and 1939 (Panel A) and between 1900 and 1910 (Panel B). All specifications include province fixed effects and control for a quadratic polynomial in latitude and longitude, for a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River, and for log population (in 1910 for regressions in Panel A, in 1900 for regressions in Panel B). Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Manufacturing Outcomes

Manuf. Outcomes: y_c	(1) Value of Products	(2) Capital Exp.	(3) Employment	(4) Materials	(5) Productivity	(6) *Decomposition* TFPR	(7) *Decomposition* AE
				Panel A: Δy_c 1939-1910			
$\Delta Ln(MA_{1920})$	0.089**	0.070*	0.091***	0.084**	0.131**	0.009	0.122*
	(0.035)	(0.037)	(0.027)	(0.037)	(0.053)	(0.022)	(0.066)
Observations	217	217	217	217	217	217	217
r2	0.236	0.248	0.340	0.240	0.200	0.128	0.230
				Panel B: Pre-Trends Δy_c 1910-1900			
$\Delta Ln(MA_{1920})$	-0.000	-0.003	-0.005	0.015	0.005	-0.042*	0.047
	(0.029)	(0.028)	(0.027)	(0.031)	(0.050)	(0.022)	(0.044)
Observations	217	217	217	217	217	217	217
r2	0.445	0.502	0.529	0.439	0.137	0.328	0.378

Note: Coefficients are standardized for moving from the 25th to the 75th percentile in terms of gains in market access in 1920. Outcomes from Manufacturing Census data are, in order: Value of production, capital expenditures, employment, materials, and productivity. Columns (6) and (7) use outcomes from a productivity decomposition based on production function estimates at the province level. Outcomes are, respectively, an estimate of Total Factor Revenue Productivity and an estimate of Allocative Efficiency. Dependent variables are in log-changes between 1910 and 1939 (Panel A) and between 1900 and 1910 (Panel B). All specifications include province fixed effects and control for a quadratic polynomial in latitude and longitude, for a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River, and for log population (in 1910 for regressions in Panel A, in 1900 for regressions in Panel B). Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Potential Confounders of the Panama Canal MA Shock

Dependent Var: Δy_c	(1) Manuf. Empl. Pop. Share	(2) N Estab.	(3) Estab. Size Labor	(4) Value of Products	(5) Capital Exp.	(6) Employment	(7) Materials	(8) Productivity
Panel A: Improvements in Railroad Network								
$\Delta Ln(MA_{1920})$	0.066**	0.064***	0.028	0.091**	0.071*	0.092***	0.085**	0.137**
	(0.025)	(0.016)	(0.019)	(0.034)	(0.036)	(0.027)	(0.037)	(0.054)
New RR Km, 1910-1920	-0.051	0.051	-0.027	0.004	0.010	0.024	0.024	-0.041
	(0.056)	(0.036)	(0.043)	(0.072)	(0.094)	(0.062)	(0.069)	(0.110)
New RR Km, 1920-1940	-0.136*	-0.002	-0.043	-0.047	-0.039	-0.045	-0.045	-0.090
	(0.072)	(0.047)	(0.048)	(0.091)	(0.085)	(0.068)	(0.099)	(0.172)
Observations	217	217	217	217	217	217	217	217
r2	0.393	0.556	0.423	0.238	0.249	0.342	0.242	0.203
Panel B: Differential Effect of $\Delta Ln(MA_{1920})$ by Initial MA Level								
$\Delta Ln(MA_{1920})$	0.053**	0.066***	0.022	0.090***	0.068*	0.089***	0.086***	0.131***
	(0.024)	(0.014)	(0.021)	(0.030)	(0.037)	(0.025)	(0.032)	(0.045)
$\Delta Ln(MA_{1920}) \times MA_{1910}$	-0.013	-0.011	-0.005	-0.036*	-0.012	-0.015	-0.041*	-0.054*
	(0.013)	(0.010)	(0.010)	(0.020)	(0.020)	(0.013)	(0.021)	(0.028)
MA_{1910}	0.232***	0.068	0.098**	0.182*	0.120	0.165**	0.163	0.312**
	(0.083)	(0.049)	(0.045)	(0.099)	(0.094)	(0.068)	(0.102)	(0.143)
Observations	217	217	217	217	217	217	217	217
r2	0.419	0.560	0.432	0.266	0.256	0.365	0.268	0.239
Panel C: Within West Coast: $\Delta Ln(MA_{1920})$ Effect Gradient on Latitude and Longitude								
$\Delta Ln(MA_{1920})$	0.056*	0.064**	0.009	0.090***	0.076*	0.073**	0.090***	0.097
	(0.028)	(0.023)	(0.032)	(0.029)	(0.038)	(0.032)	(0.030)	(0.065)
$\Delta Ln(MA_{1920}) \times$ Long.	0.138***	0.070	0.051	0.191***	0.161**	0.120***	0.202***	0.259***
	(0.036)	(0.051)	(0.047)	(0.027)	(0.060)	(0.033)	(0.029)	(0.070)
$\Delta Ln(MA_{1920}) \times$ Lat.	0.104***	0.055	0.037	0.145***	0.126***	0.092***	0.151***	0.233***
	(0.025)	(0.043)	(0.035)	(0.023)	(0.037)	(0.029)	(0.026)	(0.041)
Observations	61	61	61	61	61	61	61	61
r2	0.644	0.594	0.407	0.615	0.516	0.584	0.633	0.529
Panel D: Forest Area Share, 1900-1910 Average (Continous)								
$\Delta Ln(MA_{1920})$	0.098***	0.089***	0.027	0.132**	0.091*	0.116***	0.131**	0.175**
	(0.035)	(0.026)	(0.021)	(0.052)	(0.045)	(0.038)	(0.054)	(0.068)
$\Delta Ln(MA_{1920}) \times$	-0.002	-0.001	-0.000	-0.003	-0.001	-0.002	-0.003	-0.003
. . . Forest Area %	(0.002)	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)
Forest Area %	-0.006	-0.005*	-0.000	-0.006	-0.003	-0.006	-0.008	0.002
	(0.004)	(0.003)	(0.005)	(0.007)	(0.008)	(0.005)	(0.007)	(0.008)
Observations	217	217	217	217	217	217	217	217
r2	0.385	0.561	0.419	0.247	0.250	0.347	0.253	0.204

Note: Coefficients are standardized for moving from the 25th to the 75th percentile of $\Delta Ln(MA\_1920)$. Dependent variables for specifications in each column are the change in outcomes between 1910 and 1939 detailed in the first row. Panel A controls for the construction of new railroad kilometers. Panel B controls for 1910 Market Access level and the interaction between 1910 MA and $\Delta Ln(MA_{1920})$. Panel C explores the heterogeneity of the effect of the Panama Canal shock within the West Coast. The sample is all counties west of 90 degrees West. Specifications in Panel C include interactions between latitude, longitude, and $\Delta Ln(MA_{1920})$. Panel D explores the differential effect of the Panama Canal MA shock by intensity of the lumber industry in 1910. Specifications in Panel D include an interaction between $\Delta Ln(MA_{1920})$ and a dummy equal to one if a county is above the 1910 mean of lumber industry as a share of total manufacturing revenue. All specifications include province fixed effects and control for a quadratic polynomial in latitude and longitude, for a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River, and for log population in 1910. Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Land Values

	(1)	(2)
Dep. Variable:	Log Average Land Values	
year=1901 × $\Delta Ln(MA_{1920})$	0.004	0.003
	(0.016)	(0.009)
year=1911 × $\Delta Ln(MA_{1920})$	0.012	0.012
	(0.016)	(0.007)
year=1941 × $\Delta Ln(MA_{1920})$	0.032**	0.033***
	(0.015)	(0.011)
Observations	868	868
Dep. Var. Mean	3.095	
Pop. x Year FE		x
Coord. Poly.		x
Prov. x Year FE		x
r2	0.832	0.903

Note: Coefficients are standardized for moving from the 25th to the 75th percentile in terms of gains in market access in 1920, with and without the Canal. 1921 is the omitted category. All specifications include county and year fixed effects. Column 2 controls for population in 1911 interacted with year dummies, a quadratic polynomial in latitude and longitude, and a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River, interacted with year dummies, and province x year fixed effects. Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: The Panama Canal Shock in Perspective

Country (Sector)	Scenario	GDP Change	Population Loss (With Fixed Real Wages)	Source
Canada	Close the Panama Canal in 1939	-1.86%	2.7%	Galiani, Jaramillo, and Uribe-Castro (2022)
US	Close the Panama Canal in 1940	-0.2%		Maurer and Rausch (2022)
US	Social savings from the Canal	0.12%		Maurer and Yu (2008)
US (Agriculture)	Remove RR in 1890	-3.2%	58%	Donaldson and Hornbeck (2016)
US (Manufacturing)	Remove RR in 1890	-28%	68%	Hornbeck and Rotemberg (2022)
	Fix RR network in 1880	-2.7%		
Argentina	Remove RR in 1914	-11.8%	8.6%	Fajgelbaum and Redding (2022)
	Pre-steamship freight rates in 1914	-15.6%	12.1%	
Canada	Remove international borders in 1900	32%		Liu and Meissner (2015)
		Social Savings Rates		
Country	Project	Rate		Source
United States	Panama Canal Average: 1921-1937	10.9%		Maurer and Yu (2008)
Canada	Grand Western Railroad	6.1%		Carlos and Lewis (1992)
	Grand Trunk Railroad	2.8%		

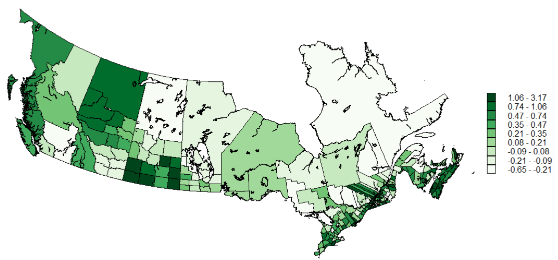
Note: The table compares the results of different counterfactual estimates of the effect of changes in trade costs on GDP and population. Results from different sources may come from models with different fundamentals. Our model is closest to that of Hornbeck and Rotemberg (2021). For the Maurer and Yu (2008) figure, we use their social savings estimate for 1926 (Table 9) and nominal GDP from the NBER Macrohistory Database.

Online Appendixes

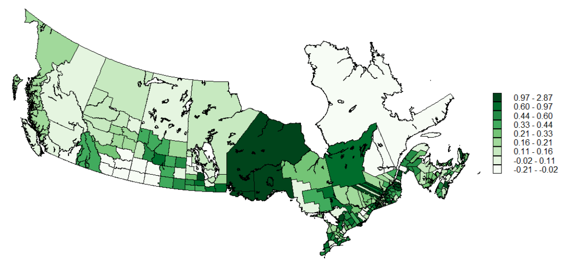
Appendix A Supporting Figures and Tables

Figure A.1: Geographic Distribution of Wedges

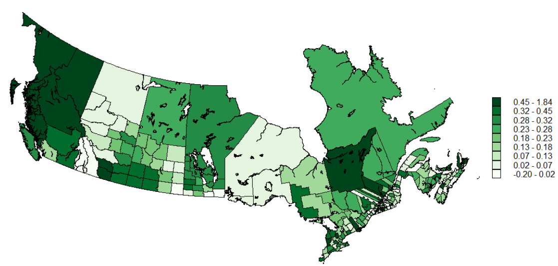
(a) Capital



(b) Labor



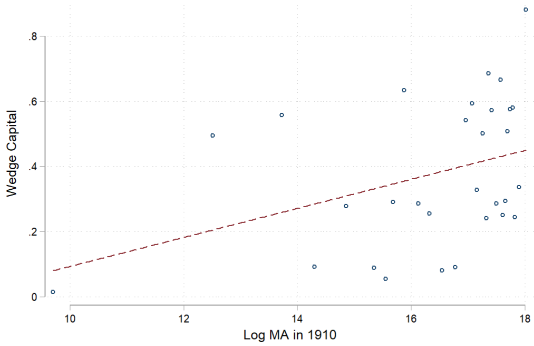
(c) Materials



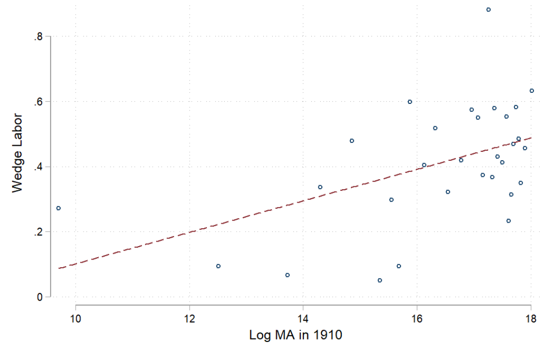
Note: The figures show the geographic distribution of input wedges. For each input and each county, we estimate the difference between the output elasticity ($\alpha_{k,c}$) and the revenue share ($s_{k,c}$). The output elasticity is defined as the cost share of input k in province p. See Section 3 for more details.

Figure A.2: Correlation Between Wedges and Market Access in 1910

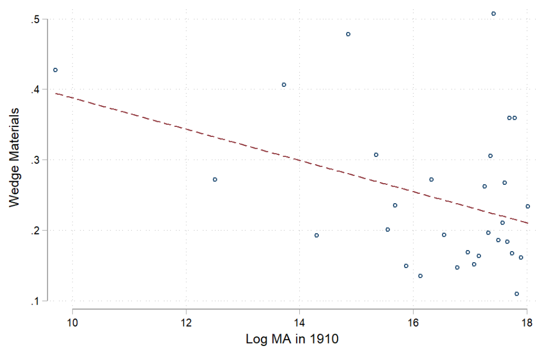
(a) Capital



(b) Labor



(c) Materials



Note: The figures show the binscatter plot for input wedges over the levels of estimated (log) Market Access in 1910. For each input and each county, we estimate wedges as the difference between the output elasticity ($\alpha_{k,c}$) and the revenue share ($s_{k,c}$). The output elasticity is defined as the cost share of input k in county c. See Section 3 for more details.

Table A.1: Change in MA induced by the Panama Canal as a Function of 1910 MA

	(1)	(2)	(3)	(4)
Dep. Var.:		$\Delta Ln(MA_{1920})$		
$Ln(MA_{1910})$	0.00069	0.00069	0.00085	0.00077
	(0.00367)	(0.00110)	(0.00113)	(0.00109)
Observations	217	217	217	217
Mean Dep. Var.	0.02165	0.02165	0.02165	0.02165
Province FE		X	X	X
Population 1900			X	
Geo Controls				X
r2	0.00056	0.88963	0.89298	0.91863

Note: The dependent variable is the change in log market access in 1920 induced by the Panama Canal. Coefficients are standardized in terms of standard deviations of the level of log market access in 1910. Geographic controls include a quadratic polynomial in latitude and longitude and a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River. Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Market Access and Railroad Construction

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln(New Km of Railroad built Between...)					
	1900-1915	1900-1920	1905-1915	1905-1920	1910-1915	1910-1920
$\Delta Ln(MA_{1920})$	0.103	0.116	0.141	0.154	0.178*	0.177*
	(0.093)	(0.094)	(0.103)	(0.104)	(0.095)	(0.097)
Observations	217	217	217	217	217	217
Mean Dep. Var.	9.493	9.843	8.688	9.174	6.533	7.225
r2	0.290	0.276	0.362	0.330	0.404	0.330

Note: Coefficients are standardized for moving from the 25th to the 75th percentile in terms of gains in market access in 1920, with and without the Canal. All specifications include province fixed effects and control for a quadratic polynomial in latitude and longitude, for a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River, and for log population in 1900. Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

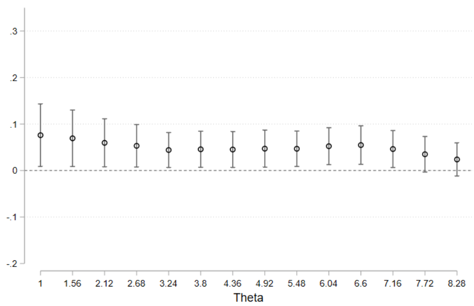
Table A.3: Manufacturing Subsectors, 1910

Subsector	Pctg. of Manufacturing Employment			Pctg. of Value of Production		
	Average	Std. Dev.	Max	Average	Std. Dev.	Max
Chemical products	0.10	0.92	10.85	0.09	0.67	7.44
Clothing	1.20	3.17	22.19	0.83	2.62	24.33
Electric light and power	0.23	0.64	4.06	0.29	1.01	11.18
Electrical apparatus and supplies	0.05	0.55	7.83	0.03	0.32	3.88
Food and beverage	14.34	18.27	88.97	25.24	21.05	91.33
Iron and steel products	2.23	5.24	45.50	2.18	5.95	52.58
Leather and leather products	0.77	4.03	34.67	1.23	6.79	67.74
Miscellaneous industries	0.18	0.80	7.80	0.15	0.70	6.86
Non-ferrous metal products	0.21	1.85	19.24	0.53	5.22	54.51
Non-metallic mineral products	2.88	6.89	35.09	1.23	3.16	18.99
Paper products	1.00	5.53	43.37	1.10	6.13	46.61
Petroleum and coal	0.04	0.43	4.44	0.05	0.53	5.49
Printing and publishing	1.52	3.45	20.73	0.87	1.94	12.94
Rubber and products	0.03	0.42	6.20	0.01	0.11	1.57
Textiles other than clothing)	0.65	4.04	54.43	0.42	2.91	40.14
Tobacco and tobacco products	0.31	1.71	16.49	0.24	1.29	13.57
Transportation equipment	2.24	6.66	62.11	1.37	3.97	26.71
Wood products	31.48	26.51	94.30	24.03	23.69	97.43
All Other Industries	40.54	26.56	100.00	40.10	26.62	100.00

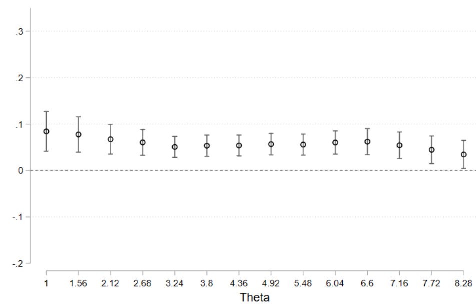
Note: The data comes from the 1911 Census of Manufactures. The table shows the county-level average share of manufacturing employment and value for different subsectors. For the average county, 40% of employment and manufacturing revenues were classified as “All Other Industries.”

Figure A.3: Robustness to Different Levels of θ for MA Calculations

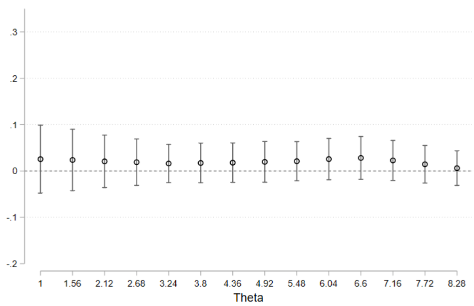
(a) Employment (Sh. Pop.)



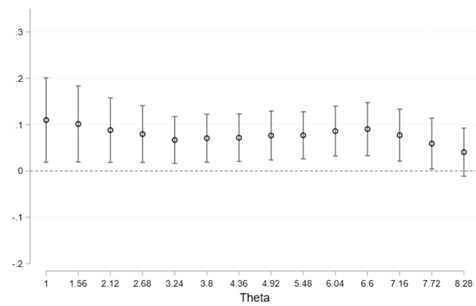
(b) N. Establishments



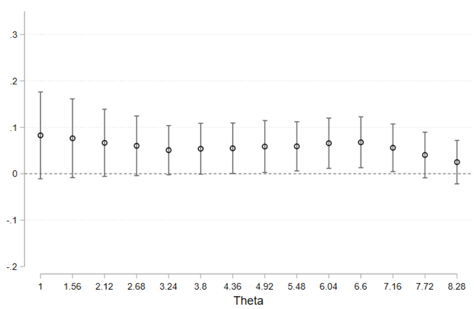
(c) Avg. Estab Size (Revenue)



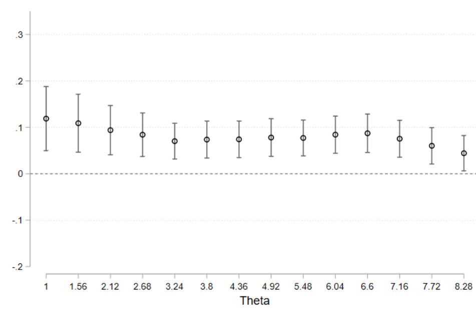
(d) Value of Products



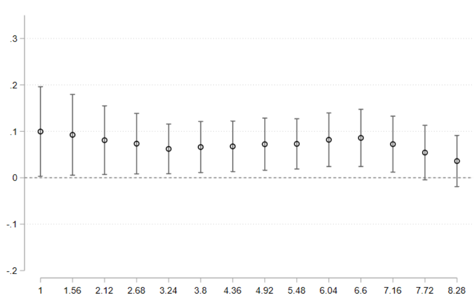
(e) Capital Expenditures



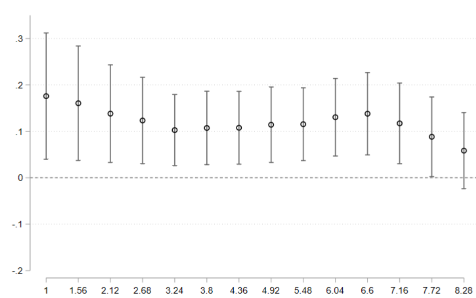
(f) Employment



(g) Materials



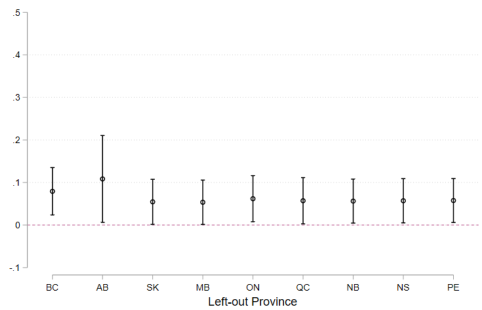
(h) Productivity



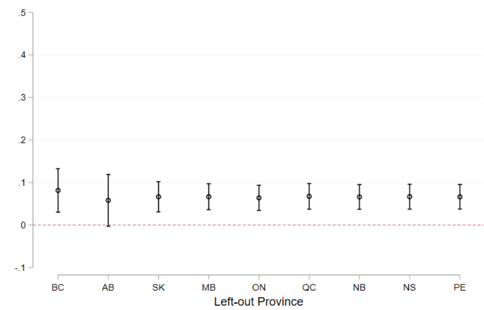
Note: Each subfigure plots the β from Equation 9 for different estimates of Market Access gains. We estimate $\Delta ln(MA_c)$ for different levels of θ, the elasticity of trade to trade costs. Each coefficient on each subfigure comes from a separate linear regression estimation. 90% confidence intervals.

Figure A.4: Robustness to Removing One Province at a Time

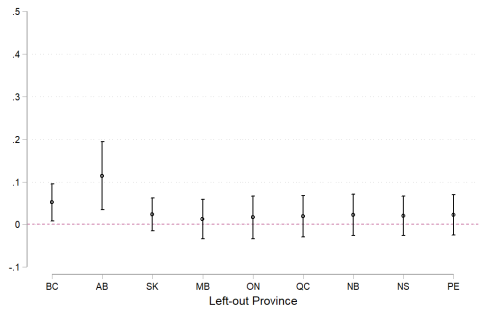
(a) Employment (Sh. Pop.)



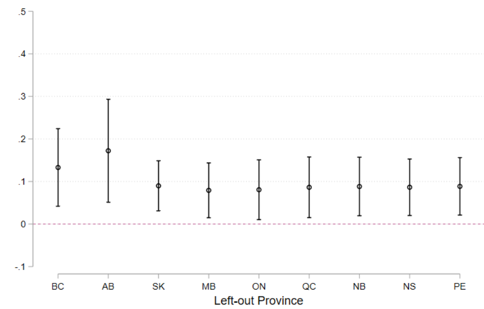
(b) N. Establishments



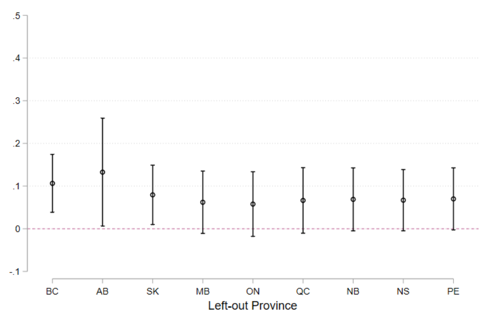
(c) Avg. Estab Size (Revenue)



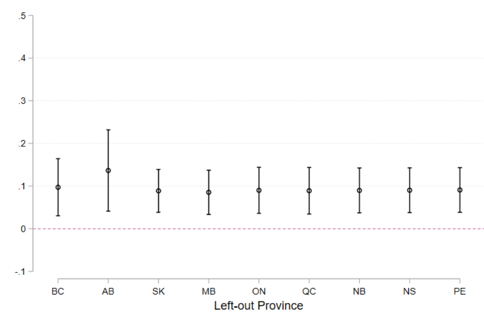
(d) Value of Products



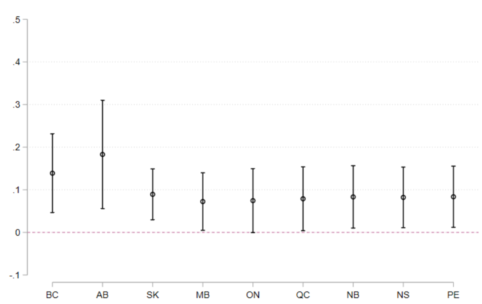
(e) Capital Expenditures



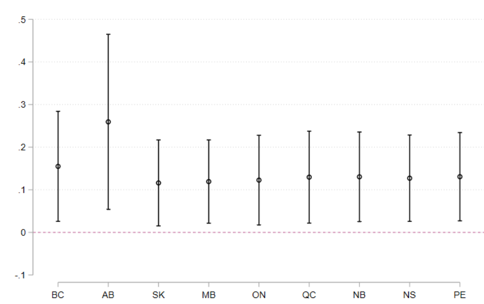
(f) Employment



(g) Materials



(h) Productivity



Note: Each subfigure plots the β from Equation 9 for different samples. We estimate the effect of $\Delta ln(MA_{1920})$ on the change in outcomes detailed in subfigures titles, removing one province at a time. For instance, the first coefficient in all graphs comes from a sample that excludes British Columbia, the second one comes from a sample without Alberta, and so on. Coefficients are organized from left to right based on each Province's west to east location. They are: British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, Quebec, New Brunswick, Nova Scotia, and Price Edwards. Each coefficient on each subfigure comes from a separate linear regression estimation. 90% confidence intervals.

Table A.4: Industry Details

Industry	(1)	(2) Revenue (% total)	(3)	(4)	(5) N. Estab.	(6)	(7)	(8) Employment (000s)	(9)	(10)	(11) Avg. Estab. Size (labor)	(12)	(13)	(14) Capital to Revenue Ratio	(15)
Year:	1901	1911	1939	1901	1911	1939	1901	1911	1939	1901	1911	1939	1901	1911	1939
Food and beverages	26.0%	21.1%	21.0%	5,594	6,985	8,465	42.40	52.73	96.15	7.6	7.5	11.4	0.46	0.54	0.56
Wood products	16.7%	15.8%	17.5%	3,034	4,999	4,319	75.70	110.05	137.36	25.0	22.0	31.8	1.12	1.41	0.90
Textiles	14.1%	11.7%	7.4%	1,684	1,444	1,200	64.19	72.67	83.73	38.1	137.9	69.8	0.89	0.80	0.80
Iron and steel products	7.3%	9.7%	3.6%	517	824	286	24.77	48.56	26.08	47.9	58.9	91.2	1.17	1.09	1.47
Leather and leather products	7.2%	5.4%	1.9%	431	399	306	19.20	22.74	21.27	44.6	57.0	69.5	0.62	0.78	0.85
Paper and printing	4.3%	4.0%	9.7%	592	773	2,357	15.41	22.89	69.57	26.0	29.6	29.5	1.30	1.35	2.13
Vehicles for land transportation	4.2%	6.0%	6.0%	425	465	147	14.87	35.78	40.12	35.0	76.9	272.9	0.80	0.71	0.89
Metals and metal products other than steel	4.1%	6.3%	13.5%	363	341	699	9.36	17.50	56.04	25.8	51.3	80.2	1.04	0.92	0.90
Tobacco and its manufactures	2.5%	2.2%	1.4%	160	173	80	6.33	8.76	8.16	39.6	50.7	102.0	0.61	0.86	1.28
Chemicals and allied products	2.4%	2.4%	6.4%	128	178	498	2.87	5.27	20.95	22.4	29.6	42.1	0.90	0.97	1.07
Liquors and beverages	1.9%	2.5%	1.3%	183	260	61	3.21	4.69	5.35	17.5	18.0	87.6	2.23	1.49	1.41
Clay, glass, and stone products	1.5%	2.2%		855	771		10.77	17.70		12.6	23.0		1.19	1.78	
Vessels for water transportation	0.4%	0.6%		57	172		2.59	4.41		45.4	25.7		1.61	1.57	
Miscellaneous industries	0.1%	9.0%	22.3%	45	1,011	6,331	0.61	38.54	173.58	13.4	38.1	27.4	1.02	2.25	1.19
Total	100%	100%	100%	14,650	19,218	24,803	313.34	471.13	658.11	21.39	24.51	26.5	0.93	1.07	1.05

Note: Definition of industries varies by Census waves, figures on this table should be treated with caution. Specifically, industry definitions may vary over time and the category "Miscellaneous industries" contains different sub-industries for all Census. Table shows descriptive statistics for different manufacturing industries, aggregated at the national level. Data from 1901, 1911, and 1939 Manufacturing Census. Table shows: industry revenue as a share of total manufacturing revenue, number of establishments, number of employees, average establishment size (workers divided by number of employees), and the capital to revenue ratio.

Table A.5: Sectoral Concentration and Input Wedges

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.:	Capital Wedge		Labor Wedge		Materials Wedge	
Total Revenue (log)	0.0508	0.0435	0.0922***	0.1354***	-0.0104	-0.0144
	(0.0323)	(0.0307)	(0.0240)	(0.0204)	(0.0158)	(0.0135)
Avg. Establishment Size (Revenue, log)	-0.1973***		-0.1975***		0.0798***	
	(0.0664)		(0.0575)		(0.0268)	
Avg. Establishment Size (Employment, log)		-0.3152***		-0.4950***		0.1623***
		(0.0858)		(0.0698)		(0.0357)
Industry HHI (Revenue)	0.8167***		0.7937***		0.0442	
	(0.2990)		(0.2499)		(0.1594)	
Industry HHI (Employment)		0.1801		0.2995**		0.2117*
		(0.2491)		(0.1458)		(0.1133)
Counties (N)	217	217	217	217	217	217
Mean Dep. Variable	0.0164	0.0164	0.0463	0.0463	0.1171	0.1171
r2	0.0098	0.0246	0.0754	0.3092	0.0437	0.1669

Note: Data for 1910. The dependent variables are the input wedges for capital (columns 1 and 2), labor (columns 3 and 4), and materials (columns 5 and 6). Wedges are defined as the difference between province-level input cost share and county-level revenue share, as explained in Section 3. Sectoral HHI indexes are calculated using industry-level revenue and employment data. Robust standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Input Wedges and Market Access

	(1)	(2)	(3)	(4)
Dep. Var.:	Input Wedge (Province-level cost share - county-level revenue share)			
Panel A: Capital				
$Ln(MA_{1910})$	0.00294	0.00487**		
	(0.00185)	(0.00242)		
RR Density, 1910 (log)			0.00385*	0.00696***
			(0.00229)	(0.00258)
Observations	217	217	217	217
Mean Dep. Var.	0.01290	0.01290	0.01290	0.01290
Province FE		X		X
R2	0.02595	0.14818	0.02431	0.16179
Panel B: Labor				
$Ln(MA_{1910})$	0.00601***	0.00210		
	(0.00191)	(0.00251)		
RR Density, 1910 (log)			0.00969***	0.00565*
			(0.00265)	(0.00333)
Observations	217	217	217	217
Mean Dep. Var.	0.05846	0.05846	0.05846	0.05846
Province FE		X		X
R2	0.04395	0.11912	0.06111	0.13079
Panel C: Materials				
$Ln(MA_{1910})$	-0.00677*	-0.00669		
	(0.00403)	(0.00499)		
RR Density, 1910 (log)			-0.00550	-0.00511
			(0.00582)	(0.00638)
Observations	217	217	217	217
Mean Dep. Var.	0.10847	0.10847	0.10847	0.10847
Province FE		X		X
R2	0.02648	0.04479	0.01540	0.03809

Note: The dependent variables are the input wedges for capital, labor, and materials. Wedges are defined as the difference between cost share and revenue share, as explained in Section 3. All specifications control for population in 1910. $Ln(MA_{1910})$ is the level of Market Access in 1910. RR Density 1910 is the meters of railroads per square kilometre. Robust standard errors are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Productivity decomposition

	(1)	(2)	(3)	(4)	(5)
Output elasticity estimate:		Province level		National level	
Outcome (y_c)	Productivity	TFPR	AE	TFPR	AE
		Panel A: Δy_c 1939-1910			
$\Delta Ln(MA_{1920})$	0.131**	0.009	0.122*	0.016	0.115*
	(0.053)	(0.022)	(0.066)	(0.020)	(0.065)
Observations	217	217	217	217	217
r2	0.200	0.128	0.230	0.127	0.244
		Panel B: Pre-Trends Δy_c 1910-1900			
$\Delta Ln(MA_{1920})$	0.005	-0.042*	0.047	-0.054**	0.058
	(0.050)	(0.022)	(0.044)	(0.024)	(0.046)
Observations	217	217	217	217	217
r2	0.137	0.328	0.378	0.353	0.375

Note: Coefficients are standardized for moving from the 25th to the 75th percentile in terms of gains in market access in 1920. Outcomes are: Productivity and estimates for Total Factor Revenue Productivity and Allocative Efficiency. Outcomes in columns (2) and (3) use output elasticities estimated at the province level to estimate the productivity decomposition. Outcomes in columns (4) and (5) use output elasticities obtained from national level data. Dependent variables are in log-changes between 1910 and 1939 (Panel A) and between 1900 and 1910 (Panel B). All specifications include province fixed effects and control for a quadratic polynomial in latitude and longitude, for a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River, and for log population (in 1910 for regressions in Panel A, in 1900 for regressions in Panel B). Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Effect of MA Gains on Input Wedges

	(1)	(2)	(3)	(4)	(5)	(6)
Output elasticity estimate (α^k):		Province level			National level	
Dependent Variable			$\Delta_{1939-1910}$ Wedge for...			
Input	Capital	Labor	Materials	Capital	Labor	Materials
$\Delta Ln(MA_{1920})$	0.002	-0.003	0.001	0.002	-0.003	0.001
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Observations	217	217	217	217	217	217
Mean Dep. Var	-0.002	-0.017	-0.022	0.002	0.009	-0.051
r2	0.184	0.160	0.172	0.178	0.410	0.237

Note: Coefficients are standardized for moving from the 25th to the 75th percentile in terms of gains in market access in 1920, with and without the Canal. Dependent variables are the change between 1910 and 1939 of the wedges for each of production input, respectively: capital, labor and materials. Wedges are defined as the difference between the output elasticity α and the revenue share of each input. Columns (1), (2), and (3) estimate α using province level input cost shares. Columns (4), (5), and (6) estimate α using national level input cost shares. All specifications include county and year fixed effects and control for a quadratic polynomial in latitude and longitude, for a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River, and for log population. Standard errors clustered at the county level are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Robustness Checks: $\Delta Ln(MA_{1920})$ coefficients for different specifications

Specification:	(1) Main Spec.	(2) No Prov. FE, No Geo. Controls	(3) Geo. Controls	(4) Prov. FE
Dep. Variable: Δy_c				
Pop.	0.033***	0.032**	0.033***	0.033***
	(0.008)	(0.012)	(0.009)	(0.008)
Manuf. Emp. Pop. Share	0.058**	0.059	0.057**	0.059
	(0.026)	(0.042)	(0.028)	(0.038)
N Estab.	0.066***	0.065*	0.066***	0.066**
	(0.015)	(0.033)	(0.022)	(0.025)
Estab. Size (Labor)	0.025	0.025	0.024	0.025
	(0.020)	(0.026)	(0.020)	(0.019)
Estab. Size (Revenue)	0.022	0.023	0.022	0.023
	(0.029)	(0.033)	(0.027)	(0.029)
K Share	-0.019	-0.018	-0.019	-0.019
	(0.013)	(0.015)	(0.014)	(0.014)
Skill Ratio	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.004)	(0.003)	(0.003)
Ratio Salary to Wage	0.012	0.007	0.010	0.009
	(0.010)	(0.013)	(0.012)	(0.011)
Value of Products	0.089**	0.088*	0.088**	0.089**
	(0.035)	(0.045)	(0.036)	(0.043)
Capital Exp.	0.070*	0.070	0.069*	0.070
	(0.037)	(0.048)	(0.040)	(0.045)
Employment	0.091***	0.091**	0.090***	0.091**
	(0.027)	(0.041)	(0.032)	(0.036)
Materials	0.084**	0.083*	0.083**	0.084*
	(0.037)	(0.046)	(0.037)	(0.045)
Productivity	0.131**	0.131*	0.130**	0.132*
	(0.053)	(0.072)	(0.058)	(0.071)
TFPR	0.009	0.009	0.009	0.009
	(0.022)	(0.026)	(0.021)	(0.024)
AE	0.122*	0.122	0.121*	0.123
	(0.066)	(0.087)	(0.069)	(0.086)

Note: Table shows $\Delta Ln(MA_{1920})$ coefficients for different dependent variables (rows) and specifications (columns). Coefficients are standardized for moving from the 25th to the 75th percentile of the independent variables. Dependent variables for specifications in each cell are the change in manufacturing outcomes between 1910 and 1939 detailed in the first column. All specifications control for log total population in 1910. Main specification includes 1) province fixed effects and 2) controls for a quadratic polynomial in latitude and longitude, for a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River. Other specifications are different combinations of 1) and 2). Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Robustness Checks: Different values of MA parameters

	(1) Fixed Pop.	(2) Low θ	(3) High θ	(4) Low $\bar{P}$	(5) High $\bar{P}$
Population	1910	1920	1920	1920	1920
θ	-5	-1	-7.5	-5	-5
$\bar{P}$	9.4	9.4	9.4	5	20
Dep. Variable: Δy_c					
Pop.	0.030***	0.042***	0.033***	0.034***	0.028***
	(0.007)	(0.011)	(0.011)	(0.010)	(0.007)
Manuf. Emp. Pop. Share	0.054**	0.090**	0.059*	0.064**	0.052**
	(0.024)	(0.042)	(0.030)	(0.030)	(0.024)
N Estab.	0.061***	0.094***	0.068***	0.071***	0.058***
	(0.013)	(0.025)	(0.021)	(0.021)	(0.012)
Estab. Size (Labor)	0.023	0.038	0.024	0.027	0.022
	(0.018)	(0.031)	(0.020)	(0.022)	(0.018)
Estab. Size (Revenue)	0.021	0.027	0.023	0.025	0.019
	(0.026)	(0.045)	(0.031)	(0.033)	(0.025)
K Share	-0.018	-0.025	-0.024	-0.022	-0.014
	(0.012)	(0.021)	(0.017)	(0.017)	(0.011)
Skill Ratio	-0.001	-0.002	-0.001	-0.001	-0.001
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Ratio Salary to Wage	0.011	0.025	0.010	0.011	0.014
	(0.009)	(0.015)	(0.010)	(0.011)	(0.008)
Value of Products	0.082**	0.121**	0.091**	0.096**	0.077**
	(0.031)	(0.055)	(0.041)	(0.042)	(0.030)
Capital Exp.	0.064*	0.096	0.068*	0.074*	0.063*
	(0.034)	(0.058)	(0.038)	(0.040)	(0.033)
Employment	0.084***	0.132***	0.092***	0.098***	0.081***
	(0.024)	(0.042)	(0.030)	(0.031)	(0.024)
Materials	0.078**	0.110*	0.085*	0.090**	0.072**
	(0.033)	(0.058)	(0.044)	(0.044)	(0.032)
Productivity	0.120**	0.190**	0.135**	0.143**	0.116**
	(0.048)	(0.082)	(0.065)	(0.066)	(0.046)
TFPR	0.007	0.026	0.019	0.014	0.009
	(0.020)	(0.039)	(0.029)	(0.029)	(0.019)
AE	0.112*	0.163	0.116*	0.130*	0.108*
	(0.060)	(0.098)	(0.068)	(0.072)	(0.058)

Note: Table shows $\Delta Ln(MA_{1920})$ coefficients for different dependent variables (rows) and estimates of MA (columns). Coefficients are standardized for moving from the 25th to the 75th percentile of the independent variables. Dependent variables for specifications in each row are the change in manufacturing outcomes between 1910 and 1939 detailed in the first column. All specifications control for log total population in 1910. Regressions in each column differ in terms of the measurement of change in market access that they use. For instance, Column (1) measure holds population fixed at 1910 levels. Measures in Columns (2) to (5) use different levels of the key parameters described in Section 6 and detailed at the top of the table. Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Main Results: Different estimates of standard errors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Coeff.	Standard Errors estimated:					
		Clustered, Grid of Size:			Conley with Dist. cutoff:		
	$\Delta Ln(MA_{1920})$	300km	150km	200km	150km	300km	600km
Dep. Variable: Δy_c							
Pop.	0.033	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
Manuf. Emp. Pop. Share	0.058	(0.026)	(0.028)	(0.027)	(0.024)	(0.024)	(0.019)
N Estab.	0.066	(0.015)	(0.019)	(0.019)	(0.019)	(0.012)	(0.018)
Estab. Size (Labor)	0.025	(0.020)	(0.018)	(0.019)	(0.016)	(0.018)	(0.008)
Estab. Size (Revenue)	0.022	(0.029)	(0.029)	(0.028)	(0.026)	(0.027)	(0.020)
K Share	-0.019	(0.013)	(0.012)	(0.013)	(0.012)	(0.010)	(0.012)
Skill Ratio	-0.001	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Ratio Salary to Wage	0.012	(0.010)	(0.008)	(0.008)	(0.008)	(0.009)	-
Value of Products	0.089	(0.035)	(0.039)	(0.036)	(0.035)	(0.029)	(0.034)
Capital Exp.	0.070	(0.037)	(0.037)	(0.036)	(0.031)	(0.031)	(0.031)
Employment	0.091	(0.027)	(0.029)	(0.028)	(0.025)	(0.023)	(0.025)
Materials	0.084	(0.037)	(0.043)	(0.041)	(0.039)	(0.031)	(0.035)
Productivity	0.131	(0.053)	(0.057)	(0.054)	(0.053)	(0.052)	(0.057)
TFPR	0.009	(0.022)	(0.022)	(0.021)	(0.020)	(0.021)	(0.014)
AE	0.122	(0.066)	(0.068)	(0.066)	(0.063)	(0.062)	(0.056)

Note: Column (1) shows $\Delta Ln(MA_{1920})$ coefficients for specifications with dependent variables shown on each row. Columns (2) to (6) show different estimates of standard errors. Columns (2), (3), and (4) show standard errors clustered using an arbitrary grid of sizes detailed at the top. Columns (4), (5), and (6) show Conley adjusted std. errors with different distance cutoffs. Coefficients are standardized for moving from the 25th to the 75th percentile of the independent variables. Dependent variables for specifications in each row are the change in manufacturing outcomes between 1910 and 1939 detailed in the first column. All specifications control for log total population in 1910. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: Alternative Measures for pre-existing Lumber Potential and Activities

Dependent Var: Δy_c	(1) Manuf. Empl. Pop. Share	(2) N Estab.	(3) Estab. Size Labor	(4) Value of Products	(5) Capital Exp.	(6) Employment	(7) Materials	(8) Productivity
Panel A: Lumbering Measured as Wood Products Value of Products Share, 1910 (Over/under Average)								
$\Delta Ln(MA_{1920})$	0.049*	0.079***	0.005	0.066*	0.045	0.084***	0.056	0.072*
	(0.028)	(0.017)	(0.018)	(0.035)	(0.036)	(0.027)	(0.037)	(0.042)
$\Delta Ln(MA_{1920}) \times$	0.029	-0.038	0.063**	0.072	0.084	0.025	0.086	0.173
... $\mathbb{1}$(Lumber Int.>Mean)	(0.059)	(0.054)	(0.024)	(0.074)	(0.059)	(0.063)	(0.079)	(0.130)
$\mathbb{1}$(Lumber Int.>Mean)	-0.086	-0.002	-0.146	-0.075	-0.243	-0.147	-0.116	0.137
	(0.145)	(0.103)	(0.128)	(0.206)	(0.212)	(0.146)	(0.210)	(0.286)
Observations	217	217	217	217	217	217	217	217
r2	0.374	0.555	0.432	0.242	0.259	0.344	0.248	0.215
Panel B: Lumbering Measured as Wood Products Value of Products Share, 1910 (Continuous)								
$\Delta Ln(MA_{1920})$	0.053	0.100***	-0.005	0.061	0.032	0.095**	0.048	0.043
	(0.040)	(0.029)	(0.022)	(0.050)	(0.047)	(0.041)	(0.054)	(0.072)
$\Delta Ln(MA_{1920}) \times$	0.024	-0.141	0.129	0.118	0.166	-0.012	0.153	0.365
... Wood VP Share	(0.172)	(0.139)	(0.087)	(0.211)	(0.199)	(0.188)	(0.226)	(0.366)
Wood VP Share	-0.364	0.033	-0.483	-0.238	-0.582	-0.450	-0.289	0.312
	(0.303)	(0.223)	(0.296)	(0.464)	(0.471)	(0.337)	(0.477)	(0.671)
Observations	217	217	217	217	217	217	217	217
r2	0.376	0.558	0.435	0.239	0.257	0.347	0.245	0.209
Panel C: Lumbering Measured as Forest Area Share, 1900 and 1910 Avg. (Over/under Average)								
$\Delta Ln(MA_{1920})$	0.079***	0.083***	0.021	0.112**	0.080**	0.104***	0.108**	0.148**
	(0.029)	(0.023)	(0.016)	(0.045)	(0.036)	(0.032)	(0.047)	(0.061)
$\Delta Ln(MA_{1920}) \times$	-0.063	-0.047	0.009	-0.068	-0.031	-0.038	-0.071	-0.048
... $\mathbb{1}$(Forest Sh.>Mean)	(0.081)	(0.056)	(0.054)	(0.117)	(0.111)	(0.084)	(0.125)	(0.182)
$\mathbb{1}$(Forest Sh.>Mean)	-0.252	-0.197*	-0.061	-0.332	-0.361	-0.258	-0.466*	-0.007
	(0.151)	(0.115)	(0.159)	(0.226)	(0.236)	(0.169)	(0.240)	(0.299)
Observations	217	217	217	217	217	217	217	217
r2	0.381	0.561	0.420	0.246	0.254	0.346	0.254	0.201

Note: Coefficients are standardized for moving from the 25th to the 75th percentile of $\Delta Ln(MA_{1920})$. Dependent variables for specifications in each column are the change in outcomes between 1910 and 1939 detailed in the first row. Specifications in Panels A and C include an interaction between $\Delta Ln(MA_{1920})$ and a dummy equal to one if a county is above the mean of two lumber industry measures. Panels B and D include an interaction between MA gains measure and the continuous version of the lumber industry measures. Those measures are: share of manufacturing revenue from Wood Products industry (A and B), and percentage of total area in forest (C and D). All specifications include province fixed effects and control for a quadratic polynomial in latitude and longitude, for a quadratic polynomial in distance to the coast, the Great Lakes or the St. Lawrence River, and for log population in 1910. Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. Standard errors clustered at 300km x 300km cells of an arbitrary grid are shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix B Data Appendix

Table A.13 presents a summary of all primary data sources used in this paper.

Table A.13: Data Sources

Document	Source	Information	URL
1901 Census	Dominion Bureau of Statistics	Population, Agriculture, Manufactures	Link
1911 Census	Dominion Bureau of Statistics	Population, Agriculture, Manufactures	Link
1921 Census	Dominion Bureau of Statistics	Population, Agriculture	Link
1939 Census	Dominion Bureau of Statistics	Manufactures	Link
1941 Census	Dominion Bureau of Statistics	Population, Agriculture	Link
GEORIA project shapefiles	University of Toronto	Railroad lines and stations	Link
Summary of Canal Statistics	Dominion Bureau of Statistics	Canal Statistics	Link
Directory of Ports and Harbours of Canada	Department of Marine and Fisheries	Ports and Harbours	Link
The Panama Canal Records	The Panama Canal	Traffic and Port Statistics	Link
Table of distances between ports via the shortest navigable routes as determined by the Hydrographic Office	United States Navy	Ports to chokepoints distances	Link
Railway Freight Rates in Canada	Canadian Railway Comission	Railroad rates	Link
Report of the Grain Markets Commission of the Province of Saskatchewan	Saskatchewan Grain Commission	Shipping Insurance Charges	Link
Official Report of the Debates of the House of Commons of the Dominion of Canada	House of Commons of Canada	Distances between Great Lakes Ports	Link
Annual report on the statistics of railways in the United States	Interstate Commerce Commission	Railroad rates for grains	Link
Cost of Hauling Crops from Farms to Shipping Points	United States Department of Agriculture	Rates for Wagons	Link
Statistical Examination of Certain General Conditions of Transportation Bearing on the Economic Problem of the Proposed Georgian Bay Canal	Georgian Bay Canal Commission	Rates for Oceanic Transportation	Link
Panama and Suez Canals: General Comparative Statistics	The Panama Canal Company	Suez and Panama Canal tolls	Link

Market Access

We compiled information on wheat trade costs by mode of transportation, including rail, wagon, rivers and canals, and international oceanic routes. Additionally, all trips pay a 50-cent fee whenever modes are switched. These follow from Fogel (1964) and the Saskatchewan Grain Commission (1914). Table A.14 summarizes information on routes and rates.

Table A.14: Transportation Network for Trade Cost Calculation.

Mode	Information	Canada	United States
Rail	Routes	GEORIA Project, University of Toronto	Donaldson and Hornbeck (2016)
	Rate	0.514 cents per ton-mile. Method: Simple average of per-ton-mile rates for 17 routes between Fort William, Ontario to 17 destinations in Canada. Standard deviation = 0.047. Minimum: 0.43 cents per ton-mile, from Calgary (1,242 miles). Maximum: 0.6 cents per ton-mile, from Saskatoon (900 miles). Source: Railways Commission of Canada (1939)	0.626 cents per ton-mile Method: National average for grains, taken directly from the Interstate Commerce Commission (1913). Source: Interstate Commerce Commission (1913).
Wagon routes	Routes	Straight lines from county centroids to other county centroids, railroad stations, and harbours within 200km of the county centroid.	Donaldson and Hornbeck (2016)
	Rate	25.657 cents per ton-mile Method: Simple average of per-ton-mile rates for wheat of US States that shared a border with Canada. Cross-checked with Saskatchewan Grain Commission (1914). Source: United States Department of Agriculture (1906), Saskatchewan Grain Commission (1914)	22.639 cents per ton-mile Method: National average for wheat. Sources: United States Department of Agriculture (1906)
Waterways, rivers, and canals	Routes	Manually drawn from historical sources including Department of Marine and Fisheries of Canada (1922) and Canada Dominion Bureau of Statistics (1940)	Donaldson and Hornbeck (2016)
	Rate	0.238 cents per ton-mile. Method: Sum of (i) per-ton-mile rate of the Fort William, Ontario – Port Arthur, Ontario (Great Lakes) route, (ii) insurance costs assuming a cargo value of 86 cents per bushel (Great Lakes), and (iii) storage costs for up to six months. Source: Saskatchewan Grain Commission (1914), House of Commons of Canada (1908), and Ward (1994).	0.260 cents per ton-mile. Method: Sum of (i) per-ton-mile rate of the Chicago – New York all-water route, ii) insurance costs assuming a cargo value of 86 cents per bushel (Great Lakes), and (iii) storage costs for up to six months. Source: Interstate Commerce Commission (1913), Saskatchewan Grain Commission (1914), House of Commons of Canada (1908), and Ward (1994).
International oceanic routes	Routes	Donaldson and Hornbeck (2016) and United States Navy (1911, 1917, 1920, 1931, 1943)	
	Rate	0.052 cents per ton-mile + toll (if applicable) If using the Suez Canal: $1.48 toll. If using the Panama Canal: $0.95 toll. Method: Sum of per-ton-mile rate and insurance costs assuming a cargo value of 90 cents per bushel (New York). The per-ton-mile results from the simple regression of (a) historical freight costs between Liverpool and the ports of New York City, Odessa, Karachi, and Buenos Aires retrieved from the Georgian Bay Commission (1916) on (b) distance. Sources: Georgian Bay Commission (1916), Saskatchewan Grain Commission (1914), United States Navy (1911, 1917, 1920, 1931, 1943), and The Panama Canal Company (1971).	

Population

For each Canadian county, we rely on the Canadian Census of Population for 1911 and 1921. For US counties, we use the Population Census for 1910 and 1920 and adjust by

Table A.16: Ports used to build shipping routes passing through the Panama Canal

Ports	
Key West, FL	Providence, RI
Tampa, FL	Boston, MA
Mobile, AL	Portland, ME
New Orleans, LA	Halifax, NS
Port Arthur, TX	Montreal, ON
Pensacola, FL	San Diego, CA
Galveston, TX	Los Angeles, CA
Jacksonville, FL	San Francisco, CA
Charleston, SC	Astoria, OR
Savannah, GA	South Bend, WA
Wilmington, NC	Portland, OR
Alexandria, VA	Port Townsend, WA
Norfolk, VA	Seattle, WA
Newport News, VA	Tacoma, WA
Baltimore, MD	Everett, WA
Philadelphia, PA	Victoria, BC
New York, NY	Vancouver, BC

Source: The Panama Canal Records

GDP per capita relative to Canada. For destinations in the rest of the world, we rely on Maddison Project GDP and population data (Bolt and van Zanden, 2020). We use countries with nonmissing GDP per capita records for 1920 and countries with data for some point between 1910 and 1930 if data for 1920 is missing. As in the case of the US data, for each country we adjust by its GDP per capita relative to Canada in the year for which population data is available in order to adjust for the GDP-per-capita weighted population.

Adjustments to 1911 Canadian Manufacturing Outcomes

We use the Census of Manufactures for 1901, 1911, and 1939. There are three key issues when using the 1911 census. First, the results for 1911 and 1939 are not comparable due to differences in coverage. The 1911 census covered establishments with five or more employees, except in the case of flour mills, saw and shingle mills, lime, brick and tile works, butter and cheese factories, fish-curing plants, and electric light and power plants, which were all surveyed regardless of size. The 1939 census surveyed all establishments, regardless of size. To address this problem, we follow Urquhart and Dales (2007), at least in spirit, by using linear expansion factors (blow-up factors) for the industries that were not fully covered in 1911. These factors are available for each outcome (revenue, capital, employment, wages) at the province level. They are calculated from the 1906 Postal Census, which published information on these outcomes for all establishments and those employing five or more people. This is simply a linear expansion that follows from those two values.

The second issue is that the 1939 census does not include "hand-trades" whereas the

1901 and 1911 censuses did. To address this problem, we explicitly removed industries that were classified as hand-trades from the 1911 census data. Given that many times these industries were classified under "All Other Industries" in a given county, we blew down this particular industry by the proportion that hand-trades represented of the total national values for the outcomes of interest (capital, employment, wages, materials, revenue).

The third and final adjustment is related to fuel and electricity expenditures, which were recorded but not tabulated in 1911. We take the average share of revenues (value of production) that these expenditures represented for 19 Standard Industrial Classification (SIC) industry groups at the national level. To match each industry to an industry group, we use a 1948 SIC classification, which is comparable, according to Urquhart and Dales (2007), to industry classifications in 1911 and 1901. We then calculate total fuel and electricity expenses as the product of total revenues and Urquhart shares for each county/industry. For the "All Other Industries" category, we use national averages for all industries. Finally, we add these expenditures at the county level and include them in the materials expenses to make them comparable to the figures for 1901 and 1939.