TRADE, STRUCTURAL TRANSFORMATION AND DEVELOPMENT:
EVIDENCE FROM ARGENTINA 1869-1914

Pablo Fajgelbaum
Stephen J. Redding

Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914
Pablo Fajgelbaum and Stephen J. Redding
NBER Working Paper No. 20217
June 2014, Revised September 2021
JEL No. F11,F14,O13,O14

ABSTRACT

We provide new theory and evidence on the role of external and internal integration in structural transformation and economic development using Argentina's integration into the world economy in the late-19th century. Our theoretical model provides microfoundations for a spatial Balassa-Samuelson effect, in which locations closer to world markets have higher population densities, urban population shares, relative prices of non-traded goods, and land prices relative to wages, and specialize in transport-cost-sensitive traded goods. We estimate the model's parameters, provide evidence in support of this spatial Balassa-Samuelson mechanism, and find substantial effects of both external and internal integration on economic development.

Pablo Fajgelbaum
Princeton University
Department of Economics
International Economics Section
Room 294 Julis Romo Rabinowitz Bldg
Princeton, NJ 08544
and NBER
pfajgelb@princeton.edu

Stephen J. Redding
Department of Economics &
School of Public and International Affairs
Princeton University
Princeton, NJ 08544
and CEPR
and also NBER
reddings@princeton.edu

1 Introduction

The relationship between economic development and international trade is central to the fields of both international economics and development economics. To what extent does international trade promote structural transformation and economic development? What is the spatial incidence of international trade shocks? What role do internal trade costs play in the transmission of these international trade shocks?

We provide new theory and evidence on these questions using Argentina's integration into the world economy in the late-19th century as a natural experiment. We use the large-scale variation in external integration (from reductions in transatlantic freight rates) and internal integration (from the construction of the railroad network) in this empirical setting. First, we provide reduced-form evidence of a *spatial Balassa-Samuelson effect* as a key feature of the relationship between structural transformation, economic development, and trade. Using a newly-constructed and spatially-disaggregated dataset for Argentina from 1869-1914, we show that locations with better access to world markets have higher population densities, urban population shares, relative prices of non-traded goods, and land prices relative to wages, and specialize in the most trade-cost-sensitive traded goods. Therefore, these locations not only have higher overall levels of economic activity, but also experience structural transformation between the traded and non-traded sectors and across goods within the traded sector.

Second, we develop a new theoretical model of the spatial distribution of economic activity across sectors and locations that provides microeconomic foundations for this spatial Balassa-Samuelson effect. We consider a neoclassical specification of preferences and production, in which consumers have inelastic demand between traded and non-traded goods, labor is geographically mobile, locations within Argentina have a comparative advantage in agriculture, and agriculture is land-intensive relative to the non-traded sector. Within the agricultural sector, we allow for different disaggregated goods, where locations can differ in terms of their productivity and transport costs for these disaggregated goods, and the goods themselves can have different sensitivities to transport costs. We show that locations with low transport costs to world markets are attractive for the production of traded goods, which increases population density. For sufficiently high population mobility, this increase in population density bids up the reward of the immobile factor (land) relative to the mobile factor (labor). This increase in population density and reduction in the wage-rental ratio together imply an expansion in the employment share of the labor-intensive non-traded sector, which with inelastic demand requires a rise in the relative price of the non-traded good. As these locations close to world markets have high relative export prices for the most-transport-cost-sensitive goods, they also specialize in these disaggregated goods within the agricultural sector.

Third, we use our rich, spatially-disaggregated data for Argentina to structurally estimate the parameters of our model. We find parameter estimates that are in line with central values in the existing empirical literature. We estimate inelastic demand between sectors, with an elasticity of substitution between traded and non-traded goods of 0.49. We find substantial population mobility, with estimated elasticities of population with respect to real income of 4.73 across locations within Argentina and 2.02 between Argentina and the rest of the world. We estimate substantial heterogeneity in idiosyncratic productivity across disaggregated goods within the traded sector, with an estimated elasticity of revenue shares within the traded sector with respect to relative prices of 3.18. We show that the model has good within-sample fit for the targeted moments. We also report a number of overidentification checks, in which we demonstrate that the model also has predictive power for non-targeted moments.

Fourth, we use our estimated model to undertake counterfactuals for the impacts of external integration (reductions in transatlantic freight rates) and internal integration (the construction of the railroad network). We find that reductions in transatlantic freight rates from 1869-1914 raised Argentina's gross domestic product (GDP), population and welfare by 17.7, 13.8 and 7.1 percent, respectively. By comparison, the construction of the railroad network increased GDP, population and welfare by 12.8, 9.4 and 4.8 percent respectively. This expansion of economic activity from railroad construction raises land income by around 6.5 percent of 1914 gross domestic product (GDP), including effects on both the agricultural and non-traded sectors. We find that the resulting increase in the net present value of land income substantially exceeds historical estimates of the railroad's construction costs. Therefore, these large-scale investments in transport infrastructure during the 19th-century can be rationalized in terms of their impact on economic activity. We find higher ratios of net present values of land income to construction costs at the levels of external integration in 1914 than at those in 1869. Intuitively, while the railroad construction costs are fixed, the absolute increase in the level of economic activity from the construction of the railroad network is larger at the higher levels of external integration in 1914.

Our empirical setting has a number of attractive features for examining the relationship between economic development and international trade. First, Argentina's integration into the world economy was driven by late-19th-century reductions in transatlantic transport costs following the invention of the steamship. This new technology was first developed for river transport in Europe and the United States, and hence is plausibly exogenous to Argentina's peripheral location. Second, we have disaggregated data on economic activity across regions and sectors within Argentina over a long historical time period, which enables us to quantify the role of structural transformation in economic development. We find that Argentina's 19th-century export boom was characterized by a high level of commodity specialization, as in many developing countries today, with agriculture accounting for over 99 percent of export value. We show that its rapid economic development involved large-scale structural transformation, both between the agricultural and non-traded sectors, and within the agricultural sector. This structural transformation within the traded sector is reflected in the emergence of major new sources of comparative advantage, with the share of cereals in exports rising from zero in 1869 to around 50 percent in 1914. Third, the invention of steam railroads lowered inland transport costs, which enables us to examine the interaction between internal and external integration in shaping economic development. In estimating the impact of this internal integration, we use the historical context of Spanish colonial rule and Argentina's late-19th-century integration into world markets to construct instruments for the railroad network to address the non-random placement of transport infrastructure.

Our findings of welfare gains from the large-scale external integration that occurred in late-19th century Argentina of 7.1 percent are comparable with conventional estimates of the welfare gains from trade. For example, Bernhofen and Brown (2005) estimates an upper bound for the welfare gains from trade in 19th-century Japan of 8-9 percent. Our findings of welfare gains from the construction of the railroad network of 4.8 percent are also in line with the range of existing findings in the literature. For example Fogel (1964) and Donaldson and Hornbeck (2016) estimate increases in the value of agricultural land of 2.7 and 3.2 percent of GDP respectively, while Donaldson (2018) finds an increase in agricultural real income of 16 percent, and Hornbeck and Rotemberg (2019) estimates an increase of up to 28 percent in real GDP, once changes in manufacturing productivity from reduced misallocation are taken into account. Our framework highlights the role of international population mobility, which dampens the impact of the construction of the railroad network on welfare, because the induced population inflow bids up the price of land. Therefore, while

welfare rises by 4.8 percent, real GDP increases by 12.8 percent, in part because of this population inflow, and land income increases by 6.5 percent of 1914 GDP (including effects through both the agricultural and non-traded sectors). More generally, our framework emphasizes the role of structural transformation, both between the agricultural and non-traded sectors, and within the agricultural sector.

Our paper is related to a number of different strands of research. First, our work contributes to the macroeconomic literature on structural transformation, including Matsuyama (1992), Caselli and Coleman (2001), Ngai and Pissarides (2007), Matsuyama (2009), Herrendorf, Schmitz, and Teixeira (2012), Uy, Yi, and Zhang (2012), Michaels, Rauch, and Redding (2012), Lagakos and Waugh (2013), Gollin and Rogerson (2014), Herrendorf, Rogerson, and Valentinyi (2014), Bustos, Caprettini, and Ponticelli (2016), Gollin, Jedwab, and Vollrath (2016), Bustos, Garber, and Ponticelli (2020), McMillan, Rodrik, and Sepulveda (2017), Eckert and Peters (2018), and Karádi and Koren (2018). Another related strand of the literature in macro-development considers the impact of population changes on the local economic structure, including Burstein, Hanson, Tian, and Vogel (2020) and Peters (2019). We make two main contributions relative to this line of work. First, whereas most existing macroeconomics research focuses on the aggregate economy, our analysis emphasizes the role of internal geography and transport costs in shaping structural transformation and the Balassa-Samuelson effect. Second, we use the natural experiment of Argentina's late-19th-century integration into world markets and disaggregated data by sector and region over a long historical time period to provide quantitative evidence on the role of this structural transformation in the process of economic development.

Second, our paper is related to research on economic geography in the international trade literature, including Hanson (1996), Fujita, Krugman, and Venables (1999), Venables and Limao (2002), Davis and Weinstein (2002), Redding and Sturm (2008), Allen and Arkolakis (2014), Coşar and Fajgelbaum (2016), Ramondo, Rodriguez-Clare, and Saborio (2016), Redding (2016), Nagy (2017), Morten and Oliveira (2017), Caliendo, Parro, Rossi-Hansberg, and Sarte (2018), Caliendo, Dvorkin, and Parro (2019), Davis and Dingel (2019), and Sotelo (2020), as reviewed in Redding (2020). While most of this research focuses on the overall level of economic activity, our work highlights the role of internal geography in shaping the composition of economic activity between traded and non-traded sectors and across disaggregated goods within the traded sector.

Third, a growing empirical literature has examined the relationship between economic activity and transport infrastructure, including Chandra and Thompson (2000), Baum-Snow (2007), Michaels (2008), Berlinski, Galiani, and Jaitman (2011), Banerjee, Duflo, and Qian (2021), Duranton and Turner (2012), Duranton, Morrow, and Turner (2014), Faber (2014), Atkin and Donaldson (2015), Coşar and Demir (2016), Storeygard (2016), Baum-Snow, Brandt, Henderson, Turner, and Zhang (2017), Donaldson (2018), and Heblich, Redding, and Sturm (2020), as reviewed in Redding and Turner (2015). Using a market-access based approach, Donaldson and Hornbeck (2016) quantifies the effect of the U.S. railroad network on the aggregate value of U.S. agricultural land in 1890, while Hornbeck and Rotemberg (2019) provides evidence on its impact on manufacturing productivity through changes in misallocation. In contrast to these studies, our focus is on the reallocation of economic activity from the agricultural to the non-traded sector, and within the agricultural sector. We rationalize our empirical findings using a new theoretical model of structural transformation across sectors and locations, which provides microeconomic foundations for our spatial Balassa-Samuelson effect.

Fourth, we build on the historical literature on Argentine economic development, including Scobie (1971), Taylor (1992), Cortés Conde (1993), Adelman (1994), and Francis (2017). We construct a new spatially-disaggregated dataset

on economic activity in Argentina over our long historical time period. We combine this new dataset with a quantitative spatial general equilibrium model to provide new evidence on the role of the spatial Balassa-Samuelson effect for the relationship between trade, structural transformation and economic development.

The remainder of the paper is structured as follows. Section 2 provides some historical background. Section 3 introduces our data sources and definitions. Section 4 presents reduced-form evidence on a number of stylized facts about patterns of economic development in 19th-century Argentina. Section 5 develops our theoretical model and derives its key prediction of the spatial Balassa-Samuelson effect that we use to rationalize these stylized facts. Section 6 structurally estimates the model's parameters. Section 7 undertakes counterfactuals for external and internal integration and compares the change in the net present value of land income from the expansion of the railroad network to historical estimates of its construction costs. Finally, Section 8 concludes. A separate web appendix collects together technical derivations, theoretical extensions, and supplementary empirical results.

2   Historical Background

The area that makes up present-day Argentina was first settled by Europeans in the early-sixteenth century. During this period of Spanish colonial rule, economic activity was centered around the silver mines in neighboring Bolivia.[1] Reflecting this orientation, official trade routes ran towards the Northwest through Panama, and trade was monopolized by Spanish merchants. In contrast, the Eastern coastal regions of Argentina, including Buenos Aires and the River Plate (*Río de la Plata*), were peripheral outposts for illegal trade with Brazil, Portugal and Britain.[2]

In response to the growth of this illegal trade and threats from encroaching Portuguese settlement, the Viceroyalty of the *Río de la Plata* was established in 1776 in Buenos Aires. With the decline in Spanish imperial power during the Napoleonic Wars, a local junta seized political power in 1810, which led to the first opening of direct trade with other foreign countries. After the failure of attempts to reassert Spanish colonial authority, full Argentinian independence was achieved in 1816. In the ensuing decades, there followed a gradual process of political consolidation, with the first national constitution agreed in 1853, the first constitutional government of all provinces meeting in 1862, and Buenos Aires absorbed into the federal structure of Argentina in 1880. Over these decades, successive military campaigns against native populations culminated in the "Conquest of the Desert" of 1879-80, which opened up the hinterland of Buenos Aires to economic development.[3] Following the election of Julio Roca to the Presidency in 1880, liberal policies were pursued towards international flows of trade, capital and migrants, which were maintained until the outbreak of the First World War in 1914 drastically reduced these flows.[4]

During the late-19th century, a series of technological improvements centered on steam power dramatically reduced international transport costs. The steamship was first developed for river transportation in Europe and North America, with regular crossings of the North Atlantic by steamship beginning in 1838. Following improvements in the speed, reliability and capacity of these steamships, international freight rates across the North Atlantic fell by around 1.5 percent per annum from 1840 onwards, with a cumulative decline of around 70 percentage points from

---

[1]For historical discussions of Argentine development, see for example Adelman (1994) and Scobie (1971).

[2]Early settlement patterns were heavily influenced by the availability of passive native Indian populations under the feudal *encomienda* system. Interior towns were established at Asunción (1537), Santiago del Estero (1553), Mendoza (1561), San Juan (1562) and San Miguel de Tucumán (1565). In contrast, the establishment of coastal towns lagged by several decades, including Santa Fe (1573), Buenos Aires (1580), Concepción del Bermejo (1585), and Corrientes (1588).

[3]Until 1880, there were periodic incursions from hostile native populations, as examined in Droller (2018).

[4]We end our sample period in 1914 to abstract from the effects of the First World War and subsequent more interventionist government policies, as discussed for example in Taylor (1992).

1840-1914, as documented in North (1958), Harley (1988) and Pascali (2017).[5]

In Figure 1, we show the implications of these reductions in international transport costs for Argentina's export prices, import prices and terms of trade. Each of these series is the trade-weighted average of prices for individual goods and is expressed as an index that takes the value one in 1914.[6] In the 1870s and 1880s, both export and import prices fell over time, consistent with rapid technological change during this period, which resulted in relatively stable terms of trade. From the 1890s, import prices remained relatively constant over time. In contrast, export prices, and hence the terms of trade, rose sharply, consistent with reductions in international transport costs inducing price convergence between the new and old worlds. In Section A.4.1 of the online appendix, we provide further evidence on reductions in transatlantic freight rates between Argentina and European markets for a number of different commodities.

Figure 1: Argentina's Export Prices, Import Prices and Terms of Trade 1869-1914
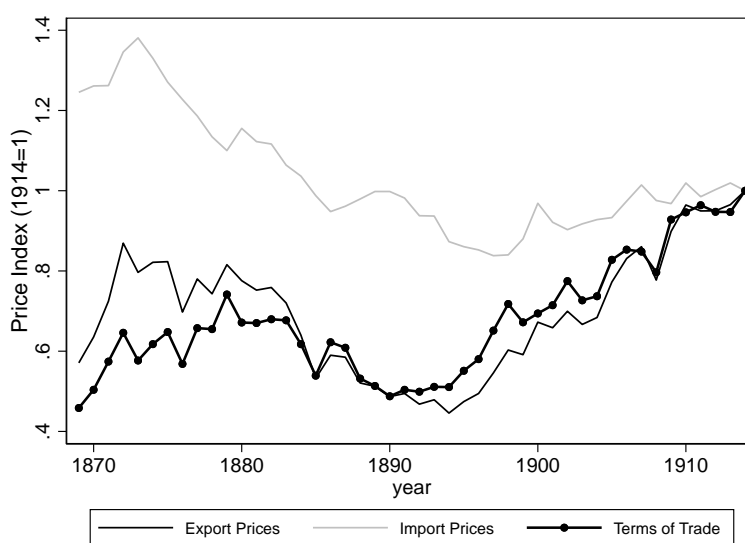
Note: Aggregate export price index; aggregate import price index; and the terms of trade (the ratio of the export price index to the import price index); price indexes computed as chained Laspeyres indexes, as discussed in Section A.6.13 of the data appendix; each series expressed as an index that is one in 1914; Source: Francis (2017).

Steam technology also revolutionized domestic transportation through the construction of railroads. The first commercial use of mobile steam locomotives was to haul freight from mines at the Stockton and Darlington railroad in the U.K. in 1825. The first railroad constructed in Argentina was the Buenos Aires Western Railroad in 1857, with around 700 kilometers of track completed by 1869. From this point onwards, the railroad network expanded rapidly to grow to around 13,000 kilometers in 1895 and 30,000 kilometers in 1914.[7] Whereas previously it had taken several months to transport goods by oxcart from Buenos Aires to an interior city such as Salta in the Northwest, the same journey could now be made in a matter of days (as discussed in Scobie 1971). Much of this railroad network was operated by private companies, which were predominantly British owned. However, these private companies operated alongside state-owned railroads, and the state influenced the development of the overall railroad network,

---

[5]These declines in freight rates were reflected in a convergence of commodity prices, with the gap between wheat prices in Liverpool and Chicago falling from 57.6 percent in 1870 to 17.8 percent in 1895 and 15.6 percent in 1913 (Harley 1980). See O'Rourke and Williamson (1999) for the seminal study of this increasing integration of the Atlantic economy.

[6]We use the export and import price indexes from Francis (2017), which use direct data on customs export prices in Argentina rather than estimates based on prices in importing countries.

[7]This rate of railroad expansion is comparable to that in the United States: between 1880 and 1913, railroad kilometers per 10,000 people rose from 9-42 in Argentina, compared to 29-44 in the United States.

through both land sales and the financing of railroads in remote areas.

With these reductions in international and domestic transport costs, Argentina experienced one of the largest booms in international trade in recorded history. Between 1869 and 1914, Argentina's real exports and imports increased by more than 500 and 200 percent, respectively. In contrast to the Spanish colonial period, this trade was now centered on the Eastern coastal regions. Following its emergence as the seat of political power, Buenos Aires rapidly developed into Argentina's main trade hub, even though its site was not particularly well suited for a port.[8] Together, Buenos Aires and the three surrounding ports of La Plata (immediately adjacent to Buenos Aires), Rosario (directly upstream) and Bahia Blanca (developed as a satellite port to alleviate congestion in Buenos Aires port) accounted for more than 75 percent of the value of exports throughout our sample period.

As in many developing countries today, Argentina's exports were characterized by a high level of commodity specialization, with agriculture accounting for over 99 percent of the value of exports throughout our sample period.[9] Historically, agriculture in the hinterland of Buenos Aires had been based on cattle ranching on large estates (*estancias*), with sheep ranching becoming more important from the late-eighteenth century onwards. The reductions in international and domestic transport costs in the late 19th-century saw large-scale changes in patterns of export specialization, as it became profitable to trade goods that were previously not tradeable. For example, the mechanical refrigeration of meat was first developed in Australia in 1861 with a view to supplying the U.K. market, and made it possible for the first time to export frozen and chilled meat from Argentina to European and US markets.

In Figure A.4 in the online appendix, we illustrate these changes in patterns of export specialization, by displaying export price indexes and export share weights for the individual goods that enter the aggregate export price index. Entirely new commodities begin to be exported over time, including in particular cereals and refrigerated and frozen beef and mutton. As a result, the export share of cattle and sheep hides falls from around 40 percent to less than 15 percent from 1869-1914. In contrast, the export share of cereals rises from zero to around 50 percent, and the export share of frozen beef rises from zero to around 10 percent.

This boom in agricultural production and exports was facilitated in part by large-scale international immigration, with Argentina's total population rising from 1.8 to 7.9 million between 1869 and 1914. Despite this substantial increase in labor supply, income per capita grew at average annual rates of 1.1 and 2.5 percent from 1869-95 and 1895-1914, respectively, based on the estimates of Taylor and Williamson (1997).[10] This rapid economic growth was accompanied by structural transformation, as the share of agriculture in employment fell by around 7 percentage points between 1869 and 1914, and the share of the population living in towns and cities rose by about 20 percentage points over the same period. By 1914, Argentina was the eighth richest country in the world, with Buenos Aires accounting for around one fifth of its overall population.

---

[8]As noted in Scobie (1971), "Ironically, the sixteenth-century Spaniards, searching for an anchorage for their tiny ships, selected one of the poorest sites imaginable in terms of 19th-century sailing vessels and steamships" (p.95). As late as the 1880s, ships had to anchor several miles from shore in the open roads, until the construction of the Madero docks in 1897.

[9]As discussed in Rocchi (2008), the limited amount of domestic manufacturing activity involved either the processing of agricultural goods for export or was orientated towards the domestic consumer goods market.

[10]Argentina is the fastest-growing country in GDP per worker in the sample of 17 countries in Taylor and Williamson (1997), which includes the richest countries of the period, such as the U.S., U.K., Australia and Canada.

3   Data

We construct a new spatially-disaggregated dataset for Argentina from 1869-1914. We combine six main sources of separate data.[11] First, we use the population censuses of 1869, 1895 and 1914 to measure the spatial distribution of economic activity across locations within Argentina. We observe total population, rural population, urban population, and geographical land area.[12] We associate rural population with employment in the agricultural sector and urban population with employment in the non-traded sector, including services and manufacturing for the local market.[13] Across the three population censuses, there are changes in the boundaries of districts and provinces, both with the geographical expansion of Argentina's frontiers from 1869-1895 and the subdivision of districts from 1895-1914. Therefore, we construct time-invariant districts and provinces based on the boundaries in the 1895 census, using the maps and concordance in Cacopardo (1967).

Second, we use detailed data on the organization of economic activity within the agricultural sector from the 1895 and 1914 population censuses. We construct agricultural land area for the following six categories of disaggregated agricultural goods: (i) Cereals as the major new export crop (Barley, Linseed, Maize (Corn), Oats, Rice, Rye, Sorghum and Wheat); (ii) Other crops; (iii) Pure and mixed-breed cattle, used predominantly for chilled and frozen meat; (iv) Native-breed cattle, used mainly for hides and skins, bones, fat and tallow; (v) Pure and mixed-breed sheep, used largely for wool or chilled and frozen mutton; (vi) Native-breed sheep, typically used for hides and skins, bones, fat and tallow. For the first two goods, we use crop cultivated area for each district, as reported in the population censuses. For the last four goods, we use the number of each type of animal for each district, and an assumed grazing area for each type of animal, as discussed further in Section A.6 of the online appendix. Additionally, we observe the number of different types of agricultural machines for each district in both 1895 and 1914, and the value of different types agricultural machines for each district in 1914.

Third, we have data on internal shipments by rail for 1895 and 1914 from the records of the Argentine railroads. We observe the total quantities loaded at each railroad station for fifteen disaggregated products: Alfalfa, Cattle, Corn, Flax, Flour, Leather, Other Live Animals, Sand and Stone, Sheep, Sugar, Sugar Cane, Wheat, Wine, Wood and Wool. We allocate railroad stations to districts using their latitude and longitude coordinates and compute the total quantity of each product loaded in each district.

Fourth, we use a variety of international trade data, including export and import values, export and import prices, and estimates of transatlantic freight rates, as measured by the ratio of "cost inclusive of freight (cif)" to "free on board" (fob) prices for 1870, 1895 and 1914.[14]

Fifth, we combine our production and trade data with a range of other geographical information. We constructed geographical information systems (GIS) shapefiles of the Argentinian railroad network in 1869, 1895 and 1914, the routes of navigable rivers, and Spanish colonial postal routes from the maps in Randle (1981). We use these shapefiles to measure a district's connection to the railroad network, and to construct two different instruments for a railroad connection, one based on connecting the centroid of every district to Argentina's trade hub, and the other based on

---

[11]See Section A.6 of the online appendix for further discussion of the data definitions and sources.

[12]We use the definition of urban population from the population census, which corresponds to the population of all cities and towns. We find similar results with an alternative definition of urban population based on the population of cities with more than 2,000 inhabitants.

[13]In 1914, the population census reports both rural population and employment in agriculture. We find a correlation coefficient of more than 0.8 between these variables, which is consistent with most rural economic activity being related to agriculture, particularly as many of those living in rural areas but not employed in agriculture are likely to have been employed in closely-related activities (such as basic agricultural processing).

[14]We use export and import prices from Francis (2017) and transatlantic freight rates from Tena-Junguito and Willebald (2020).

colonial postal routes, as discussed further in Section 4.3 below. We also use these shapefiles to measure bilateral travel times for our structural estimation in Section 6 below. We compute least-cost path measures of bilateral travel time between the centroids of districts, assuming the following weights of 1 for railroads, 3 for the coast/navigable rivers and 4.5 for land, and assuming that agents can only connect to the railroad network at stations.[15] We also construct an instrument for these bilateral travel times, in which we assume a weight of 1 for colonial postal routes, and the same weights as above for the coast/navigable rivers and land.

Sixth, we use a variety of additional sources of data, including the value of land per hectare in each district in 1895 from Dirección General de Estadística (1895), the wages of agricultural laborers for a number of districts in 1913 from República Argentina (1913), the prices of traded goods for a number of districts in 1905 from Alsina (1905), and expenditure shares on traded and non-traded goods from the aggregate household survey data reported in Bunge (1918). Using these data on prices, land values and household expenditure shares, we construct overall and tradeables consumption price indexes, as discussed further in Section A.6 of the online appendix.

The number of districts for which population census data are reported increases over time with the expansion of Argentina's geographical boundaries from 298 in 1869 to 380 in 1914. To ensure that our results are not driven by extreme values for the rural population share, we focus in our quantitative analysis of the model on a subsample of districts with rural population shares of more than 5 percent and less than 95 percent, which reduces the number of districts to 164 in 1869 and 318 in 1914. We find a similar pattern of reduced-form results, whether we consider this model subsample or the full sample of districts.

4   Reduced-Form Evidence

In this section, we provide reduced-form evidence of a large-scale change in the organization of economic activity within Argentina following the external and internal integration of the late-19th century. In Section 4.1, we establish a reorientation of economic activity away from the Spanish colonial cities of the North-West, and towards the agricultural hinterland of Buenos Aires and its surrounding ports. In Section 4.2, we provide further regression evidence on these changing gradients of economic activity with respect to geographical access to world markets. We find evidence of a spatial Balassa-Samuelson effect, where locations closer to world markets have higher population densities, urban population shares, relative prices of non-traded goods, and land prices relative to wages, and specialize in the most trade-cost-sensitive traded goods. In Section 4.3, we report instrumental variables (IV) estimates of the impact of internal integration from the construction of the railroad network on levels of economic activity.

4.1   Spatial Pattern of Economic Development

We begin by documenting the large-scale changes in the spatial distribution of economic activity within Argentina that occurred from 1869-1914. In Figure 2, we show the distribution of population density across our constant-boundary Argentinian districts in each of our census years. We divide the population density distribution in each year into the same five discrete cells, with darker shading indicating higher values. We show the railroad network in green (darker lines), the main navigable rivers (the Paraná, Plate and Uruguay rivers) in blue (lighter lines), and the customs (ports) by the red dots (solid circles).

---

[15]We use the weights estimated with 19th-century data in Donaldson (2018).

At the beginning of our sample period in 1869 (panel (a)), the main population concentrations were the Spanish colonial towns that served the mining region of Upper Peru (in the North-west) and the areas along the Paraná and Uruguay rivers and the River Plate estuary. Most of the remainder of Argentina was sparsely populated. The railroad network consisted of only 700 kilometers of track, including two lines radiating from Buenos Aires in the River Plate estuary and one line connecting the port of Rosario with the interior city of Córdoba.

Figure 2: Spatial Distribution of Population Density from 1869-1914
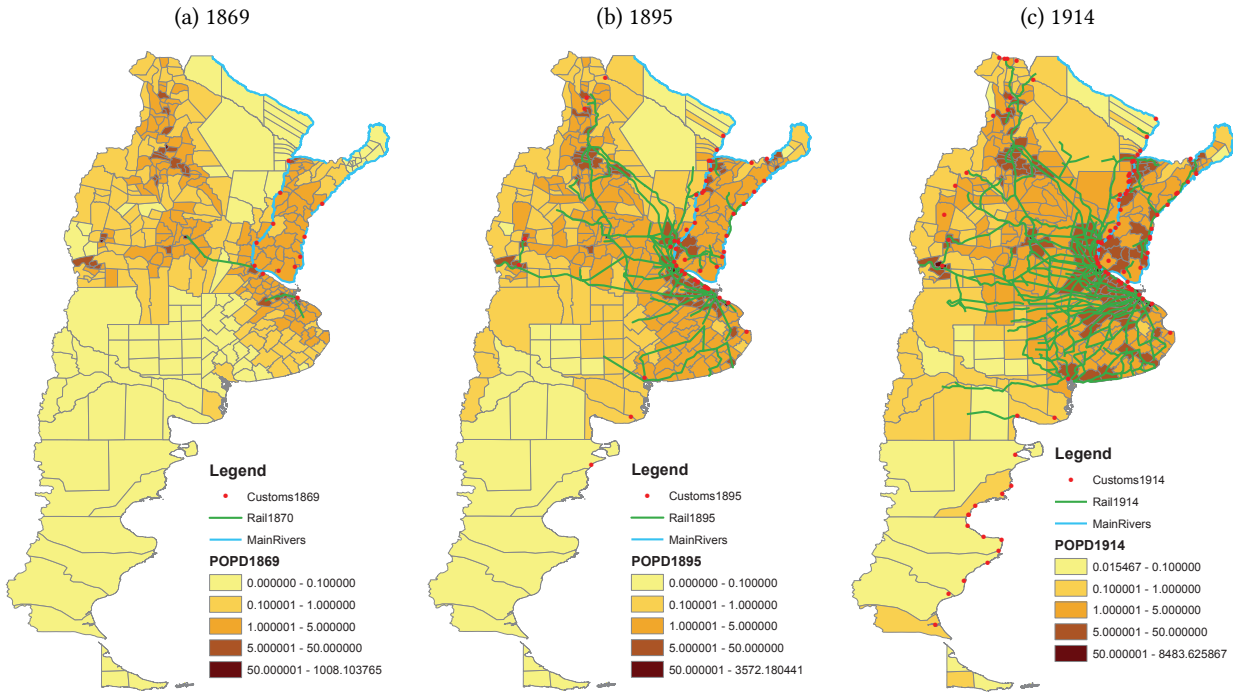
Notes: Map of population density distribution in 1869; 1895 and 1914. Railroad network shown in green (darker lines); main navigable rivers (the Paraná, Plate and Uruguay rivers) shown in blue (lighter lines); and customs (ports) shown by the red dots (solid circles).

Between 1869 and 1895 (comparing panels (a) and (b)), we observe a substantial increase in overall population density, and a reorientation of the population density distribution towards the agricultural hinterland surrounding Buenos Aires and its neighboring ports. Over this period, there is a large-scale expansion in the railroad network to connect the agricultural hinterland with these ports and to link together the Spanish colonial towns. Between 1895 and 1914 (comparing panels (b) and (c)), we see a continuation of this pattern, with a further increase in population density, which now diffuses further inland from Buenos Aires and its surrounding ports. The railroad network now radiates further into the interior, with an increase in the density of lines serving the agricultural hinterland.

In Figure A.5 in Section A.4.2.1 of the online appendix, we show that we find a similar pattern for urbanization, as measured by the share of the population living in towns and cities. Therefore, we find that an increase in the overall level of economic activity (as reflected in population density) is accompanied by urbanization (a reallocation of economic activity from rural to urban areas).

4.2 Geographical Access to World Markets

We now provide regression evidence on this change in the organization of economic activity with respect to geographical access to world markets. In our baseline specification, we measure geographical access to world markets using the distance from the centroid of each district to Argentina's trade hub, as captured by the top-four ports of

Buenos Aires, La Plata, Rosario and Bahía Blanca, which together account for more than 75 percent of export value throughout our sample period. In robustness checks, we show that we find a similar pattern of results across a range of alternative measures of geographical access to world markets. Our baseline regression specification is:

$$Y_t(\ell) = a_t + b_t \ln\left(\text{distance}(\ell)\right) + u_t(\ell), \tag{1}$$

where ℓ indexes districts and t corresponds to time; $Y_t(\ell)$ is an economic outcome of interest (e.g. log population density); $\ln\left(\text{distance}(\ell)\right)$ is a measure of geographical distance from world markets; and $u_t(\ell)$ is a stochastic error.

In Columns (1) and (2) of Table 1, we report the results of estimating equation (1) using OLS for log population density and the urban population share, respectively, at the beginning of our sample period in 1869. We find that locations close to world markets have not only higher population densities but also higher urban population shares. As distance from world markets doubles (a 100 percent increase), log population density falls by around 41 percent and the share of the population living in cities and towns falls by about 5 percentage points.[16] In Columns (3)-(4) of Table 1, we present the analogous results for the growth in log population density (which equals log population growth) and the change in urban population shares over our sample period from 1869-1914. Again we find negative and statistically significant gradients, such that as distance from world markets doubles, log population growth falls by around 45 percent and the increase in the urban population shares declines by about 6 percentage points. Therefore, following Argentina's external integration into world markets in the late-19th century, we find the largest increases in both population density and urbanization in the areas closest to world markets.

Table 1: Population Density, Structural Transformation and Geographical Access to World Markets

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log Population Density 1869	Urban Population Share 1869	Log Growth Population Density 1869-1914	Change Urban Population Share 1869-1914	Log Wage-Rental Ratio 1895	Log Relative Price of Tradeables 1895	Share Cereals Cultivated Area 1914
Log Distance Top-Four Port	-0.414***	-0.046***	-0.445***	-0.059***	1.158***	0.153***	-0.085***
	(0.089)	(0.014)	(0.048)	(0.013)	(0.168)	(0.049)	(0.019)
Observations	298	298	298	298	80	64	240
R-squared	0.085	0.041	0.238	0.068	0.420	0.177	0.118

Notes: Observations are a cross-section of Argentinian districts for a given year. Log population density is the log of the population per unit of land area. Urban population share is the share of the population living in cities and towns, as measured by the population census. Log wage-rental ratio is the log of the wages of agricultural laborers minus the log of the value of land per hectare for 1895, as discussed in Section A.6 of the online appendix. Log relative price of tradeables is the log tradeables consumption price index minus the log overall consumption price index, as constructed using data on the prices of traded goods, aggregate household expenditure shares and the value of land per hectare for 1895, as discussed in Section A.6 of the online appendix. Share cereals cultivated area 1914 is the share of agricultural land in each district used for Barley, Linseed, Maize (Corn), Oats, Rice, Rye, Sorghum and Wheat in 1914, as discussed in Section A.6 of the online appendix. The data on agricultural wages, traded goods prices and agricultural land are only available for a subset of districts, which explains the smaller number of observations in Columns (5)-(7). Distance Top-Four Port is the geographic (Great Circle) distance from the centroid of each district to the nearest top-four port (Buenos Aires, Rosario, La Plata and Bahia Blanca). Heteroskedasticity robust standard errors in parentheses. *** denotes significance at the 1 percent level; ** denotes significance at the 5 percent level; * denotes significance at the 10 percent level.

In Columns (5)-(7), we report results for the log wage-rental ratio, the log relative price of tradeables (the tradeables consumption price index relative to the overall consumption price index), and the share of the major new export of cereals in agricultural land area, respectively. The wage-rental ratio and relative price data are only available for a single cross-section (years around 1895) and the cultivated area data are only available for the last two years of our

---

[16]At the province level, we also have data on employment by occupation. Consistent with these district-level results for the urban population share, we find that provinces closer to Argentina's trade hub have larger shares of employment in service occupations.

sample (we report results for 1914). The smaller number of observations in Columns (5)-(7) reflects the fact that data on the wage-rental ratio, the relative price of tradeables and agricultural land use are only available for a subset of districts. As distance from world markets doubles, we find that the wage-rental ratio more than doubles, the relative price of tradeables increases by around 15 percent, and the share of cereals cultivated area falls by 8.5 percentage points.[17] Therefore, locations closer to world markets have higher land prices relative to wages and higher relative prices for non-traded goods, and specialize in trade-cost-sensitive goods such as cereals, consistent with the predictions of our spatial Balassa-Samuelson effect.

In Section A.4.2.2 of the online appendix, we demonstrate that we find similar results across a range of alternative specifications, including measuring geographical access to world markets using distance to Buenos Aires alone or distance to the nearest port on the coast or a navigable river; excluding districts in the immediate vicinity of the federal capital; and restricting attention to the subsample of districts used below for the quantitative analysis of the model.[18] Taken together, the findings of this section establish systematic gradients in the level and composition of economic activity with geographical access to world markets.

4.3 Railroad Access

We now provide regression evidence on the impact of the railroad network on the spatial distribution of economic activity within Argentina. The main empirical challenge is that the placement of railroads is unlikely to be random and could be targeted towards interior regions that would have experienced different growth trajectories even in the absence of a railroad connection. On the one hand, much of the railroad network was operated by private-sector companies, whose search for profits could have led them to select regions that otherwise would have grown more rapidly. On the other hand, the Argentine state promoted the development of railroads in rural areas that were unattractive to private-sector companies, which could have targeted locations that otherwise would have grown less rapidly. To address these concerns about non-random placement, we construct two instrumental variables that exploit quite different sources of variation, one based on least-cost paths to Argentina's 19th-century trade hub, and the other based on Spanish colonial postal routes. In the interests of brevity, we concentrate in our reduced-form specifications on the impact of the railroad network on overall population density, and explore its impact on structural transformation in our quantitative analysis of the model below.

Our instrumental variables (IV) estimation uses the following second-stage regression for long-differenced population growth over our sample period:

$$\Delta \ln Y_{1914-1869}(\ell) = a + c \left(\text{rail}_{1914}(\ell) \right) + d_1 \ln \left(\text{area}(\ell) \right) + d_2 \text{lat}\left(\ell \right) + d_3 \text{long}\left(\ell \right) + d_4 \ln Y_{1869}(\ell) + u(\ell), \quad (2)$$

where ℓ again indexes districts; $\Delta \ln Y_{1914-1869}(\ell)$ is log population growth from 1869-1914; $\text{rail}_{1914}(\ell)$ is a measure of whether a district has a railroad connection in 1914;[19] $\ln \left(\text{area}(\ell) \right)$ is the log geographical area of each district; $\text{lat}\left(\ell \right)$ and $\text{long}\left(\ell \right)$ are the latitude and longitude of the centroid of a district, respectively; $\ln Y_{1869}(\ell)$ is initial log population in 1869; and $u(\ell)$ is a stochastic error.

---

[17]While we focus on the consumption price index for tradeables relative to the overall consumption price index, because it plays a key role in our theoretical model below, we also find a positive and statistically significant gradient of the log consumption price index for tradeables with respect to distance from world markets.

[18]Although the natural experiment of Argentina's late-19th-century integration into world markets provides an attractive empirical setting, this property that locations with better access to world markets are both more densely populated and more urbanized is also found in other settings, such as in the Belt and Road Initiative (BRI) in Central Asia, as examined in the subsequent work by Lall and Lebrand (2020).

[19]The railroad network was negligible in 1869, and hence this measure captures the expansion of the railroad network from 1869-1914.

This second-stage regression specification (2) allows for a fixed effect in the level of log population for each district that has been differenced out. We thus allow for time-invariant unobserved heterogeneity in location characteristics that affects population levels in each year that has been differenced out. The constant a captures any common time effect that affects population growth across all Argentinian districts from 1869-1914, such as common macro shocks. The corresponding first-stage regression is:

$$\text{rail}_{1914}(\ell) = e + f_1\left(\text{port}(\ell)\right) + f_2\left(\text{colonialpost}(\ell)\right) + g_1 \ln\left(\text{area}(\ell)\right) + g_2 \text{lat}\left(\ell\right) \tag{3}$$
$$+ g_3 \text{long}\left(\ell\right) + g_4 \ln Y_{1869}(\ell) + h(\ell),$$

where $\text{port}(\ell)$ is our first instrument based on least cost paths to a top-four port; $\text{colonialpost}(\ell)$ is our second instrument based on Spanish colonial postal routes; and $h(\ell)$ is a stochastic error.

Our first instrument exploits the fact that the top-four ports are all clustered around Buenos Aires, which had already developed into Argentina's trade hub in the aftermath of the Napoleonic Wars, before the invention of the railroad in 1825. Once railroads were invented, we exploit the fact that interior regions were likely to be connected to this pre-existing trade hub, regardless of the economic characteristics of those interior regions. Therefore, our first instrument mechanically predicts a railroad connection based on constructing least-cost paths between the centroid of each district and the top-four ports. We measure the fraction of each district's surface area covered by these least-cost paths to the top-four ports, as discussed further in Section A.4.2.3 of the online appendix. Crucially, this instrument uses no information about the economic characteristics of districts, and hence cannot be influenced by some districts being economically more desirable destinations than others.

Our second instrument uses historical exploration and trade routes following Duranton and Turner (2012) and Duranton, Morrow, and Turner (2014). We use the fact that economic activity was orientated in a very different way in the Spanish colonial period than in the late-19th century. In particular, official trade routes ran towards the North-West through Panama, instead of towards the Eastern coastal areas around Buenos Aires. Despite this very different orientation of economic activity, once existing population centers had formed, they were likely to be connected to the railroad after it had been invented. Hence, locations along the route between these historical centers were also likely to be connected. To implement this idea, we georeference a map of Spanish colonial postal routes from the 18th century from Randle (1981). For each district, we construct our instrument as the length of colonial postal routes within its boundaries. We expect this instrument to have power in predicting the railroad network, because paths that are convenient for colonial postal routes using horses are also likely to be convenient for railroads.

To address the concern that a district's size could affect the likelihood that it lies along a least-cost path to a top-four port or a Spanish colonial postal route, we control separately for log district land area. To control for the fact that some geographic regions within Argentina could have higher or lower population growth than others for reasons unrelated to the railroad network (e.g. proximity to Argentina's trade hub or agroclimatic conditions), we also control separately for latitude and longitude. Finally, to control for potential heterogeneity in initial levels of economic development, we control separately for initial population in 1869. Conditional on these controls, our IV estimation assumes that there is no direct effect on economic activity of lying along a least-cost path to a top-four port or Spanish colonial postal route, other than through the probability of a railroad connection.

Table 2 presents the results of estimating the second-stage regression (2) for population growth. In Column (1), we report the results of estimating this specification using OLS and measuring whether a district is connected to the

railroad network in 1914 by a dummy variable that equals one if a district has one or more railroad stations and zero otherwise. We find a positive and statistically significant relationship between population growth and connection to the railroad network.[20] One limitation of this measure of a railroad connection is that it does not take into account ease of access to the railroad network within each district. In Column (2), we report the results of estimating the same specification using the length of railroad lines within the district's boundaries as an alternative measure of a railroad connection, where we control separately for the geographical area of the district. Again we find a positive and statistically significant estimated coefficient.[21]

Table 2: Population Growth and Railroad Access

	(1)	(2)	(3)	(4)	(5)
	Log Growth Population 1869-1914	Log Growth Population 1869-1914	Log Growth Population 1869-1914	Log Growth Population 1869-1914	Log Growth Population 1869-1914
Rail Connection 1914	0.650*** (0.092)	–	–	–	–
Rail length 1914	–	0.514*** (0.036)	0.633*** (0.073)	0.717*** (0.144)	0.608*** (0.073)
Latitude	-0.070*** (0.012)	-0.045*** (0.011)	-0.035*** (0.013)	-0.028 (0.017)	-0.037*** (0.012)
Longitude	0.041*** (0.012)	0.044*** (0.011)	0.043*** (0.011)	0.042*** (0.012)	0.043*** (0.011)
Log Land Area	-0.016 (0.039)	-0.204*** (0.036)	-0.245*** (0.044)	-0.274*** (0.061)	-0.237*** (0.044)
Log Population 1869	-0.431*** (0.069)	-0.393*** (0.058)	-0.393*** (0.057)	-0.394*** (0.057)	-0.393*** (0.057)
Estimation	OLS	OLS	IV	IV	IV
Instruments	–	–	Both	Port	Colonial Post
First-stage F-statistic	–	–	34.39	21.02	59.31
Overidentification test (p-value)	–	–	0.465	–	–
Observations	298	298	298	298	298
R-squared	0.437	0.574	–	–	–

Notes: In Columns (1)-(5), observations are a cross-section of Argentinian districts for a single difference from 1869-1914. Log population growth is the log growth of the total population of each district. Rail Connection 1914 is an indicator that is one if a district has a rail connection (one or more railroad stations) in 1914 and zero otherwise. Rail length 1914 is the length of railroads in each district in 1914. Port instrument is the fraction of the surface area of each district that lies along the least-cost paths from the centroids of all Argentinian districts to the top-four ports (Buenos Aires, Rosario, La Plata and Bahia Blanca). Colonial post is the length of Spanish colonial postal routes in each district. Latitude is the latitude of the centroid of a district. Longitude is the longitude of the centroid of a district. Land Area is the total geographical land area of each district. First-stage F-statistic is a test of the statistical significance of the instruments in the first-stage regression. Overidentification test is a Hansen-Sargan test of the model's overidentifying restrictions. In the IV specifications, the second-stage R-squared is not reported, because it does not have a meaningful interpretation. Heteroskedasticity robust standard errors in parentheses. *** denotes significance at the 1 percent level; ** denotes significance at the 5 percent level; * denotes significance at the 10 percent level.

In Column (3), we estimate the same specification using two-stage least squares and both our port and colonial post instruments. Consistent with a causal effect of the expansion of the railroad network on population growth, we find that the estimated railroad coefficient remains positive and statistically significant. The IV estimate is marginally larger but not statistically significantly different from the OLS estimate. This pattern of results is consistent with the two counteracting effects discussed above, with the profit-seeking incentives of the private-sector railroad companies to target regions that otherwise would have grown more rapidly being offset by the public-sector promotion of regions

---

[20]In principle, one could distinguish between being connected to the railroad network, and being connected to a top-four port by the railroad network. In practice, almost all of the districts that have a railroad station in a given year are also connected to one of the top-four ports in that year, because the railroad network in Argentina radiated outwards from Buenos Aires and the other top-four ports.

[21]We also find similar results if we use the number of stations within a district's boundaries as an alternative measure of ease of access to the railroad network within each district, with an estimated coefficient (standard error) in Column (2) of 0.075 (0.005).

that otherwise would have grown more slowly. We find that the instruments have power in the first-stage regression, with the F-statistic for the significance of the instruments in the first stage equal to 34.39 (above the conventional threshold of 10), as reported at the bottom of the column. In a Hansen-Sargan overidentification test, we are unable to reject the null hypothesis of the model's overidentifying restrictions (p-value $= 0.465$), as also reported at the bottom of the column. Therefore, assuming that one of the instruments is valid, we are unable to reject the null hypothesis that the other instrument only matters for population growth through a railroad connection.

In Columns (4) and (5) we report exactly-identified specifications using each instrument separately. We find that each instrument has power, with a first-stage F-statistic in each case above the conventional threshold of 10. In both specifications, the IV estimates are marginally larger than the OLS estimate in Column (2), although the differences are again not statistically significant. This similarity of the estimates using instruments that exploit quite different sources of variation provides support for our identifying assumptions. If only one of the instruments were invalid, we would expect to find a quite different pattern of results using that instrument. To explain the similarity of the results using all combinations of the instruments, we need either both instruments to be valid or an improbable pattern of correlation to exist between the instruments and the error term in the second-stage regression. Finally, we interpret the fact that all three IV estimates are close to the OLS estimate as supporting the idea that, conditional on our controls, the expansion of the railroad network within Argentina was indeed mainly driven by connecting interior regions with the top-four ports, and connecting existing colonial centers, rather than targeting interior regions that would have grown more rapidly for other reasons, even in the absence of the railroad.[22]

Taken together, the results of this subsection provide empirical support for a causal impact of the railroad network on the spatial distribution of economic activity within Argentina.

5 Theoretical Model

In this section, we develop our theoretical model of the distribution of economic activity within Argentina across sectors and locations.[23] We show that the model provides microeconomic foundations for the *spatial Balassa-Samuelson effect*, such that locations closer to world markets have higher population densities, urban population shares, relative prices of non-traded goods and land prices relative to wages, and specialize in the most transport-sensitive goods within the traded sector.

We consider an economy that consists of a set of locations $\ell \in \mathcal{L}$. There are three sectors: manufacturing (M), agriculture (A), and non-tradeables (N). Motivated by the overwhelming concentration of Argentinian exports in agriculture, we assume that all locations within Argentina have a comparative advantage in agriculture. To rationalize the observed differences in the composition of agricultural production across these locations, we assume that this agricultural sector consists of a discrete number of disaggregated goods indexed by $g = 1, \cdots, G$. Unless otherwise indicated, we suppress time subscripts to simplify notation, but we take it as understood that we allow all variables to change over time.

Each location is connected through an internal transport network to world markets at a trade hub. We denote the trade cost for good g between each location ℓ and the trade hub ℓ^* by $\delta_g\left(\ell, \ell^*\right)$, where this trade cost changes over

---

[22]As a final robustness check, Table A.2 in Section A.4.2.3 of the online appendix shows that we find a similar pattern of results for the subsample of districts used in the quantitative analysis of the model.

[23]See Section A.2 of the online appendix for the derivation of the theoretical results and the proofs of the propositions in this section.

time with the expansion of the railroad network. The trade hub in the model corresponds to the four leading ports centered on Buenos Aires in the data. International prices for traded goods at the trade hub ($\{P_g^*\}_{g=1}^G, P_M^*$) depend on external transport costs. For our structural estimation of the model, we do not need to take a stand on whether these international prices ($\{P_g^*\}_{g=1}^G, P_M^*$) are exogenous (small open economy) or endogenous (large open economy), because we condition on observed endogenous variables that capture the impact of international prices. When we undertake counterfactuals, we assume a small open economy as our baseline specification, but report a robustness exercise allowing for endogenous changes in the international terms of trade.[24]

Each location ℓ has a land area $L(\ell)$ and a continuum of land plots $j \in [0, L(\ell)]$ that are heterogeneous in terms of their productivity for these disaggregated agricultural goods $g = 1, \cdots, G$.

5.1 Preferences

Worker preferences for each location have a common component that is the same for all workers and an idiosyncratic component. The common component ($u(\ell)$) depends on traded goods consumption ($c_T(\ell)$), and non-traded goods consumption ($c_N(\ell)$), and is assumed to take the constant elasticity of substitution (CES) form:

$$u(\ell) = \left[\beta_T^{\frac{1}{\sigma}} c_T(\ell)^{\frac{\sigma-1}{\sigma}} + (1-\beta_T)^{\frac{1}{\sigma}} c_N(\ell)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{4}$$

where $\beta_T \in (0,1)$ controls the relative weights of traded and non-traded goods; we implicitly set amenities in each location as equal to one, as they enter the model isomorphically to an increase in productivity in both sectors, given our observed data on population density and the agricultural employment share; we assume inelastic demand between sectors ($0 < \sigma < 1$), as in the macroeconomics literature on structural transformation, and consistent with our estimates of the model's parameters below.

Tradeables consumption is in turn defined over consumption of a composite manufacturing good and the set of agricultural goods $g = 1, \ldots, G$ with the following homothetic price index:

$$E_T(\ell) = E_T\left(\{P_g(\ell)\}_{g=1}^G, P_M(\ell)\right), \tag{5}$$

where $P_g(\ell)$ is the price of agricultural good g in location ℓ and $P_M(\ell)$ is the corresponding price of the composite manufacturing good.

In addition to the common component of utility ($u(\ell)$) in equation (4), we allow for an idiosyncratic component of utility that captures heterogeneity in worker tastes for locations. We assume the following timing for worker location decisions. First, workers observe idiosyncratic tastes for Argentina (B^A) and the rest of the world (B^{RW}) that are drawn independently from a Fréchet distribution with shape parameter ε^{INT}, and choose whether to live in Argentina or the rest of the world. Second, if workers choose to live in Argentina, they observe idiosyncratic tastes for each location within Argentina ($b(\ell)$) that are drawn independently from a Fréchet distribution with shape parameter ε, and choose a location within Argentina to live in.[25]

---

[24]Throughout our sample period, Argentina's production of each export good remains relatively small compared to world markets. For example, for the major export product of wheat, Bennett (1933) estimates that world production in 1895 (1914) was 2,731 (3,618) bushels, which compares with Argentinian production of 46.4 (169.2) bushels.

[25]While we make this timing decision, worker location decisions would be identical under the alternative assumption of a nested Fréchet, with the choice between Argentina and the rest of the world in the upper nest, and the choice across locations within Argentina in the lower nest.

5.2 Labor and Land Markets

Each worker is endowed with one unit of labor that is supplied inelastically with zero disutility. Since workers decide between Argentina and the rest of the world before observing their realizations for tastes for individual locations within Argentina, this decision depends on expected utility in Argentina (u^*) and the rest of the world (u^{RW}). Under our distributional assumptions, the supply of workers choosing to live in Argentina ($N^S(u^*)$) is given by:

$$N^S(u^*) = \frac{1}{1 + \left(\frac{u^{RW}}{u^*}\right)^{\varepsilon^{INT}}} N^W, \qquad \varepsilon^{INT} > 1, \tag{6}$$

where N^W is total world population, and where we have normalized the ratio of the Fréchet scale parameters for Argentina and the rest of the world to one, because this ratio enters the model isomorphically to an increase in productivity in both sectors across all locations within Argentina, given our observed data on population density and the agricultural employment share. Conditional on choosing to live in Argentina, the supply of workers that decide to live in location ℓ ($N(\ell)$) is given by:

$$N(\ell) = \left(\frac{u(\ell)}{u^*}\right)^{\varepsilon} N^S(u^*), \qquad \varepsilon > 1, \tag{7}$$

where we have again normalized the Fréchet scale parameter for each location to one, because it enters the model isomorphically to an increase in productivity in both sectors, given our observed data on population density and the agricultural employment share. The corresponding expected utility from living in Argentina (u^*) is:

$$u^* = \left[\sum_{\ell \in \mathcal{L}} u(\ell)^{\varepsilon}\right]^{\frac{1}{\varepsilon}}. \tag{8}$$

A larger Fréchet shape parameter ε implies greater internal labor mobility within Argentina, such that the supply of labor to each location within Argentina in equation (7) is more responsive to relative utility levels across locations ($u(\ell)/u^*$). Similarly, a larger Fréchet shape parameter ε^{INT} implies greater international labor mobility, such that that the total population of Argentina is more responsive to relative utility levels in Argentina and the rest of the world (u^{RW}/u^*). The labor market clearing condition for the economy as a whole is:

$$\sum_{\ell \in \mathcal{L}} L(\ell)n(\ell) = N, \tag{9}$$

where $n(\ell) = N(\ell)/L(\ell)$ is population density at location ℓ; and N is the economy's total population, which in equilibrium satisfies the international population mobility condition in equation (6): $N = N^S(u^*)$.

Land is owned by immobile landowners who consume where they live and do not own any labor.[26] Total income per unit of land equals the sum of payments to both labor and land and is denoted by $y(\ell)$.

5.3 Production Technology

Production in each sector is characterized by constant returns to scale in land and labor. For simplicity, we assume a Cobb-Douglas technology, so that output per unit of land in the non-traded sector ($q_N(\ell)$), in the manufacturing

---

[26]Under our assumption of identical and homothetic preferences, all equilibrium allocations are invariant to the number of these landowners. In Section A.3.1 of the online appendix, we report a robustness check using non-homothetic preferences, in which we assume that the number of landowners is proportional to the area of each location.

sector ($q_M(\ell)$), and for an agricultural good g ($q_{g,j}(\ell)$) in land plot j in location ℓ are respectively:

$$q_N(\ell) = z_N(\ell)n_N(\ell)^{1-\alpha_N},$$ (10)

$$q_M(\ell) = z_M(\ell)n_M(\ell)^{1-\alpha_M},$$

$$q_{g,j}(\ell) = z_{g,j}(\ell)n_{g,j}(\ell)^{1-\alpha_A},$$

where $z_N(\ell)$ is non-traded productivity; $z_M(\ell)$ is manufacturing productivity; $z_{g,j}(\ell)$ is productivity for a disaggregated agricultural good; $n_N(\ell)$ is non-traded employment per unit of land; $n_M(\ell)$ is manufacturing employment per unit of land; $n_{g,j}(\ell)$ is employment for a disaggregated agricultural good per unit of land; and $0 < \alpha_i < 1$ is the land intensity in sector $i = A, M, N$. We make the natural assumptions that agriculture is land intensive ($\alpha_A > \alpha_M$ and $\alpha_A > \alpha_N$) and that all sectors use at least some land ($\alpha_M, \alpha_N > 0$).[27]

Productivity in all three sectors ($z_M(\ell)$, $z_N(\ell)$, $z_{g,j}(\ell)$) differs across locations ℓ. In the manufacturing and non-traded sectors, productivity is the same across all land plots j within a given location ℓ. In the agricultural sector, we assume that land plots $j \in [0, L(\ell)]$ can differ in terms of their productivities for individual agricultural goods j ($z_{g,j}(\ell)$) within a location ℓ. This variation in agricultural productivity enables us to rationalize the production of a range of agricultural goods within each location in the data and captures the impact of differences in soil conditions and topography. In particular, we assume that the realizations of productivity for each agricultural good and land plot $\{z_{g,j}(\ell)\}_{g=1}^{G}$ are drawn independently from the following Fréchet distribution:

$$\text{Prob}\left[z_{g,j}(\ell) < z\right] = e^{-T_g(\ell)z^{-\theta}},$$ (11)

where $T_g(\ell)$ controls the average productivity of good g in location ℓ; θ controls the dispersion of agricultural productivity across land plots, which we assume is the same for all goods. For our estimation and calibration of the model, we are not required to take a stand on whether productivity ($z_M(\ell)$, $z_N(\ell)$, $\{z_{g,j}(\ell)\}$) is exogenous or endogenous, because we use the equilibrium conditions of the model to recover the productivities implied by the observed data on the endogenous variables. When we undertake counterfactuals, we assume exogenous productivities as our baseline specification, but also report a robustness exercise allowing for agglomeration forces.

In the international trade literature following Eaton and Kortum (2002), the properties of the Fréchet distribution are used across a continuum of *goods* to determine patterns of production for each country. In contrast, we use these properties across a continuum of *land plots* to characterize patterns of production for each good. This formulation enables us to consider a discrete number of goods, as observed in the data, and yet still obtain determinate predictions for production patterns for each good (by using the law of large numbers across the continuum of land plots). Our specification also allows us to accommodate zero agricultural land shares for some goods in some locations, as observed in the data, because the Fréchet scale parameter that determines average productivity ($T_g(\ell)$) can vary by both good g and location ℓ. Therefore, we rationalize a zero agricultural land share for good g in location ℓ by taking the limit as this productivity parameter converges to zero ($\lim_{T_g(\ell)\to 0}$). Finally, our framework allows for zero populations in some locations in equilibrium, as also observed for some locations and years, which can be rationalized in the model by zero productivities in both traded sectors ($\lim_{T_g(\ell)\to 0}$ for all g and $\lim_{z_M(\ell)\to 0}$).[28]

---

[27]For simplicity, we assume the same factor intensity across all agricultural goods. Although this assumption can be relaxed, we show below that our parsimonious parameterization has substantial predictive power for the data.

[28]In Section A.3.3 of the web appendix, we consider an extension of the model in which landowners make an endogenous decision whether to leave land wild or convert it to productive use. In this extension, zero population in a location in equilibrium also can be rationalized by it not being profitable to convert land to productive use.

5.4 Profit Maximization

Production in each sector is perfectly competitive. Firms choose employment density (employment per unit of land) to maximize profits, taking as given goods and factor prices and the location decisions of other firms and workers. In equilibrium, firms make zero profits in each sector and location with positive production. Therefore, if a plot of land in location ℓ is used for manufacturing or non-tradeables $i = M, N$, land rents ($r_i(\ell)$) are equal to revenue per unit of land minus labor costs per unit of land at the equilibrium value of employment density:

$$r_i(\ell) = \max_{n_i(\ell)} \{P_i q_i(n_i(\ell)) - w(\ell)n_i(\ell)\} \qquad \text{for } i = M, N, \tag{12}$$

where $w(\ell)$ is the wage. Alternatively, if a plot of land j in location ℓ is used in agriculture, it is allocated to the agricultural good that offers the highest land rent, and these land rents are determined in an analogous way:

$$\begin{aligned} r_j(\ell) &= \max_{g=1,..,G} \{r_{g,j}(\ell)\}, \\ r_{g,j}(\ell) &= \max_{n_{g,j}(\ell)} \{P_g(\ell)q_{g,j}(n_{g,j}(\ell)) - w(\ell)n_{g,j}(\ell)\}. \end{aligned} \tag{13}$$

We assume that the decision whether to allocate a land plot to agriculture, manufacturing or non-tradeables is made before observing the realizations for agricultural productivities $\{z_{g,j}(\ell)\}_{g=1}^{G}$, which captures the role of idiosyncratic shocks to agricultural productivity, such as weather shocks. Therefore, this land use decision depends on the comparison of expected land rents in agriculture ($r_A(\ell) = \mathbb{E}[r_j(\ell)]$) to land rents in the other two sectors ($r_M(\ell), r_N(\ell)$), where expected land rents in agriculture are determined by the probability distribution for agricultural productivities in equation (11). After a landowner has allocated a land plot to agriculture, she observes the realizations for productivity for each agricultural good, and decides which of these agricultural goods to produce.

5.5 Sectoral Employment and Wage-Rental Ratio

Using profit maximization and zero profits, equilibrium variables in each sector and location can be written in terms of the equilibrium wage-rental ratio $\omega_i(\ell) = w(\ell)/r_i(\ell)$, which in turn depends on wages ($w(\ell)$), productivity ($z_i(\ell)$) and prices ($P_i(\ell)$). For the manufacturing and non-traded sectors $i \in \{M, N\}$, employment per unit of land and the wage-rental ratio in each location with positive production must satisfy:

$$n_i(\ell) = \frac{1 - \alpha_i}{\alpha_i} \frac{1}{\omega_i(\ell)}, \tag{14}$$

$$\omega_i(\ell) = \left(\frac{w(\ell)}{P_i(\ell)z_i(\ell)}\right)^{\frac{1}{\alpha_i}}. \tag{15}$$

For the agricultural sector, once a plot of land j in location ℓ has been assigned to the production of a given agricultural good g, the equilibrium values of employment per unit of land ($n_{g,j}(\ell)$) and the wage-rental ratio ($\omega_{g,j}(\ell)$) take the same form as above, except with price $P_g(\ell)$ and productivity $z_{g,j}(\ell)$.

We now establish a key aggregation property of the model. Under our assumption of a Fréchet distribution for agricultural productivity, there exists an aggregate measure of agricultural revenue productivity ($z_A(\ell)$) that is a sufficient statistic for the impact of the prices and productivity of the disaggregated agricultural goods on aggregate employment and output in the agricultural sector:

$$z_A(\ell) = \Gamma\left(\frac{\alpha_A\theta - 1}{\alpha_A\theta}\right)^{\alpha_A} \left[\sum_{g=1}^{G} T_g(\ell) P_g(\ell)^{\theta}\right]^{1/\theta}, \tag{16}$$

where $\Gamma\left(\cdot\right)$ is the Gamma function; and we assume that θ is sufficiently large that $\alpha_A \theta > 1$.

An implication of this aggregation result in equation (16) is that we can treat the agricultural sector $i = A$ as if it consisted of a single good with a common revenue productivity $z_A\left(\ell\right)$ across all land plots in location ℓ. Using this result, employment density ($n_A\left(\ell\right)$) and the wage-rental ratio ($w_A\left(\ell\right)$) in the aggregate agricultural sector take the same form as for the manufacturing and non-traded sectors in equations (14) and (15), but using the expected land rent ($r_A\left(\ell\right) = \mathbb{E}\left[r_j(\ell)\right]$) and the ratio of wages to expected land rents ($\omega_A\left(\ell\right) \equiv w(\ell)/r_A\left(\ell\right)$), and replacing $P_i(\ell)z_i\left(\ell\right)$ with aggregate revenue productivity $z_A\left(\ell\right)$.

5.6   Definition of Equilibrium

Under our neoclassical assumptions, the definition of general equilibrium takes a standard form, in which workers maximize utility and choose their location optimally, firms maximize profits and zero profits are made in each location with positive production, and markets clear.

Definition 1. *A general equilibrium consists of an expected utility u^*; a total population N; allocations of population density $n(\ell)$, land $\{L_i(\ell)\}_{i=N,M,A}$, and employment density $\{n_i(\ell)\}_{i=N,M,A}$; wages $w\left(\ell\right)$; land rents $r\left(\ell\right)$; and prices $\{P_g\left(\ell\right)\}_{g=1}^{G}$, $P_M\left(\ell\right)$, $P_N\left(\ell\right)$ for all $\ell \in \mathcal{L}$ such that:*

(i) *workers maximize utility and choose their location optimally; i.e., $n(\ell)$ is given by (7) and N is given by (6);*

(ii) *producers maximize profits and land is allocated optimally across sectors, $r(\ell) = \max\{r_A(\ell), r_M(\ell), r_N(\ell)\}$;*

(iii) *the land market clears in each location: $\sum_{i=M,N,A} L_i(\ell) = L(\ell)$;*

(iv) *the labor market clears in each location: $\sum_{i=M,N,A} \frac{L_i(\ell)}{L(\ell)} n_i(\ell) = n(\ell)$;*

(v) *the non-traded goods market clears in each location: $c_N(\ell) = \frac{L_N(\ell)}{L(\ell)} q_N(n_N(\ell))$;*

(vi) *traded goods' prices are determined by no arbitrage: under our assumption that all locations ℓ within Argentina have a comparative advantage in agriculture and export the disaggregated agricultural goods, the local price of each exported agricultural good g ($P_g\left(\ell\right)$) equals its price on international markets (P_g^*) less transport costs, $P_g\left(\ell\right) = P_g^*/\delta_g\left(\ell, \ell^*\right)$, and the local price of the imported manufacturing good M ($P_M\left(\ell\right)$) equals its price on international markets (P_M^*) adjusted for transport costs, $P_M\left(\ell\right) = \delta_M\left(\ell, \ell^*\right) P_M^*$;*

(vii) *expected utility u^* adjusts to clear the labor market for the economy as a whole, i.e. condition (9) holds.*

We establish the existence and uniqueness of the equilibrium, such that there exists a unique common expected utility across locations (u^*) and a unique set of prices $\{w\left(\ell\right),\, r\left(\ell\right),\, \{P_g\left(\ell\right)\}_{g=1}^{G},\, P_M\left(\ell\right),\, P_N\left(\ell\right)\}$ and allocations $\{n(\ell),\, \{L_i(\ell)\}_{i=N,M,A},\, \{n_i(\ell)\}_{i=N,M,A}\}$ for each location $\ell \in \mathcal{L}$ that satisfies the above equilibrium conditions.

Proposition 1. *Assuming that agriculture is land intensive ($\alpha_A > \alpha_N$), traded and non-traded goods are complements ($\sigma < 1$), and population is more mobile within Argentina than internationally ($\varepsilon > \varepsilon^{INT}$), there exists a unique general equilibrium given the exogenous location characteristics ($T_g\left(\ell\right), z_M\left(\ell\right), z_N\left(\ell\right), L\left(\ell\right)$, and $\delta\left(\ell, \ell^*\right)$ for all $\ell \in \mathcal{L}$).*

Proof. See the online appendix $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

5.7 Specialization Across Sectors

We now use the equilibrium conditions of the model to characterize specialization across sectors. We show that labor mobility and constant returns to scale imply complete specialization in the traded sector between agriculture and manufacturing, unless a location is in autarky. Therefore, under our assumption that all locations within Argentina have a comparative advantage in agriculture, these locations all produce and export agricultural goods, and import the manufacturing good. The model is thus consistent with the extreme concentration of Argentinian exports in agriculture observed in the data. Finally, our specification of CES preferences between traded and non-traded goods implies that the utility function satisfies the Inada conditions, which ensures that all populated locations produce and consume the non-traded good.

To establish these results, we begin by expressing the common component of utility ($u(\ell)$) in equation (4) in the indirect utility form:

$$u(\ell) = \frac{w(\ell)}{E(\ell)} = \frac{w(\ell)}{\left[\beta_T E_T(\ell)^{1-\sigma} + (1 - \beta_T) P_N(\ell)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}}. \tag{17}$$

A landowner's decision over how to use a land plot is determined by a comparison of the rental rate across the three sectors or, equivalently, by a comparison of the wage-rental ratios across these three sectors ($\omega_M(\ell), \omega_A(\ell), \omega_N(\ell)$). As all populated locations produce the non-traded good and at least one traded good, factor mobility across sectors ensures that there is a common equilibrium wage-rental ratio between the non-traded sector and the traded sector(s) with positive production: $\omega_N(\ell) = \omega_i(\ell)$ for $i = A, M$ if $n_i(\ell) > 0$. Using equation (17) and profit maximization and zero profits from equation (15), this equilibrium wage-rental ratio must satisfy,[29]

$$\left[\beta_T \left(\frac{P_i(\ell)}{E_T(\ell)} z_i(\ell) \omega_i(\ell)^{\alpha_i} \right)^{\sigma-1} + (1 - \beta_T) \left(z_N(\ell) \omega_i(\ell)^{\alpha_N} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}} = u(\ell). \tag{18}$$

Under autarky, there is positive production in all three sectors $\{N, M, A\}$, and hence a common wage-rental ratio across them. Using equations (15) and (18), we can solve in closed-form for this autarkic wage-rental ratio ($\omega^a(\ell) = \omega_N^a(\ell) = \omega_A^a(\ell) = \omega_M^a(\ell)$) for each location ℓ,

$$\omega^a(\ell) = \left(\frac{P_M(\ell) z_M(\ell)}{z_A(\ell)} \right)^{1/(\alpha_A - \alpha_M)}. \tag{19}$$

In contrast, when a location is open to trade, it produces the non-traded good and only one of the two traded goods. The reason is that under our assumptions of labor mobility and constant returns to scale, the equilibrium wage-rental ratio in each traded sector in equation (15) depends solely on prices, productivity and the common wage across sectors, and does not depend on the scale of production in any sector. Therefore, depending on the values of prices and productivities, one of the two traded sectors in general will have a lower wage-rental ratio than the other in a given location, which implies that this location will produce only one of the two traded goods. We summarize this complete specialization result within the traded sector as follows.

Proposition 2. *If location ℓ trades, it is fully specialized in agriculture if $\omega_A(\ell) < \omega^a(\ell)$. Assume that $\alpha_N < \alpha_M < \alpha_A$ and $\sigma < 1$. Then, complete specialization in agriculture occurs for sufficiently high values of agricultural productivity ($z_A(\ell)$) relative to manufacturing productivity ($z_M(\ell)$).*

---

[29]To obtain equation (18), first rewrite equation (17) as $u(\ell) = \left[\beta_T \left(\frac{w(\ell)}{E_T(\ell)} \right)^{\sigma-1} + (1 - \beta_T) \left(\frac{w(\ell)}{P_N(\ell)} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}$ and then eliminate $w(\ell)$ using the expressions $w(\ell) = P_i(\ell) z_i(\ell) \omega(\ell)^{\alpha_i}$ and $\frac{w(\ell)}{P_N(\ell)} = z_N(\ell) \omega(\ell)^{\alpha_N}$ implied by (ii) in Definition 1 and equation (15).

Proof. See the online appendix. □

As discussed above, based on the overwhelming concentration of exports in agriculture, we assume that all locations within Argentina have a comparative advantage relative to the rest of the world in agriculture (i.e., $z_A(\ell)$ is sufficiently large in each location that $\omega_A(\ell) < \omega^a(\ell)$). Whether any given location is closed or open to trade is determined by comparative advantage and transport costs. In particular, trade occurs if the relative price of the imported manufacturing good net of transport costs is less than the relative price of the manufacturing good under autarky. Finally, for a given value of transport costs, a location is open to trade for a sufficiently large comparative advantage in agricultural goods (a high enough value of $z_A(\ell)/z_M(\ell)$). We assume that all populated interior locations have a large enough $z_A(\ell)/z_M(\ell)$ so as to be open to trade.[30]

Under these assumptions on comparative advantage, each location that is open to trade specializes in agriculture and non-traded goods, which implies that the common component of utility (18) can be re-written as:

$$u(\ell) = \left[\beta_T \left(\widetilde{z}_A(\ell) \, \omega(\ell)^{\alpha_A} \right)^{\sigma-1} + (1 - \beta_T) \left(z_N(\ell) \omega(\ell)^{\alpha_N} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}, \tag{20}$$

where we have defined $\widetilde{z}_A(\ell)$ as a measure of adjusted agricultural productivity:

$$\widetilde{z}_A(\ell) = \frac{z_A(\ell)}{E_T(\ell)}, \tag{21}$$

which includes the tradeables consumption price index ($E_T(\ell)$). We show below that adjusted agricultural productivity ($\widetilde{z}_A(\ell)$) and non-traded productivity ($z_N(\ell)$) are sufficient statistics for population density, the agricultural employment share, and the aggregate values of all other endogenous variables in each location.

5.8 Specialization Within the Agricultural Sector

We now examine patterns of specialization across disaggregated goods within the agricultural sector, as determined by relative productivity and trade costs for these disaggregated agricultural goods.

With a continuum of land plots within each location, the share of agricultural land allocated to good g equals the probability that an individual land plot is allocated to that good. Therefore, using the properties of the Fréchet distribution of agricultural productivities, the share of agricultural land allocated to each good depends on relative productivities $\{T_g(\ell)\}$, relative local prices $\{P_g(\ell)\}$, and the Fréchet shape parameter θ:

$$l_g(\ell) = \frac{T_g(\ell) \, P_g(\ell)^\theta}{\sum_{g'} T_{g'}(\ell) \, P_{g'}(\ell)^\theta}. \tag{22}$$

Combining this result for patterns of agricultural production with an assumption over the functional form for the tradeables price index $E_T(\ell)$, we can solve for patterns of trade in the disaggregated agricultural goods. In particular, under the assumption that the tradeables price index is Cobb-Douglas, a constant share (γ_g) of overall spending on agriculture is allocated to each agricultural good:

$$E_T(\ell) = P_M(\ell)^{1-\gamma_A} \prod_{g=1}^{G} P_g(\ell)^{\gamma_g}, \qquad \text{where} \qquad \sum_{g=1}^{G} \gamma_g = \gamma_A. \tag{23}$$

---

[30]For sufficiently large transport costs, the model features a "trade frontier" beyond which regions further inland are in autarky. As transport costs fall, this frontier expands further inland as additional regions are integrated into world markets. Coşar and Fajgelbaum (2016) study this property in a model without a non-traded sector.

Using this constant Cobb-Douglas expenditure share (γ_g) together with our expression for the share of agricultural land allocated to each good in equation (22), we obtain the following closed-form solution for exports of each disaggregated agricultural good ($x_g(\ell)$) as a share of overall agricultural exports ($x_A(\ell)$):

$$\frac{x_g(\ell)}{x_A(\ell)} = \frac{l_g(\ell) - \gamma_g}{1 - \gamma_A}. \tag{24}$$

Therefore, each location is a net exporter of an individual disaggregated agricultural good ($x_g(\ell) > 0$) if the share of agricultural land that it allocates to the production of that good is greater than its share of expenditure on that good ($x_g(\ell) > \gamma_g$). We thus obtain a neoclassical prediction for chains of comparative advantage within the agricultural sector, such that if location ℓ exports good g, it necessarily exports all goods g' such that $l_{g'}(\ell)/l_g(\ell) > \gamma_{g'}/\gamma_g$:

$$\frac{x_{g'}(\ell)}{x_g(\ell)} = \frac{l_{g'}(\ell) - \gamma_{g'}}{l_g(\ell) - \gamma_g}. \tag{25}$$

Although each location is a net exporter of agricultural goods, and a net importer of manufacturing goods, in general there can be internal bilateral trade in the disaggregated agricultural goods between locations within Argentina. For simplicity, we focus on the case in which each location ℓ exports all of the disaggregated agricultural goods g through the trade hub ℓ^*, such that local prices are pinned down through no-arbitrage by the price at the trade hub (P_g^*) and transport costs ($\delta_g(\ell, \ell^*)$): $P_g(\ell) = P_g^*/\delta_g(\ell, \ell^*)$. This corresponds to the case in which the share of agricultural land exceeds the share of consumer expenditure in equation (24) for each disaggregated good, which can be true for all disaggregated goods, because $\sum_{g=1}^{G} l_g(\ell) = 1$ and $\sum_{g=1}^{G} \gamma_g = \gamma_A < 1$. In this case, there is no incentive for internal trade in the disaggregated goods within Argentina, because the export price for each good in each location is greater than or equal to the price from shipping to another location within Argentina.[31]

From this no-arbitrage relationship, if some exported goods have higher elasticities of transport costs ($\delta_g(\ell, \ell^*)$) to distance from world markets than others, these transport-cost sensitive goods will have higher relative prices in locations closer to world markets. Therefore, from the relationship between agricultural land shares and relative prices in equation (22), locations closer to world markets will specialize in these transport-cost sensitive goods.

5.9 Spatial Balassa-Samuelson Effect

We now use the results from the previous sections to establish the spatial Balassa-Samuelson effect as a key feature of the relationship between structural transformation, economic development and international trade. We start by combining the population mobility condition (17), the requirement of zero-profits and profit maximization in each sector (15), and the labor market clearing condition (9) to obtain the following closed-form solutions for population density ($n(\ell)$) and the agricultural employment share ($\nu_A(\ell)$) as a function of the wage-rental ratio ($\omega(\ell)$) and the relative price of traded goods ($E_T(\ell)/E(\ell)$):

$$n(\ell) = \frac{N(\ell)}{L(\ell)} = \left(\frac{1}{\alpha_N + (\alpha_A - \alpha_N)\beta_T \left(\frac{E_T(\ell)}{E(\ell)} \right)^{1-\sigma}} - 1 \right) \frac{1}{\omega(\ell)}, \tag{26}$$

$$\nu_A(\ell) = \frac{N_A(\ell)}{N(\ell)} = \frac{(1 - \alpha_A)\beta_T \left(\frac{E_T(\ell)}{E(\ell)} \right)^{1-\sigma}}{1 - \alpha_N - (\alpha_A - \alpha_N)\beta_T \left(\frac{E_T(\ell)}{E(\ell)} \right)^{1-\sigma}}. \tag{27}$$

---

[31]It is straightforward to instead consider cases with internal trade, as long as the local price of each disaggregated agricultural good in each location is ultimately pinned down through no-arbitrage with the port. For example, Mendoza could import a disaggregated agricultural good from Córdoba, as long as Córdoba also exports this good to the port of Rosario. In this case, the local price of the good in Mendoza is pinned down by transport costs and the local price in Córdoba, and this local price in Córdoba is pinned down by transport costs and the price at the port of Rosario.

These closed-form solutions summarize the equilibrium relationships in the model between population density ($n(\ell)$), the agricultural employment share ($\nu_A(\ell)$), the wage-rental ratio ($\omega(\ell)$), and the relative price of tradeables ($E_T(\ell)/E(\ell)$). Both closed-form solutions have an intuitive interpretation. A higher population density must be accommodated through some combination of both sectors using more labor-intensive production techniques (which requires a lower wage-rental ratio, $\omega(\ell)$) or a higher share of employment in the labor-intensive non-traded sector (which with $0 < \sigma < 1$ requires a lower relative price of traded goods, $E_T(\ell)/E(\ell)$).

We now connect these endogenous variables to exogenous changes in external integration (changes in international prices ($\{P_g^*\}_{g=1}^G$, P_M^*) from, for example, reductions in transatlantic freight rates) and internal integration (reductions in domestic transport costs ($\{\delta_g(\ell,\ell^*)\}_{g=1}^G$, $\delta_M(\ell,\ell^*)$)) from, for example, the construction of the railroad network). We begin by totally differentiating equations (26) and (27), and using profit maximization and zero profits from equation (15) and labor market clearing from equation (9). Using these total derivatives, we obtain the following system of four equations that link the four endogenous variables of changes in population density ($\widehat{n}(\ell)$), the agricultural employment share ($\widehat{\nu}_A(\ell)$), the wage-rental ratio ($\widehat{\omega}(\ell)$) and the relative price of traded goods ($\widehat{\frac{E_T(\ell)}{E(\ell)}}$) to changes in (i) adjusted-agricultural productivity ($\widehat{\widetilde{z}_A}(\ell)$), (ii) non-traded productivity ($\widehat{z_N}(\ell)$), and (iii) local utility $u(\ell)$:

$$\widehat{n}(\ell) = -\frac{(\alpha_A - \alpha_N)\nu_A(\ell)}{\alpha_N(1-\alpha_A) + (\alpha_A - \alpha_N)\nu_A(\ell)}\widehat{\nu}_A(\ell) - \widehat{\omega}(\ell), \tag{28}$$

$$\widehat{\nu}_A(\ell) = \left(1 + \frac{\alpha_A - \alpha_N}{1-\alpha_A}\nu_A(\ell)\right)(1-\sigma)\left(\widehat{\frac{E_T(\ell)}{E(\ell)}}\right), \tag{29}$$

$$\widehat{\omega}(\ell) = \frac{(1-\alpha_N)\nu_A(\ell)\left(\widehat{u(\ell)} - \widehat{\widetilde{z}_A}(\ell)\right) + (1-\alpha_A)(1-\nu_A(\ell))\left(\widehat{u(\ell)} - \widehat{z_N}(\ell)\right)}{\alpha_A(1-\alpha_N)\nu_A(\ell) + \alpha_N(1-\alpha_A)(1-\nu_A(\ell))}, \tag{30}$$

$$\left(\widehat{\frac{E_T(\ell)}{E(\ell)}}\right) = \frac{(1-\alpha_A)(1-\nu_A(\ell))\left[\alpha_A\widehat{z_N}(\ell) - \alpha_N\widehat{\widetilde{z}_A}(\ell) - (\alpha_A - \alpha_N)\widehat{u(\ell)}\right]}{\alpha_A(1-\alpha_N)\nu_A(\ell) + \alpha_N(1-\alpha_A)(1-\nu_A(\ell))}, \tag{31}$$

where a hat above a variable denotes a proportional change, such that $\widehat{n}(\ell) \equiv dn(\ell)/n(\ell)$.

We next connect these changes in adjusted agricultural productivity ($\widehat{\widetilde{z}_A}(\ell)$), non-traded productivity ($\widehat{z_N}(\ell)$), and local utility $\widehat{u(\ell)}$ to exogenous changes in (i) international prices ($\{P_g^*\}_{g=1}^G$, P_M^*), (ii) domestic transportation costs ($\{\delta_g(\ell,\ell^*)\}_{g=1}^G$, $\delta_M(\ell,\ell^*)$), (iii) agricultural technology ($\{T_g(\ell)\}_{g=1}^G$), (iv) expected utility in the rest of the world (u^{RW}), and (v) world population (N^W). First, we totally differentiate domestic population mobility (7), international population mobility (6) and labor market clearing (9) to obtain:

$$\widehat{n}(\ell) = \widehat{N} + \varepsilon\left(\widehat{u}(\ell) - \widehat{u}^*\right), \tag{32}$$

$$\widehat{N}^S(\widehat{u}^*) = \left(1 - \frac{N}{N^W}\right)\left(\widehat{u}^* - u^{RW}\right) + \frac{N}{N^W}\widehat{N}^W, \tag{33}$$

$$\sum_\ell \nu(\ell)\widehat{n}(\ell) = \widehat{N}^S(\widehat{u}^*), \tag{34}$$

where $\nu(\ell) = N(\ell)/N$ is the share of location ℓ in the economy's total population.

Second, we totally differentiate adjusted agricultural productivity (21), agricultural productivity (16), the Cobb-

Douglas tradeables consumption price index (23), and land shares (22) to obtain:

$$\widehat{\widetilde{z}_A}(\ell) = \widehat{z}_A(\ell) - \widehat{E}_T(\ell),\tag{35}$$

$$\widehat{z_A}(\ell) = \sum_{g=1}^{G} l_g(\ell)\left(\frac{\widehat{T}_g(\ell)}{\theta} + \widehat{P}_g^* - \widehat{\delta}_g(\ell,\ell')\right),\tag{36}$$

$$\widehat{E_T(\ell)} = (1-\gamma_A)\left[\widehat{P}_M^* + \widehat{\delta}_M(\ell,\ell')\right] + \sum_{g=1}^{G}\gamma_g\left[\widehat{P}_g^* - \widehat{\delta}_g(\ell,\ell')\right],\tag{37}$$

$$\widehat{l}_g(\ell) = \left[\widehat{T}_g(\ell) - \sum_{g'=1}^{G} l_{g'}(\ell)\widehat{T}_{g'}(\ell)\right] + \theta\left[\widehat{P}_g(\ell) - \sum_{g'=1}^{G} l_{g'}(\ell)\widehat{P}_{g'}(\ell)\right],\tag{38}$$

where we have used no-arbitrage between local prices and prices at the trade hub for both exports ($P_g(\ell) = P_g^*/\delta_g(\ell,\ell^*)$) and imports ($P_M = P_M^*\delta_M(\ell,\ell^*)$).

Given exogenous changes in (i) international prices ($\{P_g^*\}_{g=1}^{G}$, P_M^*), (ii) domestic transport costs ($\{\delta_g(\ell,\ell^*)\}_{g=1}^{G}$, $\delta_M(\ell,\ell^*)$), (iii) technology ($\{T_g(\ell)\}_{g=1}^{G}$, $z_N(\ell)$), (iv) expected utility in the rest of the world (u^{RW}), and (v) world population (N^W), we can solve for the response of the endogenous variables of the model ($n(\ell)$, $\nu_A(\ell)$, $\omega(\ell)$, $E_T(\ell)/E(\ell)$) using the system of equations (28)-(38). We are now in a position to establish our Spatial Balassa-Samuelson effect, in which we evaluate the impact of a change in transport costs ($\{\delta_g(\ell,\ell^*)\}_{g=1}^{G}$, $\delta_M(\ell,\ell^*)$), holding constant the other exogenous determinants of economic activity ($\{P_g^*\}_{g=1}^{G}$, P_M^*, $\{T_g(\ell)\}_{g=1}^{G}$, $z_N(\ell)$, u^{RW}, N^W).

Proposition 3. *(**Spatial Balassa-Samuelson Effect**) Assume that traded and non-traded goods are complements ($\sigma < 1$), agriculture is land-intensive relative to non-tradeables ($\alpha_N < \alpha_A$), and population is mobile within Argentina (ε sufficiently large). Under these assumptions, low trade-cost locations (locations ℓ with lower transport costs $\{\delta_g(\ell,\ell^*)\}_{g=1}^{G}$, $\delta_M(\ell,\ell^*)$) have **(i)** higher adjusted-agricultural productivity ($\widetilde{z}_A(\ell)$), **(ii)** higher relative prices of non-traded goods (lower $E_T(\ell)/E(\ell)$), **(iii)** higher population density ($n(\ell)$), **(iv)** lower agricultural employment shares ($\nu_A(\ell)$), and **(v)** lower wage-rental ratios.*

Proof. See the online appendix. ◻

This proposition is related to the conventional Balassa-Samuelson effect in macroeconomics, in which higher productivity in tradeables at the level of the economy as a whole causes a rise in the relative price of the non-traded good. In this conventional specification, with inelastic demand between sectors, higher productivity in the traded sector can either raise or reduce employment in that sector, depending on whether the economy is open or closed to international trade, as in Matsuyama (1992) and Uy, Yi, and Zhang (2012). In contrast, our spatial Balassa-Samuelson effect operates across locations within an open economy that are linked through factor mobility, and arises because internal trade costs induce endogenous differences across these locations in price-adjusted productivity in the traded sector.

The economic intuition for our spatial Balassa-Samuelson effect is as follows. Locations with good access to world markets are attractive for the production and consumption of traded goods, which increases population density. As long as this population increase is sufficiently strong (ε sufficiently large), it bids up the reward of the immobile factor (land) relative to that of the mobile factor (labor). Together, the increase in population and the reduction in wages

relative to land rents induce an expansion in the employment share of the labor-intensive non-traded sector, which requires a higher relative price for the non-traded good given inelastic demand between sectors ($0 < \sigma < 1$).

Therefore, population mobility plays a central role in the spatial Balassa-Samuelson effect. In contrast, if population is immobile (ε sufficiently small), parts (i)-(iv) of the proposition hold, but we obtain the opposite prediction for the wage-rental ratio. In Section 6 below, we structurally estimate the degree of population mobility across locations within Argentina (ε). We estimate $\varepsilon > \varepsilon^{INT} > 1$, such that the wage-rental ratio falls with adjusted agricultural productivities, which is also consistent with the reduced-form evidence presented in Section 4 above, where we find that locations closer to world markets have lower wage-rental ratios.

Through this spatial Balassa-Samuelson effect, our model provides a microeconomic rationale for our earlier reduced-form evidence in Section 4 above. As locations closer to Argentina's trade hub face lower trade costs in accessing world markets, the model rationalizes the higher population densities, urban population shares, relative prices of non-traded goods, and land rents relative to wages. As we derive this spatial Balassa-Samuelson effect from a neoclassical production structure, it captures a generic feature of patterns of economic development that applies in settings with population mobility, specialization according to comparative advantage, labor-intensive non-traded goods, and inelastic demand between sectors.

6 Parameter Estimation

In this section, we structurally estimate the model's parameters. Our quantitative analysis has a sequential structure, such that we undertake our analysis in a number of steps, where each step uses results from the previous one and imposes the minimal set of additional assumptions relative to the previous step. In Section 6.1, we estimate the production cost share parameters (α_A, α_N) (Step 1). In Section 6.2, we estimate the elasticity of substitution between sectors (σ) and the weight of tradeables in consumer expenditure (β_T) (Step 2). In Section 6.3, we estimate the population mobility parameters across locations within Argentina (ε) and between Argentina and the rest of the world (ε^{INT}) (Step 3). In Section 6.4, we recover the implied values of expected utility across locations within Argentina (u^*) and for the rest of the world (u^{RW}) (Step 4).

In Section 6.5, we invert the calibrated model to recover the unobserved values of adjusted agricultural productivity ($\widetilde{z}_A(\ell)$) and non-traded productivity ($z_N(\ell)$) that exactly rationalize the observed data on population density ($n(\ell)$) and the agricultural employment share ($\nu_A(\ell)$) as an equilibrium of the model (Step 5). In Section 6.6, we estimate the relationship between adjusted agricultural productivity ($\widetilde{z}_A(\ell)$) and internal and external integration, which yields an estimate of the parameter determining the dispersion of productivity across the disaggregated agricultural goods (θ) (Step 6). As part of this parameter estimation, we provide evidence on the model's within-sample fit using targeted moments, and report overidentification checks using non-targeted moments.

In Section 7 below, we use the estimated model to undertake counterfactuals to quantify the impact of Argentina's external integration (reductions in transatlantic freight rates) and internal integration (the expansion of the railroad network) on the spatial distribution of economic activity. We compare the estimated impact of the railroad network on the net present value of GDP and land income to historical estimates of its construction costs.

6.1 Production Cost Share Parameters (Step 1)

We estimate the production cost share parameters (α_A, α_N) by comparing the model's predictions for the wage-rental ratio to observed data on this variable. From equations (26) and (27), the model implies the following prediction for the wage-rental ratio, given observed population density $(n(\ell))$ and the agricultural employment share $(\nu_A(\ell))$:

$$\omega(\ell) = \frac{w(\ell)}{r(\ell)} = \frac{(1 - \alpha_A)(1 - \alpha_N)}{\alpha_N(1 - \alpha_A) + (\alpha_A - \alpha_N)\nu_A(\ell)} \frac{1}{n(\ell)}. \tag{39}$$

We measure wages in the data using the wages of agricultural laborers, which are reported for a number of districts in each Argentinian province in the statistical abstract for 1913, as discussed in Section 3 above. We measure land rents in the data using the value of land per hectare, which is reported for each district in the 1895 statistical yearbook, and assuming a constant proportional relationship between land rents and land values.[32]

In general, there are several reasons why the model's predictions need not exactly equal the observed data on the wage-rental ratio. In particular, our measure of land rents need not perfectly control for land quality; the observed wages are for workers in the single occupation of agricultural laborers; and our wages and land rents data are for slightly different years. We assume that this measurement error is independently distributed, and estimate the production cost share parameters (α_A, α_N) using a minimum distance estimator, which minimizes the sum of squared log deviations between the model's predictions and the observed data.[33]

Consistent with the non-traded sector being labor intensive, we estimate labor shares of $(1 - \alpha_A) = 0.39$ and $(1 - \alpha_N) = 0.58$ for agriculture and the non-traded sector respectively. These parameter values imply a mean share of labor in district income of 0.45, which is close to the aggregate labor share of 0.48 in 1913 reported in Frankema (2010). Despite the several sources of potential discrepancies between the model's predictions and the data, we show in Section A.4.3 of the online appendix that the estimated model has substantial explanatory power. In Figure A.6, we find a strong, positive and statistically significant correlation between the model's predictions and the observed wage-rental ratio of 0.64. In Figure A.7, we show that the spatial Balassa-Samuelson effect in the model generates a similar gradient in the wage-rental ratio with respect to distance from Argentina's trade hub as observed in the data and reported in Section 4.2 above.

6.2 Demand Parameters (Step 2)

We estimate the elasticity of substitution between tradeables and non-tradeables (σ) and the weight of tradeables in expenditure (β_T) using the relationship in the model between the tradeables expenditure share $(s_T(\ell))$ and the relative tradeables price index $(E_T(\ell)/E(\ell))$. Although we do not directly observe the tradeables expenditure share $(s_T(\ell))$ for each district, the model yields a closed-form solution for this variable in terms of the observed agricultural employment share $(\nu_A(\ell))$. In particular, given the production cost share parameters (α_A, α_N) from the previous step, equation (27) implies the following relationship between the agricultural employment share $(\nu_A(\ell))$ and the relative tradeables price index $(E_T(\ell)/E(\ell))$:

$$\ln s_T(\ell) = \ln\left[\frac{(1 - \alpha_N)\nu_A(\ell)}{(1 - \alpha_A) + (\alpha_A - \alpha_N)\nu_A(\ell)}\right] = \kappa_0 + \kappa_1 \ln\left(\frac{E_T(\ell)}{E(\ell)}\right) + \ln h_T(\ell), \tag{40}$$

---

[32]The only year for which comprehensive data on land values are reported is 1895. In contrast, only the distribution of agricultural establishments across a number of discrete land value bins is reported in 1914.

[33]This log specification allows for the proportional relationship between land rents and land values per hectare (captured by an additive constant in logs) and for proportional changes in the overall price level between the two different years of 1895 and 1913 for which the land value and wage data are reported (again captured by an additive constant in logs).

where $\kappa_0 = \ln(\beta_T)$; $\kappa_1 = (1 - \sigma)$; and the regression error ($h_T(\ell)$) captures measurement error in the relative tradeables price index ($E_T(\ell)/E(\ell)$) and local preference shocks for tradeables.

We estimate this relationship using our data on the agricultural employment share ($\nu_A(\ell)$) and the relative tradeables price index ($E_T(\ell)/E(\ell)$). We measure both the tradeables price index ($E_T(\ell)$) and the overall price index ($E(\ell)$) by combining data on aggregate household expenditure shares for Argentina as a whole, traded goods prices by district and the value of land per hectare by district as a measure of housing costs. We construct these price indexes by weighting district-level prices by our aggregate household expenditure shares, using a methodology similar to that used for the U.S. consumer price index by the Bureau of Labor Statistics (BLS), as discussed further in Section A.6.7 of the online appendix. Food accounts for 30 percent of household expenditure; Other Household Expenses (including clothing, household equipment, and tools) make up for 50 percent; and Housing is responsible for the remaining 20 percent. We compute these price indexes for a cross-section of 63 districts for which data on traded goods prices and the value of land per hectare are available for years around 1895.

In Column (1) of Table 3, we report the results of estimating equation (40) using OLS. We find a positive and statistically significant relationship between the agricultural employment share and the relative price of tradeables. Consistent with the model's assumptions, we find an implied elasticity of substitution of less than one of $\sigma = 0.65$, and an implied weight of tradeables in consumer expenditure of $\beta_T = 0.75$. One potential concern about this specification is that unobserved local preference shocks for tradeables in the error term ($h_T(\ell)$) could affect both the agricultural employment share ($\nu_A(\ell)$) and the relative tradeables price index ($E_T(\ell)/E(\ell)$). From the spatial Balassa-Samuelson effect in the model, transport costs to world markets are a valid instrument for the relative tradeables price index. Therefore, we instrument the relative tradeables price index using a measure of travel time to the nearest top-four port. To address the concern that travel time could be influenced by the non-random placement of railroads, we use a measure of travel time to the nearest top-four port based on our colonial postal routes instrument, which assigns weights of 1 for colonial postal routes, 3 for the coast/navigable rivers, and 4.5 for land.[34]

In Column (2), we report these instrumental variables (IV) estimates of equation (40). We continue to find a positive and statistically significant coefficient on the relative tradeables price index. We find a smaller elasticity of substitution of $\sigma = 0.49$ and a larger weight of tradeables in consumer expenditure of $\beta_T = 0.77$. In Column (3), we report the corresponding first-stage regression. We find that travel time to the nearest top-four port based on colonial postal routes is a powerful predictor of the relative tradeables price index, with a first-stage F-statistic of 38.38, above the conventional threshold of 10.

In Figure A.9 in Section A.4.4 of the online appendix, we show that there is a strong and approximately log linear relationship between the model's prediction for the share of tradeables ($s_T(\ell)$) and the data on the relative tradeables price index ($E_T(\ell)/E(\ell)$), as implied by equation (40). In Figure A.10, we show that the spatial Balassa-Samuelson effect in the model generates a similar gradient in the relative tradeables price index with respect to distance from Argentina's trade hub as in the data, and as reported in Section 4.2 above.

Our IV estimate of the elasticity of substitution between tradeables and non-tradeables of $\sigma = 0.49$ compares closely to central estimates of this substitution parameter across aggregate industries in the existing empirical literature. For example, using our assumption of constant elasticity of substitution (CES) preferences and data for a number

---

[34]Our second port instrument for a railroad connection from Section 4.3 above is a measure of the fraction of a district's surface area covered by least-cost paths, and hence cannot be as easily converted into a conventional transport network as our colonial postal routes instrument.

of different countries and time periods, Bah (2007), Rogerson (2008), Duarte and Restuccia (2010), and Üngör (2017) obtain values for this elasticity of substitution of 0.44, 0.45, 0.40, and 0.47, respectively. Throughout the remainder, we use our instrumental variables estimate of $\sigma = 0.49$ and $\beta_T = 0.77$ as our baseline specification.

Table 3: Tradeables Expenditure Share ($s_T(\ell)$) and Relative Tradeables Price Index ($E_T(\ell)/E(\ell)$)

	(1)	(2)	(3)
	Log Share of Tradeables in Expenditure ($s_T(\ell)$)	Log Share of Tradeables in Expenditure ($s_T(\ell)$)	Log Relative Price of Tradeables ($E_T(\ell)/E(\ell)$)
Regression Constant (κ_0)	-0.293***	-0.263***	-3.576***
	(0.042)	(0.055)	(0.535)
Log Relative Price of Tradeables ($E_T(\ell)/E(\ell)$) (κ_1)	0.350***	0.506*	–
	(0.124)	(0.272)	
Log IV Travel Time Top-four Port	–	–	0.255***
			(0.041)
Implied Elasticity of Substitution (σ)	0.650	0.494	–
Implied Weight of Tradeables in Expenditure (β_T)	0.746	0.768	–
Estimation	OLS	IV	OLS
		(Second-Stage)	(First-Stage)
Observations	63	63	63
R-squared	0.11	–	0.381
First-stage F-Statistic	–	–	38.38

Notes: Observations are a cross-section of Argentinian districts for which data on the tradeables price index ($E_T(\ell)$) and overall price index ($E(\ell)$) are available. Log Share of Tradeables in Expenditure ($s_T(\ell)$) is measured using the model's predictions and the observed agricultural employment share ($\nu_A(\ell)$), as in equation (40). Log Relative Price of Tradeables ($E_T(\ell)/E(\ell)$) is measured using aggregate household expenditure shares, prices for traded goods, and land values per hectare as a measure of housing costs. IV travel time top-four port is the lowest-cost travel time to the nearest top-four port (Buenos Aires, Rosario, La Plata and Bahia Blanca) using a transport network consisting of colonial postal routes (weight 1), coast/navigable rivers (weight 3), and land (weight 4.5). First-stage F-statistic is a test of the statistical significance of the instruments in the first-stage regression. In the IV specification, the second-stage R-squared is not reported, because it does not have a meaningful interpretation. Heteroskedasticity robust standard errors in parentheses. *** denotes significance at the 1 percent level; ** denotes significance at the 5 percent level; * denotes significance at the 10 percent level.

6.3 Population Mobility Parameters (Step 3)

We estimate the domestic population mobility parameter (ε) that captures the dispersion of idiosyncratic preferences across locations within Argentina using the domestic population mobility condition (7). Using the general equilibrium relationships in equations (26) and (27), we can re-write this domestic population mobility condition in terms of the agricultural employment share ($\nu_A(\ell)$), lands rents ($r(\ell)$), and the consumption price index ($E(\ell)$):

$$\ln n(\ell) = \kappa_n + \frac{\varepsilon}{1+\varepsilon} \ln \left[\frac{r(\ell)}{[\alpha_N(1-\alpha_A)+(\alpha_A-\alpha_N)\nu_A(\ell)]E(\ell)} \right] + h_n(\ell), \tag{41}$$

where the constant is given by $\kappa_n = N^{\frac{1}{1+\epsilon}}\left((1-\alpha_A)(1-\alpha_N)/u^*\right)^{\frac{\epsilon}{1+\epsilon}}$; the term inside square parentheses corresponds to real income in each location; and the stochastic error ($h_n(\ell)$) captures measurement error in the overall consumption price index ($E(\ell)$).

We estimate this population mobility condition (41) using the same measure of the overall price index ($E(\ell)$) as used in Step 2 above for the 63 districts for which these data are available. We again use the value of land per hectare as a measure of land rents, assuming a constant proportional relationship between land rents and land values, which is absorbed into the constant κ_n. In Column (1) of Table 4, we report the results of estimating equation (41) using OLS. As predicted by the model, we find a positive and statistically significant relationship between population density and real income, with an implied preference dispersion parameter of $\epsilon = 2.81$.

Table 4: Population Density and Real Income

	(1)	(2)	(3)
	Log Population Density $(n(\ell))$	Log Population Density $(n(\ell))$	Log Real Income
Log Real Income	0.738***	0.826***	–
	(0.118)	(0.153)	
Log IV Travel Time Top-four Port	–	–	-1.078***
			(0.163)
Implied preference dispersion (ε)	2.811	4.733	–
Estimation	OLS	IV	OLS
		(Second-Stage)	(First-Stage)
Observations	63	63	63
R-squared	0.52	–	0.39
First-stage F-Statistic	–	–	43.69

Notes: Observations are a cross-section of Argentinian districts in 1895 for which data on the overall price index ($E(\ell)$) are available. Population Density ($n(\ell)$) is observed in the data. Real Income is a model prediction based on the observed agricultural employment share ($\nu_A(\ell)$), value of land per hectare ($r(\ell)$) and overall price index ($E(\ell)$). IV Travel Time Top-four Port is the lowest-cost travel time to the nearest top-four port (Buenos Aires, Rosario, La Plata and Bahia Blanca) using a transport network consisting of colonial postal routes (weight 1), coast/navigable rivers (weight 3), and land (weight 4.5). First-stage F-statistic is a test of the statistical significance of the instruments in the first-stage regression. In the IV specification, the second-stage R-squared is not reported, because it does not have a meaningful interpretation. Heteroskedasticity robust standard errors in parentheses. *** denotes significance at the 1 percent level; ** denotes significance at the 5 percent level; * denotes significance at the 10 percent level.

As in the previous subsection, one potential concern with this specification is that the measurement error ($h_n(\ell)$) could be correlated with the expression for real income in square parentheses. Again the spatial Balassa-Samuelson effect in the model implies that transport costs to world markets are a valid instrument for real income. Therefore, we instrument real income using our measure of travel time to the nearest top-four port based on colonial postal routes, as in Step 2 above. As shown in Column (2), we again find a positive and statistically significant relationship, with a higher implied preference dispersion parameter of $\varepsilon = 4.73$. In Column (3), we report the corresponding first-stage regression. Consistent with the predictions of the model, we find a negative and statistically significant relationship between real income and travel time to the nearest top-four port based on our colonial postal routes instrument, with a first-stage F-statistic of 43.69, again above the conventional threshold of 10.

Our estimates of the preference dispersion parameter (ε) are comparable to estimates in the existing empirical literature. For example, using internal migration data for U.S. states and Indonesian regions, Bryan and Morten (2019) estimates dispersion parameters of 2.7 and 3.2 respectively. Similarly, using internal migration data for U.S. commuting zones (CZs), Galle, Yi, and Rodriguez-Clare (2020) estimates dispersion parameters ranging from 1.42-2.79. Although our estimates are towards the high end of the range of existing estimates, and hence imply relatively high labor mobility across locations, this is consistent with the large-scale population movements observed across districts in late-19th century Argentina.

We estimate the international population mobility parameter (ϵ^{INT}) using the international mobility condition (6), which relates Argentina's share of the world population (N_t/N_t^W) to relative expected utility in Argentina and the rest of the world (u_t^*/u_t^{RW}). We measure Argentina's share of the world population using the international historical data from Maddison (2003), which are available for the years 1870 and 1913, close to the beginning and end of our sample period. We proxy relative expected utility in Argentina and the rest of the world by relative real gross domestic

product (GDP) per capita from the same data source. Taking log differences in equation (6) between 1913 and 1870, we choose the parameter ε^{INT} such that the observed change in Argentina's share of the world population on the left-hand side is equal to the predicted change based on the observed changes in relative real GDP per capita on the right-hand side. We obtain an estimated value of $\varepsilon^{INT} = 2.02$. Several caveats are relevant here: historical estimates of population and GDP necessarily depend on a number of assumptions and real GDP per capita is an approximation to expected utility. Nevertheless, our findings are consistent with both substantial international immigration in the late-19th century, and with lower levels of population mobility internationally than domestically.

6.4   Expected Utility (Step 4)

We calibrate expected utility in Argentina in each year of our sample (u_t^*) such that the model is consistent with the observed data on GDP in Argentina from Cortés Conde (1994) (constant 1914 prices). Using the population mobility condition (7) together with equations (26) and (27), we can express aggregate real income as the following sum of real income across all locations within Argentina:

$$
Y_t = \sum_{\ell \in \mathcal{L}} \frac{y_t(\ell) L(\ell)}{E_t(\ell)} = \frac{u_t^*}{N_t^{\frac{1}{\varepsilon}}} \sum_{\ell \in \mathcal{L}} \frac{(1 - \alpha_A) + (\alpha_A - \alpha_N) \nu_{At}(\ell)}{(1 - \alpha_A)(1 - \alpha_N)} L(\ell)^{\frac{\epsilon+1}{\epsilon}} n_t(\ell)^{\frac{\varepsilon+1}{\varepsilon}} , \tag{42}
$$

where recall that $y_t(\ell) = w_t(\ell) [n_t(\ell) + 1/\omega_t(\ell)]$ is nominal income per unit of land; $E_t(\ell)$ is the overall price index; and from now onwards we make explicit the time subscript.

From Steps 1-3 above, we have estimates of the production cost share (α_A, α_N) and domestic population mobility parameters (ε). Additionally, we observe population density ($n_t(\ell)$), the agricultural employment share ($\nu_{At}(\ell)$) and land area ($L(\ell)$) at the district-level and aggregate population (N_t) and GDP (Y_t) for Argentina as a whole. Therefore, given these parameter estimates and observed data, we solve for the unique value for expected utility in Argentina in each year of our sample (u_t^*) for which equation (42) holds.

Having recovered the level of expected utility in Argentina in each year (u_t^*), we use the international population mobility condition (6) and our estimate of ε^{INT} to solve for the level of expected utility in the rest of the world in each year (u_t^{RW}). From our estimation of the international mobility parameter (ε^{INT}) in Step 3, the increase in relative expected utility in Argentina and the rest of the world (u_t^*/u_t^{RW}) from 1869-1914 is necessarily equal to the observed relative growth in real GDP per capita in the Maddison (2003) data.

6.5   Model Inversion (Step 5)

Using the parameter estimates from Steps 1-4, (α_A, α_N, σ, β_T, ε, ε^{INT}, u_t^*, u_t^{RW}), we now invert the model to recover the values of adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) and non-traded productivity ($z_{Nt}(\ell)$) that exactly rationalize the observed data on population density ($n_t(\ell)$) and the agricultural employment share ($\nu_{At}(\ell)$) in each location as an equilibrium of the model. Using the equilibrium conditions of the model, we can uniquely determine these productivities, without taking a stand on whether they are exogenous (e.g. determined by agroclimatic conditions) or endogenous (e.g. through agglomeration forces). Adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) provides a sufficient statistic through which external and internal integration in goods markets affect the spatial distribution of economic activity (via international prices ($\{P_{gt}^*\}_{g=1}^{G}$, P_{Mt}^*) and domestic transport costs ($\{\delta_{gt}(\ell, \ell^*)\}_{g=1}^{G}$, $\delta_{Mt}(\ell, \ell^*)$)).

First, we recover the relative price of tradeables ($E_{Tt}(\ell)/E_t(\ell)$), using the general equilibrium equation (27) for

the agricultural employment share ($\nu_{At}(\ell)$):

$$\frac{E_{Tt}(\ell)}{E_t(\ell)} = \left(\frac{1}{\beta_T}\frac{(1-\alpha_N)\,\nu_{At}(\ell)}{(1-\alpha_A)+(\alpha_A-\alpha_N)\,\nu_{At}(\ell)}\right)^{\frac{1}{1-\sigma}}. \tag{43}$$

Second, we solve for the wage-rental ratio ($\omega_t(\ell)$) by combining equation (26) for population density ($n_t(\ell)$) and equation (27) for the agricultural employment share ($\nu_{At}(\ell)$), as shown in equation (39).

Third, we determine aggregate adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) and non-traded productivity ($z_{Nt}(\ell)$) using these solutions for the relative price of traded goods ($E_{Tt}(\ell)/E_t(\ell)$) and the wage-rental ratio ($\omega_t(\ell)$) together with profit maximization and zero-profits in equation (15) and population mobility in equation (17):

$$\widetilde{z}_{At}(\ell) = \frac{z_{At}(\ell)}{E_{Tt}(\ell)} = \frac{u_t^*}{\omega_t(\ell)^{\alpha_A}}\left(\frac{N_t(\ell)}{N_t}\right)^{1/\varepsilon}\frac{1}{(E_{Tt}(\ell)/E_t(\ell))}, \tag{44}$$

$$z_{Nt}(\ell) = \frac{u_t^*}{\omega_t(\ell)^{\alpha_N}}\left(\frac{N_t(\ell)}{N_t}\right)^{1/\varepsilon}\left(\frac{1-\beta_T}{1-\beta_T\left(\frac{E_{Tt}(\ell)}{E_t(\ell)}\right)^{1-\sigma}}\right)^{\frac{1}{1-\sigma}}. \tag{45}$$

6.6 Agricultural Specialization (Step 6)

In the remainder of this section, we use our data on agricultural land shares for the disaggregated goods ($l_{gt}(\ell)$) to connect adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) to external integration (changes in relative prices at Argentina's trade hub) and internal integration (the expansion of the railroad network). From equations (35), (36) and (37), the change in adjusted agricultural productivity is related to initial land shares ($l_{gt}(\ell)$) and changes in external integration ($\widehat{P}_{gt}^*, \widehat{P}_{Mt}^*$), internal integration ($\widehat{\delta}_{gt}(\ell,\ell^*), \widehat{\delta}_{Mt}(\ell,\ell^*)$) and technology ($\widehat{T}_{gt}(\ell)$) as follows:

$$\widehat{\widetilde{z}}_{At}(\ell) = \sum_{g=1}^{G}\left[l_g(\ell)\left(\underbrace{\frac{\widehat{T}_{gt}(\ell)}{\theta}}_{\text{Tech.}} + \left(\underbrace{\widehat{P}_{gt}^*}_{\text{Ext.}} - \underbrace{\widehat{\delta}_{gt}(\ell,\ell^*)}_{\text{Int.}}\right)\right) - \gamma_g\left(\underbrace{\widehat{P}_{gt}^*}_{\text{Ext.}} - \underbrace{\widehat{\delta}_{gt}(\ell,\ell^*)}_{\text{Int.}}\right)\right] - (1-\gamma_A)\left(\underbrace{\widehat{P}_{Mt}^*}_{\text{Ext.}} + \underbrace{\widehat{\delta}_{gt}(\ell,\ell^*)}_{\text{Int.}}\right), \tag{46}$$

where recall that a hat above a variable denotes a proportional change, such that $\widehat{\widetilde{z}}_{At}(\ell) = d\widetilde{z}_{At}(\ell)/\widetilde{z}_{At}(\ell)$.

We use this relationship to estimate the productivity dispersion parameter (θ) and to quantify the impact of changes in external integration, internal integration and technology. A key empirical challenge is that our district-level data on agricultural land shares ($l_{gt}(\ell)$) for each disaggregated good are only reported for 1895 and 1914, whereas for 1869 we only have data on aggregate agricultural land shares for each disaggregated good for Argentina as a whole. To overcome this challenge, we proceed in four steps. First, we estimate the impact of the railroad network on relative technology-adjusted prices for the disaggregated goods using equation (22) and our cross-section data on district-level agricultural land shares for 1914.

Second, we use these estimates and the observed change in the railroad network going backwards in time to 1895 and 1869 to generate predictions for the impact of the removal of the railroad network on relative technology-adjusted prices and agricultural land shares in these earlier years. We combine these predictions with our aggregate data on agricultural land shares for Argentina as a whole in 1895 and 1869 to back out the implied change in relative technology-adjusted prices at Argentina's trade hub in order for the aggregate agricultural land shares for each disaggregated good in the model to match the observed values in the data. Using these intermediate predictions for changes in relative technology-adjusted prices and agricultural land shares, we generate predicted changes in aggregate agricultural productivity ($\widehat{z}_{At}(\ell)$).

Third, we use these predictions for changes in agricultural productivity ($\widehat{z}_{At}(\ell)$) and our solutions for changes in adjusted agricultural productivity ($\widehat{\widetilde{z}}_{At}(\ell)$) from Step 5 above to estimate the productivity dispersion parameter (θ), and back out the estimated contributions of external and internal integration to these changes in adjusted agricultural productivity ($\widehat{\widetilde{z}}_{At}(\ell)$).

Agricultural Land Shares Estimation in 1914 (Step 6(i)). We begin by estimating the cross-section relationship between district-level agricultural land shares and travel time to Argentina's trade hub in the year $t = 1914$. Defining technology-adjusted prices as $\mathbb{P}_{gt}(\ell) \equiv T_{gt}(\ell)^{\frac{1}{\theta}} P_{gt}(\ell)$, we can re-write equation (22) as:

$$l_{gt}(\ell) = \mathbb{P}_{gt}(\ell)^{\theta} \mathbb{P}_{At}(\ell)^{-\theta}, \qquad \mathbb{P}_{At}(\ell) \equiv \left[\sum_{k=1}^{G} \mathbb{P}_{kt}(\ell)^{\theta} \right]^{\frac{1}{\theta}}, \tag{47}$$

where $\mathbb{P}_{At}(\ell)$ is a technology-adjusted price index that is defined across the disaggregated agricultural goods.

We model both the technology-adjusted price for each agricultural good ($\mathbb{P}_{gt}(\ell)$) and the technology-adjusted price index ($\mathbb{P}_{At}(\ell)$) as constant elasticity functions of (i) an intercept that is common to all districts and captures technology-adjusted prices at Argentina's trade hub; (ii) travel time to Argentina's trade hub using the transport network, and (iii) controls for latitude and longitude to capture geographical location within Argentina and local agroclimatic conditions. We thus obtain the following specification for the share of agricultural land allocated to each disaggregated good ($l_{gt}(\ell)$) in year $t = 1914$:

$$l_{gt}(\ell) = \mu_{gt} \tau_t(\ell, \ell^*)^{\phi_g} \, lat(\ell)^{\kappa_g} \, long(\ell)^{\vartheta_g} \, h_{gt}(\ell), \tag{48}$$

where μ_{gt} is the intercept which we allow to vary by good; $\tau_t(\ell, \ell^*)$ is the travel time from location ℓ to the nearest top-four port ℓ^* using the transport network in year $t = 1914$; ϕ_g (our key coefficient of interest) is the exponent on travel time, which we allow to vary across the disaggregated agricultural goods g to capture differences in transport costs for these goods; $lat(\ell)$ and $long(\ell)$ are controls for the latitude and longitude of the centroid of district ℓ, where we also allow the exponents on these controls (κ_g and ϑ_g) to vary across the disaggregated agricultural goods to capture differences in the impact of agroclimatic conditions; and $h_{gt}(\ell)$ is a stochastic error.

We estimate equation (48) separately for each good using the Poisson Pseudo Maximum Likelihood (PPML) estimator, which allows for zero agricultural land shares. As the left-hand side of this equation is a share ($l_{gt}(\ell)$), we expect a mixture of positive and negative coefficients across the different disaggregated goods, and the estimated coefficients (ϕ_g) capture the impact of travel times on relative technology-adjusted prices and hence the shares of agricultural land allocated to these disaggregated goods. In Column (1) of Table 5, we report the estimated coefficients on travel time for each good, where each cell of the table corresponds to a separate regression. We find a pattern of estimated coefficients that is consistent with the existing historical literature on the impact of the railroad network on agricultural specialization. We estimate negative and statistically significant coefficients for cereals and pure/mixed-breed cattle, which had relatively high transport costs, with pure/mixed-breed cattle used predominantly for chilled or frozen meat. We estimate positive and statistically significant coefficients for native-breed cattle and native-breed sheep, which had relatively low transport costs, because they largely were used for hides, skins, bones, fat and tallow. Finally, we find statistically insignificant coefficients for other crops and pure/mixed-breed sheep, which is consistent with these goods having intermediate transport costs, with pure/mixed-breed sheep mainly used for wool.[35]

---

[35]See Adelman (1994) for a discussion of the role of the railroads in opening up the hinterland for cereals cultivation; Perren (2017) for a discussion

A concern about this specification is that the expansion of the railroad network could have been influenced by the potential for agricultural development of interior locations. We address this concern in three ways. First, as our specification uses agricultural land shares, any effect on the overall level of agricultural development differences out from the numerator and denominator. Second, we include controls for latitude and longitude and allow the coefficients on these controls to vary across goods to capture the differential impact of geographical location and agroclimatic conditions. Third, we report an instrumental variables specification in Column (2), in which we instrument travel time to the nearest top-four port using our travel time instrument based on colonial postal routes, as used in Steps 2 and 3 above. Even when we focus solely on the variation in travel time induced by our instrument, we continue to find the same pattern of results across the disaggregated agricultural goods.

Historical Agricultural Land Share Predictions in 1869 and 1896 (Step 6(ii)). We next generate intermediate model predictions for the impact of the removal of the railroad network on agricultural land shares in 1869 and 1895. Comparing these model predictions to the data on aggregate shares of agricultural land for Argentina as a whole in 1869 and 1895, we estimate the change in relative technology-adjusted prices at Argentina's trade hub.

In particular, we generate historical predictions for agricultural land shares using an "exact-hat algebra" approach, in which we rewrite land shares in an earlier year $\chi < t$ in terms of the observed land shares in our baseline year $t = 1914$ and the relative change of variables between those years (where we now use a hat to denote the relative change in a variable, such that $\widehat{\tau}_\chi(\ell, \ell^*) = \tau_\chi(\ell, \ell^*) / \tau_t(\ell, \ell^*)$):

$$l_{g\chi}(\ell) = \frac{l_{gt}(\ell)\, \widehat{\mu}_{g\chi} \widehat{\tau}_\chi(\ell, \ell^*)^{\phi_g}}{\sum_{k=1}^{G} l_{kt}(\ell)\, \widehat{\mu}_{k\chi} \widehat{\tau}_\chi(\ell, \ell^*)^{\phi_k}}, \tag{49}$$

where we have estimated the travel time parameters for each agricultural good (ϕ_g) in Step 6(i) and we can compute the relative change in travel times ($\widehat{\tau}_\chi(\ell, \ell^*)$) from the observed change in the railroad network.

To compute these predicted land shares for earlier years in equation (49), we require an estimate of the common change in technology-adjusted prices at Argentina's trade hub ($\widehat{\mu}_{g\chi} = \mu_{g\chi}/\mu_{gt}$). In our baseline specification, we estimate these common changes in technology-adjusted prices by requiring that the predicted aggregate land shares for each disaggregated good in 1869 and 1895 for Argentina as a whole in the model are equal to their observed values in the data. There are three main advantages of this approach relative to the alternative of using observed international prices or transatlantic freight rates at Argentina's trade hub. First, this approach allows us to compute $\widehat{\mu}_{g\chi} = \mu_{g\chi}/\mu_{gt}$ without having to take a stand on the value of the productivity dispersion parameter θ, and hence to use the resulting solutions for $\widehat{\mu}_{g\chi}$ to estimate this productivity dispersion parameter below. Second, we are able to solve for $\widehat{\mu}_{g\chi} = \mu_{g\chi}/\mu_{gt}$, and recover changes in technology-adjusted prices for each good, without having to observe changes in technology for each good. Third, this approach ensures that the model matches the observed aggregate agricultural land shares in earlier years. Intuitively, we use the observed changes in these aggregate agricultural land shares to reveal the implied changes in common technology-adjusted prices ($(\mathbb{P}_{gt}^*)^\theta = T_{gt}^*(P_{gt}^*)^\theta$).

---

of the use of specialized railroad cars for the shipment of live cattle or refrigerated or frozen meat; and Amaral (1998) for a discussion of the use of native-breed animals for traditional products such as hides, skins, bones, fat and tallow.

Table 5: Agricultural Specialization and the Transport Network in 1914

	(1)	(2)
	Share of Agricultural Land Area ($s_g(\ell)$)	Share of Agricultural Land Area ($s_g(\ell)$)
Cereal cultivation		
Travel Time Top-Four Port	-0.421***	-0.253***
	(0.083)	(0.082)
Other Crop Cultivation		
Travel Time Top-Four Port	0.088	0.093
	(0.262)	(0.286)
Cattle Grazing (Pure/Mixed Breed)		
Travel Time Top-Four Port	-0.355***	-0.340***
	(0.059)	(0.063)
Cattle Grazing (Native Breed)		
Travel Time Top-Four Port	0.473***	0.298**
	(0.121)	(0.145)
Sheep Grazing (Pure/Mixed Breed)		
Travel Time Top-Four Port	-0.106	-0.030
	(0.125)	(0.121)
Sheep Grazing (Native Breed)		
Travel Time Top-Four Port	1.544***	1.412***
	(0.268)	(0.302)
Latitude and Longitude	Yes	Yes
Estimation	PPML	IV PPML
Observations	380	380
First-stage F-Statistic	-	3017.24

Notes: Observations are a cross-section of Argentinian districts in 1914. Each cell of the table corresponds to a separate regression. All specifications estimated using Poisson Pseudo Maximum Likelihood (PPML) to allow for zero agricultural land shares for each good. Share of Agricultural Land Area ($s_g(\ell)$) is the share of agricultural land area allocated to each of the six disaggregated goods, as defined in the data appendix: cereals; other crops; pure/mixed-breed cattle; native-breed cattle; pure/mixed-breed sheep; and native-breed sheep. Travel time top-four port is measured as the lowest-cost travel time using the 1914 transport network to the nearest top-four port (Buenos Aires, Rosario, La Plata and Bahia Blanca) and the following weights for each mode of transport: railroads (1); navigable rivers/coast (3), and land (4.5). In Column (2), travel time top-four port is instrumented with the lowest-cost travel time to the nearest top-four port using a transport network consisting of colonial postal routes (weight 1), coast/navigable rivers (weight 3), and land (weight 4.5). Latitude and longitude are the latitude and longitude of the centroid of each district, respectively. First-stage F-statistic is a test of the statistical significance of the instruments in the first-stage regression. Heteroskedasticity robust standard errors in parentheses. *** denotes significance at the 1 percent level; ** denotes significance at the 5 percent level; * denotes significance at the 10 percent level.

Recall that the agricultural land shares in equation (22) are homogenous of degree zero in technology-adjusted prices, which implies that only relative changes in these technology-adjusted prices at Argentina's trade hub ($\widehat{\mu}_{g\chi}$) are identified from the observed land shares. We choose native cattle as our numeraire and set the relative change in the technology-adjusted price for this good as equal to one. Since the change in the absolute level of technology-adjusted prices at Argentina's trade hub takes the same value across all locations, we capture it in the regression constant in our estimation of the productivity dispersion parameter (θ) later in this section. In our counterfactuals for external integration in Section 7 below, we capture the change in the absolute level of prices at Argentina's trade hub using direct data on transatlantic freight rates.

We report four overidentification checks on this estimation procedure, which are discussed in further detail in Section A.4.5 of the online appendix. First, we compare our estimated relative changes in technology-adjusted prices at Argentina's trade hub ($\widehat{\mu}_{g\chi} = \mu_{g\chi}/\mu_{gt}$) to separate data on relative changes in transatlantic freight rates over the period 1869-1914 from Tena-Junguito and Willebald (2020). Although we have a relatively small number of agricultural goods, and the transatlantic freight rates do not capture common changes in technology that are included in our estimates of $\widehat{\mu}_{g\chi}$, we find the expected negative correlation, with lower relative transatlantic freight rates implying higher relative export prices. Over the entire period from 1869-1914, we find a correlation of -0.44, while over the later part of our sample period from 1895-1914, we find a correlation of -0.81. Second, we use our exact-hat algebra pro-

cedure to generate predicted district agricultural land shares for 1895, and compare these predictions to the observed data on district agricultural land shares for 1895. Although our model is necessarily an abstraction, we find a positive and statistically significant correlation between the predicted and observed land shares for each of our disaggregated agricultural goods, with an average correlation across these goods of 0.64.

Third, we compare our model predictions for the quantity of cereals produced in each district in 1895 and 1914 to separate data on railroad shipments of cereals, which are not used in any part of our estimation. Although there are several reasons why the observed data on railroad shipments need not exactly equal our model's predictions for quantities produced, including local consumption and shipments using other modes of transport, we find a strong, positive, statistically significant and approximately log linear relationship between the two variables. Fourth, we compare our model's predictions for the value of cereals production in each district in 1914 to separate data on the total value of cereals machinery used in each district in 1914. Again we find a strong, positive, statistically significant and approximately log linear relationship between the two variables. Although there are a number of idiosyncratic factors that could affect the relationship between the value of cereals production and cereals machinery used in individual districts that are not captured by the model, these empirical results again provide further evidence that the model has predictive power for separate data not used in the estimation of its parameters.

Estimating θ and Agricultural Productivity Growth (Step 6(iii))   We now estimate the productivity dispersion parameter (θ), the change in agricultural productivity ($\widehat{z}_{A\chi}(\ell)$), and the contribution of changes in external and internal integration to these changes in agricultural productivity. From equation (21), we have a first equation for the change in agricultural productivity ($\widehat{z}_{A\chi}(\ell)$) in terms of the change in adjusted agricultural productivity ($\widehat{\widetilde{z}}_{A\chi}(\ell)$) and the change in the tradeables price index ($\widehat{E}_{T\chi}(\ell)$):

$$\ln \widehat{z}_{A\chi}(\ell) = \ln \widehat{\widetilde{z}}_{A\chi}(\ell) + \ln \widehat{E}_{T\chi}(\ell). \tag{50}$$

We have already solved for the change in adjusted agricultural productivity ($\widehat{\widetilde{z}}_{At\chi}(\ell)$) from our model inversion in Step 5 above. We model the level of the tradeables price index ($E_{Tt}(\ell)$) as a constant elasticity function of (i) an intercept that captures international prices (E_{Tt}^*), (ii) travel time to the nearest top-four port using the transport network ($\tau_t(\ell, \ell^*)$), (iii) controls for latitude and longitude to capture geographical location within Argentina and agroclimatic conditions, and (iii) an error term ($e_{Tt}(\ell)$):

$$E_{Tt}(\ell) = E_{Tt}^* \tau_t(\ell, \ell^*)^{\phi_T} \, lat(\ell)^{\kappa_T} \, long(\ell)^{\vartheta_T} \, e_{Tt}(\ell). \tag{51}$$

In Column (1) of Table 6, we report the results of estimating this relationship by OLS, using our observed data on the tradeables price index (as used in Step 2 above) for the 63 districts for which these data are available. We find a positive and statistically significant relationship between the tradeables price index and travel time from the nearest top-four port. Therefore, while our previous estimates established an increasing relationship between the relative tradeables price index ($E_{Tt}(\ell)/E_t(\ell)$) and remoteness from world markets, we now find the same increasing relationship for the level of the tradeables price index ($E_{Tt}(\ell)$). To address the potential concern of non-random railroad placement, we again instrument travel time to the nearest top-four port using our instrument based on colonial postal routes, as in Steps 2 and 3 above. As shown in Column (2), we continue to find a positive and statistically significant relationship, with an estimated coefficient (standard error) for ϕ_T of 0.061 (0.023). Using this estimate (ϕ_T) and the observed change

in the railroad network ($\widehat{\tau}_\chi(\ell, \ell^*)$), we can recover the change in the tradeables price index ($\widehat{E}_{T\chi}(\ell)$) up to a constant ($\kappa_T$) that is the same across all locations and captures the change in international prices at Argentina's trade hub ($E_{T\chi}^*$): $\ln \widehat{E}_{T\chi}(\ell) = \kappa_T + \phi_T \ln \widehat{\tau}_\chi(\ell, \ell^*)$. Using this relationship, our first equation for agricultural productivity (50) can be re-written as follows:

$$\ln \widehat{z}_{A\chi}(\ell) = \kappa_T + \ln \widehat{\widetilde{z}}_{A\chi}(\ell) + \phi_T \ln \widehat{\tau}_\chi(\ell, \ell^*) \tag{52}$$

From equations (16) and (22), and our land shares estimation in Steps 6(i)-(ii) above, we have a second equation for the change in agricultural productivity ($\widehat{z}_{A\chi}(\ell)$) in terms of the agricultural land shares ($l_{gt}(\ell)$) in our baseline year of $t = 1914$ and our estimated changes in relative technology-adjusted prices for the disaggregated agricultural goods ($\widehat{\mu}_{g\chi}\widehat{\tau}_\chi(\ell, \ell^*)^{\phi_g}$):

$$\ln \widehat{z}_{A\chi}(\ell) = \kappa_A + \frac{1}{\theta}\ln\left[\sum_{g=1}^{G} l_{gt}(\ell)\,\widehat{\mu}_{g\chi}\widehat{\tau}_\chi(\ell, \ell^*)^{\phi_g}\right], \tag{53}$$

where recall that only relative changes in technology-adjusted prices at Argentina's trade hub ($\widehat{\mu}_{gt\chi}$) are identified from our land shares estimation. Therefore, the constant κ_A controls for changes in the absolute level of technology-adjusted prices for the disaggregated agricultural goods at Argentina's trade hub.

Combining equations (52) and (53), we obtain the following relationship for agricultural productivity growth that can be used to estimate the productivity dispersion parameter (θ):

$$\ln\left(\widehat{\widetilde{z}}_{A\chi}(\ell)\,\widehat{\tau}_\chi(\ell, \ell^*)^{\phi_T}\right) = \kappa_Z + \xi_Z \ln\left[\sum_{g=1}^{G} l_{gt}(\ell)\,\widehat{\mu}_{g\chi}\widehat{\tau}_\chi(\ell, \ell^*)^{\phi_g}\right] + \widehat{e}_{T\chi}(\ell), \tag{54}$$

where $\xi_Z = 1/\theta$; the constant κ_Z captures common changes in the absolute level of prices for both the tradeables price index (κ_T) and the disaggregated agricultural goods (κ_A); $\widehat{e}_{T\chi}(\ell)$ is a stochastic error that captures measurement error that is assumed to be uncorrelated with the predictions from our land shares estimation.

In Column (3) of Table 6, we report the results of estimating equation (54) using OLS, for districts that are in our model sample in both 1869 and 1914, and have positive agricultural land shares for at least one disaggregated agricultural good in 1914.[36] The terms on the left and right-hand of this equation use two quite different sources of information. Adjusted agricultural productivity growth scaled by changes in travel times ($\widehat{\widetilde{z}}_{A\chi}(\ell)\,\widehat{\tau}_\chi(\ell, \ell^*)^{\phi_T}$) on the left-hand side comes from our model inversion, which uses observed population density ($n_t(\ell)$) and the agricultural employment share ($\nu_{At}(\ell)$), and our data on the tradeables price index ($E_{Tt}(\ell)$). In contrast, the term for ($1/\theta$) times predicted agricultural productivity growth on the right-hand side uses the land shares ($l_g(\ell)$) for the disaggregated agricultural goods. In general, there are several reasons why these two sets of predictions can differ, including the fact that our model of changes in the tradeables price index ($\widehat{E}_{T\chi}(\ell) = \kappa_T\widehat{\tau}_\chi(\ell, \ell^*)^{\phi_T}$) is necessarily an abstraction. Nonetheless, we find strong, positive and statistically significant relationship between them.

From this estimated coefficient in Column (3), we obtain an implied agricultural productivity dispersion parameter of $\theta = 3.176$, which implies that the requirement $\alpha_A\theta > 1$ from equation (16) is satisfied, even though we did not impose this restriction. This estimated productivity dispersion parameter is in line with the range of values in the existing empirical literature. Using variation across land plots, Sotelo (2020) finds an elasticity of 1.658. Using variation

---

[36] As discussed in Section 3 above, there are 164 districts in 1869 and 318 districts in 1914 that have rural population shares of more than 5 percent and less than 95 percent, and 152 districts that satisfy this condition in both years. Of these 152 districts, 93 have positive disaggregated agricultural land shares for at least one disaggregated good in 1914, as shown in Column (3) of Table 6.

across goods, Eaton and Kortum (2002) obtains elasticities from 2-12; Simonovska and Waugh (2014) estimates a value of 4; and Costinot and Rodríguez-Clare (2014) uses a central value of 5.

Table 6: Estimates for the Log Tradeables Price Index ($E_T(\ell)$) and Productivity Dispersion Parameter (θ)

	(1)	(2)	(3)
	Log Tradeables Price Index ($E_T(\ell)$)	Log Tradeables Price Index ($E_T(\ell)$)	Log Growth of Adjusted Agricultural Productivity Scaled by the Tradeables Price Index
Log Travel Time Top-Four Port	0.063**	0.061***	–
	(0.025)	(0.023)	
Log Agricultural Productivity Predicted by Land Shares	–	–	0.315***
			(0.062)
Implied Productivity Dispersion Parameter (θ)	–	–	3.176
Latitude and Longitude	Yes	Yes	–
Estimation	OLS	IV	OLS
First-stage F-statistic	–	129.83	–
Observations	63	63	93
R-squared	0.33	–	0.24

Notes: In Columns (1) and (2), observations are a cross-section of Argentinian districts in 1895 for which data on the tradeables price index ($E_T(\ell)$) and overall price index ($E(\ell)$) are available. In Column (3), observations are a cross-section of Argentinian districts from 1869-1914 that have rural population shares of above 5 percent and less than 95 percent in both years and that have positive agricultural land shares for at least one disaggregated agricultural good in 1914. Log Growth of Adjusted Agricultural Productivity scaled by the tradeables price index is recovered from the observed population density ($n(\ell)$) and agricultural employment share ($\nu_A(\ell)$) using our model inversion, and from our data on the tradeables price index ($E_T(\ell)$), as shown on the left-hand side of equation (54). Log Agricultural Productivity Predicted by Land Shares is the land-share weighted-average of changes in relative technology-adjusted prices at Argentina's trade hub and changes in travel times to Argentina's trade hub, as shown on the right-hand side of equation (54). Travel time top-four port is measured as the lowest-cost travel time using the transport network to the closest top-four port (Buenos Aires, Rosario, La Plata and Bahia Blanca) and the following weights for each mode of transport: railroads (1); navigable rivers/coast (3), and land (4.5). In Column (2), travel time top-four port is instrumented with the lowest-cost travel time to the closest top-four port using a transport network consisting of colonial postal routes (weight 1), coast/navigable rivers (weight 3), and land (weight 4.5). Latitude and longitude are the latitude and longitude of the centroid of each district. First-stage F-statistic is a test of the statistical significance of the instruments in the first-stage regression. In the IV specification in Column (2), the second-stage R-squared is not reported, because it does not have a meaningful interpretation. Heteroskedasticity robust standard errors in parentheses. *** denotes significance at the 1 percent level; ** denotes significance at the 5 percent level; * denotes significance at the 10 percent level.

7 Counterfactuals

We now use our estimates from the previous section to evaluate the impact of external and internal integration on macroeconomic aggregates and the spatial distribution of economic activity in Argentina. We undertake our counterfactuals starting from the observed equilibrium in our baseline year of $t = 1914$, for which we have district-level data on land shares ($l_{gt}(\ell)$) for each of the disaggregated agricultural goods. Therefore, starting from the observed data in 1914, we reverse external integration (raising transatlantic freight rates) and reverse internal integration (removing the railroad network), going backwards in time to year 1869. We focus on the set of districts in our model sample for which we have data in both 1869 and 1914, so that we can compute adjusted productivity for both years.

In these counterfactuals, we assume population mobility between Argentina and the rest of the world, where the elasticity of Argentina's total population with respect to expected utility in Argentina is determined by our estimated international population mobility parameter (ε^{INT}). In our baseline specification, we treat the international terms of trade ($\{P_{gt}^*\}_{g=1}^{G}, P_{Mt}^*$) as exogenous, which implicitly assumes that Argentina is a small open economy. In Section A.5.2 of the online appendix, we report a robustness exercise in which we allow for endogenous changes in the international terms of trade in response to changes in economic activity within Argentina. In Sections A.5.3-A.5.4 of the online appendix, we report additional robustness exercises to allow for endogenous productivity because of agglomeration forces and non-homothetic CES preferences.

Our counterfactuals use the property of the model that adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$), non-traded productivity ($z_{Nt}(\ell)$), total world population (N_t^W) and expected utility in the rest of the world (u_t^{RW}) are sufficient statistics for all aggregate variables, including population density ($n_t(\ell)$) and the agricultural employment share ($\nu_{At}(\ell)$). First, we make assumptions about external and internal integration, which determine our sufficient statistics $\{\widetilde{z}_{At}(\ell), z_{Nt}(\ell), N_t^W, u_t^{RW}\}$. Second, given these four sufficient statistics, we solve for the counterfactual values of all aggregate variables of the model $\{n_t(\ell), \nu_{At}(\ell), \omega_t(\ell), E_{Tt}(\ell), E_t(\ell)\}$.

We report two sets of counterfactuals for external and internal integration. First, we directly change the four aggregate sufficient statistics $\{\widetilde{z}_{At}(\ell), z_{Nt}(\ell), N_t^W, u_t^{RW}\}$, which allows us to isolate the role of different mechanisms in the model, using the property that changes in traded goods prices only directly affect the spatial distribution of economic activity through adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$). Second, we examine the impact of external and internal integration, by either changing agricultural productivity ($z_A(\ell)$) based on the predictions from our land shares estimation, transatlantic freights as a direct measure of external integration, travel times based on the construction of the railroad network as a direct measure of internal integration, or changing both transatlantic freights and travel times to capture the combined impact of external and internal integration.

In Section A.5 of the online appendix, we show that we can solve for a counterfactual equilibrium by solving for the equilibrium wage-rental ratio ($\omega_t(\ell)$) at which the demand for labor ($N_t^D(\ell)$) equals the supply of labor ($N_t^S(\ell)$) in each location. From Proposition 1, there exists a unique counterfactual wage-rental ratio ($\omega_t(\ell)$) at which the demand for labor ($N_t^D(\ell)$) equals the supply of labor ($N_t^S(\ell)$) for each location. For each of our three sets of counterfactuals, we report predicted changes in expected utility (using equation (8)), total population (using the international population mobility condition (6)), and real GDP (using equation (42)).

Counterfactuals for Adjusted Agricultural Productivity   We begin with our counterfactuals in which we directly change the aggregate sufficient statistics $\{\widetilde{z}_{At}(\ell), z_{Nt}(\ell), N_t^W, u_t^{RW}\}$. From our model inversion in Section 6.5 above, we exactly rationalize the observed data on population density ($n_t(\ell)$) and the aggregate employment shares ($\nu_{At}(\ell)$) using our solutions for adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$), non-traded productivity ($z_{Nt}(\ell)$), world population (N_t^W) and expected utility in the rest of the world (u_t^{RW}). Therefore, if we start from our baseline year of 1914 and undertake a counterfactual in which we simultaneously change all of these variables ($\widetilde{z}_{At}(\ell)$, $z_{Nt}(\ell)$, N_t^W, u_t^{RW}) back to 1869, we exactly reproduce the observed data for that year. In Row (1) in Panel A of Table 7, we report the results of this first exercise, in which real GDP, total population and expected utility fall to 13.6, 26.7 and 51.7 percent of their 1914 values, respectively.[37]

In our second counterfactual, we set adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) equal to its value in 1869, and hold non-traded productivity ($z_{Nt}(\ell)$), total world population (N_t^W) and expected utility in the rest of the world (u_t^{RW}) constant at their 1914 values. As shown in Row (2) in Panel B, we find that expected utility in Argentina falls substantially to 66.3 percent of its 1914 value. Given our assumption of constant expected utility in the rest of the world, this decline in expected utility in Argentina leads to a substantial population outflow, with total population falling to 43.7 percent of its 1914 value. We find that real GDP declines to 33.8 percent of its 1914 value.

Since adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) captures the net effect of all changes in prices and technology

---

[37]This fall in real GDP of 13.6 percent from 1869-1914 for the districts in our model sample in both years is close to the fall in real GDP across all districts of 10.2 percent, which from our model inversion in Section 6.4 equals the fall in Argentina's aggregate real GDP in Cortés Conde (1994).

Table 7: Counterfactual Predictions for Real GDP, Total Population and Expected Utility in Argentina

	Counterfactual Exercise	Real GDP 1869/1914	Total Population 1869/1914	Expected Utility 1869/1914
	Panel A: Observed Data			
(1)	All sufficient statistics back to 1869	0.136	0.267	0.517
	Panel B: Adjusted Agricultural Productivity			
(2)	Adjusted agricultural productivity ($z_A(\ell)/E_T(\ell)$) back to 1869	0.338	0.437	0.663
	Panel C: External and Internal Integration			
(3)	Agricultural productivity ($z_A(\ell)$) back to 1869	0.660	0.747	0.865
(4)	Transatlantic freights back to 1869	0.823	0.862	0.929
(5)	Railroad network back to 1869	0.872	0.906	0.952
(6)	Transatlantic freights and railroad back to 1869	0.720	0.782	0.885

Notes: Table reports counterfactual values in 1869 divided by actual values in 1914; Rows (1)-(6) report counterfactuals starting from the observed equilibrium in the data in our baseline year of 1914; Row (1) changes all our sufficient statistics ($\widetilde{z}_{At}(\ell)$, $z_{Nt}(\ell)$, N_t^W, u_t^{RW}) back to their 1869 values from our model inversion, which reproduces the observed equilibrium in the data in 1869; Row (2) changes only adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) back to its 1869 value from our model inversion; Row (3) changes adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) by the change in agricultural productivity ($\widetilde{z}_{A\chi}^{\text{Relative}}(\ell)$) back to 1869 from our land shares estimation in equation (55); Row (4) changes adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) by changes in transatlantic freight rates back to 1869 using equations (56) and (57); Row (5) changes adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) by removing the railroad network back to 1869 using equations (58) and (59); Row (6) changes adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) by both changes in transatlantic freight rates and the removal of the railroad network back to 1869 using equations (56), (57), (58) and (59).

in the traded sector, this first counterfactual establishes substantial aggregate effects from the transformation that occurred within the traded sector over our sample period. Comparing the first two rows, this transformation within the traded sector in Row (2) makes up 70 percent of the overall change in expected utility, 77 percent of the overall change in population, and 77 percent of the change in real GDP in Row (1).[38]

Counterfactuals for the Agricultural Transformation   In our third counterfactual, we use our land shares estimation from Section 6.6 to explore the relative importance of different mechanisms for these changes in adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$). In particular, we examine the importance of changes in agricultural productivity ($z_{At}(\ell)$) relative to changes in the tradeables consumption price index ($E_{Tt}(\ell)$). Using equation (53), we compute the change in agricultural productivity ($\widehat{z}_{A\chi}(\ell)$) induced by changes in relative technology-adjusted prices from external integration (as captured by our estimated intercepts ($\widehat{\mu}_{g\chi}$)) and internal integration (as captured by our estimated impact of changes in travel times from the construction of the railroad network ($\widehat{\tau}_\chi(\ell,\ell^*)^{\phi_g}$)):

$$\widehat{z}_{A\chi}(\ell) = \left[\sum_{g=1}^{G} l_{gt}(\ell)\,\widehat{\mu}_{g\chi}\,\widehat{\tau}_\chi(\ell,\ell^*)^{\phi_g}\right]^{\frac{1}{\theta}}. \tag{55}$$

Recall that only relative movements in technology-adjusted prices at Argentina's trade hub ($\widehat{\mu}_{g\chi}$) are identified in our land shares estimation, because the agricultural land shares are homogenous of degree zero in prices. Additionally, this counterfactual focuses only on the disaggregated products within the agricultural sector that are *exported*, and hence abstracts from any impact of external integration on *import* prices. While we present additional counterfactuals below that incorporate changes in the absolute level of both export and import prices, we use this second counterfactual to

[38]Since $((1-0.663)/(1-0.517)) \times 100 = 70$, $((1-0.437)/(1-0.267)) \times 100 = 77$, and $((1-0.338)/(1-0.136)) \times 100 = 77$. The larger percentage shares for population and real GDP than for expected utility reflect the fact that we hold expected utility in the rest of the world constant in this second counterfactual, which induces a population outflow, as we reduce adjusted agricultural productivity in Argentina. In contrast, in our first counterfactual, we reduce expected utility in the rest of the world to its 1869 value.

quantify the role of reallocation across the disaggregated goods within the agricultural sector.

In Row (3) in Panel C, we report the results of changing adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) by this change in agricultural productivity ($\widehat{z}_{A\chi}(\ell)$) from equation (55), holding constant our other sufficient statistics ($z_{Nt}(\ell)$, N_t^W, u_t^{RW}) at their 1914 values. We find that expected utility, total population and real GDP fall to 86.5, 74.7 and 66.0 percent of their 1914 values, respectively. Comparing Rows (2) and (3), we find that these counterfactual predictions make up 40, 45 and 51 percent of the changes in expected utility, population and real GDP induced by the overall change in adjusted agricultural productivity, respectively.[39]

Spatial Distribution of Economic Activity   We now show that these counterfactual predictions of our estimated model have substantial explanatory power for the observed reorientation of the spatial distribution of economic activity during our sample period and for Argentina's present-day distribution of economic activity.

We construct four measures of population growth for each Argentina district in our model sample: (i) Observed log population growth from 1869-1914; (ii) Observed log population growth from 1869-2001; (iii) Predicted log population growth from 1869-1914 in our counterfactual changing only adjusted agricultural productivity ($\widehat{\widetilde{z}}_{A\chi}(\ell)$) (Counterfactual (2)); (iv) Predicted log population growth from 1869-1914 in our counterfactual changing only agricultural productivity ($\widehat{z}_{A\chi}(\ell)$) based on the predictions from our land shares estimation (Counterfactual (3)). In each case, we compute relative log population growth by subtracting the mean of each variable.

In Figure 3, we display the fitted values from locally-weighted linear least squares regressions of these population growth measures on distance from the nearest top-four port. Consistent with the reduced-form evidence in Section 4, we find a strong negative gradient in population growth from 1869-1914 (solid black line). We find that this negative gradient in population growth is even stronger from 1869-2001 (solid gray line), highlighting that the reorientation of economic activity within Argentina during our sample period from 1869-1914 is predictive for the emergence of the present-day distribution of economic activity, centered on Argentina's trade hub, and quite different from the North-West orientation under Spanish colonial rule.

We find a similar strong negative gradient in counterfactual population growth from 1869-1914 based on only changing adjusted agricultural productivity ($\widehat{\widetilde{z}}_{A}(\ell)$) (black dashed line). Therefore, the transformation that occurs within the traded sector over our sample period plays an important role in explaining the observation reorientation in economic activity within Argentina. We continue to find this strong negative gradient even when we only change agricultural productivity ($\widehat{z}_{A\chi}(\ell)$) by the amount predicted by our land shares estimation (gray dashed line). This pattern of results highlights the role played by both external integration (common changes in relative technology-adjusted prices at Argentina's trade hub, $\widehat{\mu}_{g\chi}$) and internal integration (changes in travel time from the construction of the railroad network, $\widehat{\tau}_{\chi}(\ell,\ell^*)^{\phi_g}$) in driving the observed reorientation of economic activity.

---

[39]Since $((1 - 0.865)/(1 - 0.663)) \times 100 = 40$, $((1 - 0.747)/(1 - 0.437)) \times 100 = 45$ and $((1 - 0.660)/(1 - 0.338)) \times 100 = 51$.

Figure 3: Actual and Counterfactual Gradients in Population Growth in Geographical Access to World Markets
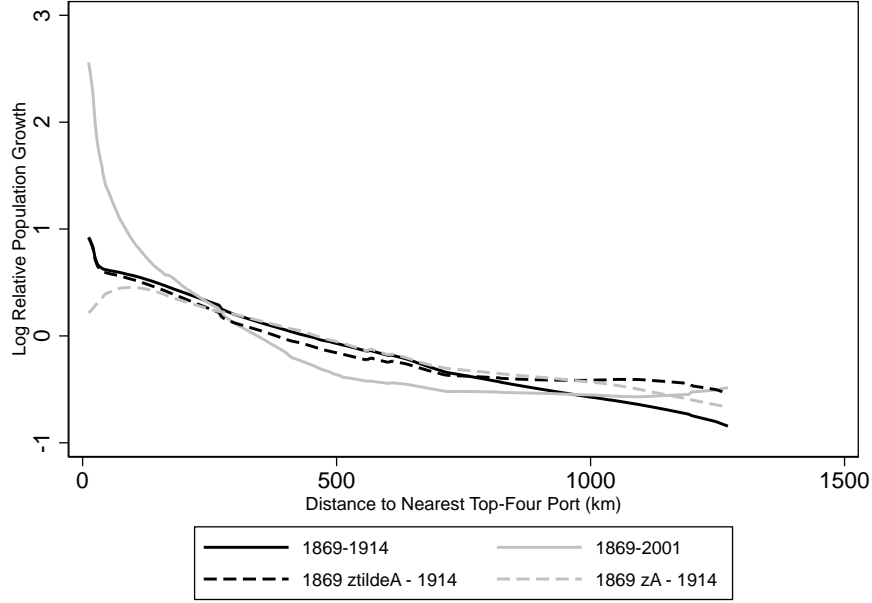
Note: Fitted values from locally-weighted linear least squares regressions of population growth on geographical (Great Circle) distance to the nearest top-four port (Buenos Aires, Rosario, La Plata and Bahia Blanca) for districts in our model sample; black solid line shows fitted values for actual population growth from 1869-1914; gray solid line shows fitted values for actual population growth from 1869-2001; black dashed line shows counterfactual population growth from 1869-1914 from changing only adjusted agricultural productivity ($\widehat{\widetilde{z}}_{At}(\ell)$), holding constant our other sufficient statistics ($z_{Nt}(\ell)$, N_t^W, u_t^{RW}); gray dashed line shows counterfactual population growth from 1869-1914 from changing agricultural productivity based on our land shares estimation ($\widehat{\widetilde{z}}_{A\chi}(\ell)$ from equation (55)), holding constant our other sufficient statistics ($z_{Nt}(\ell)$, N_t^W, u_t^{RW}); all log population growth variables normalized to have a mean of zero.

Counterfactuals for External Integration   To incorporate changes in the absolute level of export and import prices, and explore further the roles of external and internal integration, we next report counterfactuals using direct data on transatlantic freight rates from Tena-Junguito and Willebald (2020). Under our baseline assumption of a small open economy, there is complete passthrough of changes in transport costs into changes in prices, which implies that the changes in the prices of the exported agricultural goods ($\widehat{P}_{g\chi}^*$) are equal to the inverse of the changes in transatlantic freight rates ($(\widehat{\tau}_{g\chi}^*)^{-1}$). Thus, the implied change in agricultural productivity is:

$$\widehat{z}_{A\chi}^{\text{External}}(\ell) = \left[\sum_{g=1}^{G} l_{gt}(\ell) \left(\widehat{\tau}_{g\chi}^*\right)^{-\theta} \right]^{\frac{1}{\theta}}. \tag{56}$$

We measure the impact of external integration on the tradeables price index at Argentina's trade hub by combining these estimates of changes in transatlantic freight rates for each exported agricultural good with corresponding estimates for imported manufacturing goods, as discussed further in Section A.6 of the online appendix:

$$\widehat{E}_{T\chi}^{\text{External}}(\ell) = \left(\widehat{\tau}_{M\chi}^*\right)^{1-\gamma_A} \prod_{g=1}^{G} \left(\left(\widehat{\tau}_{g\chi}^*\right)^{-1} \right)^{\gamma_g}, \qquad \text{where} \qquad \sum_{g=1}^{G} \gamma_g = \gamma_A, \tag{57}$$

where lower transatlantic freights for imported goods ($\widehat{\tau}_{M\chi}^*$) imply lower import prices ($\widehat{P}_{M\chi}^*$).

In Row (4) in Panel C, we report the results of changing adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) by the amount implied by these changes in agricultural productivity and the tradeables price index ($\widehat{\widetilde{z}}_{A\chi}(\ell) = \widehat{z}_{A\chi}^{\text{External}}(\ell) / \widehat{E}_{\chi}^{\text{External}}(\ell)$), holding constant our other sufficient statistics ($z_{Nt}(\ell)$, N_t^W, u_t^{RW}). We find that reversing external integration reduces expected utility, total population and real GDP to 92.9, 86.2 and 82.3 percent of their 1914 values, respectively.

These results imply welfare gains from the external integration in late 19th-century Argentina of 7.1 percent, which is broadly in line with existing estimates of the welfare gains from trade in the quantitative international trade literature. For example, Bernhofen and Brown (2005) estimate an upper bound to the welfare gains from Japan's emergence from autarky in the 19th century of 8-9 percent of GDP.[40] Our framework incorporates international population mobility, which implies that an increase in expected utility from external integration induces a population inflow that bids up the price of land, and hence dampens welfare gains from external integration relative to a setting in which population is exogenous. Taking into account the changes in both expected utility and population, we find an increase in Argentina's real GDP of 17.7 percent from external integration in the late-19th century.

Counterfactuals for Internal Integration   In our fifth counterfactual for internal integration, we focus on the estimated impact of the changes in travel times from the construction of the railroad network from our land shares estimation. For comparability with external integration above, we focus on railroad lines constructed from 1869-1914, leaving unchanged the small length of lines already completed in 1869. In particular, we compute the change in agricultural productivity ($\widehat{z}_{A\chi}(\ell)$) induced by the changes in travel times ($\widehat{\tau}_{\chi}(\ell, \ell^*)$) from removing all railroad lines constructed from 1869-1914:

$$\widehat{z}_{A\chi}^{\text{Internal}}(\ell) = \widehat{\tau}_{\chi}(\ell, \ell^*)^{\phi_X} \left[\sum_{g=1}^{G} l_{gt}(\ell) \, \widehat{\tau}_{\chi}(\ell, \ell^*)^{\phi_g} \right]^{\frac{1}{\theta}}, \tag{58}$$

where the term before the square parentheses captures the impact of the change in travel times on the absolute level of prices for exported goods.[41] We compute the impact of the railroad network on the overall tradeables price index ($\widehat{E}_{Tt}(\ell)$) using our estimates from Step 6(iii) above:

$$\widehat{E}_{T\chi}^{\text{Internal}}(\ell) = \widehat{\tau}_{\chi}(\ell, \ell^*)^{\phi_T}. \tag{59}$$

In Row (5) in Panel C, we report the results of changing adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) by these predicted changes in agricultural productivity and the tradeables price index ($\widehat{\widetilde{z}}_{A\chi}(\ell) = \widehat{z}_{A\chi}^{\text{Internal}}(\ell) / \widehat{E}_{T\chi}^{\text{Internal}}(\ell)$), holding constant our other sufficient statistics ($z_{Nt}(\ell)$, N_t^W, u_t^{RW}). We find that removing the railroad network reduces expected utility, total population and real GDP to 95.2, 90.6 and 87.2 percent of their 1914 values, respectively. These results for the impact of the railroad network in 19th-century Argentina are also in line with the range of findings in the existing empirical literature. For example, Fogel (1964) and Donaldson and Hornbeck (2016) estimate that the impact of the construction of the railroad network in the 19th-century United States was to raise the value of agricultural land by the equivalent of 2.7 and 3.2 percent of real GDP, respectively. By comparison, Donaldson (2018) finds that the construction of the Indian railroad network raised agricultural real income by 16 percent, while Hornbeck and Rotemberg (2019) obtain estimates of up to 28 percent on real GDP, once changes in manufacturing productivity from reduced misallocation are taken into account.

In principle, the welfare gains from the construction of the railroad network in 19th-century Argentina again could be either higher or lower than in these other settings. On the one hand, our estimates are based on both the

---

[40]In principle, welfare gains from trade could be either higher or lower for Argentina than for Japan. On the one hand, Argentina by 1914 was a small open economy with a relatively high share of trade in GDP, which implies relatively large welfare gains from trade. On the other hand, Argentina was already more open to trade in 1869 than Japan was at the time of its integration into the world economy.

[41]We regress the log of traded goods prices ($P_{gt}(\ell)$) on log travel times ($\tau_t(\ell, \ell^*)$), including an intercept to control for international prices at Argentina's trade hub (P_{gt}^*), and instrumenting actual travel times using travel times implied by our colonial postal routes instrument, as in previous steps. Using the estimated coefficient (ϕ_X) and the observed change travel times ($\widehat{\tau}_{\chi}(\ell, \ell^*)$), we generate predicted changes in the absolute level of traded goods prices from the removal of the railroad network ($\widehat{\tau}_{\chi}(\ell, \ell^*)^{\phi_X}$).

agricultural and non-traded sectors, which implies larger increases in real income than for the agricultural sector by itself. On the other hand, economic activity in the tradeables sector was overwhelmingly concentrated in agriculture, thereby reducing the scope for impacts on manufacturing activity. Additionally, as for external integration above, our framework incorporates international population mobility, which dampens the impact of the construction of the railroad network on welfare, because of the induced population inflow, which bids up the price for land. Therefore, we find a welfare gain in terms of expected utility of 4.8 percent, but an increase in real GDP of 12.8 percent, which lies in the middle of the range of existing estimates above. The corresponding increase in land income equals 6.5 percent of 1914 GDP (including effects through both the agricultural and non-traded sectors). We compare these estimated impacts of the railroad network to its construction costs further below.

Counterfactuals for External and Internal Integration   In our sixth counterfactual, we evaluate the combined effect of external and internal integration, by simultaneously changing transatlantic freight rates and travel times from the construction of the railroad network using equations (56)-(59). As reported in Row (6) in Panel C, we find substantial estimated impacts from reversing both external and internal integration, with expected utility, population and real GDP falling to 88.5, 78.2 and 72.0 percent of their 1914 values, respectively.

Comparing across rows, we find that our direct measures of external and internal integration in Row (6) can account for sizeable amount of the overall impact of changes in adjusted agricultural productivity in the traded sector in Row (2). The fact that the estimated impact of changes in adjusted agricultural productivity in Row (2) remains above the combined impact of external and internal integration in Row (6) is consistent with secular productivity growth from increases in the absolute level of technology over our long historical time period. We find that our estimates using direct measures of external and internal integration in Row (6) lie relatively close to those using the predictions from our land shares estimation in Row (3).

Robustness   We demonstrate the robustness of our quantitative results across a range of different specifications. First, our baseline specification assumes that Argentina is a small open economy that faces exogenous prices on world markets ($\left\{ P_{gt}^* \right\}_{g=1}^{G}, P_{Mt}^*$), which implies complete passthrough of changes in transatlantic freight rates into export and import prices. In Section A.5.2 of the online appendix, we allow for endogenous changes in the international terms of trade, which dampens the impact of changes in external integration, because it gives rise to incomplete passthrough. For central values of trade elasticities from the existing empirical literature ranging from 3-5, we show that allowing for an endogenous terms of trade has only modest effects on our quantitative conclusions.

Second, in Section A.5.3 of the online appendix, we report a robustness exercise in which we allow for agglomeration forces in both the non-traded and agricultural sectors. Using conventional values for the elasticity of productivity with respect to employment density from the existing agglomeration literature of up to 0.10, we also find that the introduction of these agglomeration forces has only modest effects on our quantitative conclusions.

Third, in Section A.3.1 of the online appendix, we generalize our baseline homothetic CES specification to allow for non-homothetic CES preferences. In Section A.5.4 of the online appendix we implement this non-homothetic CES specification, using a range of parameterizations consistent with the estimates in Comin, Lashkari, and Mestieri (2021). We find that the introduction of non-homotheticities has relatively little impact on the distribution of adjusted agricultural productivity from our model inversion. Undertaking counterfactuals for external and internal integration,

we demonstrate a similar pattern of results as for our baseline homothetic specification, with the counterfactual predictions for changes in population and welfare changing by less than one percentage point.

Comparison with Railroad Construction Costs   We now compare the counterfactual predictions of our model for the economic impact of the construction of the railroad network to historical estimates of its construction costs. First, we evaluate this economic impact at the level of external integration in 1914, which corresponds to undertaking a counterfactual for the removal of all railroad lines constructed from 1869-1914, starting from the observed equilibrium in our baseline year of $t = 1914$. Second, we evaluate this economic impact at the level of external integration in 1869, which involves first undertaking a counterfactual for reversing the external integration that occurred from 1869-1914, and then undertaking a counterfactual for the removal of all railroad lines constructed from 1869-1914. The difference between these last two counterfactuals corresponds to the economic impact of the railroad network starting from 1869 levels of external integration.

We evaluate the economic impact of the construction of the railroad network using the net present values of changes in real GDP (which captures the income of both the mobile and immobile factors) and real land income (which focuses on the income of the immobile factor, as in the classical approach to valuing public goods following George 1879), and as discussed further in Section A.5.5 of the online appendix. We compute net present values assuming an infinite lifetime, and either 5 or 3 percent discount rates, which are standard values in cost-benefit analyses. We compare these changes in net present values to historical estimates of construction costs, based on the total capital issued for each railroad line, as discussed in further detail in Section A.6 of the data appendix. For simplicity, we assume that these construction costs are not directly affected by the level of external integration.[42]

In Table A.6 of the online appendix, we report the results of both sets of counterfactuals. Starting at the observed equilibrium in the data in 1914, we find that the removal of the railroad network reduces the net present value of real GDP and real land income by 4,976-8,293 and 2,537-4,228 million pesos, respectively, depending on whether we assume a 5 or 3 percent discount rate. Starting from 1869 levels of external integration, we find that the removal of the railroad network reduces the net present value of real GDP and real land income by 3,993-6,655 and 2,039-3,398 million pesos, again depending on whether we assume a 5 or 3 percent discount rate. Total construction costs as measured by the capital issued by all railroad lines were 1,308 million pesos in 1914 prices. Therefore, in both cases, we find ratios of changes in the net present value of income and land payments to construction costs that are greater than one using either discount rate.

However, the ratios of net present values to construction costs are substantially larger starting from 1914 levels of external integration than from those in 1869, which is intuitive. A uniform percentage reduction in internal transport costs leads to the same *percentage* increase in aggregate real GDP and land income in the model, regardless of the value of international prices. Therefore, although the reduction in internal transport costs from the construction of the railroad network are not uniform, we find only small differences in the percentage changes in aggregate real GDP and land income, depending on whether we start from 1869 or 1914 levels of external integration. Nevertheless, the absolute values of the changes in aggregate real GDP and land income are larger for higher levels of external

---

[42]In practice, almost all of Argentina's railroad equipment was imported, which implies that the construction costs of the railroad network likely would have been higher at levels of external integration in 1869 than at those in 1914. Therefore, our finding below that greater external integration magnifies the ratio of benefit to cost is conservative, because this magnification would be even greater if we allowed the construction costs of the railroad network to be higher in a more closed economy.

integration, relative to the fixed costs of the construction of the railroad network.

8 Conclusions

We provide new theory and evidence on the role of external integration (reductions in international transport costs) and internal integration (reductions in domestic transport costs) in the process of structural transformation and economic development, using Argentina's late-19th-century integration into the world economy as a natural experiment. We introduce the new *spatial Balassa-Samuelson effect* as a key feature of economic development: Locations closer to world markets have higher population densities, urban population shares, relative prices of non-traded goods, and land prices relative to wages, and specialize in the most trade-cost-sensitive traded goods. Therefore, these locations not only have higher overall levels of economic activity, but also experience structural transformation, both between the traded and non-traded sectors, and across goods within the traded sector.

We develop a novel theoretical model of the spatial distribution of economic activity across sectors and locations that provides microeconomic foundations for this spatial Balassa-Samuelson effect. We show that locations with low transport costs to world markets are attractive for the production and consumption of traded goods, which increases population density, and bids up the reward of the immobile factor (land) relative to the mobile factor (labor). This increase in population density and reduction in the wage-rental ratio together imply an expansion in the employment share of the labor-intensive non-traded sector, which with inelastic demand requires a rise in the relative price of the non-traded good. As these locations close to world markets have high relative export prices for the most-transport-cost-sensitive goods, they also specialize in these disaggregated goods within the agricultural sector.

We use our rich, spatially-disaggregated data for Argentina to estimate the model's parameters. We find inelastic demand between sectors, with an elasticity of substitution between traded and non-traded goods of 0.49, consistent with the macroeconomics literature on structural transformation. We estimate substantial population mobility, with elasticities of population with respect to real income of 4.73 across locations within Argentina and 2.02 between Argentina and the rest of the world, in line with the range of existing estimates. We also find substantial heterogeneity in idiosyncratic productivity across disaggregated goods within the traded sector, with an estimated elasticity of revenue shares within the traded sector with respect to relative prices of 3.18, comparable to existing findings in international trade. We show that the model has good within-sample fit for targeted moments, based on the observed data for wage-rental ratios and the relative price of traded goods. We also show that the model has predictive power for non-targeted moments, including separate data on railroad shipments and machinery use that are not used in the estimation of the model's parameters.

We use our estimated model to undertake counterfactuals for external and internal integration. We find that estimated changes in transatlantic freight rates for the period 1869-1914 imply increases in Argentina's GDP, population and welfare of 17.7, 13.8 and 7.1 percent, respectively. By comparison, our estimated reductions in internal trade costs from the construction of the railroad network imply increases in Argentina's GDP, population and welfare of 12.8, 9.4 and 4.8 percent, respectively. This expansion in economic activity from railroad construction raises land income by around 6.5 percent of 1914 gross domestic product (GDP), including effects on both the agricultural and non-traded sectors. The resulting increase in the net present value of land income exceeds historical estimates of the railroad's construction costs. Therefore, these large-scale investments in transport infrastructure during the 19th-century can

be rationalized in terms of their impact on economic activity. We find higher ratios of net present values of land income to construction costs at the levels of external integration in 1914 than at those in 1869. Whereas the railroad construction costs are fixed, the absolute increase in the level of economic activity from the construction of the railroad network is larger at the higher levels of external integration in 1914.

Although we focus on 19th-century Argentina because it provides an attractive empirical setting, we show that the spatial Balassa-Samuelson effect is a generic implication of the neoclassical assumptions of population mobility, labor-intensive non-tradeables and inelastic demand between sectors, and hence is relevant for other settings in which these assumptions apply. In many developing countries today, areas close to large ports have high population density, high shares of employment in the non-traded sector, high land rents relative to wages, and high relative prices of non-traded goods, consistent with our spatial Balassa-Samuelson effect. While we derive the spatial Balassa-Samuelson effect in a static general equilibrium model, exploring its implications in dynamic models of migration and factor accumulation is an interesting area for further research.

References

ADELMAN, J. (1994): *Frontier Development: Land, Labor and Capital on the Wheatlands of Argentina and Canada 1890-1914*. Oxford University Press, Oxford.

ALLEN, T., AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129(3), 1085–1140.

ALSINA, J. A. (1905): *El Obrero en la Repuública Argentina*. Imprenta, Calle de Mexico, Buenos Aires.

AMARAL, S. (1998): *The Rise of Capitalism on the Pampas: The Estancias of Buenos Aires, 1785-1870*. Cambridge University Press, Cambridge.

ATKIN, D., AND D. DONALDSON (2015): "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," *NBER Working Paper*, 21439.

BAH, E.-H. M. (2007): "A Three-sector Model of Structural Transformation and Economic Development," *Munich Personal RePEc Working Paper*, 10654.

BANERJEE, A., E. DUFLO, AND N. QIAN (2021): "On the Road: Transportation Infrastructure and Economic Development," *Journal of Development Economics*, forthcoming.

BAUM-SNOW, N. (2007): "Did Highways Cause Suburbanization?," *Quarterly Journal of Economics*, 122(2), 775–780.

BAUM-SNOW, N., L. BRANDT, J. HENDERSON, M. A. TURNER, AND Q. ZHANG (2017): "Roads, Railroads and Decentralization of Chinese Cities," *Review of Economics and Statistics*, 99(2), 435–448.

BENNETT, M. K. (1933): "World Wheat Crops, 1885-1932: New Series, with Areas and Yields, by Countries," *Wheat Studies*, 9(7), Stanford: Food Research Institute.

BERLINSKI, S., S. GALIANI, AND L. JAITMAN (2011): "Impact of Railways: Expansion on Spatial Population Patterns," *Lincoln Institute of Land Policy Working Paper*, WP11SB1.

BERNHOFEN, D. M., AND J. C. BROWN (2005): "An Empirical Assessment of the Comparative Advantage Gains from Trade: Evidence from Japan," *American Economic Review*, 95(1), 208–225.

BRYAN, G., AND M. MORTEN (2019): "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 127(5), 2229–2268.

BUNGE, A. E. (1918): "Costo de la vida en la Argentina, de 1910 a 1917," *Revista de EconomÃa Argentina*, 1(1), 39–63.

BURSTEIN, A., G. HANSON, L. TIAN, AND J. VOGEL (2020): "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence From the United States," *Econometrica*, 88(3), 1071–1112.

BUSTOS, P., B. CAPRETTINI, AND J. PONTICELLI (2016): "Agricultural Productivity and Structural Transformation: Evidence from Brazil," *American Economic Review*, 101(1), 304–340.

BUSTOS, P., G. GARBER, AND J. PONTICELLI (2020): "Capital Accumulation and Structural Transformation," *Quarterly Journal of Economics*, 135(2), 1037–1094.

CACOPARDO, M. C. (1967): *Republica Argentina: Cambios en Los Limites Nacionales, Provinciales y Departmentales, A Traves de Los Censos Nacionales de Poblacion*. Instituto Torcuato Di Tella, Buenos Aires.

CALIENDO, L., M. DVORKIN, AND F. PARRO (2019): "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 87(3), 741–835.

CALIENDO, L., F. PARRO, E. ROSSI-HANSBERG, AND P.-D. SARTE (2018): "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy," *Review of Economic Studies*, 85(4), 2042–2096.

CASELLI, F., AND W. J. I. COLEMAN (2001): "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, 109(3), 584–616.

CHANDRA, A., AND E. THOMPSON (2000): "Does Public Infrastructure affect Economic Activity? Evidence from the Rural Interstate Highway System," *Regional Science and Urban Economics*, 30(4), 457–490.

Comin, D., D. Lashkari, and M. Mestieri (2021): "Structural change with long-run income and price effects," *Econometrica*, 89(1), 311–374.

Cortés Conde, R. (1993): "The Growth of the Argentine Economy, 1870-1914," in *Argentina Since Independence*, ed. by L. Bethell, vol. 1, chap. 2. Cambridge University Press, Cambridge.

——— (1994): "Estimaciones del Producto Bruto Interno de Argentina 1875-1935," *Documento de Trabajo*, 3, Universidad San Andres, Buenos Aires.

Coşar, K., and B. Demir (2016): "Domestic Road Infrastructure and International Trade: Evidence from Turkey," *Journal of Development Economics*, 118, 232–244.

Coşar, K., and P. Fajgelbaum (2016): "Internal Geography, International Trade, and Regional Specialization," *American Economic Journal: Microeconomics*, 8(1), 24–56.

Costinot, A., and A. Rodríguez-Clare (2014): "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in *Handbook of International Economics*, ed. by E. Helpman, G. Gopinath, and K. Rogoff, vol. 4, pp. 197–262. Elsevier North Holland, Amsterdam.

Davis, D. R., and J. Dingel (2019): "A Spatial Knowledge Economy," *American Economic Review*, 109(1), 153–70.

Davis, D. R., and D. E. Weinstein (2002): "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92(5), 1269–1289.

Dirección General de Estadística (1895): *Anuario de la Dirección General de Estadística*. Dirección General de Estadística, Buenos Aires.

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

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

Droller, F. (2018): "Migration and Long-run Development: Evidence from Settlements in the Pampas," *Economic Journal*, 128(614), 2321–2352.

Duarte, M., and D. Restuccia (2010): "The Role of the Structural Transformation in Aggregate Productivity," *Quarterly Journal of Economics*, 125, 129–173.

Duranton, G., P. Morrow, and M. A. Turner (2014): "Roads and Trade: Evidence from the US," *Review of Economic Studies*, 81(2), 681–724.

Duranton, G., and M. A. Turner (2012): "Urban Growth and Transportation," *Review of Economic Studies*, 79(4), 1407–1440.

Eaton, J., and S. Kortum (2002): "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

Eckert, F., and M. Peters (2018): "Spatial Structural Change," Yale University, mimeograph.

Faber, B. (2014): "Integration and the Periphery: The Unintended Effects of New Highways in a Developing Country," *Review of Economic Studies*, 81(3), 1046–1070.

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

Francis, J. A. (2017): "Globalisation, the Terms of Trade, and Argentina's Expansion in the Long Nineteenth Century," *Journal of Latin American Studies*, 49, 709–738.

Frankema, E. (2010): "Reconstructing Labor Income Shares in Argentina, Brazil and Mexico, 1870-2000," *Revista de Historia Económica*, 28(2), 343–374.

Fujita, M., P. Krugman, and A. J. Venables (1999): *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge MA.

Galle, S., M. Yi, and A. Rodriguez-Clare (2020): "Slicing the Pie: Quantifying the Aggregate and Distributional Consequences of Trade," University of Calirfornia, Berkeley, mimeograph.

GEORGE, H. (1879): *Progress and Poverty*. Doubleday, Page & Company, New York.

GOLLIN, D., R. JEDWAB, AND D. VOLLRATH (2016): "Urbanization With and Without Industrialization," *Journal of Economic Growth*, 21(1), 35–70.

GOLLIN, D., AND R. ROGERSON (2014): "Productivity, Transport Costs and Subsistence Agriculture," *Journal of Development Economics*, 107, 38–48.

HANSON, G. H. (1996): "Localization Economies, Vertical Organization, and Trade," *American Economic Review*, 86(5), 1266–1278.

HARLEY, C. K. (1980): "Transportation, World Wheat Trade, and the Kuznets Cycle, 1850-1913," *Explorations in Economic History*, 17, 218–250.

――― (1988): "Ocean Freight Rates and Productivity, 1740-1913: The Primacy of Mechanical Invention Reaffirmed," *The Journal of Economic History*, 48(4), 851–876.

HEBLICH, S., S. REDDING, AND D. STURM (2020): "The Making of the Modern Metropolis: Evidence from London," *Quarterly Journal of Economics*, 135(4), 2059–2133.

HERRENDORF, B., R. ROGERSON, AND A. VALENTINYI (2014): "Growth and structural transformation," in *Handbook of economic growth*, vol. 2, pp. 855–941. Elsevier.

HERRENDORF, B., J. A. SCHMITZ, AND A. TEIXEIRA (2012): "Transportation and Development: Insights from the U.S., 1840-1860," *International Economic Review*, 53(3), 693–716.

HORNBECK, R., AND M. ROTEMBERG (2019): "Railroads, Reallocation, and the Rise of American Manufacturing," *NBER Working Paper*, 26594.

KARÁDI, P., AND M. KOREN (2018): "Cattle, Steaks and Restaurants: Development Accounting when Space Matters," Central European University, mimeograph.

LAGAKOS, D., AND M. E. WAUGH (2013): "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, 103(2), 948–980.

LALL, S. V., AND M. LEBRAND (2020): "Who Wins, Who Loses? Understanding the Spatially Differentiated Effects of Belt and Road Within Central Asia," *Journal of Development Economics*, 146(102496).

MADDISON, A. (2003): *The World Economy: Historical Statistics*. Organisation for Economic Cooperation and Development (OECD), Paris.

MATSUYAMA, K. (1992): "Agricultural Productivity, Comparative Advantage and Economic Growth," *Journal of Economic Theory*, 58, 317–334.

――― (2009): "Structural Change in an Interdependent World: A Global View of Manufacturing Decline," *Journal of the European Economic Association*, 7, 478–486.

MCMILLAN, M., D. RODRIK, AND C. SEPULVEDA (2017): "Structural Change, Fundamentals and Growth: A Framework and Case Studies," *NBER Working Paper*, 23378.

MICHAELS, G. (2008): "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *Review of Economics and Statistics*, 90(4), 683–701.

MICHAELS, G., F. RAUCH, AND S. J. REDDING (2012): "Urbanization and Structural Transformation," *Quarterly Journal of Economics*, 127, 535–586.

MORTEN, M., AND J. OLIVEIRA (2017): "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *NBER Working Paper*, 22158.

NAGY, D. (2017): "Hinterlands, City formation and Growth: Evidence from the U.S. Westward Expansion," CREI, mimeograph.

NGAI, R., AND C. PISSARIDES (2007): "Structural Change in a Multisector Model of Growth," *American Economic Review*, 97(1), 429–443.

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

O'Rourke, K. H., and J. G. Williamson (1999): *Globalization and History: The Evolution of a Nineteenth-Century Atlantic Economy*. MIT Press, Cambridge MA.

Pascali, L. (2017): "The Wind of Change: Maritime Technology, Trade and Economic Development," *American Economic Review*, 107(9), 2821–54.

Perren, R. (2017): *Taste, Trade and Technology: The Development of the International Meat Industry since 1840*. Routledge, New York.

Peters, M. (2019): "Market Size and Spatial Growth: Evidence from Germany's Post-war Population Expulsions," .

Ramondo, N., A. Rodriguez-Clare, and M. Saborio (2016): "Scale Effects and Productivity: Does Country Size Matter?," *American Economic Review*, 106(10), 3159–84.

Randle, P. H. (1981): *Atlas del Desarrollo Territorial de la Argentina*. Oikos, Buenos Aires.

Redding, S. J. (2016): "Goods Trade, Factor Mobility and Welfare," *Journal of International Economics*, 101, 148–167.

——— (2020): "Trade and Geography," *NBER Working Paper*, 27821.

Redding, S. J., and D. M. Sturm (2008): "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98(5), 1766–1797.

Redding, S. J., and M. A. Turner (2015): "Transportation Costs and the Spatial Organization of Economic Activity," in *Handbook of Regional and Urban Economics*, ed. by G. Duration, J. V. Henderson, and W. Strange, chap. 20, pp. 1339–1398. Elsevier, Amsterdam.

República Argentina (1913): *Boletin del Departamento Nacional de Trabajo*, vol. 25. República Argentina, Buenos Aires.

Rocchi, F. (2008): *Chimneys in the Desert: Industrialization in Argentina During the Export Boom Years 1870-1930*. Stanford University Press, Stanford.

Rogerson, R. (2008): "Structural Transformation and the Deterioration of European Labor Market Outcomes," *Journal of Political Economy*, 116, 235–59.

Scobie, J. R. (1971): *Argentina: A City and a Nation*. Oxford University Press, New York.

Simonovska, I., and M. E. Waugh (2014): "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 92(1), 34–50.

Sotelo, S. (2020): "Domestic Trade Frictions and Agriculture," *Journal of Political Economy*, 128(7), 2690–2738.

Storeygard, A. (2016): "Farther on Down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa," *Review of Economic Studies*, 83, 1263–1295.

Taylor, A. (1992): "External Dependence, Demographic Burdens, and Argentine Economic Decline After the Belle Époque," *Journal of Economic History*, 52(4), 907–936.

Taylor, A., and J. G. Williamson (1997): "Convergence in the Age of Mass Migration," *European Review of Economic History*, 1, 27–63.

Tena-Junguito, A., and H. Willebald (2020): "On the Accuracy of Export Growth in Argentina, 1870-1913," *Economic History of Developing Regions*, 28(1), 28–68.

Üngör, M. (2017): "Productivity Growth and Labor Reallocation: Latin America versus East Asia," *Review of Economic Dynamics*, 24, 25–42.

Uy, T., K.-M. Yi, and J. Zhang (2012): "Structural Change in an Open Economy," *Journal of Monetary Economics*, 60(6), 667–682.

Venables, A., and N. Limao (2002): "Geographical Disadvantage: a Heckscher-Ohlin-von-Thünen Model of International Specialization," *Journal of International Economics*, 58, 239–263.