Economic Geography Aspects of the Panama Canal

Stephan Maurer* and Ferdinand Rauch†

October 31, 2019

Abstract

This paper studies how the opening of the Panama Canal in 1914 changed market access and influenced the economic geography of the United States. We compute shipment distances with and without the canal from each US county to each other US county and to key international ports and compute the resulting change in market access. We relate this change to population changes in 20-year intervals from 1880 to 2000. We find that a 1 percent increase in market access led to a total increase of population by around 6 percent. We compute similar elasticities for wages, land values and immigration from out of state. When we decompose the effect by industry, we find that tradable (manufacturing) industries react faster than non-tradable (services), with a fairly similar aggregate effect.

JEL classification: F1, R1, O1, N72

Keywords: Market access; Panama Canal; trade shock; gravity

We thank many colleagues and participants at multiple seminars and conferences for valuable comments and suggestions. We thank Luigi Pascali for sharing his dataset of major ports with us.

*University of Konstanz and CEP, stephan.maurer@uni-konstanz.de.

†University of Oxford, CEP and CEPR, ferdinand.rauch@economics.ox.ac.uk.

1 Introduction

The effect of changes in market access on the spatial equilibrium is an old question in the economic literature. It is of practical importance to policy makers that consider investing in transportation infrastructure. Typically, this question is addressed in case studies that consider the effects of railroads, highways, ports and other changes in transport infrastructure. Here we use the opening of the Panama Canal in 1914 and see how it influenced the economic geography of the United States in following decades. The opening of the canal was one of the largest changes to international shipment distances, leading to big changes of market access for every US county, while at the same time giving much variation of the degree of this change within the US. The opening took place at a time when international trade overwhelmingly happened by ship, which makes this change in distances a more precise measure or trade than it would be in later decades.

Contributions of this project include the following four: First, we build a dataset of international and domestic market access for US counties for around 1900 that may be useful for other studies. We also measure the change in market access induced by the Panama canal for each US county. Second, we show that there is a strong positive causal effect of market access change on population growth throughout the 20^{th} century. Our main magnitude implies that an increase of market access of one percent leads to a population that is around 6 percent larger in 1940, and 7 percent larger in 2000. This is an elasticity estimate that may be of use elsewhere. We also provide related numbers for manufacturing wages, agricultural land prices and immigration from out of state, and show that all these react positively to increased market access. Third, we show that this effect seems to be fairly similar for tradable and non-tradable industries overall, with tradable workers reacting faster initially, and services catching up a little slower. Workers in agriculture react only by little. The long-term average effect is similar for tradable and non-tradable workers. Finally, we use economic theory to provide a cost-benefit analysis of the canal that suggests that the benefits from the canal easily outweigh the costs.

The basis of our dataset is an existing 20-year frequency county panel for the US from 1880-2000 (Michaels et al 2012). This dataset also allows us to consider total population growth and employment growth in agriculture, manufacturing and services separately. We combine these data with US domestic transport costs in 1890 (Donaldson and Hornbeck 2016) and data on the location and population of major international ports at the turn of the centuries (Pascali 2017). We compute shipment distances with and without the canal from each US county to each other US county and to each international port. We combine these data by computing the minimum distance from every county to every major international port, and to every other county with and without the canal. We then use a gravity-type framework to compute market access measures that are distance weighted population measures for every county.[1]

A first set of results describes the impact of the canal. Using our preferred set of parameters, about three percent of US county pairs have improved domestic market access as a result of the Canal and all have some improved international market access, with much variation of the magnitude of the change throughout the country.[2] On average, US counties experience a 5.8 percent total market access improvement, which consists of a 1.6 percent domestic gain and a 6.3 percent international one.

Looking at results over time we find that the canal had no effect on population growth of counties in the Placebo period before its opening, from 1880-1900. It has a significant positive effect in all the 20-year intervals after, particularly in the

[1] We use distance weighted population measures as proxy for trade flows that we don't measure directly. Typically policy makers that consider to invest in transport infrastructure also observe distance weighted population.

[2] In a specification where we restrict ocean based trade to go via three main US ports, this reduces to two percent of county pairs.

period of the opening 1900-1920. After that we continue to observe a positive effect that declines monotonically over time. This continued positive effect on growth reflects that the canal becomes more valuable over time as globalization intensifies. All this is consistent with our priors. The sample is large enough that we can consider heterogeneous effects. When we non-parametrically decompose the effect by initial density deciles we find that there is a fairly linear and monotonic relationship between improved market access and population growth, with no difference for small or large counties. We also find that the effect rises linearly with treatment intensity. Our results are robust to including additional control variables, to various ways of parametrizing cost-distance routes, and to instrumenting the market access change due to the canal by the cost-distance to the canal.

This paper relates to a large literature on the relationship between trade and growth (Frankel and Romer 1999, Redding and Venables 2004, Pascali 2017, Bakker et al 2018, Donaldson 2018). Our setting is particularly close to Feyrer (2009), who studies the impact of the closure of the Suez Canal following the Six Day War, and to Hugot and Umana Dajud (2016), who examine the effects of the Panama and Suez canals for international trade. Our paper differs in that we consider effects within a country rather than across countries, and that we study a permanent change over a longer time horizon. In this sense, this paper also relates to the literature on location fundamentals and cities' long-term development (Davis and Weinstein 2002, Bleakley and Lin 2012, Bosker and Buringh 2017, Hanlon 2017, Michaels and Rauch 2018). Particularly related in that literature are papers that examine the role of market access in determining city growth. Redding and Sturm (2008) study the effect of the Iron Curtain on the development of towns in West Germany. What we add to their findings is a more precise measurement of market access changes in our setting. A paper by Donaldson and Hornbeck (2016) on the effects of railroads is closely connected in data, econometric setup and research question. In a broader sense, our paper is also connected to several other papers that have looked at the growth implications of infrastructure measures that enhance market access. In this literature, railroads have received particular attention[3], but so have roads and highways (Banerjee et al 2012, Duranton and Turner 2012, Faber 2014, Baum-Snow et al 2019), and air links (Campante and Yanagizawa-Drott 2017). Our findings and the magnitudes of the effects we report may contribute to a recent theoretical literature on the spatial effects of trade (for example Fajgelbaum and Redding 2014, Cosar and Fajgelbaum 2016, Fajgelbaum and Gaubert 2018, Potlogea 2018, Bakker 2019). Finally, we also contribute to a literature on the economic effects of the Panama Canal (Huebner 1915, Hutchinson and Ungo 2004, Maurer and Yu 2008, Umana Dajud 2017) by adding our own measures and estimates.

The next section will give a brief overview of the history of the Panama Canal, and present some facts on its usage today. Section 3 describes the dataset we assemble for this project. Section 4 presents the main regression results. Section 5 presents a few robustness checks on the main results. Section 6 uses the results in combination with a theoretical model to produce an estimate of the welfare contribution to the US by the canal, and uses it in a cost-benefit analysis. Section 7 concludes.

2 The Panama Canal

The idea to connect the Atlantic and the Pacific to facilitate trade is old. A priest by the name of Francisco López de Gómara drew an optimistic plan to dig a canal in the area for the King of Castile already in 1552. A glance at a world map shows that the obvious place to dig is in the area of today's Nicaragua, Costa Rica or Panama, where the oceans are separated only by a small strip of land. The current canal in Panama is in fact close to the shortest possible passage. Nicaragua was frequently considered as a viable alternative, offering a longer distance but with lower heights

[3]See, for example, Atack et al 2010, Hornung 2015, Jedwab and Moradi 2016, Berger and Enflo 2017, Bogart et al 2018, Donaldson 2018, Buechel and Kyburz 2019, Braun and Franke 2019

to cross. Alexander von Humboldt wrote a study on a canal project in this area in 1811, and likely discussed it with US president Jefferson, another early proponent of this idea.[4]

An old Spanish trading route existed in the area of today's canal from possibly the 16th century. This heavily used path was developed by private entrepreneurs into a railway line connecting the oceans that opened in 1855. This railway line consisted of 76 kilometers of track and connected with ships at either end. This railway line benefitted from the gold rush in California, and helped transport people to California, as well as gold back. In the 19^{th} century it had the heaviest volume of freight of any railroad in the world. At some point its parent company was the highest priced stock listed on the New York Stock exchange. Yet despite its great success, the railway line had severe shortcomings, essentially excluding bulky or heavy goods trade. It was not a useful substitute for a proper canal, and the idea to dig remained a consideration for the US government and others. President Ulysses S. Grant remarked in 1881 "To Europeans the benefits of and advantages of the proposed canal are great, to Americans they are incalculable" (McCullough 1977, p. 26).

Yet despite the great importance of the canal to the United States, it was a Frenchman who pioneered this project. After playing the central role in the construction of the Suez Canal, French diplomat Ferdinand de Lesseps founded the 'Panama Canal Company', obtained the rights to dig from the Colombian government (at the time Panama was part of Colombia), raised private funds and started digging. Construction started in 1882. This project relied primarily on workers from the West Indies as well as French engineers, but also sourced moderate amounts of supplies and workers from the United States. The company underestimated the difficulties that a combination of yellow fever, malaria, tropical climate and remoteness presented. It

[4] The historic information in this section draws mainly on Cameron (1971), McCullough (1977) and Maurer and Yu (2010).

also may have made an error in insisting on a canal at sea level. The company went bankrupt in 1889. About 20.000 workers, mainly from the West Indies, died while working for the French company, primarily from malaria and yellow fever. Many French families lost money following the bankruptcy.

The project was abandoned, until President Theodore Roosevelt made it a priority and revived it. He declared in his first message to congress in 1901: “No single great material work [...] is of such consequence to the American people”. The US government bought the remains of the French company, and continued construction from 1904. The US encouraged a revolution in Panama, and prevented the Colombian government from interfering. This created the country of Panama and secured control over the canal for the US government in its first decades. The canal was completed and opened in 1914. 800 ships used it in 1916.

By 2018, the number of ships had increased to 15,000. Ships crossing the canal are lifted and then lowered about 26 meters in several locks. In total it takes around 8-10 hours to cross. The United States are the main user of the canal: Around 67.7 percent of shipments have either their origin or destination in the United States. This represents 175 million tons in 2018. The next beneficiaries are China (16 percent), Mexico, Chile and Japan (each around 12 percent), Peru and Colombia (9 percent). The main European beneficiaries are Spain and the Netherlands (1.9 and 1.8 percent respectively). Decomposing shipments by broad routes in 2018 reveals that 78 million tons are shipped from the US East Coast to Asia, which is by far the most important connection. Next comes the US East Coast and West Coast of South America (37 million tons) followed by East Coast US and West Coast Central America (17 million tons). Europe and the US West Coast exchange another 17 million tons through the canal. US intercoastal accounts for 8 million tons, as does Asia and East-Central America and South America Intercoastal. Europe and the US West Coast exchange 7 million tons (Panama Canal Authority 2018a, 2018b).

In this paper, we compare a world with the canal to one without the canal. When do we expect to see a difference between the two? There could have been some small effect of distance to the Panama Canal on trade from the time of the railway in 1850. The years of construction of the Canal from 1882-1914 drew some resources and people from the United States to Panama and may account for some small effect. The opening of the canal in 1914 marks the greatest change to transport costs we observe, and we expect big effects over this period. There is a question whether people's expectations of the canal lead to measurable effects before 1914, but we have reasons to doubt it. The closest year before the opening we use here is the year 1900, before any construction or planning attempts by the US government. At that time, construction of this project was of sufficient risk and difficulty that it was uncertain after the French failure if it would be taken up again and if so when it would be completed. In some of our regressions, we use 1880 as a base year, where the impact of the Panama Canal likely had an even lesser effect on people's location choices within the US. It is also worth noting that politicians at the time, as well as academic analysts of the canal, typically stress the military importance of the canal first, and commercial impacts only second (Huebner 1915). The canal continues to become more important throughout the 20^{th} century as international trade increases, as shipping technology improves, shipping volumes increase, and as the destination markets grow, first Europe after the wars, followed by the rise of Asian countries. For these reasons we expect a continued effect after 1914. The effect we report for the years 1900-1920 and 1900-1940 are less influenced by these other factors, and isolate the gains through the change in distances, the later effects adds additional treatments through technology and the growth of destinations.

3 Data

Our dataset aims to construct transport costs and destination market sizes as they were before the opening of the Canal, around 1900. We do not rely on information on either destination markets or domestic transport costs in the US after the opening, since both are endogenous to the new transport cost matrix. This implies that our measurement of the market access induced by the canal is more precise in the earlier decades of the 20^{th} century than in the later ones.

The Panama Canal facilitates commerce between US coasts, but also between US and international ports. We measure both effects separately. To calculate how it changed international market access, we draw on a dataset on major ports in the 19th century assembled by Pascali (2017). For every country, this dataset identifies the primary ports in 1850, and assigns them the country's respective population and GDP data for 1900. We then calculate seaborne least-cost paths and distances from every coastal county in the US to every major international port, using a 20x20km grid of the world and ArcGIS's "Cost Distance" tool. We calculate these distances under two scenarios: Once with the Panama Canal being closed or not existent, once with it being open.[5]

For every mainland US county, we then calculate the distance from this county's centroid to every international port. This distance consists of two components: The distance from the US coast to the international port calculated before, and the distance from the respective county to the coast. For the latter, we draw on a 1890 cost-distance matrix provided by Donaldson and Hornbeck (2016). This matrix takes into account railroads and canals with different cost parameters and therefore

[5] Given the global scale of our analysis, and our interest in distances, we use a Azimuthal Equidistant Projection of the world, centered around 39.83N 98.58W, the geographic center of the United States. The maps we use are based on Manson et al (2018) for the US, Bjorn Sandvik's public domain map on world borders (available from `http://thematicmapping.org/downloads/world_borders.php`), and the Rivers and lake centerlines dataset available from Natural Earth (`https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-rivers-lake-centerlines/`).

gives a precise picture of the effective distance between counties in that period.

To add the within-US cost distances (measured in US$) and the distances to international harbors (measured in km), we first transform the domestic distance matrix from monetary units to kilometers by scaling them such that the distances expressed in kilometers match the great-circle distances for the average county pair[6]. Total distance between county c and international port p^{int} via domestic port p^{dom} is calculated as the minimum distance of all possible routes via all domestic ports

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

One of the crucial parameters of this exercise is to define the relative cost of shipment of one kilometer inland against one kilometer by sea, parameterized as α. This parameter is such that $\alpha = 1$ would imply that a kilometer of trade over sea costs the same as a kilometer inland, while $\alpha = 1/2$ implies that trading over sea is half the cost of trading over land.[7] Data on freight rates between Cardiff and Port Royal in Pascali (2017) suggest that transporting one ton over a straight-line mile over the ocean cost around 0.15 cents in 1890. This can then be compared to the cost of land-based transportation. Donaldson and Hornbeck (2016) assume that the cost of transporting a ton-mile via railroad was 0.63 cents in 1890. This would imply a value of α of around 0.25. However, this assumes that any land-based transportation could be done via railroads and thus seems overly conservative. For a more realistic comparison, we calculated the average Donaldson-Hornbeck cost for a straight-line mile, which is roughly 1.7 cents per ton-mile in 1890. Comparing this with the ocean transportation costs from above suggests an α of around $\frac{1}{11}$.

[6]We also show a robustness check where we instead express all cost distances in dollars, assuming a cost of 0.1 cent per km over the sea.

[7]Glaeser (2011) gives an estimate of exactly this parameter when he writes "In 1816, it cost as much to ship goods thirty miles overland as it did for those goods to cross the Atlantic". Taking the distance to ship over the Atlantic as at least 3000 miles, this would imply a parameter of $\alpha = 1/100$. This is however smaller than the corresponding value for 1914, since with the introduction of the railway transport costs inland fell more than over sea in the 19^{th} century.

Even this might be on the conservative end, as Maurer and Yu (2008) estimate that the variable cost for a ton-mile over the ocean was only 0.045 cents in 1890 dollars, with implied αs of $\frac{1}{13.7}$ when using railroad costs as the comparison, and $\frac{1}{37.7}$ when using the average Donaldson-Hornbeck values.[8] Given these estimates, we use a value of $\alpha = 1/10$ in our preferred specification as a plausible, yet conservative estimate for the cost advantage of ocean-based transportation.We also present an alternative specification where we calculate transportation costs simply based on the Donaldson-Hornbeck land-based costs, and an assumed cost of 0.1 cent per ton-km over the ocean. Additionally, we also show robustness to including a fixed transshipment cost for sea routes, a fixed tariff rate for international routes, or a fixed toll cost to using the Panama Canal.

In our baseline specification, we allow every coastal US county to act as an international port. In a robustness check we restrict the paths from US counties to international ports to go through one of the three major US ports from the international ports database we use (New York, New Orleans, San Francisco). These ports constituted the most important trade gateways on the Atlantic, Gulf and Pacific coast, respectively. Results remain fairly similar, but the market access gains are more geographically clustered and hence less robust to including state fixed effects. The same is true for a set of ports that uses the 11 most important American ports in 1900.

Under these assumptions, we then calculate the minimum distance of every US county to every international port, with and without the Panama Canal. Calculations of minimum distances can be computationally intensive, but our set of coastal counties is small enough that we can calculate every possible route and select the minimum from all possible ones, even in the case where we allow every coastal county to connect to international trade. International market access for county c is then

[8] Donaldson and Hornbeck (2016) also include a sea-based path from San Diego to Florida to represent the route around Cape Horn. They set the cost for this at 8 dollars, which would imply αs of similar magnitudes to those based on Maurer and Yu (2008).

defined as

$$MA_c^{int} = \sum_{p^{int}} \left(d_{cp^{int}}\right)^{\theta} pop_{p^{int}}$$

where $d_{cp^{int}}$ is the distance between county c and international port p^{int}, $pop_{p^{int}}$ is the population of international port p^{int}, which is defined as the population of the port's country divided by the number of major ports in the country. In most countries this is inconsequential, as they only have one major internatinal port. For those countries with more ports, it means that we spread the population evenly across all ports. θ is the distance elasticity of market access and is the second key parameter. The unit of measurement of MA_c^{int} is in terms of population. Irrespective of parameter θ, this variable will increase by one if one person is added at a distance of one kilometer.

In our baseline model, we set parameter θ to -1. This is in line with the tradition of the *market access* literature (Harris 1954), with standard estimates of the distance coefficient from gravity equations for international trade (Disdier and Head 2008) and domestic trade within the US (Dingel 2016) and with theory on the geometry of trade flows (Rauch 2016, Chaney 2018). Empirical gravity equations are estimates of the impact of distance weighted populations on trade, and so estimate the parameter we need directly. It is also consistent with the estimates obtained by Donaldson (2018) when pooling across various commodities.[9] Finally, Redding (2016) simulates the effect of a change in transportation costs in a model economy with different locations and also chooses his parameters such that the elasticity of trade flows with respect to effective distances equals -1.

The opening of the Panama Canal changes the distances in the market value calculations, so that we can calculate $[MA_c^{int}|Canal]$ and $[MA_c^{int}|No\ Canal]$. We then

[9]Market access measures can alternatively be derived from structural models, see Arkolakis et al (2012) for references and examples.

define the change in international market access due to the canal, ΔMA_c^{int} as the ratio of the two. We use the ratio, which gives us a percentage change in market access and delivers coefficients that are straightforward to interpret. A second advantage of using a ratio here is that it does not change with the arbitrary unit of distance measurement. Besides international market access, the Panama Canal also affected domestic market access in the US, mainly by facilitating commerce between the West and East Coasts. We capture this by calculating domestic market access analogously to international one as:

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

where c^d are all potential destination counties in the US, and *pop* is their 1880 population. The change in domestic market access, ΔMA_c^{dom} is defined as a ratio similar to the international one. Finally, we also calculate the change in total market access:

$$\Delta MA_c^{tot} = \left[MA_c^{int} + MA_c^{dom}|Canal\right] / \left[MA_c^{int} + MA_c^{dom}|No\ Canal\right].$$

These changes in county-level market access form our main explanatory variable. We use the log of this ratio in the regressions so that the left-hand side and right hand side of our main specification are both in terms of log difference, and can be interpreted as time differences. We then merge this with county-level data on population and employment in three broad industry categories (agriculture, services, manufacturing) compiled by Michaels et al (2012). These data are based on the US census and are available at 20-year intervals from 1880 to 2000. They span the whole mainland United States, excluding only North Dakota, Oklahoma, South Dakota, and Wyoming, which had not obtained statehood by 1880. In addition, we also add data on manufacturing wages and land values in 1900 and 1940 from Haines and ICPSR (2010). Land values are measured as the average value of farmland and buildings per acre, manufacturing wages as the ratio of a county's total manufacturing wage sum and total manufacturing employment. Finally, to shed lights on the potential reallocation of populations across space, we also include data on migration. Unfortunately, the historical US census does not allow us to identify all migrants. However, based on birthplace information we can at least identify all the people that live in a state different from the one they were born in. Drawing on the full count census records available from IPUMS (Ruggles et al 2019) for 1900-1940, we therefore create the share of a county's population that was born in a different state. This at least allows us to analyze long-distance migration. We merge the ICPSR and IPUMS data with our main dataset by obtaining centroid coordinates for each county using maps provided by NHGIS. We then use these centroids to link data to the main dataset. In addition, we also collect basic geographic control variables such as the longitude and latitude of a county's centroid.

For our outcome variables, we use average growth rates over all 20-year periods. These are calculated as the difference in the log of the respective variable at time t and t-20, divided by the 20 years elapsed in between, such that for example $\ln \Delta pop_{1880-1900} = (\ln pop_{1900} - \ln pop_{1880})/20$. Regressions are then of the form:

$$\ln \Delta pop_{t_1 - t_0 c} = \alpha + \beta_1 \ln \Delta MA_c + \beta_2 \ln MA_c + \ln pop_{t_0 c} + \epsilon_c.$$

Note that the main elements of this equation, the left-hand side and the term $\ln \Delta MA_c$ can both be interpreted in terms of time difference. This means that time-invariant fixed effects, such as a time-invariant county or state fixed effects as well as any locational fundamental that does not vary over the period of interest, are implicitly controlled for in every version of the regression we show. We occasionally include state fixed effects additionally. Given our difference specification, these are essentially state $\times$ year fixed effects and allow us to abstract from population reallocation towards certain states. If the effect of market access decays logarithmically, there is no reason to expect that within-state effects should be different from overall effects. Yet the meaning of the coefficient is slightly different, and it should be noted that this specification removes a lot of potentially useful variation from the right-hand side of our regression equation. As a further control, we always include the level of market access (MA_c), measured in a world without the canal, since we are interested only in the effects of changes in market access due to the Panama Canal, and not in effects from generally better market access levels. We also typically include linear controls for latitude and longitude to level the board, and also initial log population to adjust for the potential impact of different starting positions. We cluster standard errors at the level of a grid of five by five degrees following Bester, Conley and Hanson (2011) to account for spatial correlation in all our regressions. Figure 1 shows the market access gain due to the Panama Canal conditional on latitude, longitude, market access level, log population in 1900 and state fixed effects. The map interpolates information from county centroids, darker pixels show areas that benefit less. The map shows the right hand side variation we use, and demonstrates that we have healthy variation of the effect across the country.

Market access expressed in units of distance weighted population is a proxy for actual trade flows, which we do not observe directly. Yet, it is a useful proxy: Policy makers that evaluate the consequences of new infrastructure programs such as new railway lines or highways typically can more easily measure the implied market access changes in units similar to ours than the implied actual trade flow changes. In our discussion of welfare effects below, we show how policy makers can translate the effect into welfare estimations.

Table 1 provides summary statistics for these main variables of interest, computed using our preferred parameters. Domestic market access is considerably lower than international one, which is consistent with the US being still relatively lightly populated country in 1900. The Panama Canal changed domestic market access by around 2 percent for the average county. International and total market access both change by more, both by around six percent for the average county. Population numbers for the average county reflect population growth of the US over this period.

4 Results

We start by analyzing the differential impact of the opening of the canal. In Table 2, we show results from a difference specification, where we use annualized population growth between 1900 and 1940 (Columns 1 and 2) 1900 and 2000 (Columns 3 and 4) or as outcome variables. The starting year 1900 is 14 years prior to the opening, and years before the United States started the project. 1940 is an end year that gives the canal over two decades to establish its main effect. We also report results with 1920 as end year below. The alternative end year 2000 shows the aggregate long-term effect. This longer term coefficient is harder to interpret than the one to 1940 for at least two reasons: First, domestic transport infrastructure in the United States likely adjusted to the world with the Panama Canal after its opening, which makes the pre-treatment measures of domestic transport costs that don't condition on these endogenous adjustments, increasingly noisy. Second, developments such as increased global trade and growth in Europe and Asia over the 20th century make the Canal more valuable, and thus add to the observed treatment. Consequently, the treatment effect is harder to define over this longer period. The main coefficient of interest is the coefficient on ΔMA^{tot}. All four coefficients have a positive sign and are statistically significant, which suggests that counties with an improved market access due to the canal indeed experience higher overall population growth. Looking at the total coefficient of Column (4), increasing total market access by 1% would increase the annualized population growth rate by 0.0007, which translates to a population that is about 7 percent larger after 100 years. This implies a long-run elasticity of population with respect to market access of well above 1. The same holds for the medium-run results in 1940, where a 1% increase in market access is associated with a population gain of 5.8%.

Several recent contributions show that the effect of market access on subsequent city growth can vary with initial city size. For example, Baum-Snow et al (2019) find that the benefits of highway expansion in China accrued mostly to more important cities. A natural question thus is whether the effects of the Panama Canal also varied with initial county population. To analyze this, we divide the counties into deciles according to their 1880 population. We then perform long-difference regressions for 1880 to 2000 on interactions of total market access changes with these deciles, controlling for the direct population effects. The resulting coefficients by initial decile are shown in Figure 2. We do not find much evidence for heterogeneity along this dimension. Coefficients do not appear statistically different from one another.

A related concern is that the effect may be non-linear in treatment intensity, which would be interesting to know, and imply a miss-specified empirical model. To address this concern, we run another non-parametric specification, in which we replace the total market access change variable by ten decile indicator variables for the intensity of treatment. Coefficients and 95 percent confidence intervals of this exercise are displayed in Figure 3. The figure suggests a linear, increasing effect of treatment effect with treatment intensity, with no clear structural break.[10]

In Table 3 we repeat the exercise, separating the total market access effect into a domestic and international component. We find that all coefficients on the market access change variables are positive and significant when we separate domestic and internation in Columns (1) to (4). Coefficients for international markets are larger,

[10]We compute versions of this graph using different sets of weights to balance deciles. They tend to show similar smooth upward trends.

which suggests that international market access was a more important driver of population reallocation than domestic market access. This is particularly borne out by columns 5 and 6, which introduce both domestic and international market access changes at the same time. The fact that international market access change is more important is consistent with the observation that US inter-coastal trade is a relatively minor part of traffic through the canal. It is also worth reminding here that domestic market access changed considerably less than international one.

As a next step, we estimate separate regressions for each of the 20-year intervals from 1880 to 2000 to decompose the long-difference effect from above into its different sub-periods. These are shown in Table 4, again separately with and without state fixed effects. We find that there is no significant effect in the 'placebo' period 1880-1900, before the US construction and before the opening of the canal, in both specifications. Across states we see the biggest impact in the interval during which the canal was opened, 1900-1920. We also find a continued effect for the remaining intervals, which is monotonically decreasing towards the end of the 20^{th} century. Within states, we also find no effect in the period before the opening 1880-1900. The biggest effect is again in the interval during which the canal was opened. Coefficients show a decreasing trend and become insignificant after 1960.[11]

In Table 5, we analyze how the opening of the Panama Canal affected the sectoral composition of local economies. For this, we use data from Michaels et al (2012) on employment in three broad economic sectors: agriculture, manufacturing, and services. We show the results for every 20-year interval between 1880 and 1940, with and without state fixed effects. Here we focus on total market access on the

[11] It is worth reminding that the market access measures are computed using information from around 1900 in all specifications in this table, since we do not want to compute measures based on endogenous updates to transport infrastructure. This means that our main independent variable becomes increasingly imprecise over time. This increased measurement error could contribute to a bias in either direction for the later coefficients in this table. If it is classic measurement error it could lead to estimates that are too low. If on the other hand infrastructure investment after 1900 takes the Panama Canal into account, this would lead to actual market access gains due to Panama that are larger than the ones we measure. If so, coefficient estimates would be biased in the other direction.

right hand side. The only change to regressions we make is that we additionally add control variables for agricultural and services sectoral share in 1880, since initial industry shares are likely to influence sectoral developments. Columns (1) and (4) show that all three sectors do not have a positive relationship with market changes in the period before the opening of the canal, 1880-1900, at five percent level. Both services and manufacturing react strongly in the period of the opening, manufacturing about twice as much as services. While manufacturing shows a weaker but positive effect for 1920-1940, services shows an increased effect for this later period. Agriculture shows a modest and marginally significant effect during the opening, and is insignificant else.

Figure 4 plots these coefficients for manufacturing and services with their 95 percent confidence intervals, also for the periods after 1940. For comparison we add the coefficients for the total population effect. Taken together, these sectors react stronger than the total effect in the period of the main treatment, from 1900-1920. Manufacturing seems to react faster than services, and services catch up thereafter. The trends of both employment sectors are broadly similar: strongest in the period of the opening, with a gentle decline thereafter and insignificant coefficients later.

A simplification that may have some justification in the earlier parts of the 20^{th} century would be to call manufacturing industries the tradable sector, and the service sector the non-tradable sector. This simplification is less justified in later years, when services become increasingly traded. This may explain why the services sector coefficients are above those for manufacturing in the later years. But why would non-tradable jobs react to changes in market access? One explanation might be that tradable workers cause local demand, which in turn attracts workers in the non-tradable sector. Estimates suggest that one additional tradable job creates 0.8 (Van Dijk 2015), 1.5 (Moretti 2010) or 1.6 (Van Dijk 2017) non-tradable jobs. Given these estimates, we would expect non-tradable employment to react in a way that is similar, or perhaps even stronger than employment in the tradable sector. The finding that the tradable and non-tradable sector react to market access changes in a somewhat similar way may be of interest to theories of spatial economies. Frequently such models nest separate CES indices for tradable and non-tradable sectors in a utility function using a Cobb-Douglas function as aggregator. This would imply constant expenditures for each sector, and might for many sets of assumptions on production lead indeed to similar reactions for both sectors due to market access changes. Our results could give some empirical validation to such a modeling choice.

We consider other outcome variables in Table 6. Regressions here are exactly the same as those in Table 2, except that we use different dependent variables here. All outcome variables in this table are also annualized growth rates. In Columns (1) and (2) we use average manufacturing wage growth as reported by ICPSR. Without state fixed effects there is a positive and significant relationship, with state fixed effects it is insignificant. We caution against overstating this result, since the wage measure suffers from a few shortcomings. It represents average manufacturing wages, without adjusting for occupation, education, age or any other control variable. It also shows nominal wage growth, without taking into account real wage growth that could vary due to changes in house prices. In Columns (3) and (4) we use the annual growth rate of agricultural land values, and find a positive and significant relationship with and without state fixed effects. The coefficient is smaller than the market access to land value coefficient reported by Donaldson and Horneck (2016), but the longer run effect implied by accumulating the annual effect over multiple years is larger. Another difference is that our measure of land values includes buildings and other improvements. Columns (5) and (6) use the growth rate of the share of population born outside of the state on the left-hand side. Again, the relationship is positive and significant with and without state fixed effects, which suggests that immigration from outside the state contributes to the population growth we find. The magnitude is larger without state fixed effects, which means that the effect of immigration from outside the state is more important across states than within states. The results for land values are broadly similar in magnitude to findings for population, manufacturing wages grow slower, and immigration from out of state grows faster.

Our analysis so far pools coastal states with inland states. In Table 7 we repeat the analysis for counties located in coastal states only, and also for counties in non-coastal inland states only. The coastal sample may provide cleaner measures for market access changes due to Panama, since it does not rely on the assumptions made for the domestic transport cost matrix. This sample also addresses the point that coastal states had particularly strong population growth over the 20^{th} century, and this different development could influence regression results in ways not captured by the state-level trends. The cost of the exercise is that the reduced sample comes with reduced statistical power. Coefficients remain positive, and significant at 5 percent in the long run versions. When we consider inland states only in Columns (5)-(8) we also find positive coefficient, that are significant except for the long run estimate with state fixed effects in Column (8). Magnitudes are larger for inland than coastal counties, but in both cases not orders of magnitude different from the overall effect.

5 Robustness checks

The computation of the market access variables in this paper relies on several assumptions. In Table 8, we assess the robustness of our results by varying a few of these assumptions. For brevity, we focus on total market access change results. For ease of comparison, Panel A repeats our baseline results. We always show specifications without state trends in Column (1), and those including state fixed effects in Column (2).

In Panel B, we include other potential drivers of population growth as additional control variables. Firstly, we include soil quality to account for differences in agricultural productivity. To do so, we focus on three major crops: cotton, maize, and wheat. For each crop, we draw on the FAO-GAEZ database and calculate the counties' average attainable yield under intermediate inputs and rain-fed irrigation. An additional confounder could be weather. As shown by Rappaport (2007), nice weather has been an important factor in explaining population movements within the US over the last century. We therefore use the data provided by Rappaport (2007) and, following his specification, control for linear and squared terms of January daily maximum temperature, the July daily maximum heat index, July average daily mean relative humidity, average annual precipitation, and the average number of days per year on which there was at least 0.01 inch of precipitation. When we include these additional controls, our point estimate decreases in the specification without state fixed effects, but remains very stable in the specification with state fixed effects.

Throughout the paper, we have so far allowed every costal county in the US to be a port for ocean-trade. This might be too generous, and allow for travel routes that were in fact not used. In Panel C, we take a different approach and impose the restriction that any ocean-based trade has to go through 3 major ports- New York on the Atlantic Coast, San Francisco on the Pacific one, and New Orleans in the Gulf of Mexico. These three ports represent the most important port on every coast during the early 20th century (see for example Secretary of the Treasury 1880, Table 138, or Department of Commerce (1922), Table 293). These are also the three US ports featuring in the Pascali (2017) database we use for our international ports. Panel D takes a more generous approach by allowing the eleven most used ports in 1900 to serve as points of entry and exit.[12] As can be seen, these restrictions change

[12] We used the top 10 ports by tonnage cleared, and the top 10 ports by tonnage entered, resulting in 11 distinct ports: New York, New Orleans, San Francisco, Baltimore, Philadelphia, Newport News (Virginia), Boston, Galveston (Texas), Pensacola (Florida), Mobile (Alabama) and Puget

little in the specification without state fixed effects, but become insignificant in the specification without. A likely reason for this is that the resulting market access changes in these specifications are clustered around the admissible harbors, so that state fixed effects remove a lot of useful variation.

Panel E shows an alternative robustness check, where we impose an additional fixed cost for loading goods from domestic modes of transport to ocean ships. Following Fogel (1964) and Donaldson and Hornbeck (2016), we set this fixed costs for transshipment from one mode of transport to another at 50 cent per ton, which expressed in the units of our preferred specification corresponds to a cost of 56.9 land-based km. As Panel E indicates, adding such a fixed cost to every ocean-based route changes our results little.

In Panel F, we add a fixed cost to all international routes in order to assess how tariffs could affect our results. Actual tariffs during our period of analysis depended on the country of origin and the type of good imported, which we both don't observe, and which might be endogenous to trade distances. They also were typically assessed on an ad valorem basis, making it difficult to assign a precise value per average traded ton. We assign a fixed cost of 5$ per ton of traded good, which corresponds to 569.5 additional land-based km. While this is a simplified approach, it captures the basic idea that tariffs made international trade more costly than domestic one. A rate of 5$ is similar to the tariffs levied on the few goods that were assessed on a per-ton basis.[13] Judging from examples listed by the Treasury Department (1913) we think this parameter is at the upper end of possible choices, and so it is reassuring that results remain qualitatively and quantitatively similar to our baseline case.

In Panel G, we add a toll cost to all routes via the Panama Canal. According to Huebner (1915), initial canal tolls were such that a ton of cargo cost around 80-90

Sound, which we assigned to Seattle. For the ports and tonnages see Secretary of Treasury (1900), Table 162.

[13] For example, the tariff of 1913 stipulated a rate of 50 cent to $1.50 per ton for clay and earth, 2$ for hay, and 7$ for nitrate of saltpeter (Treasury Department 1913).

cents. We therefore add a fixed cost of 90 cents, corresponding to 102.51 additional land-based km. The results are again similar to our previous ones, both with and without state fixed effects.

In Panel H, we use a simpler cost-distance calculation that relies on geometric distances between county centroids in the US, instead of the more elaborate domestic cost distance matrix. This makes our results less precise, and also leads to negative point estimates that are statistically significant in the specification with state fixed effects. This shows that the detailed transport matrix we use for the inland US adds important information, and that our results depend on it. For example, counties that connect to the coast via railway might benefit much more from the Panama Canal than those geographically closer, but less connected in terms of infrastructure. Column H highlights the importance of using better measurement.

In our cost calculations so far, we transform the units in the domestic trade cost matrix from dollars to distances and then compared them to ocean distances, taking into account our assumed cost advantage of ocean-borne transportation. Alternatively, in Panel I, we calculate a dollar value of ocean-based transportation, assuming a cost of 0.1 cent per km. This cost assumption is consistent with data on actual shipping rates for 1890 from Pascali (2017). In this approach, the distance weights are expressed in terms of dollars and not kilometers. This change of unit does not affect regression coefficients. However, in this check the weighting between shipment over land and ocean is calibrated slightly differently, following the parameters expressed in dollars. Conceptually this specification remains fairly similar to our baseline and it produces very similar results.

Finally, in the next section, we will perform a welfare analysis. For this welfare analysis, we calibrate international market access such that the population-weighted ratio of international market access (before the canal) to 1900 population equals the US import penetration ratio of 4.6% (Lipsey 1994) in 1900. This requires penalizing international market access by a factor of 0.0024. In Panel J, we therefore show results when we use this penalized international market access in our calculation for market access changes. Point estimates become lower, but remain positive and both economically and statistically significant. An alternative way to match the 1900 import penetration ratio is to decrease θ to -1.82. Results for this are shown in Panel K. They remain positive and significant, but are not robust to including state fixed effects.

Conditional on market access levels, the measure of additional market access gain due to Panama is an abstract enough variable that we find it reasonable to treat it as exogeneous in this paper. Still, as a final robustness check we run an instrumental variable version of our main regression specification. In this specification, we use the (cost-) distance to the Panama Canal as an instrument for the market access change induced by the Panama Canal. This approach addresses concerns that the trade cost matrix we compute might correlate with endogeneous changes that occur later. It also corrects for potential bias that could arise due to any remaining measurement error in the market access variable. We think that the distance to the city of Panama is unlikely to influence economic developments in US counties apart from their effect via the canal, given the latitude controls we always include, and even more so when we include state trends. The first stages for this test with and without state fixed effects are reported in Columns (1) and (2) of Table 9. As expected, greater cost-distance to the canal correlates with lower market access change. The F-statistics of the first stage are 19.6 and 95.7. In the second stage we find coefficients that are positive, significant, and similar in magnitude to the ones reported in the OLS equivalent in Table 2.

6 Welfare and cost-benefit analyses

To assess the overall welfare impact of the Panama Canal on the US economy requires the use of a general equilibrium model that gives us a way to translate our estimates into aggregated welfare changes for the US as a whole. To do this in a broad way, we rely on a result by Arkolakis et al. (2012) that shows that on the question of how trade impacts welfare, a large class of trade models delivers essentially the same welfare calculation. In their notation, for real income $\hat{W}$ the share of within-country expenditure $\hat{\lambda}$ and the trade elasticity ϵ it holds that:

$$\hat{W} = \hat{\lambda}^{1/\epsilon}. \tag{1}$$

This welfare equation does not depend on the population elasticities we compute, but relies simply on the market access measures themselves. This model measures welfare effects for the country as a whole, and does not take within-country readjustments, such as we estimate in this paper, into account. We think that this is not a big limitation in this context, since in the cost-benefit analysis we are interested in the overall welfare effect for the country. We take weighted averages of market access changes and trade shares to translate our measures into representative values for the whole country. We take parameter ϵ to correspond to the same distance elasticity that we denote by θ in our empirical model.[14] We also need the change in the domestic trade share. Market access changes resulting from the Panama Canal as shown in Table 1 are closely related. We use population as a proxy for trade twice in this calculation, once to compute the domestic trade share, and once when we

[14] Arkolakis et al. (2012) define this parameter as: 'elasticity of imports with respect to variable trade costs', which seems to correspond to θ, which is -1 in our preferred specification. Redding (2016) distinguishes between the elasticity of trade with respect to effective distance (which he sets to -1), and the elasticitiy of trade with respect to trade costs (which he sets to -4, in line with Simonovska and Waugh (2014)). If we take a parameter of -4 in this calculation while leaving everything else as it is, welfare gains reduce to 0.2 percent in the unweighted and 0.09 percent in the weighted specification.

compute the change caused by the canal. The assumption that distance weighted population is proportional to trade flows, is substantiated by the large literature on the gravity equation in trade. So, we define:

$$\lambda^{NoCanal} = 1 - \frac{MA^{int,NoCanal}}{pop^{1900}}$$

and

$$\lambda^{Canal} = 1 - \frac{MA^{int,Canal}}{pop^{1900}}$$

from which we infer the domestic expenditure share change due to the Panama Canal using the welfare change equation above.[15] The unit of measurement for distance becomes important here. So far our only concern was that we measure domestic and international distances in the same unit. Whether that unit is expressed in terms of dollar, km or cm was of no importance for our regression results, since this unit cancels out when we compute our main variable, the market access change measures. When comparing distance weighted market access population to the population of a county, the unit of measurement of distance becomes important, since it influences this comparison. We can normalize this arbitrary distance parameter by requiring that the resulting population weighted average international market access implies an import penetration ratio for 1900 that is the actual one, which we take to be 4.6 percent, following the import-to-GDP ratio reported by Lipsey (1994).

Using our preferred market access measures, computed with the parameters of $\theta = -1$ and $\alpha = 1/10$ we obtain a mean welfare change of 0.8 percent of real income. When we weight counties by their 1900 population, to get a more representative measure for the average person rather than the average county, the weighted mean becomes 0.4 percent. The cost of the Panama Canal was 352 million nominal USD

[15] The same equation is used in a similar context by Cosar and Thomas (2019).

in 1914.[16] The GDP of the US in 1913 was around 40,000 million nominal USD (Maddison 2007). The annual welfare gains implied by our main estimate of 0.4 percent correspond to 160 million nominal USD in 1914. This suggests that about a 45 percent of the cost of the canal was offset by the welfare benefits obtained by the canal in a single year. Even a strongly declining social discount rate would imply that the costs were fully offset after only a few years of using the canal. We suggest that the construction of the Canal passes a cost-benefit test easily when we use the specification with our preferred set of parameters.[17]

Two caveats to this exercise should be noted. First, the total costs of the canal may have been substantially larger than the measure we use here if the price the US government paid for the French company was less than the value of the French excavations. This could have been the case, since the bankrupt French company was not in a strong bargaining position at the time of the sale. In response, we note that the US congress had a long debate on whether they should start from scratch in Nicaragua rather than continue in Panama. This debate was only narrowly decided in favor of Panama, which implies that this accounting mismatch, if it exists, can't have substantially influenced the economic costs for the US government. Second, the loss of over 20,000 workers under the French command, and another 5,000 under the US leadership, mainly due to malaria and yellow fever, implies a heavy welfare cost of canal construction not factored in this calculation. Despite these workers being well-paid volunteers who knew about the risks, we may want to factor this in beyond the monetary cost from today's perspective. Yet the cost-benefit calculation is so strongly in favor of the canal, that even a most generous adjustment for the losses of these workers and their families would not change the conclusion of this cost-benefit comparison. The monetary consideration for the US government, and their internal

[16]Using conventional inflation indices, this roughly corresponds to 10 billion USD in 2018 USD.

[17]As mentioned previously, setting $\theta = -1.82$ also produces a population-weighted ratio of international market access to 1900 population of roughly 4.6%. If we use market access calculated with $\theta = -1.82$ and a trade elasticity of -1.82 we get population-weighted welfare gains of 0.3%.

cost-benefit calculations, are not affected by either of these caveats.[18]

7 Conclusion

This paper presents three main contributions. First, it provides an estimate of the elasticity of population with respect to market access change. This is an important parameter for policy makers that try to evaluate the potential benefits of transport infrastructure such as a new railway or highway. We show that a one percent increase in market access led to an increased annualized population growth of 0.14 percent in the medium and 0.07 percent in the long run. This coefficient implies that a 1 percent change of market access in 1914 led to a population that is around 6 percent larger in 1940. This relatively large elasticity, which is well above one, can help explain the well-established fact (Rappaport and Sachs 2003) that the US population is disproportionately located near the coasts. We also show similar magnitudes for the growth of manufacturing wages, agricultural land values and immigration from out of state.

Second, we show that this effect seems to be fairly similar for tradable and non-tradable industries overall, with tradable workers reacting faster initially, and services catching up a little slower. These results could imply that the tradable sector reacts to market access changes instantly, while the non-tradable workers follow the local demand shifts caused by the movement in the tradable sector.

Finally, we use a general equilibrium framework to provide a cost-benefit analysis for the Panama Canal, one of the largest infrastructure project in the history of the United States. It suggests that the benefits from the canal easily outweigh the costs.

[18] Another caveat is that the Arkolakis et al (2012) framework does not take into account population movements within the US. In the Appendix, we apply a different welfare calculation based on Redding (2016) that allows for welfare changes within a country.

References

[1] Arkolakis, Costas, Arnaud Costinot, and Andrś Rodrǵuez-Clare. 2012. New Trade Models, Same Old Gains? *American Economic Review* 102: 94-130.

[2] Atack, Jeremy, Fred Bateman, Michael Haines and Robert A. Margo. 2010. Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth inthe American Midwest, 1850-1860. *Social Science History* 34: 171-197.

[3] Bakker, Jan. 2019. International Trade and Regional Inequality. Working Paper.

[4] Bakker, Jan David, Stephan Maurer, Jörn-Steffen Pischke, and Ferdinand Rauch. 2018. Of Mice and Merchants: Trade and Growth in the Iron Age. NBER Working Paper No. w24825.

[5] Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2012. On the Road: Access to Transportation Infrastructure and Economic Growth in China. NBER Working Paper17897.

[6] Baum-Snow, Nathaniel, J. Vernon Henderson, Matthew A. Turner, Qinghua Zhang, and Loren Brandt. 2019. Does Investment in National Highways Help or Hurt Hinterland City Growth? *Journal of Urban Economics*, forthcoming.

[7] Berger, Thor, and Kerstin Enflo. 2017. Locomotives of Local Growth: The Short- and Long-Term Impact of Railroads in Sweden. *Journal of Urban Economics* 98: 124-138.

[8] Bester, C. Alan, Timothy G. Conley, and Christian B. Hansen. 2011. Inference with Dependent Data Using Cluster Covariance Estimators. *Journal of Econometrics* 165:137-151.

[9] Bleakley, Hoyt, and Jeffrey Lin. 2012. Portage and Path Dependence. *The Quarterly Journal of Economics* 127: 587-644.

[10] Bogart, Dan, Xuesheng You, Eduard Alvarez, Max Satchell, and Leigh Shaw-Taylor. 2018. Were Railways Indispensable for Urbanization? Evidence from England and Wales. Working Paper.

[11] Bosker, Maarten, Eltjo Buringh, and Jan Luiten Van Zanden. 2013. From Baghdad to London: Unraveling Urban Development in Europe, the Middle East, and North Arica, 800-1800. *Review of Economics and Statistics* 95: 1418-1437.

[12] Braun, Sebastian Till, and Richard Franke. 2019. Railways, Growth, and Industrialisation in a Developing German Economy, 1829-1910. Working Paper.

[13] Büchel, Konstantin, and Stephan Kyburz. 2019. Fast Track to Growth? Railway Access, Population Growth and Local Displacement in 19th Century Switzerland. *Journal of Economic Geography*, forthcoming.

[14] Cameron, Ian. 1971. *The Impossible Dream: The Building of the Panama Canal.* Hodder and Stoughton.

[15] Campante, Filipe, and David Yanagizawa-Drott. 2018. Long-Range Growth: Economic Development in the Global Network of Air Links. *Quarterly Journal of Economics* 133: 1395-1458.

[16] Chaney, Thomas. 2018. The Gravity Equation in International Trade: An Explanation. *Journal of Political Economy* 126: 150-177.

[17] Cosar, Kerem and Benjamin Thomas. 2019. The Geopolitics of International Trade in Southeast Asia. Working Paper.

[18] Cosar, A. Kerem, and Pablo D. Fajgelbaum. 2016. Internal Geography, International Trade, and Regional Specialization. *American Economic Journal: Microeconomics* 8: 24-56.

[19] Davis, Donald, and David Weinstein. 2002. Bones, Bombs, and Break Points: The Geography of Economic Activity. *American Economic Review* 92: 1269-1289.

[20] Department of Commerce 1922. *Statistical Abstract of the United States 1921.* Washington: Government Printing Office.

[21] Dingel, Jonathan I. 2016. The determinants of Quality Specialization. *Review of Economic Studies* 84: 1551-1582.

[22] Disdier, Anne-Celia and Keith Head. 2008. The Puzzling Persistence of the Distance Effect on Bilateral Trade.*Review of Economics and Statistics* 90: 37-48.

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

[24] Donaldson, Dave, and Richard Hornbeck. 2016. Railroads and American Economic Growth: a "Market Access" Approach. *Quarterly Journal of Economics* 131: 799-858.

[25] Duranton, Gilles, and Matthew A. Turner. 2012. Urban Growth and Transportation. *Review of Economic Studies* 79: 1407-1440.

[26] Eaton, Jonathan, and Samuel Kortum. 2002. Technology, Geography and Trade. *Econometrica* 70: 1741-79.

[27] Feyrer, James. 2009. Distance, Trade, and Income-the 1967 to 1975 Closing of the Suez Canal as a Natural Experiment. NBER Working Paper No. 15557.

[28] Faber, Benjamin. 2014. Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies* 81: 1046-1070.

[29] Fajgelbaum, Pablo, and Cecile Gaubert. 2018. Optimal Spatial Policies, Geography and Sorting. NBER Working Paper No. w24632.

[30] Fajgelbaum, Pablo, and Stephen J. Redding. 2014. External Integration, Structural Transformation and Economic Development: Evidence from Argentina. 1870-1914. NBER Working Paper No. w20217.

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

[32] Frankel, Jeffrey A., and David H. Romer. 1999. Does Trade Cause Growth? *American Economic Review* 89: 379-399.

[33] Glaeser, Edward. 2011. *Triumph of the City*. Pan.

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

[35] Hanlon, W. Walker. 2017. Temporary Shocks and Persistent Effects in Urban Economies: Evidence from British Cities after the U.S. Civil War. *Review of Economics and Statistics* 99: 67-79.

[36] Harris, Chauncy D. 1954. The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers* 44: 315-348.

[37] Hornung, Erik. 2015. Railroads and Growth in Prussia. *Journal of the European Economic Association* 13: 699-736.

[38] Huebner, G. G. 1915. Economic aspects of the Panama Canal. *The American Economic Review* 5: 816-829.

[39] Hugot, Jules, and Camilo Umana Dajud. 2016. Trade Costs and the Suez and Panama Canals. CEPII Working Paper No 2016-29.

[40] Hutchinson, William K., and Ricardo Ungo. 2004. Social Saving of the Panama Canal. University of Vanderbilt Working Paper No. 04-W23, 2004.

[41] Jedwab, Remi, and Alexander Moradi. 2016. The Permanent Effects of Transportation Revolutions in Poor Counties: Evidence from Africa. *Review of Economics and Statistics* 98: 268-284.

[42] Lipsey, Robert E. 1994. US Foreign Tade and the Balance of Payments, 1800-1913. NBER Working Paper No. 4710.

[43] Maddison, Angus. 2007. *Contours of the world economy 1-2030 AD: Essays in macro-economic history*. Oxford University Press.

[44] Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles. 2018. IPUMS National Historical Geographic Information System: Version 13.0 [Database]. Minneapolis: University of Minnesota. `http://doi.org/10.18128/D050.V13.0`

[45] Maurer, Noel, and Carlos Yu. 2010. *The Big Ditch: How America Took, Built, Ran, and Ultimately Gave Away the Panama Canal*. Princeton University Press.

[46] Maurer, Noel, and Carlos Yu. 2008. What TR Took: The Economic Impact of the Panama Canal, 1903-1937. *Journal of Economic History* 68: 686-721.

[47] McCullough, David G., and Edward Herrmann. 1977. *The Path Between the Seas: the Creation of the Panama Canal, 1870-1914*. New York: Simon and Schuster.

[48] Michaels, Guy, Ferdinand Rauch, and Steven Redding. 2012. Urbanization and Structural Transformation. *Quarterly Journal of Economics* 127: 535-586.

[49] Michaels, Guy, and Ferdinand Rauch. 2017. Resetting the Urban Network: 117-2012. *Economic Journal* 128, no. 608: 378-412.

[50] Moretti, Enrico. 2010. Local Multipliers. *American Economic Review: Papers & Proceedings*. Vol. 100.

[51] Panama Canal Authority, 2018a, 'Top 15 Countries by Origin and Destination of Cargo', report obtained from https://www.pancanal.com/ on January 20, 2019.

[52] Panama Canal Authority, 2018b, 'Panama Canal Traffic Along Principal Trade Routes', report obtained from https://www.pancanal.com/ on January 20, 2019.

[53] Pascali, Luigi. 2017. The Wind of Change: Maritime Technology, Trade and Economic Development. *American Economic Review* 107: 2821-2854.

[54] Potlogea, Andrei V. 2018. Globalization and the Skilled City. *Journal of Urban Economics* 107: 1-30.

[55] Rappaport, Jordan, and Jeffrey D. Sachs. 2003. The United States as a Coastal Nation. *Journal of Economic Growth* 8: 5-46.

[56] Rappaport, Jordan. 2007. Moving to Nice Weather. *Regional Science and Urban Economics* 37: 375-398.

[57] Rauch, Ferdinand. 2016. The geometry of the Distance Coefficient in Gravity Equations in International Trade. *Review of International Economics* 24: 1167-1177.

[58] Redding, Stephen J. 2016. Goods Trade, Factor Mobility and Welfare. *Journal of International Economics* 101: 148-167.

[59] Redding, Stephen J., and Daniel M. Sturm. 2008. The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review* 98: 1766-97.

[60] Redding, Stephen, and Anthony J. Venables. 2004. Economic geography and international inequality. *Journal of international Economics* 62: 53-82.

[61] Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. IPUMS USA: Version 9.0 [dataset]. Minneapolis, MN: IPUMS, 2019. https://doi.org/10.18128/D010.V9.0

[62] Secretary of Treasury. 1880. *Statistical Abstract of the United States 1880*. Washington: Government Printing Office.

[63] Secretary of Treasury. 1900. *Statistical Abstract of the United States 1900*. Washington: Government Printing Office.

[64] Simonovska, Ina, and Michael E. Waugh. 2014. The Elasticity of Trade: Estimates and Evidence. *Journal of International Economics* 92: 34-50.

[65] Treasury Department. 1913. *The Tariff Act of October 3, 1913 on Imports into the United States*. Washington: Government Printing Office.

[66] Umana Dajud, Camilo 2017. Domestic transport costs, Canada, and the Panama Canal. CEPII Working Paper, No 2017-02.

[67] van Dijk, Jasper Jacob. 2015. Local multipliers in United States Cities: a Replication of Moretti (2010). University of Oxford, Department of Economics Working Papers, 771.

[68] van Dijk, Jasper Jacob. 2017. Local Employment Multipliers in US Cities. *Journal of Economic Geography* 17: 465-487.

		Mean	Standard deviation	N
$MA^{dom}\|NoCanal$	Domestic MA, no canal	57,744	18,065	2,425
$MA^{dom}\|Canal$	Domestic MA, canal	58,141	17,614	2,425
$MA^{int}\|NoCanal$	International MA, no canal	594,439	111,235	2,425
$MA^{int}\|Canal$	International MA, canal	635,756	144,176	2,425
ΔMA^{dom}	Domestic MA change	1.016	0.046	2,425
ΔMA^{int}	International MA change	1.063	0.046	2,425
ΔMA^{tot}	Total MA change	1.058	0.043	2,425
pop_{1880}	1880 population	20,598	43,188	2,425
pop_{1900}	1900 population	31,073	80,524	2,425
pop_{1920}	1920 population	42,093	122,008	2,425
pop_{1940}	1940 population	52,671	166,068	2,425
pop_{1960}	1960 population	71,932	245,521	2,425
pop_{1980}	1980 population	90,641	301,078	2,425
pop_{2000}	2000 population	113,039	378,513	2,425

Table 1: Summary statistics for the county dataset. Market access values are computed using parameters $\theta = -1$ and $\alpha = 1/10$. In this table, the abbreviation MA stands for *market access, Δ indicates annualized growth rates.*

	(1) 1900-1940	(2) 1900-1940	(3) 1900-2000	(4) 1900-2000
$\ln \Delta MA^{tot}$	0.144***	0.154***	0.115***	0.071***
	(0.026)	(0.042)	(0.023)	(0.026)
$\ln MA_{No\ Canal}$	-0.016***	-0.016***	0.000	0.003
	(0.006)	(0.006)	(0.004)	(0.003)
State FE	No	Yes	No	Yes
Observations	2,425	2,425	2,425	2,425
Clusters	47.000	47.000	47.000	47.000

Table 2: Difference results. $\Delta pop_{1900-1940}$ denotes annualized population growth from 1900 to 1940. The right hand side variables show change of total market access due to the Panama Canal as well as market access levels. Regressions control for longitude, latitude, log population in 1900 and state fixed effects as indicated. Robust standard errors are clustered using a chessboard grid.

	(1)	(2)	(3)	(4)	(5)	(6)
	1900-1940	1900-2000	1900-1940	1900-2000	1900-1940	1900-2000
$\ln \Delta MA_{dom}$	0.012***	0.011**			0.006	-0.005
	(0.003)	(0.004)			(0.008)	(0.006)
$\ln MA_{dom}$	-0.010***	0.001				
	(0.003)	(0.002)				
$\ln \Delta MA_{int}$			0.155***	0.071**	0.104**	0.067**
			(0.044)	(0.027)	(0.049)	(0.030)
$\ln MA_{int}$			-0.016***	0.003	-0.007	0.005
			(0.006)	(0.003)	(0.008)	(0.005)
Observations	2,425	2,425	2,425	2,425	2,425	2,425

Table 3: Separate for domestic and international. $\Delta pop_{1900-1940}$ denotes annualized population growth from 1900 to 1940. The right hand side variables show change of domestic and international market access due to the Panama Canal as well as market access levels. Regressions control for longitude, latitude, log population in 1900 and state fixed effects. Robust standard errors are clustered using a chessboard grid.

	(1)	(2)	(3)	(4)	(5)	(6)
			$\ln \Delta$ pop			
	1880-1900	1900-1920	1920-1940	1940-1960	1960-1980	1980-2000
$\ln \Delta$ popPanel A: No state FE						
$\ln \Delta MA^{tot}$	-0.032	0.167***	0.120***	0.092***	0.094***	0.082***
	(0.060)	(0.038)	(0.024)	(0.034)	(0.029)	(0.030)
$\ln \Delta$ popPanel A: State FE						
$\ln \Delta MA^{tot}$	0.033	0.229***	0.089*	0.099**	0.032	-0.028
	(0.077)	(0.057)	(0.045)	(0.054)	(0.036)	(0.036)
Observations	2,425	2,425	2,425	2,425	2,425	2,425
Clusters	47	47	47	47	47	47

Table 4: Results at 20 year intervals. Each coefficient is from a different regression. All regressions control for longitude, latitude, and market access levels. Standard errors, clustered on a 5x5 degree grid, in parentheses.

	1880-1900	1900-1920	1920-1940	1880-1900	1900-1920	1920-1940
Panel A: Service sector employment growth						
$\ln \Delta MA^{tot}$	-0.143**	0.167***	0.180***	-0.011	0.148**	0.170***
	(0.063)	(0.055)	(0.028)	(0.116)	(0.068)	(0.064)
Panel B: Manufacturing sector employment growth						
$\ln \Delta MA^{tot}$	0.082	0.358***	0.113**	-0.086	0.356***	-0.081
	(0.086)	(0.096)	(0.051)	(0.115)	(0.138)	(0.105)
Panel C: Agricultural sector employment growth						
$\ln \Delta MA^{tot}$	0.019	0.090**	-0.048	0.124	0.143*	-0.012
	(0.057)	(0.043)	(0.030)	(0.081)	(0.072)	(0.052)
State FE	No	No	No	Yes	Yes	Yes
N	2,417	2,416	2,416	2,424	2,424	2,404
Clusters	47	47	47	47	47	47

Table 5: Results for different sectors at 20 year intervals. Each coefficient is from a different regression. All regressions control for longitude, latitude, log initial population, and market access levels. These regressions also control for employment shares by sector in the initial year. Standard errors, clustered on a 5x5 degree grid, in parentheses.

	Manuf wage growth		Land value growth		Immigrant growth	
	(1)	(2)	(3)	(4)	(5)	(6)
	1900-1940	1900-1940	1900-1940	1900-1940	1900-1940	1900-1940
$\ln \Delta MA^{tot}$	0.036***	-0.018	0.117***	0.106***	0.270***	0.056**
	(0.012)	(0.023)	(0.028)	(0.039)	(0.038)	(0.025)
$\ln MA_{no\ canal}$	-0.001	0.000	-0.009	-0.011**	-0.012**	-0.000
	(0.002)	(0.002)	(0.006)	(0.005)	(0.005)	(0.004)
State FE	No	Yes	No	Yes	No	Yes
Observations	2,425	2,425	2,425	2,425	2,425	2,425

Table 6: The left hand side variable is annualized population growth for manufacturing wages (Columns (1) and (2)), land values (Columns (3) and (4)) and the share of population born outside of the state (Columns (5) and (6)). The right hand side variables show change of total market access due to the Panama Canal as well as market access levels. Regressions control for longitude, latitude, log population in 1900 and state fixed effects as indicated. Robust standard errors are clustered using a chessboard grid.

	Coastal states				Inland states			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1900-1940	1900-1940	1900-2000	1900-2000	1900-1940	1900-1940	1900-2000	1900-2000
$\ln \Delta MA^{tot}$	0.010	0.100	0.060**	0.063**	0.246***	0.317***	0.190***	0.071
	(0.067)	(0.059)	(0.028)	(0.030)	(0.052)	(0.071)	(0.047)	(0.066)
$\ln MA_{no\ canal}$	0.003	-0.007	0.011**	0.005	-0.030***	-0.026***	-0.010**	-0.002
	(0.007)	(0.007)	(0.004)	(0.004)	(0.008)	(0.008)	(0.005)	(0.003)
State FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1,104	1,104	1,104	1,104	1,321	1,321	1,321	1,321

Table 7: Results for coastal states (Columns 1-4) and non-coastal states (Columns 5-8) only. All regressions control for longitude, latitude, log initial population, and market access levels. Standard errors, clustered on a 5x5 degree grid, in parentheses.

	(1)	(2)
	Δ ln(pop) 1900-1940	
Panel A: Baseline specification		
$\ln \Delta MA^{tot}$	0.144***	0.154***
	(0.026)	(0.042)
Panel B: Additional Controls		
$\ln \Delta MA^{tot}$	0.082**	0.142***
	(0.037)	(0.045)
Panel C: International trade via three major ports		
$\ln \Delta MA^{tot}$	0.111***	0.026
	(0.029)	(0.059)
Panel D: International trade via ten major ports		
$\ln \Delta MA^{tot}$	0.126***	0.019
	(0.028)	(0.060)
Panel E: Adding a fixed transshipment cost		
$\ln \Delta MA^{tot}$	0.148***	0.158***
	(0.025)	(0.040)
Panel F: Adding a fixed tariff cost (for international routes)		
$\ln \Delta MA^{tot}$	0.152***	0.160***
	(0.023)	(0.038)
Panel G: Adding a fixed toll cost (for using Panama Canal routes)		
$\ln \Delta MA^{tot}$	0.163***	0.185***
	(0.029)	(0.048)
Panel H: Euclidean distance domestic cost matrix		
$\ln \Delta MA^{tot}$	-0.049	-0.181**
	(0.038)	(0.073)
Panel I: Cost-based approach		
$\ln \Delta MA^{tot}$	0.166***	0.181***
	(0.028)	(0.045)
Panel J: Penalizing international market access		
$\ln \Delta MA^{tot}$	0.029***	0.013***
	(0.010)	(0.004)
Panel K: $\theta = -1.82$		
$\ln \Delta MA^{tot}$	0.071**	-0.018
	(0.028)	(0.039)
State FE	No	Yes

Table 8: Robustness checks. Each regression is for 2,425 observations and uses 47 clusters. All regressions control for longitude, latitude, log market access levels, and log population in 1900. Standard errors, clustered on a 5x5 degree grid, in parentheses.

	(1)	(2)	(3)	(4)
	First stage		Second stage	
	$\ln \Delta MA^{tot}$	$\ln \Delta MA^{tot}$	$\Delta pop_{1900-1940}$	$\Delta pop_{1900-1940}$
$\ln P_{dist}$	-0.077***	-0.087***		
	(0.017)	(0.009)		
$\ln \Delta MA^{tot}$			0.179***	0.112**
			(0.061)	(0.054)
$\ln MA_{nocanal}$	-0.011	-0.099***	-0.021**	-0.013**
	(0.042)	(0.022)	(0.010)	(0.006)
State FE	No	Yes	No	Yes
F-test of excluded instrument	19.6	95.7		
Observations	2,425	2,425	2,425	2,425

Table 9: IV results. First stage in Columns (1) and (2), second stage in Columns (3) and (4). All regressions control for longitude, latitude, log initial population, and market access levels. Standard errors, clustered on a 5x5 degree grid, in parentheses.

Figure 1: Market access impact of the Panama Canal conditional on latitude, longitude, market access level and state fixed effect. The scale is in terms of 20 bins of equal size. Darker pixels indicate areas that benefit less from the Panama Canal.

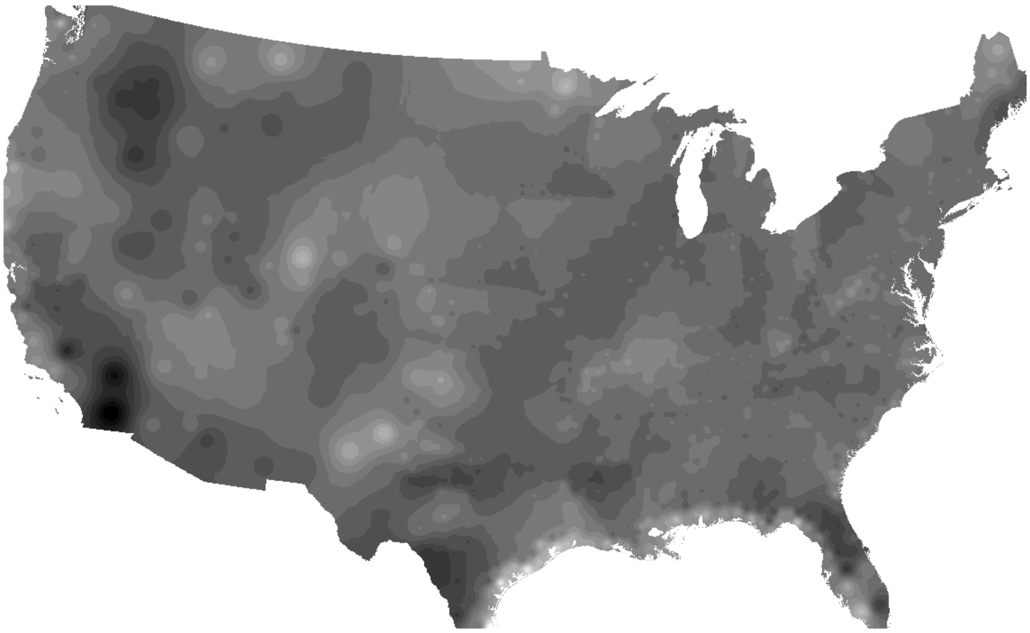


Figure 2: Effect by initial population. The graph shows regression coefficient of ten indicator variables for ten initial population deciles interacted with the main treatment effect. The gray lines show 95 percent confidence intervals.

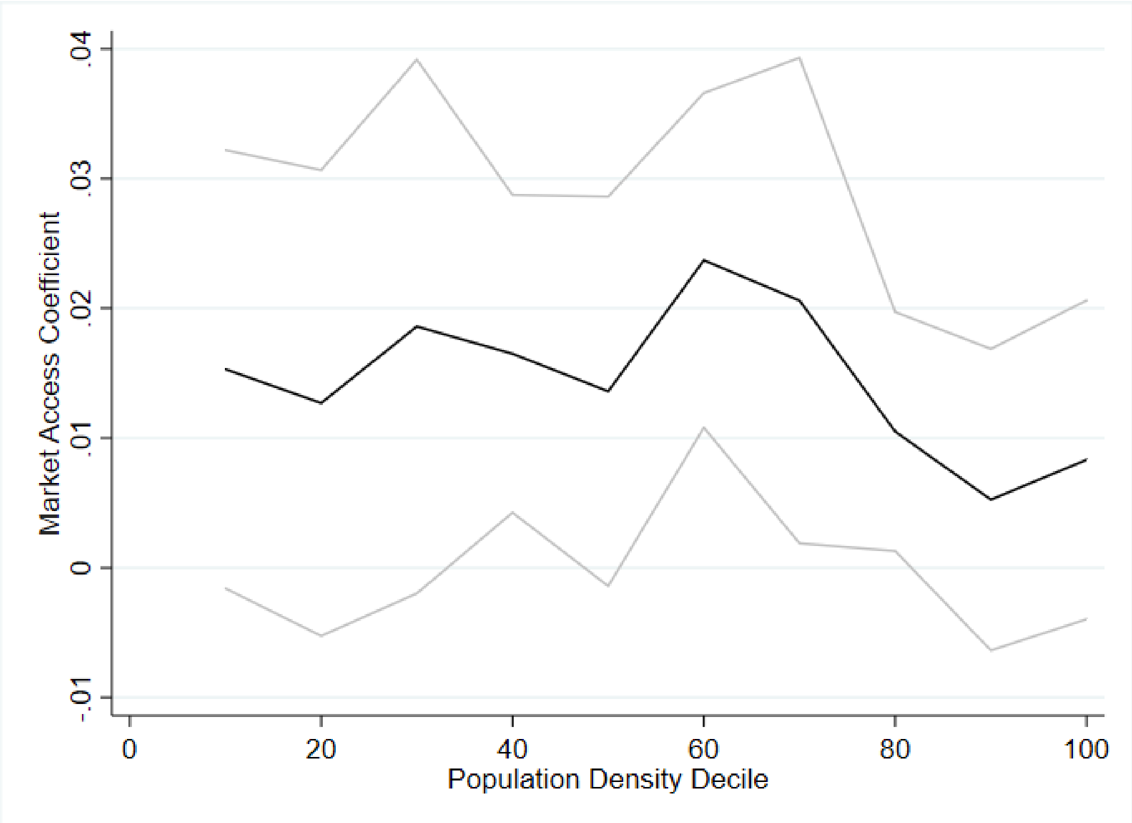


Figure 3: Effect by market access change. The graph shows regression coefficient of ten indicator variables for ten treatment intensity deciles interacted with the main treatment effect. The gray lines show 95 percent confidence intervals.

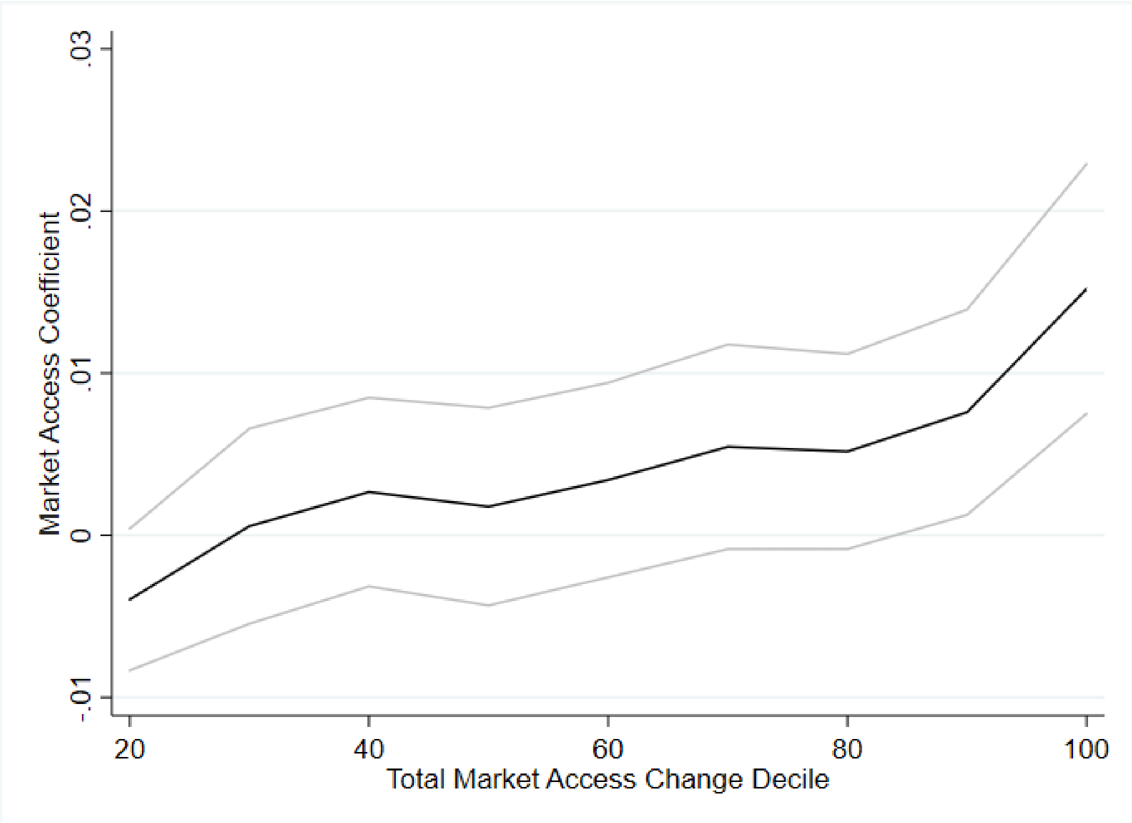
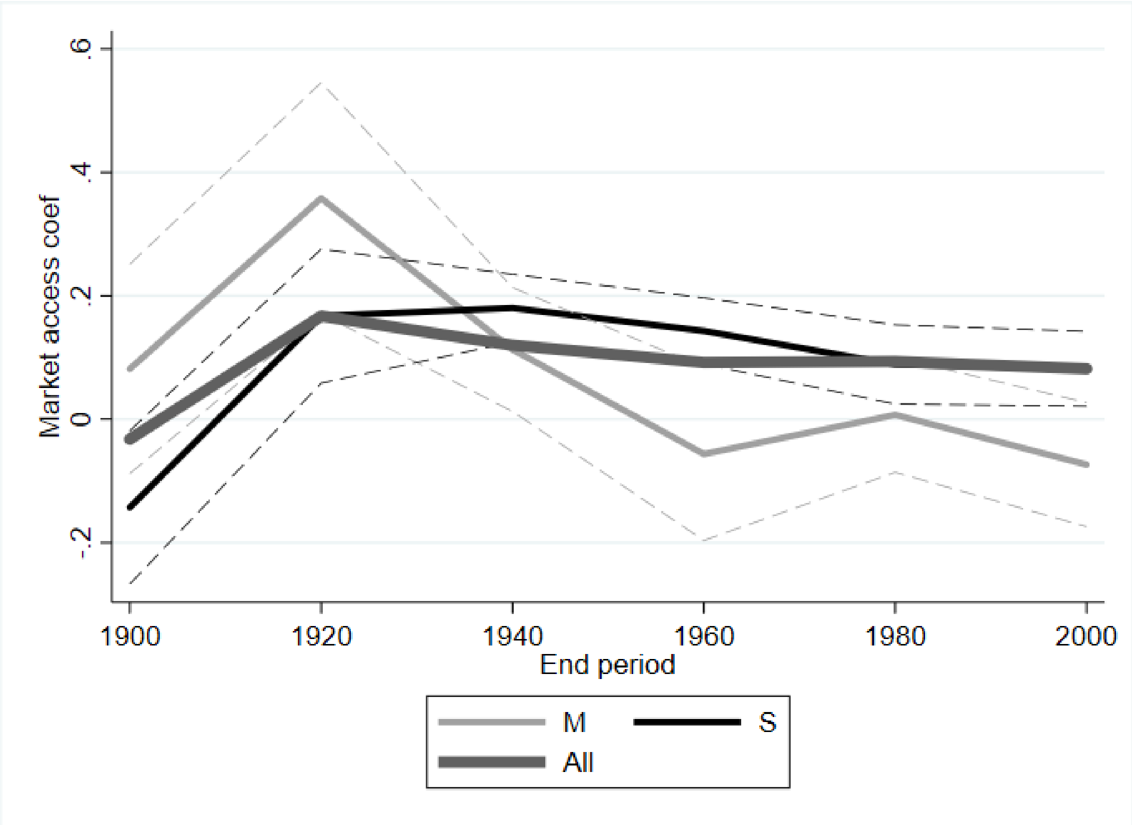


Figure 4: Effect by industry. The graph shows annualized growth rates for 20 year intervals for total population, manufacturing and services. The dashed lines indicate 95 percent confidence intervals for the two sectors.



8 Appendix: Alternative Welfare Calculations

In this paper, we use the framework by Arkolakis et al (2012) to calculate a rough estimate of the canal's welfare gains. One shortcoming of this approach is that it does not account for welfare changes from reallocation within the country, such as we document in this paper. Such population reallocations are included in the model of Redding (2016), who provides the following formula for the welfare gains from trade in a country with n subregions (e.g. counties):

$$\frac{U^T}{U^A} = \left[\sum_n \frac{L_n}{L} \left(\frac{\pi_{nn}^A}{\pi_{nn}^T} \right)^{\frac{\alpha\epsilon}{\theta}} \left(\frac{L_n^A}{L_n^T} \right)^{\epsilon(1-\alpha)} \right]^{\frac{1}{\epsilon}}.$$

In this equation, A indicates variables under autarky, T in the world with trade, π_{nn} is the share of a county's expenditures that are spent within the county, L is population. $1-\alpha$ is the share of consumption that goes to land, ϵ denotes the elasticity of population with respect to real income, and θ the elasticity of trade flows with respect to trade costs (not to be confused with the elasticity of trade flows with respect to effective distance).

While this equation allows to incorporate welfare effects due to population reallocation, it is more difficult to take to our data. Firstly, as before, we need an estimate of the within-county expenditure share. As in the main part of the paper, we assume that we can use market access values to back these out and define the within-county expenditure shares as $\pi_{nn}^A = 1 - \frac{MA_{int}^{nocanal} + MA_{dom}^{nocanal}}{pop1900}$ and $\pi_{nn}^T = 1 - \frac{MA_{int}^{canal} + MA_{dom}^{canal}}{pop1900}$, respectively.[19]

Secondly, we need an estimate of $\frac{L_n^A}{L_n^T}$, the inverse of the population growth due to the market access change. As in our empirical analysis, we consider the year 1940 as the post-canal year of observation. However, we cannot simply use the

[19] As in our previous welfare calculation, throughout this exercise, we penalize international market acess such that 1 minus the ratio of international market access to population on average matches the historical import-to-GDP ratio of 4.6%.

ratio of the 1900 to the 1940 population, as population might have changed for reasons unrelated to market access. Instead we use two different approaches to estimate the counterfactual population in 1940 that isolate the effect of the Panama Canal on population changes. Firstly, we use our estimated relationship between market access changes (conditional on controls) and annualized growth rates between 1900 and 1940, denoted by $\hat{\beta}_1^{growth}$. For each county c, we can then calculate the counterfactual population, had only market access changed, as

$$L_c^T = L_c^{1900} \cdot exp(\hat{\beta}_1^{growth} \cdot \ln \Delta MA_c \cdot 40)$$

As a second, more direct approach, we estimate the effect of market access changes (conditional on controls) on 1940 population levels ($\hat{\beta}_1^{level}$)and then set the counterfactual 1940 population as

$$L_c^T = L_c^{1900} + \hat{\beta}_1^{level} \cdot \ln \Delta MA_c$$

For the parameter values, we follow Redding (2016) and set α=0.75, $\epsilon = 3$ and $\theta = 4$.

One final complication is that the welfare calculation in Redding (2016) allows for internal, but not international migration, such that the national population remains constant. In our setting, this is not the case, counterfactual 1940 populations by this approach will be larger than the 1900 ones. To adress this, we normalize the counterfactual 1940 populations such that their sum equals the sum of the 1900 populations. When we plug the resulting numbers into the welfare formula, an interesting result is that the welfare changes are sensitive to how we predict the 1940 population. If we use the level regression, we find that the canal led to a 1.23% increase in welfare by 1940. If we use growth rate regressions instead, the welfare change is only 0.33%. While both numbers are positive, they are below the ones found under the Arkolakis et al (2012) framework, which we estimated to be 0.3-0.4% per year. This could reflect the fact that population reallocation and the resulting land price changes decrease the welfare gains relative to the non-migration scenario. However, there are also several caveats to this exercise. Firstly, as before, we have used market access at 1900 population levels to estimated within-county expenditure shares. This was relatively inconsequential for the Arkolakis et al (2012) welfare calculation, as it can be interpreted as a short-run gain. In the present case, it becomes more grave, as we use 1940 as the post-canal benchmark. By not accounting for the population (and hence market access) growth between 1900 and 1940, we underestimate the change in within-county expenditures due to the canal, and thus welfare. Secondly, it also has to be borne in mind that this calculation allows migration within the US, but not into the United States. Finally, the model compares a world of autarky with a world of trade, while we approximate it with a world of two different trade intensities, corresponding to the data.