Online Appendix for "Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914" (Not for Publication)

Pablo Fajgelbaum

Princeton University and NBER

Stephen J. Redding

Princeton University, NBER and CEPR

October 2021

Table of Contents

A.1 Introduction

Section A.2 reports the technical derivations of expressions reported in the paper and the proofs of propositions. Equations in the main paper are referenced by their number (for example (1)), while equations in this web appendix are referenced by the letter A and their number (for example (A.1)). Section A.3 reports theoretical extensions and additional theoretical results that are discussed in the paper. Subsection A.3.1 introduces our extension with non-homothetic preferences; Subsection A.3.2 provides further results for the extensive margin of specialization; and Subsection A.3.3 presents our extension with the endogenous conversion of land to productive use.

Section A.4 contains additional empirical results that are discussed in the paper. Subsection A.4.1 presents further evidence on transatlantic freight rates; Subsection A.4.2 reports additional reduced-form empirical results; Subsection A.4.3 provides further empirical results for our estimation of the production cost share parameters (α_A, α_N); Subsection A.4.4 presents additional evidence for the estimation of the demand parameters (σ, β_T); and Subsection A.4.5

reports further empirical results for our land shares estimation of the productivity dispersion parameter (θ), including a number of overidentification checks on our estimation procedure.

Section A.5 provides further results for the counterfactuals from Section 7 of the paper. Section A.6 reports additional information about the data sources and definitions.

A.2 Theoretical Model

Subsection A.2.1 presents the formal full definition of equilibrium in the model. Subsection A.2.2 solves for a local equilibrium in each location $\ell \in \mathcal{L}$ for given prices of traded goods $\{P_g(\ell), P_M(\ell)\}$ and common expected utility in Argentina (u^*). Subsection A.2.3 characterizes the comparative statics of a local equilibrium with respect to agricultural productivity. Subsections A.2.4 to A.2.6 provides proofs of Proposition 1 to 3.

A.2.1 Definition of Equilibrium

Definition 1. A general equilibrium consists of a real wage 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 $w(\ell)$; land rents $r(\ell)$; and prices $\{P_g(\ell)\}_{g=1}^{G}$, $P_M(\ell)$, $P_N(\ell)$ for all locations $\ell \in \mathcal{L}$ such that

(i) workers maximize utility and choose their location optimally, i.e.,

$$n(\ell)\, L(\ell) = N \left(\frac{u(\ell)}{u^*}\right)^{\varepsilon} \tag{A.1}$$

where $u(\ell) = w(\ell)/E(\ell)$, and where the population of the economy is

$$\frac{N}{N^W} = \frac{(u^*)^{\varepsilon^{INT}}}{(u^{RW})^{\varepsilon^{INT}} + (u^*)^{\varepsilon^{INT}}}; \tag{A.2}$$

(ii) land is allocated optimally across sectors,

$$r(\ell) = \max\{r_A(\ell), r_M(\ell), r_N(\ell)\},$$

where, from producer optimization, zero profit conditions hold:

$$P_i(\ell) z_i(\ell) \le w(\ell)^{1-\alpha_i} r(\ell)^{\alpha_i} \text{ for } i = A, M, N, \tag{A.3}$$

with $=$ if $n_i(\ell) > 0$;

(iii) the land market clears in each location,

$$\sum_{i=M,N,A} L_i(\ell) = L(\ell); \tag{A.4}$$

(iv) the labor market clears in each location,

$$\sum_{i=M,N,A} \frac{L_i(\ell)}{L(\ell)} n_i(\ell) = n(\ell) \tag{A.5}$$

where, from producer optimization, employment to land ratios are:

$$n_i(\ell) = \frac{1-\alpha_i}{\alpha_i} \frac{r(\ell)}{w(\ell)}; \tag{A.6}$$

(v) the non-traded goods market clears in each location,

$$c_N(\ell) = \frac{L_N(\ell)}{L(\ell)} q_N(n_N(\ell)), \tag{A.7}$$

where, from consumer optimization, non-traded consumption is:

$$c_N(\ell) = \beta_T \left(\frac{P_N(\ell)}{E(\ell)}\right)^{-\sigma} \frac{y(\ell)}{E(\ell)}; \tag{A.8}$$

(vi) traded goods prices are determined by no arbitrage, i.e.

- if a location ℓ exports an agricultural good g to the rest of the world, its price equals the price at the trade hub less transport costs, $P_g(\ell) = P_g^*/\delta_g(\ell)$, where $\delta_g(\ell) = \min_{\ell' \in \mathcal{L}_C} \{\delta(\ell, \ell')\}$,
- if the location ℓ imports the manufacturing good M from the rest of the world its price equals $P_M(\ell) = \delta_M(\ell) P_M^*$, where $\delta_M(\ell) = \min_{\ell' \in \mathcal{L}_C} \{\delta_M(\ell, \ell')\}$; and

(vii) the common real wage u^* adjusts to clear the labor market for the economy as a whole, i.e. condition (9) in the paper holds.

A.2.2 Local Equilibrium in Levels

Step 1: we use equilibrium conditions (iii) to (v) from Definition 1 to solve for $\{n(\ell), \nu_A(\ell)\}$ as a function of the wage-rental ratio $\omega(\ell)$ and the share of expenditures in tradeables, $s_T(\ell)$. Combining producer optimization in non-tradeables from equation (14) in the paper and non-tradeable market clearing (A.7), we obtain:

$$\frac{N_N(\ell)}{L(\ell)} = (1-\alpha_N)\left(n(\ell) + \omega(\ell)^{-1}\right)(1 - s_T(\ell)), \tag{A.9}$$

$$\frac{L_N(\ell)}{L(\ell)} = \alpha_N \omega(\ell)\left(n(\ell) + \omega(\ell)^{-1}\right)(1 - s_T(\ell))). \tag{A.10}$$

Using these two expressions and producer optimization in agriculture from equation (14) in the paper, the factor market clearing conditions (iii) and (iv) of Definition 1 can be re-written as follows:

$$\alpha_N (1 - s_T(\ell)) \left(n(\ell) + \frac{1}{\omega(\ell)}\right) + \frac{\lambda_A(\ell)}{\omega(\ell)} = \frac{1}{\omega(\ell)}, \tag{A.11}$$

$$(1-\alpha_N)(1 - s_T(\ell)) \left(n(\ell) + \frac{1}{\omega(\ell)}\right) + \frac{1-\alpha_A}{\alpha_A}\frac{\lambda_A(\ell)}{\omega(\ell)} = n(\ell), \tag{A.12}$$

where $\lambda_A(\ell) = L_A(\ell)/L(\ell)$ is the land share in agriculture. Adding up these two equations, we obtain

$$\lambda_A(\ell) = \alpha_A s_T(\ell)(1 + \omega(\ell) n(\ell)).$$

Replacing this back in equation (A.11) we obtain population density:

$$n(\ell) = \left(\frac{1}{\alpha_N + (\alpha_A - \alpha_N) s_T(\ell)} - 1\right)\frac{1}{\omega(\ell)} \tag{A.13}$$

as well as the agricultural labor share:

$$\nu_A(\ell) = \frac{(1-\alpha_A) s_T(\ell)}{1 - \alpha_N - (\alpha_A - \alpha_N) s_T(\ell)}. \tag{A.14}$$

Step 2: Solve for $\{w(\ell), r(\ell), P_N(\ell), s_T(\ell)\}$ as function of the wage rental-ratio and local utility, $\{\omega(\ell), u(\ell)\}$. First, using (ii) and equation (15) in the paper evaluated at $i = A$ gives the prices as function of the wage-rental ratio,

$$w(\ell) = z_A(\ell)\,\omega(\ell)^{\alpha_A}, \tag{A.15}$$

$$r(\ell) = z_A(\ell)\,\omega(\ell)^{\alpha_A - 1}, \tag{A.16}$$

$$P_N(\ell) = \frac{z_A(\ell)}{z_N(\ell)}\omega(\ell)^{\alpha_A - \alpha_N}. \tag{A.17}$$

Second, we combine these expressions with $w(\ell) = u(\ell)\,E(\ell)$ from (i) to write the ratio of the price index in the traded sector to the total price index as follows:

$$\frac{E_T(\ell)}{E(\ell)} = \frac{\omega(\ell)^{-\alpha_A}}{\widetilde{z}_A(\ell)}u(\ell), \tag{A.18}$$

$$\frac{P_N(\ell)}{E(\ell)} = \frac{\omega(\ell)^{-\alpha_N}}{\widetilde{z}_N(\ell)}u(\ell). \tag{A.19}$$

Finally, combining $\frac{E_T(\ell)}{E(\ell)}$ with the definition of the tradeable expenditure share we obtain:

$$s_T(\ell) \equiv \beta_T\left(\frac{\omega(\ell)^{-\alpha_A}}{\widetilde{z}_A(\ell)}u(\ell)\right)^{1-\sigma}. \tag{A.20}$$

Step 3: Following the steps in Section 5.7 in the paper we reach expression (20) in the main text, which can be written as follows:

$$\beta_T u(\ell)^{1-\sigma}\left(\widetilde{z}_A(\ell)\,\omega(\ell)^{-\alpha_A}\right)^{\sigma-1} + (1-\beta_T)\,u(\ell)^{1-\sigma}\left(\widetilde{z}_N\omega(\ell)^{-\alpha_N}\right)^{\sigma-1} = 1. \tag{A.21}$$

This expression gives a unique solution for $\omega(\ell)$ given $u(\ell)$.

In sum, conditions (A.13) to (A.20) give a solution for $\{n(\ell), \nu_A(\ell), w(\ell), r(\ell), P_N(\ell), s_T(\ell), \omega(\ell)\}$ given $u(\ell)$. In turn, equation (A.1) gives a second relationship between $n(\ell)$ and $u(\ell)$ that we use to solve the local equilibrium given u^* and N.

A.2.3 Comparative Statics of the Local Equilibrium with Respect to Agricultural Productivity

We now derive the changes in local equilibrium variables of an economy that is completely specialized in agriculture. We consider the changes resulting from to an infinitesimal shock to agricultural productivity $z_A(\ell)$ and tradeable prices $E_T(\ell)$ keeping national utility and population $\{u^*, N\}$ constant. Therefore, the following results correspond to a comparative static across locations that are completely specialized in agriculture in a given equilibrium. In the following expressions we let $\widehat{x} \equiv d\ln x$. In response to $\{\widehat{z_A}(\ell), \widehat{E}_T(\ell)\}$, the local equilibrium consists of $\left\{\widehat{u}(\ell), \widehat{w}(\ell), \widehat{\omega}(\ell), \widehat{n}_i(\ell), \widehat{L}_i(\ell), \widehat{P}_N(\ell), \widehat{n}(\ell)\right\}$ such that

(i) workers maximize utility and choose their location optimally, i.e.,

$$\widehat{n}(\ell) = \varepsilon\widehat{u}(\ell) \tag{A.22}$$

where

$$\widehat{u}(\ell) = \widehat{w}(\ell) - \widehat{E}(\ell) \tag{A.23}$$

where the change in the price index is

$$\widehat{E}(\ell) = (1 - s_T(\ell))\,\widehat{P}_N(\ell) - s_T(\ell)\,\widehat{E}_T(\ell);$$

(ii) zero-profit conditions hold in A and N,

$$\widehat{z}_A(\ell) = \widehat{w}(\ell) - \alpha_A \widehat{\omega}(\ell), \tag{A.24}$$

$$\widehat{P}_N(\ell) = \widehat{w}(\ell) - \alpha_N \widehat{\omega}(\ell); \tag{A.25}$$

(iii) the land market clears in each location,

$$(1-\lambda_A(\ell))\,\widehat{L}_N(\ell) + \lambda_A(\ell)\,\widehat{L}_A(\ell) = 0; \tag{A.26}$$

(iv) the labor market clears in each location,

$$(1-\nu_A(\ell))\left(\widehat{n}_N(\ell) + \widehat{L}_N(\ell)\right) + \nu_A(\ell)\left(\widehat{n}_A(\ell) + \widehat{L}_A(\ell)\right) = \widehat{n}(\ell), \tag{A.27}$$

where, from producer optimization, employment to land ratios are:

$$\widehat{n}_i(\ell) = \widehat{r}(\ell) - \widehat{w}(\ell); \text{ and} \tag{A.28}$$

(v) the non-traded goods market clears in each location,

$$\widehat{c}_N(\ell) = \widehat{q}_N(\ell), \tag{A.29}$$

where, from consumer optimization,

$$\widehat{c}_N(\ell) = \left(\widehat{y}(\ell) - \widehat{E}(\ell)\right) - \sigma\left(\widehat{P}_N(\ell) - \widehat{E}(\ell)\right). \tag{A.30}$$

To solve the system, first note that, using trade balance, the non-tradeables market clearing condition (v) implies:[1]

$$\sigma\left(\widehat{P}_N(\ell) - \widehat{E}_T(\ell)\right) = \widehat{\widehat{z}_A}(\ell) + (1-\alpha_A)\,\widehat{n}_A(\ell) + \widehat{L}_A(\ell) - (1-\alpha_N)\,\widehat{n}_N(\ell) - \widehat{L}_N(\ell). \tag{A.33}$$

Second, combining the land and labor market clearing conditions (iii) and (iv):

$$\widehat{L}_N(\ell) = \frac{\lambda_A(\ell)}{\lambda_A(\ell) - \nu_A(\ell)}\left(\widehat{n}(\ell) + \widehat{\omega}(\ell)\right) \tag{A.34}$$

$$\widehat{L}_A(\ell) = -\frac{1-\lambda_A(\ell)}{\lambda_A(\ell) - \nu_A(\ell)}\left(\widehat{n}(\ell) + \widehat{\omega}(\ell)\right) \tag{A.35}$$

Then, we substitute equations (A.24), (A.25), (A.28), (A.34) and (A.35) into equation (A.33) to get a first relationship between ω and density:

$$\widehat{\omega}(\ell) = \frac{(1-\sigma)\left(\lambda_A(\ell) - \nu_A(\ell)\right)}{1-(1-\sigma)\left(\lambda_A(\ell) - \nu_A(\ell)\right)(\alpha_A - \alpha_N)}\widehat{\widehat{z}_A}(\ell) - \frac{1}{1-(1-\sigma)\left(\lambda_A(\ell) - \nu_A(\ell)\right)(\alpha_A - \alpha_N)}\widehat{n}(\ell). \tag{A.36}$$

[1] To obtain this expression, note that tradeable expenditures in changes are

$$\widehat{c}_T(\ell) = \left(\widehat{y}(\ell) - \widehat{E}(\ell)\right) - \sigma\left(\widehat{E}_T(\ell) - \widehat{E}(\ell)\right). \tag{A.31}$$

In addition, trade balance implies that, in levels, tradeable expenditure equals tradeable income: $E_T(\ell)\,C_T(\ell) = \sum_g P_g(\ell)\,Q_g(\ell)$. Combining this expression with equations (A.29), (A.30), and (A.31), relative non-traded sales change according to:

$$\widehat{P}_N(\ell) + \widehat{q}_N(\ell) - \sum_g \widehat{P_g(\ell)\,Q_g}(\ell) = (1-\sigma)\left(\widehat{P}_N(\ell) - \widehat{E}_T(\ell)\right). \tag{A.32}$$

Moreover, from the solution to the producer optimization problem, agricultural sales are: $\sum_g P_g(\ell)\,Q_g(\ell) = z_A(\ell)\left(\frac{1}{\alpha_A}\right)^{\alpha_A}\left(\frac{n_A(\ell)}{1-\alpha_A}\right)^{1-\alpha_A} L_A(\ell)$. Therefore, in changes, $\sum_g \widehat{P_g(\ell)\,Q_g}(\ell) = \widehat{z}_A(\ell) + (1-\alpha_A)\,\widehat{n}_A(\ell) + \widehat{L}_A(\ell)$. Combining this expression with equation (A.32) and using $\widehat{q}_N = (1-\alpha_A)\,\widehat{n}_N(\ell) + \widehat{L}_N(\ell)$ gives equation (A.33).

Next, we substitute equation (A.24) and (A.25) into equation (A.22) to obtain:

$$\widehat{n}(\ell) = \varepsilon s_T(\ell)\widehat{\widetilde{z}_A}(\ell) + \varepsilon\left(s_T(\ell)\alpha_A + (1 - s_T(\ell))\alpha_N\right)\widehat{\omega}(\ell) \tag{A.37}$$

Combining the last two expressions we obtain the solution for the wage-rental ratio,

$$\begin{aligned}\widehat{\omega}(\ell) &= \frac{1}{\alpha_A - \alpha_N}\left(\frac{1 + \varepsilon\alpha_N}{K(\ell)} - 1\right)\widehat{\widetilde{z}_A}(\ell) \\ &= \frac{(1-\sigma)(\lambda_A(\ell) - \nu_A(\ell)) - \varepsilon s_T(\ell)}{K(\ell)}\widehat{\widetilde{z}_A}(\ell),\end{aligned} \tag{A.38}$$

and for population density,

$$\widehat{n}(\ell) = \frac{\varepsilon\left[s_T(\ell)\left(1 + (\varepsilon - 1)(s_T(\ell)\alpha_A + (1 - s_T(\ell))\alpha_N)\right) + \alpha_N(1-\sigma)(\lambda_A(\ell) - \nu_A(\ell))\right]}{K(\ell)}\widehat{\widetilde{z}_A}(\ell), \tag{A.39}$$

where

$$K(\ell) \equiv 1 - (1-\sigma)(\lambda_A(\ell) - \nu_A(\ell))(\alpha_A - \alpha_N) + \varepsilon\left(s_T(\ell)\alpha_A + (1 - s_T(\ell))\alpha_N(\ell)\right). \tag{A.40}$$

For future reference, we note that $K(\ell) > 0$ if $\sigma < 1$.

A.2.4 Proof of Proposition 1

Proposition. *If $\alpha_A > \alpha_N$ and $\sigma < 1$, there exists a unique general equilibrium in which every location specializes in agriculture.*

Proof. We assume that each location is fully specialized in sector A and provide conditions for this to be the case in Proposition 2. Similar steps apply if all locations are fully specialized in M. First, given that $0 < \sigma < 1$, the expression on the left of equation (A.21) is continuous, strictly decreasing in $\omega(\ell)$, goes to ∞ as $\omega(\ell)$ goes to 0, and goes to 0 as $\omega(\ell)$ goes to ∞. Therefore, given $u(\ell)$, there is a unique $\omega(\ell)$ consistent with equation (20) in the paper and unique local prices $\{r(\ell), w(\ell), P_N(\ell)\}$ given by equations (A.15) to (A.17). In addition, given $\{E_T(\ell), u(\ell), \omega(\ell)\}$, the variables $E_T(\ell)/E(\ell)$ and $n(\ell)$ are uniquely determined from equation (A.18) and equation (26) in the paper. Therefore, there is a unique local equilibrium given $u(\ell)$. These relationships define a local labor demand mapping $n^D(u(\ell), \ell)$ from $u(\ell)$ to population density. In general equilibrium, $\{u(\ell), u^*, N\}$ are such that:

$$n^D(u(\ell), \ell) = \frac{N^S(u^*)}{L(\ell)}\left(\frac{u(\ell)}{u^*}\right)^{\varepsilon}, \tag{A.41}$$

$$\sum_{\ell} u(\ell)^{\varepsilon} = (u^*)^{\varepsilon}. \tag{A.42}$$

The first equation is the local labor market equilibrium, where we have used that total population in Argentina equals the foreign supply $N^S(u^*)$ defined in equation (6) in the paper, and the second equation follows from the labor market clearing condition in equation (9) in the paper. Equation (A.41) has a unique solution for $u(\ell)$ given u^* because, from equations (28) to (31) in the paper, $\frac{\partial n^D(u(\ell),\ell)}{\partial u(\ell)} < 0$ if $\alpha_A > \alpha_N$ and $\sigma < 1$. Let $\upsilon(\ell, u^*)$ be the solution to this local equilibrium:

$$\frac{n^D(\upsilon(\ell, u^*), \ell)L(\ell)}{N^S(u^*)} \equiv \left(\frac{u(\ell)}{u^*}\right)^{\varepsilon}. \tag{A.43}$$

Substituting $u(\ell) = \upsilon(\ell, u^*)$ into equation (A.42), there is a unique solution whenever the equation

$$\sum_{\ell} \frac{n^D(\upsilon(\ell, u^*), \ell)L(\ell)}{N^S(u^*)} = 1 \tag{A.44}$$

has a unique solution for u^*. When $\upsilon(\ell, u^*)$ is increasing with u^*, the expression inside the sum is decreasing, implying that a unique u^* solves equation (A.44). Moreover equation (A.44) is equivalent to:

$$\sum_{\ell} \left(\frac{\upsilon(\ell, u^*)}{u^*} \right)^{\varepsilon} = 1. \tag{A.45}$$

Therefore, when $\upsilon(\ell, u^*)$ is decreasing with u^*, a unique u^* solves this equation. Therefore, equilibrium is unique if $\upsilon(\ell, u^*)$ is either strictly increasing or strictly decreasing in u^*. From equation (A.43), the sign of $\frac{\partial \upsilon(\ell,u^*)}{\partial u^*}$ equals that of $\varepsilon - \frac{d \ln N^S}{d \ln u^*}$. Therefore, using equation (6) in the paper, an equilibrium must be unique if, for all values of N, either

$$\varepsilon^{INT} \left(1 - \frac{N}{N^W} \right) < \varepsilon,$$

in which case u^* is a negative supply shifter of each location; or if, for all values of N, the opposite is true. Assuming $\varepsilon > \varepsilon^{INT}$ is sufficient for this condition to hold. □

A.2.5 Proof of Proposition 2

Proposition. *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 ($P_M(\ell) z_M(\ell)$).*

Proof. From condition (ii) of the general equilibrium, if the economy is specialized in A then $\{w(\ell), \omega(\ell)\}$ satisfy:

$$z_A(\ell) = w(\ell)\,\omega(\ell)^{-\alpha_A}$$
$$P_M(\ell)\, z_M(\ell) \leq w(\ell)\,\omega(\ell)^{-\alpha_M}$$
$$P_N(\ell)\, z_N(\ell) = w(\ell)\,\omega(\ell)^{-\alpha_N}$$

implying

$$1 < \frac{z_A(\ell)\,\omega(\ell)^{\alpha_A - \alpha_M}}{P_M(\ell)\, z_M(\ell)} \tag{A.46}$$

If this condition holds, there is an equilibrium in which the economy specializes in A. Next, we show that the right-hand side of equation (A.46) is increasing with $\frac{z_A(\ell)}{P_M(\ell) z_M(\ell)}$. For this, we must show that it increases with $z_A(\ell)\,\omega(\ell)^{\alpha_A - \alpha_M}$. Using equation (A.38), given a shock to $z_A(\ell)$, we have:

$$\widehat{z}_A(\ell) + (\alpha_A - \alpha_M)\,\widehat{\omega}_A(\ell) = \left(1 - \frac{\alpha_A - \alpha_M}{\alpha_A - \alpha_N} \left(1 - \frac{1 + \varepsilon \alpha_N}{K(\ell)} \right) \right) \widehat{z}_A(\ell).$$

Using $\alpha_A > \alpha_N$, when $\widehat{z}_A(\ell) > 0$ this expression is positive if:

$$\alpha_N - \alpha_M < (\alpha_A - \alpha_M)\, \frac{1 + \varepsilon \alpha_N}{K(\ell)}.$$

Since $K(\ell) > 0$, this condition is verified for $\alpha_N < \alpha_M < \alpha_A$. Therefore, given this condition, equation (A.46) is verified for sufficiently large $\frac{z_A(\ell)}{P_M(\ell) z_M(\ell)}$. Since $P_M(\ell)\, z_M(\ell)$ does not affect equilibrium outcomes conditional on agricultural specialization, this condition can be guaranteed for sufficiently low $P_M(\ell)\, z_M(\ell)$. □

A.2.6 Proof of Proposition 3

Proposition. *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 $\{\delta_g(\ell,\ell^*)\}_{g=1}^{G}$, $\delta_M(\ell,\ell^*)$ to the trade hub) 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. Part (i) follows from the definition of $\widetilde{z}_A(\ell)$. For part (ii), note that:

$$\widehat{E}_T(\ell) - \widehat{E}(\ell) = \widehat{E}_T(\ell) - \Big((1-s_T(\ell))\,\widehat{P}_N(\ell) + s_T(\ell)\,\widehat{E}_T(\ell)\Big).$$

In addition, from equations (A.24) and (A.25), non-traded prices are

$$\widehat{P}_N(\ell) = (\alpha_A - \alpha_N)\,\widehat{\omega}(\ell) + \widehat{z}_A(\ell).$$

Combining these expressions with equation (A.38) we have that:

$$\widehat{E}_T(\ell) - \widehat{E}(\ell) = -(1-s_T(\ell))\,\frac{1+\varepsilon\alpha_N}{K(\ell)}\widehat{\widetilde{z}}_A.$$

When $\widehat{\widetilde{z}}_A > 0$, this expression is negative because $K(\ell) > 0$. For part (iii), we have from (A.39) that $\frac{\widehat{n}(\ell)}{\widehat{\widetilde{z}}_A(\ell)} > 0$ if $\sigma < 1$, $\varepsilon > 1$, and $\lambda_A(\ell) > \nu_A(\ell)$, which is verified for $\alpha_A > \alpha_N$.[2] Part (iv) follows from the fact that $\nu_A(\ell)$ increases with $\frac{E_T(\ell)}{E(\ell)}$ (see equation (27) in the paper) and $\frac{E_T(\ell)}{E(\ell)}$ falls with $\widehat{\widetilde{z}}_A$ from part (iii) of this proposition. For part (v), using equation (A.38) and the definition of $K(\ell)$, we have that, given $\alpha_A > \alpha_N$,

$$\frac{\omega(\ell)}{\widehat{\widetilde{z}}_A(\ell)} < 0 \longleftrightarrow \frac{1+\varepsilon\alpha_N}{1-(1-\sigma)(\lambda_A(\ell)-\nu_A(\ell))(\alpha_A-\alpha_N)+\varepsilon(\alpha_A s_T(\ell)+\alpha_N(1-s_T(\ell)))} < 1$$

As $\varepsilon \to \infty$ (perfect mobility), then

$$\frac{\omega(\ell)}{\widehat{\widetilde{z}}_A(\ell)} < 0 \longleftrightarrow \frac{\alpha_N}{\alpha_A s_T(\ell)+\alpha_N(1-s_T(\ell))} < 1,$$

which holds given $\alpha_N < \alpha_A$ regardless of σ. As $\varepsilon \to 0$ (no mobility), then

$$\frac{\omega(\ell)}{\widehat{\widetilde{z}}_A(\ell)} < 0 \longleftrightarrow \frac{1}{1-(1-\sigma)(\lambda_A(\ell)-\nu_A(\ell))(\alpha_A-\alpha_N)} < 1,$$

which is necessarily violated with $\sigma < 1$ and $\alpha_N < \alpha_A$. Given $\{\lambda_A(\ell), \nu_A(\ell), s_T(\ell)\}$ and the parameters, there must exist a threshold value of ε such that $\frac{\omega(\ell)}{\widehat{\widetilde{z}}_A(\ell)} < 0$ when ε is above that threshold and $\frac{\omega(\ell)}{\widehat{\widetilde{z}}_A(\ell)} > 0$ otherwise. □

A.3 Theoretical Extensions

In this section of the online appendix, we report a number of theoretical extensions of our model. Subsection A.3.1 generalizes our baseline specification using homothetic CES preferences in the paper to allow for non-homothetic CES preferences. Subsection A.3.2 provides further results for the extensive margin of specialization, in which there is a "trade frontier" beyond which regions further inland are in autarky. Subsection A.3.3 develops our extension in which landowners must incur a fixed cost in order to convert land to productive use.

[2] From equation (27) in the paper:

$$\begin{aligned}\text{sign}(\lambda_A(\ell)-\nu_A(\ell)) &= \text{sign}(\alpha_A s_T(\ell)[1-(\alpha_N(1-s_T(\ell))+\alpha_A s_T(\ell))] - (1-\alpha_A)s_T(\ell)[(1-s_T(\ell))\alpha_N + s_T(\ell)\alpha_A])\\ &= \text{sign}((\alpha_A-\alpha_N)(1-s_T(\ell))).\end{aligned}$$

A.3.1 Non-homothetic Preferences

In this subsection, we report our extension of our baseline specification using homothetic CES preferences in equation (4) in Section 5 of the paper to allow for non-homothetic CES preferences.

A.3.1.1 Equilibrium Conditions

In this section we re-derive the equilibrium conditions from Section A.2.2 assuming non-homothetic CES preferences, as used in Hanoch (1975) and Comin, Lashkari and Mestieri (2021). We begin by introducing the changes relative to the benchmark analysis. First, instead of equation (4) in the paper, the common component of utility $u(\ell)$ is implicitly defined according to a non-homothetic CES aggregator:

$$\left[\beta_T^{\frac{1}{\sigma}} u(\ell)^{\frac{\zeta_T-\sigma}{\sigma}} C_T(\ell)^{\frac{\sigma-1}{\sigma}} + (1-\beta_T)^{\frac{1}{\sigma}} u(\ell)^{\frac{\zeta_N-\sigma}{\sigma}} C_N(\ell)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} = 1, \tag{A.47}$$

where our baseline specification in equation (4) in Section 5 of the paper corresponds to the special case with $\zeta_T = \zeta_N = 1$. Second, we assume that each land plot is owned by a different landowner. Therefore, there are $L(\ell)$ landowners in location ℓ.

The steps used in Section A.2.2 of this appendix to derive equation (26) in the paper and equation (A.17) above go through unchanged in this case. However, now the expenditure share in tradeables $s_T(\ell)$ is an income-weighted average of the tradeable shares of workers, $s_T^W(\ell)$, and landowners, $s_T^L(\ell)$:

$$s_T(\ell) = \frac{\omega(\ell)\, n(\ell)\, s_T^W(\ell) + s_T^L(\ell)}{\omega(\ell)\, n(\ell) + 1}, \tag{A.48}$$

where

$$s_T^W(\ell) = \beta_T \left(\frac{E_T(\ell)}{E(\ell)}\right)^{1-\sigma} u(\ell)^{\zeta_T-1}, \tag{A.49}$$

$$s_T^L(\ell) = \beta_T \left(\frac{E_T(\ell)}{E^L(\ell)}\right)^{1-\sigma} u^L(\ell)^{\zeta_T-1}. \tag{A.50}$$

In these expressions,

$$u(\ell) = \frac{w(\ell)}{E(\ell)}, \qquad u^L(\ell) = \frac{r(\ell)}{E^L(\ell)} \tag{A.51}$$

are the real income of workers and landowners, and $E(\ell)$ and $E^L(\ell)$ are their respective price indexes:

$$E(\ell)^{1-\sigma} \equiv \beta_T u(\ell)^{\zeta_T-1} E_T(\ell)^{1-\sigma} + \beta_N u(\ell)^{\zeta_N-1} P_N(\ell)^{1-\sigma}, \tag{A.52}$$

$$E^L(\ell)^{1-\sigma} \equiv \beta_T u^L(\ell)^{\zeta_T-1} E_T(\ell)^{1-\sigma} + \beta_N u^L(\ell)^{\zeta_N-1} P_N(\ell)^{1-\sigma}. \tag{A.53}$$

We now solve the model as follows. First, combining equations (A.49) to (A.51) with equations (A.15) to (A.17), the tradeable expenditure shares are:

$$s_T^W(\ell) = \beta_T \left(\frac{\omega(\ell)^{-\alpha_A}}{\widetilde{z}_A(\ell)}\right)^{1-\sigma} u(\ell)^{\zeta_T-\sigma}, \tag{A.54}$$

$$s_T^L(\ell) = \beta_T \left(\frac{\omega(\ell)^{-\alpha_A}}{\widetilde{z}_A(\ell)}\right)^{1-\sigma} u^L(\ell)^{\zeta_T-\sigma}. \tag{A.55}$$

Second, using equation (A.51) and (A.52) along with equations (A.15) to (A.17), the previous condition (A.21) that we used to determine the wage-rental ratio $\omega(\ell)$ as function of worker real income $u(\ell)$ now becomes:

$$\beta_T u(\ell)^{\zeta_T-\sigma} \left(\widetilde{z}_A(\ell)\, \omega(\ell)^{\alpha_A}\right)^{\sigma-1} + (1-\beta_T)\, u(\ell)^{\zeta_N-\sigma} \left(z_N(\ell)\, \omega(\ell)^{\alpha_N}\right)^{\sigma-1} = 1. \tag{A.56}$$

Finally, using equations (A.51) and (A.53) along with equations (A.15) to (A.17), we have a condition similar to equation (A.56) to determine the utility of landowners $u^L(\ell)$ as function of the wage-rental ratio $\omega(\ell)$:

$$\beta_T u^L(\ell)^{\zeta_T-\sigma}\left(\widetilde{z}_A(\ell)\,\omega(\ell)^{\alpha_A-1}\right)^{\sigma-1} + (1-\beta_T)\,u^L(\ell)^{\zeta_N-\sigma}\left(z_N(\ell)\,\omega(\ell)^{\alpha_N-1}\right)^{\sigma-1} = 1. \tag{A.57}$$

In sum, conditions (A.13) to (A.17) along with equations (A.48) and (A.54) to (A.57) give a solution for $\{n(\ell), \nu_A(\ell), w(\ell), r(\ell), P_N(\ell), s_T(\ell), s_T^W(\ell), s_T^L(\ell), \omega(\ell)\}$ given $u(\ell)$. And, as in the homothetic case, equation (A.1) gives a second relationship between $n(\ell)$ and $u(\ell)$ that we use to solve the local equilibrium given u^*.

A.3.1.2 Recovering Fundamentals

We now show how to obtain the fundamentals $\widetilde{z}_A(\ell)$ and $z_N(\ell)$ when preferences are non-homothetic. Using the definition of the tradeable expenditure share for workers in equation (A.54) and the prices from equations (A.15) to (A.17), we obtain:

$$\widetilde{z}_A(\ell) = \frac{u(\ell)^{\frac{\zeta_T-\sigma}{1-\sigma}}}{\omega(\ell)^{\alpha_A}}\left(\frac{\beta_T}{s_T^W(\ell)}\right)^{\frac{1}{1-\sigma}},$$

$$\widetilde{z}_A(\ell) = \frac{u(\ell)^{\frac{\zeta_N-\sigma}{1-\sigma}}}{\omega(\ell)^{\alpha_N}}\left(\frac{1-\beta_T}{1-s_T^W(\ell)}\right)^{\frac{1}{1-\sigma}}.$$

Therefore, to recover the fundamentals, we need information on $u(\ell)$, $\omega(\ell)$ and $s_T^W(\ell)$.

First, similarly to the homothetic case, we use the labor supply condition from equation (A.1) to obtain the distribution of utilities $u(\ell)$ up to the national real income u^*,

$$u(\ell) = u^*\left(\frac{n(\ell)\,L(\ell)}{N}\right)^{1/\varepsilon}. \tag{A.58}$$

Second, as in the homothetic case, the equation (A.13) that recovers the wage-rental ratio as function of the agricultural employment share and population density continues to hold:

$$\omega(\ell) = \frac{(1-\alpha_A)(1-\alpha_N)}{\alpha_N(1-\alpha_A)+(\alpha_A-\alpha_N)\,\nu_A(\ell)}\frac{1}{n(\ell)} \tag{A.59}$$

Finally, it only remains to obtain the tradeable expenditure share of workers, $s_T^W(\ell)$. From equation (A.48), we have that $s_T^W(\ell)$ is a function of the aggregate tradeable share in the location, $s_T(\ell)$, and that of landowners $s_T^L(\ell)$:

$$s_T^W(\ell) = s_T(\ell) + \frac{1}{\omega(\ell)\,n(\ell)}\left(s_T(\ell) - s_T^L(\ell)\right). \tag{A.60}$$

Using equation (A.14), we recover the aggregate trade share in the location from the observed agricultural share:

$$s_T(\ell) = \frac{(1-\alpha_N)\,\nu_A(\ell)}{1-\alpha_A+(\alpha_A-\alpha_N)\,\nu_A(\ell)}. \tag{A.61}$$

In addition, combining equations (A.54) with (A.55), we obtain an implicit relationship between the expenditure shares of landowners and workers:

$$\frac{\left(1-s_T^L(\ell)\right)^{\frac{\zeta_T-\sigma}{\zeta_N-\sigma}}}{s_T^L(\ell)} = \frac{\left(1-s_T^W(\ell)\right)^{\frac{\zeta_T-\sigma}{\zeta_N-\sigma}}}{s_T^W(\ell)}\,\omega(\ell)^{(\sigma-1)\left(1-\frac{\zeta_T-\sigma}{\zeta_N-\sigma}\right)}. \tag{A.62}$$

Equations (A.60) and (A.62) define a system for $\left\{s_T^L(\ell), s_T^W(\ell)\right\}$ given $\omega(\ell)$ and $s_T(\ell)$, which are in turn inferred through equations (A.59) and (A.61). Given $\zeta_T > \sigma$ and $\zeta_N > \sigma$, the first equation defines a negative relationship and the second one a positive relationship between both variables, implying a unique solution to the system.

A.3.2 Extensive Margin Specialization

If transport costs differ across goods and increase with distance to the trade hub, the model implies that more remote regions export a narrower range of products than more centrally-located regions, because transport costs are a source of comparative advantage.

To see this, consider a set of locations with identical technologies, $T_g(\ell) = T_g$ and suppose that ℓ indexes distance to the trade hub. Then, as distance ℓ increases, the share of land allocated to good g defined in equation (22) in the paper varies as follows:

$$\frac{l'_g(\ell)}{l_g(\ell)} = \theta \left(\frac{P'_g(\ell)}{P_g(\ell)} - \sum_{h=1}^{G} \frac{P'_h(\ell)}{P_h(\ell)} l_h(\ell) \right).$$

Assume iceberg trade costs τ_g for each good, order goods such that $\tau_1 < \tau_2 < .. < \tau_G$, and suppose that all goods are shipped internally. If the prices are determined by no-arbitrage with the trade hub then $P_g(\ell) = P_g^* e^{-\tau_g \ell}$. This implies that $\frac{l'_g(\ell)}{l_g(\ell)} > 0 \longleftrightarrow \tau_g < \overline{\tau_g}(\ell)$, where $\overline{\tau_g} = \sum_h \tau_h l_h(\ell)$, i.e., as we move inland, comparative advantage strengthens for goods with lower transportation costs and weakens for goods with higher transportation costs.

If all prices were determined by no-arbitrage with the port, we would have that $\overline{\tau_g}'(\ell) < 0$ and $\overline{\tau_g} \to \tau_1$. Hence, for each good $g > 1$ there would be a threshold $\overline{\ell}_g$ above which its land share is decreasing, $\frac{l'_g(\ell)}{l_g(\ell)} < 0 \longleftrightarrow \ell > \overline{\ell}_g$, where $\overline{\ell}_G < \overline{\ell}_{G-1} < .. < \overline{\ell}_2$. Each good $g > 2$ would not be exported for all $\ell > \overline{\ell}_g^* > \overline{\ell}_g$, where $l'_g\left(\overline{\ell}_g^*\right) = \gamma_g$, and there would be some $\overline{\ell}_1$ such that only $g = 1$ is exported for all $\ell > \overline{\ell}_1$. Hence, under heterogeneous iceberg costs, the number of exported goods decreases toward the interior and, as ℓ increases, goods with higher τ_g are dropped.

A.3.3 Endogenous Land Conversion

In our baseline specification of the model in the paper, all land is used productively. Therefore in our empirical analysis we use geographical land area as our measure of land area. Here, we develop 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, the amount of land used in each location is also endogenous.

In our baseline specification of the model in the paper, the zero populations observed for some locations in the data can be rationalized by zero productivities in tradeables: $z_A(\ell) = z_M(\ell) = 0$. In this extension of the model, a location also may have zero population because it is not profitable to convert land to productive use.

Now, we use $\overline{L}(\ell)$ to denote the total land area of each location, and $L(\ell)$ to denote the land area that is used productively. The only extension to the baseline model is the following: each land plot $j \in \overline{L}(\ell)$ requires a fixed cost of f_j units of labor to be opened and maintained for productive use. Once opened, the specialization across sectors and goods within agriculture is the same as in the baseline model. The cost f_j is independently and identically distributed across land plots and districts. We let $G_f(x)$ be the share of land plots in each district whose cost f is less than x.

Conditional on a land plot being open for productive use, we can solve equations (12) and (13) in the paper as before. From the solution to this problem, we obtain the expected (gross) land rents from using any land plot j for production in sector i,

$$r_i(\ell) = \bar{f}_i(\ell)\, w(\ell)\,,$$

where $\bar{f}_i(\ell)$ is defined identically to the inverse of $\omega_i(\ell)$ in equation (15) in the paper. If a location is specialized in agriculture, $\bar{f}(\ell)$ equals the inverse of $\omega(\ell)$ from equation (20) in the paper. Net rents to the landowner of plot j are then

$$r_j(\ell) = \left(\bar{f}(\ell) - f_j\right) w(\ell),$$

where land is converted to productive use if these net rents are positive. The total amount of land in productive use can be expressed as:

$$L(\ell) = G_f\left(\bar{f}(\ell)\right) \overline{L}(\ell).$$

The determination of $\bar{f}(\ell)$ is independent from the amount of land that is used, and therefore from the support of the fixed-cost distribution $G_f(x)$. Therefore, as long as the support of this distribution is bounded from below by some positive number, we will have $L(\ell) = 0$ for sufficiently low $\omega(\ell)$.

While this extension does not affect gross returns to land and labor (i.e., **Step 1** in the proof of Proposition 1), it does affect the model prediction for population density and for the agricultural labor and land share. Following steps similar to **Step 2** in the proof of Proposition 1 in Section A.2.4 of this online appendix, we reach the following solutions that generalize equations (26) and (27) in the paper:

$$n(\ell) = \frac{N(\ell)}{L(\ell)} = \left(\frac{1}{\alpha_N + (\alpha_A - \alpha_N)\beta_T\left(\frac{E_T(\ell)}{E(\ell)}\right)^{1-\sigma}} - 1 \right) \overline{f}(\ell) + \mathbb{E}\left[f \mid f < \overline{f}(\ell)\right], \tag{A.63}$$

and

$$\nu_A(\ell) = \frac{N_A(\ell)}{N(\ell)} = \frac{\beta_T\left(\frac{E_T(\ell)}{E(\ell)}\right)^{1-\sigma}\left(\alpha_A \mathbb{E}\left[f \mid f < \overline{f}(\ell)\right] + (1-\alpha_A)\overline{f}(\ell)\right)}{\left(\alpha_N + \beta_T\left(\frac{E_T(\ell)}{E(\ell)}\right)^{1-\sigma}(\alpha_A - \alpha_N)\right)\left(\mathbb{E}\left[f \mid f < \overline{f}(\ell)\right] - \overline{f}(\ell)\right) + \overline{f}(\ell)}. \tag{A.64}$$

It can be verified that these expressions nest equations (26) and (27) in the paper for the case without fixed costs, in which $\mathbb{E}\left[f \mid f < \overline{f}(\ell)\right] = 0$.

A.4 Additional Empirical Results

In this section of the online appendix, we report additional empirical results that are discussed in the main paper. In Subsection A.4.1, we present further evidence on the evolution of transatlantic freight rates over time, as discussed in Section 2 of the paper. In Subsection A.4.2, we report additional reduced-form empirical results, as discussed in Section 4 of the paper.

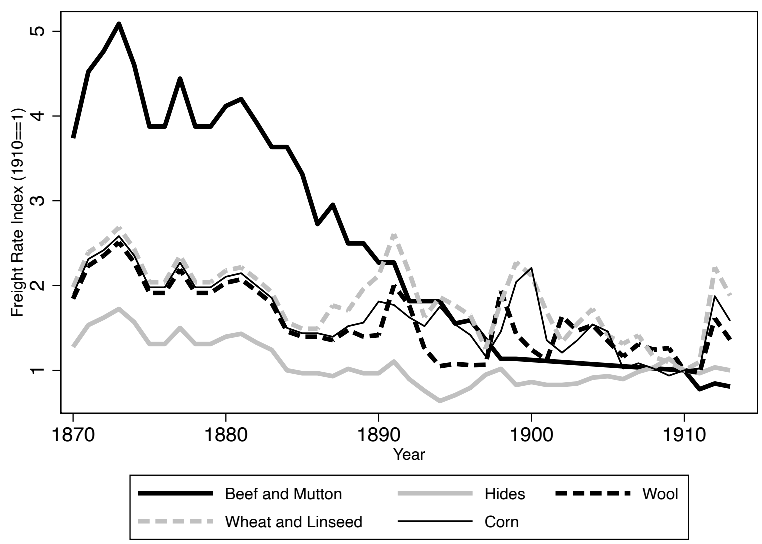
In Subsection A.4.3, we provide further results for the estimation of the production cost share parameters (α_A, α_N) from Section 6.1 of the paper (Step 1). In Subsection A.4.4, we present additional results for the estimation of the demand parameters (σ, β_T) from Section 6.2 of the paper (Step 2). In Subsection A.4.5, we report further results for our land shares estimation from Section 6.6 of the paper (Step 6), including our overidentification checks using district-level railroad shipments and machinery use.

A.4.1 International Transport Costs and Export Specialization

In this section of the online appendix, we provide further evidence on reductions in transatlantic freight rates and changes in export specialization from 1869-1914. First, we show that transatlantic freight rates fell substantially for all goods with improvements in transportation technology, including improvements in the size, speed and reliability of steam ships. Second, we show that these transatlantic freight rates fell unevenly across goods, as new technologies made transatlantic trade in some goods profitable that were previously prohibitively costly to transport (e.g. the mechanical refrigeration of meat made possible trade in chilled and frozen meat). Third, we illustrate the large-scale changes in patterns of export specialization over our sample period, with a decline in the export share of traditional goods, such as cattle and sheep hides, and the emergence of entirely new export commodities, such as cereals and refrigerated and frozen beef and mutton.

In Figure A.1, we display estimates of transatlantic freight rates from Buenos Aires to Western Europe for five different categories of disaggregated agricultural goods from Tena-Junguito and Willebald (2013). These estimates are based on the ratio of "cost inclusive of freight" (cif) to "free on board" (fob) prices. Each freight rate is shown an index that is equal to one in 1910. As apparent from the figure, we find substantial declines in these transatlantic freight rates over time, which are much larger for goods such that beef and mutton that disproportionately benefited from new transport technologies, such as the mechanical refrigeration of meat. As discussed in Section 2 of the paper, this pattern of results is consistent with a large existing literature in economic history documenting the decline in transatlantic transportation costs during the late-19th century, including in particular North (1958), Harley (1980, 1988), and O'Rourke and Williamson (1999).

Figure A.1: Transatlantic Freight Rates for Agricultural Goods



Note: Index of transatlantic freight rates for agricultural goods from Buenos Aires to Western Europe (1910=1); transatlantic freight rates estimated by Tena-Junguito and Willebald (2013) using data on "free on board" (fob) and "cost inclusive of freight" (cif) prices.

As further evidence on this decline in transatlantic transportation costs, Figure A.2 shows the freight rate for coal from Wales (in the United Kingdom) to the River Plate, as reported in Angier (1920). Although freight rates are not reported for the intermediate years from 1875-1896 inclusive, we again find a substantial decline in transportation costs between the beginning and end of our sample period. Finally, as a further cross check, Figure A.3 presents the freight rate for wheat from South America to London, as reported in North (1958). Again, we find a substantial decline in transportation costs in the closing decades of the 19th-century, before these begin to increase after 1910 in the years immediately preceding the First World War. These empirical findings are consistent with a large-scale decline in transatlantic transportation costs in the second half of the 19th-century with improvements in transportation technology, including in particular the invention of steam ships, and improvements in the size, speed and reliability of these steam ships.

Figure A.2: Transatlantic Freight Rate for Coal from Wales to the River Plate

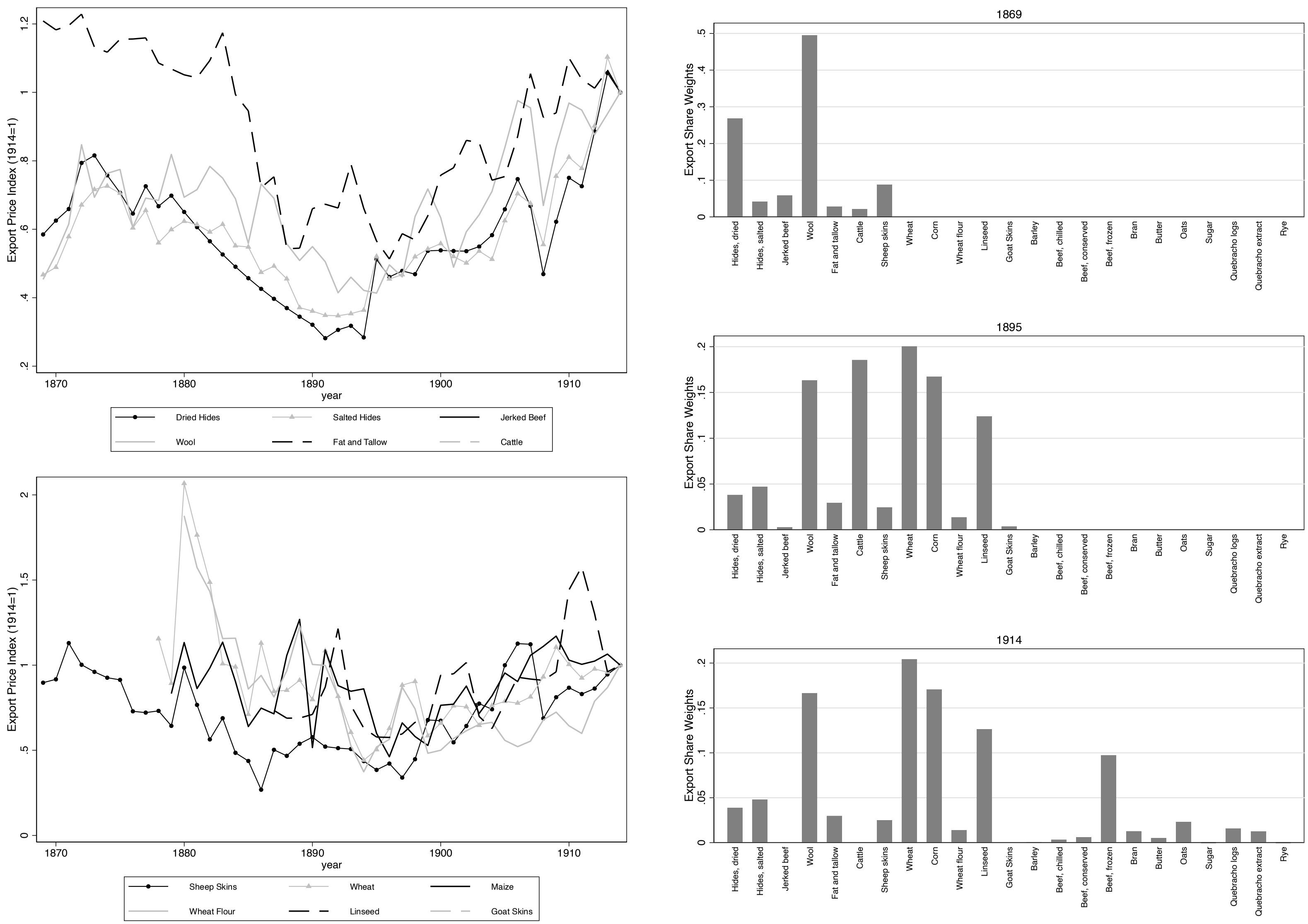
Note: Coal freight rate from Wales in the United Kingdom to the River Plate in Argentina, as reported in Angier (1920); data not reported from 1875 through 1896.

Figure A.3: Transatlantic Freight Rate for Wheat from South America to London

Note: Freight rate from South America to London, as reported in North (1958).

In Figure A.4, we illustrate the large-scale changes in export specialization that occurred over our sample period. In the left panel, we show export price indexes for the individual goods that enter the aggregate export price index in Figure 1 in the paper. In the right panel, we display the weights of these individual goods in this aggregate export price index. As apparent from Figure A.4, entirely new commodities began to be exported over time, including in particular cereals and refrigerated and frozen beef and mutton. As a result, between 1869 and 1914, the export share of cattle and sheep hides falls from around 40 percent to less than 15 percent. 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.

Figure A.4: Export Price Indices and Value Weights for Individual Export Goods from 1869-1914



Note: Export price indexes (left panels) and export value weights (right panels) for the individual goods included in the aggregate export price index shown in Figure 1 in the paper; left panels display export prices for the 12 goods with the largest export value shares over our sample period; right panels show export value shares for all of the 23 goods on which we have data from 1869-1914; Source: Francis (2017).

A.4.2 Additional Reduced-Form Empirical Results

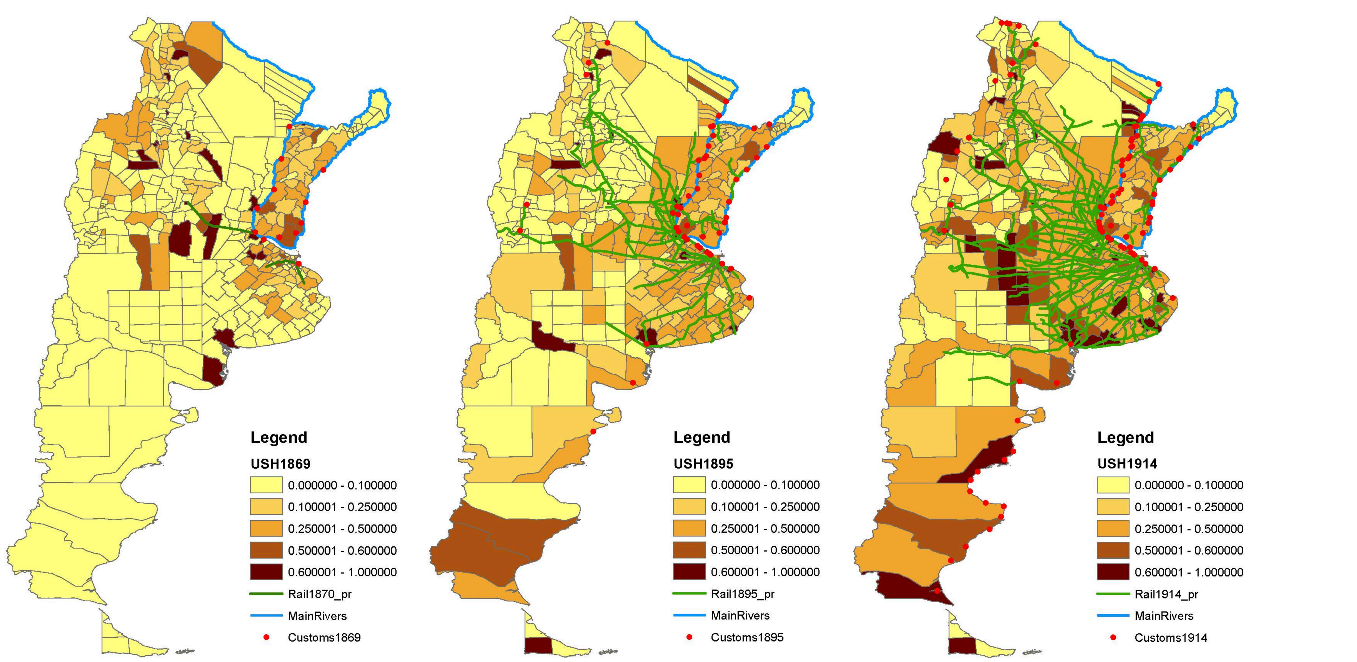
In this section of the online appendix, we report additional empirical results for the reduced-form evidence in Section 4 of the paper. In Section A.4.2.1, we present further evidence on the large-scale changes in the spatial distribution of economic activity within Argentina that occurred from 1869-1914, as considered in Section 4.1 of the paper. In Subsection A.4.2.2, we present additional evidence on the gradient in economic activity with distance from world markets, as considered in Subsection 4.2 of the paper. In Subsection A.4.2.3, we report further evidence on the impact of the railroad network on the spatial distribution of economic activity within Argentina, as discussed in Subsection 4.3 of the paper.

A.4.2.1 Spatial Patterns of Economic Development

In Figure 2 in the paper, we display the evolution of population density across Argentinian districts in each of our census years. In Figure A.5, we show that we find a similar pattern of results for urbanization, as measured by the share of the population living in towns and cities. In 1869 (shown in panel (a)), high urban population shares were concentrated around the Spanish colonial towns towards the North-West and along the main navigable rivers. Between each of the periods of 1869-95 and 1895-1914 (comparing panels (a) and (b)) and panels (b) and (c)), there is a general increase in the urban population share, which again radiates further inland from Buenos Aires and its neighboring ports. 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). Additionally, with the expansion of economic activity into more peripheral locations, some remote areas with low population densities become dominated by few cities or towns, as reflected in high urban population shares.

Figure A.5: Urban Population Share from 1869-1914

(a) 1869 (b) 1895 (c) 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).

A.4.2.2 Geographical Access to World Markets

In Subsection 4.2 of the paper, we provide regression evidence on the change in the organization of economic activity within Argentina with respect to geographical distance from world markets. We use the following reduced-form regression specification:

$$Y_t(\ell) = a_t + b_t \ln \left(\text{Distance}(\ell)\right) + u_t(\ell), \tag{A.65}$$

where ℓ indexes districts and t corresponds to time; $Y_t(\ell)$ is an economic outcome (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 Table 1 of the paper, we report the results of estimating our baseline specification, in which we measure geographical access to world markets using distance from the centroid of each district to Argentina's trade hub, as captured by the top-four ports of Buenos Aires, Rosario and Bahía Blanca, which together account for more than 75 percent of export value throughout our sample period. We find the same pattern results across a wide range of different specifications, where in the interests of brevity we focus here on the estimates in Columns (1) and (2) of Table 1 in the paper for log population density and the urban population share. For example, we obtain similar results excluding the 12 districts with centroids within 40 kilometers of the center of Buenos Aires to abstract from factors idiosyncratic to the federal capital (coefficients (standard errors) of of -0.195 (0.078) and -0.037 (0.015) respectively); using distance to the nearest port on the coast or the Paraná, Plate and Uruguay navigable rivers (coefficients (standard errors) of -0.329 (0.073) and -0.048 (0.012) respectively); or using distance to Buenos Aires itself (coefficients (standard errors) of -0.483 (0.085) and -0.042 (0.014) percent respectively).

In Table A.1 below, we present the results of estimating this baseline specification using the subsample of districts used for the quantitative analysis of the model, in which agriculture accounts for more than 5 percent of employment and less than 95 percent of employment. As apparent from the table, we find the same qualitative and quantitative pattern of results as reported for the full sample in the paper. We find that locations further from world markets have lower population density, a smaller share of the population living in cities and towns, higher wage-rental ratios, higher relative prices of traded goods, and greater specialization in the most-transport-sensitive agricultural goods. Following Argentina's growing external integration in the late-19th century, we find a steepening of the gradients in population density and the urban population share in geographical distance from world markets. Therefore, whether we consider the full sample of districts or the subsample used for the quantitative analysis of our model, we find a pattern of results consistent with the spatial Balassa-Samuelson effect in our theoretical model.

Table A.1: Structural Transformation and Geographical Access to World Markets (Quantitative Model Sample)

	(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	-0.375***	-0.033**	-0.390***	-0.048***	0.939***	0.139***	-0.059**
	(0.093)	(0.015)	(0.055)	(0.015)	(0.150)	(0.046)	(0.024)
Model Sample	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	152	152	152	152	37	41	93
R-squared	0.113	0.033	0.314	0.076	0.590	0.260	0.083

Notes: Observations are a cross-section of Argentinian districts for a given year and correspond to the sample used for our quantitative analysis of the model, in which agriculture accounts for more than 5 and less than 95 percent of employment in each district. 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 this 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 this 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 this 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). Land Area is total geographical land area of each district. 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 for the full sample of districts in Section 4.2 of the paper, we find the same pattern results across a wide range of different specifications, where in the interests of brevity we focus here on the estimates in Columns (1) and (2) for log population density and the urban population share. For example, we obtain similar results excluding the 12 districts with centroids within 40 kilometers of the center of Buenos Aires to abstract from factors idiosyncratic to the federal capital (coefficients (standard errors) of of -0.223 (0.091) and -0.034 (0.017) respectively); using distance to the nearest port on the coast or the Paraná, Plate or Uruguay navigable rivers (coefficients (standard errors) of -0.232 (0.077) and -0.024 (0.012) respectively); or using distance to Buenos Aires itself (coefficients (standard errors) of -0.436 (0.088) and -0.022 (0.013) percent respectively).

A.4.2.3 Railroad Access

In Subsection 4.3 of the paper, we provide regression evidence on the impact of the railroad network on the spatial distribution of economic activity within Argentina. In this subsection of the online appendix, we first provide further information about the construction of our instrumental variables. We next report additional instrumental variables estimation results for the impact of the construction of the railroad network on the spatial distribution of economic activity, as discussed in the paper.

Construction of the Instruments 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 start by discretizing the surface of Argentina into a fine mesh of grid points. For each district, we construct least-cost paths across this grid from its centroid to each of the top-four ports, assuming an equal cost of travel across grid points. We thus obtain four lines on a map of Argentina from the centroid of that district to each of the top-four ports. Repeating this process for all districts, we obtain a web of these least-cost paths from the centroids of every district to each of the top-four ports. Finally, we construct our instrument for a railroad connection for each district as the fraction of the district's surface area (the fraction of grid points within its geographical boundaries) that is covered by these least-cost paths.[3]

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. To address the concern that a district's size could affect the fraction of its surface area that lies on least-cost paths to top-four ports, 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 other 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. Therefore, our estimates exploit variation in the frequency with which a district lies along a least-cost path to Argentina's trade hub, conditional on geographical location. 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 first instrument assumes that there is no direct effect on economic activity of frequently lying along a least-cost path to Argentina's trade hub, other than through the probability with which a district is connected to the railroad.

Our second instrument uses historical exploration and trade routes following Duranton and Turner (2012) and Duranton et al. (2014). We use the fact that economic activity in the Spanish colonial period was orientated in a very different way from 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 the construction of railroads. To address the concern that districts along colonial postal routes could in differ in historical levels of economic activity, geographical location within Argentina or land area, we again control separately for the initial level of economic activity in 1869, latitude and longitude, and land area. After conditioning on these controls, our second instrument assumes that there is no direct effect of lying along Spanish colonial postal routes on subsequent late-19th-century economic growth, other through the probability with which a district is connected to the railroad.

Importantly, our two instruments use quite different sources of variation. Our first instrument is based on connecting the interior to the late-19th century trade hub centered on the Buenos Aires coastal region. In contrast, our second instrument uses postal routes between the Spanish colonial cities that were orientated around trade routes through the North-Western interior regions towards Panama. Therefore, we can use these two different sources of variation to provide a check on our identifying assumptions, by reporting Hansen-Sargan overidentification tests and the results of specifications using only one of the two instruments. If we find a similar pattern of results using each of

[3] As a robustness check, we constructed a version of this instrument based on constructing least-cost paths from all districts to Buenos Aires alone, and find a similar pattern of results, because the top-four ports are clustered around Buenos Aires, as discussed above.

the two instruments separately, this implies either that both instruments are valid, or that both are invalid and there exists an implausible correlation structure, such that the error term has a similar correlation with these two quite different sources of variation.

Additional Instrumental Variables Estimation Results We consider an instrumental variables specification with the following second-stage regression for long-differenced population growth from 1869-1914:

$$\Delta \ln Y_{1914-1869}(\ell) = a + c\left(\text{rail}_{1914}(\ell)\right) + d_1 \ln\left(\text{area}(\ell)\right) + d_2 \text{lat}\left(\ell\right) + d_3 \text{long}\left(\ell\right) + d_4 \ln Y_{1869}(\ell) + u(\ell), \quad \text{(A.66)}$$

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; recall that the railroad network was negligible in 1869, and hence this railroad connection measure captures the expansion of the railroad network from 1869-1914; $\ln\left(\text{area}(\ell)\right)$ 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.

This second-stage regression specification (A.66) 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 1869-14, such as common macro shocks. The corresponding first-stage regression is:

$$\begin{aligned} \text{rail}_{1914}(\ell) = & e + f_1\left(\text{port}(\ell)\right) + f_2\left(\text{colonialpost}(\ell)\right) + g_1 \ln\left(\text{area}(\ell)\right) + g_2 \text{lat}\left(\ell\right) \\ & + g_3 \text{long}\left(\ell\right) + g_4 \ln Y_{1869}(\ell) + h(\ell), \end{aligned} \quad \text{(A.67)}$$

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

In Table 2 of the paper, we report the results of estimating this specification for the full sample of districts. In Table A.2 below, we present the results of estimating the same specification using the subsample of districts used for the quantitative analysis of the model, in which agriculture accounts for more than 5 percent of employment and less than 95 percent of employment. We find the same qualitative and quantitative pattern of results as reported for the full sample in the paper. In both Columns (1) and (2), we find that a positive and statistically significant relationship between population growth and a railroad connection. Whether we use both instruments or each instrument individually in Columns (3)-(5), we obtain a similar IV estimate that is marginally larger than our OLS estimate in Column (2). Despite the reduction in sample size, we find that both instruments continue to have power in the first-stage regression, with first-stage F-statistics above the conventional threshold of 10. In the specification with both instruments, we are unable to reject the null hypothesis of the model's overidentifying restrictions. 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. Taken together, these results provide further evidence in support of a causal impact of the railroad network on the spatial distribution of economic activity within Argentina.

Table A.2: Population Growth and Railroad Access (Quantitative Model Sample)

	(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 by 1914	0.534***	–	–	–	–
	(0.106)				
Rail length 1914	–	0.386***	0.525***	0.518**	0.525***
		(0.054)	(0.115)	(0.210)	(0.119)
Latitude	-0.086***	-0.064***	-0.050***	-0.051*	-0.050***
	(0.016)	(0.014)	(0.017)	(0.027)	(0.017)
Longitude	0.016	0.034**	0.037**	0.037**	0.037**
	(0.015)	(0.014)	(0.015)	(0.016)	(0.015)
Log Land Area	-0.019	-0.143***	-0.182***	-0.180***	-0.182***
	(0.043)	(0.040)	(0.050)	(0.069)	(0.050)
Log Population 1869	-0.332***	-0.268***	-0.269***	-0.269***	-0.269***
	(0.102)	(0.084)	(0.080)	(0.080)	(0.080)
Estimation	OLS	OLS	IV	IV	IV
Instruments	–	–	Both	Port	Colonial Post
First-stage F-statistic	–	–	37.61	21.72	59.38
Overidentification test (p-value)	–	–	0.974	–	–
Observations	152	152	152	152	152
R-squared	0.439	0.517	–	–	–

Notes: In Columns (1)--(5), observations are a cross-section of Argentinian districts for a single difference from 1869-1914, and correspond to the sample used for our quantitative analysis of the model, in which agriculture accounts for more than 5 and less than 95 percent of employment in each district. Log population is the log of the total population of each district. Rail Connection by 1914 is an indicator that is one if a district has a railroad connection (one or more railroad stations) by 1914 and zero otherwise. Rail length 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 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.

A.4.3 Production Parameters (Step 1)

As discussed in Section 6.1 of the paper, 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) in the paper, the model implies the following predictions for the wage-rental ratio, given observed population density ($n(\ell)$) and the agricultural employment share ($\nu_A(\ell)$):

$$\omega(\ell) = \frac{w(\ell)}{r(\ell)} = \frac{(1-\alpha_A)(1-\alpha_N)}{\alpha_N(1-\alpha_A) + (\alpha_A - \alpha_N)\nu_A(\ell)} \frac{1}{n(\ell)}. \tag{A.68}$$

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 in the paper. 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.[4]

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

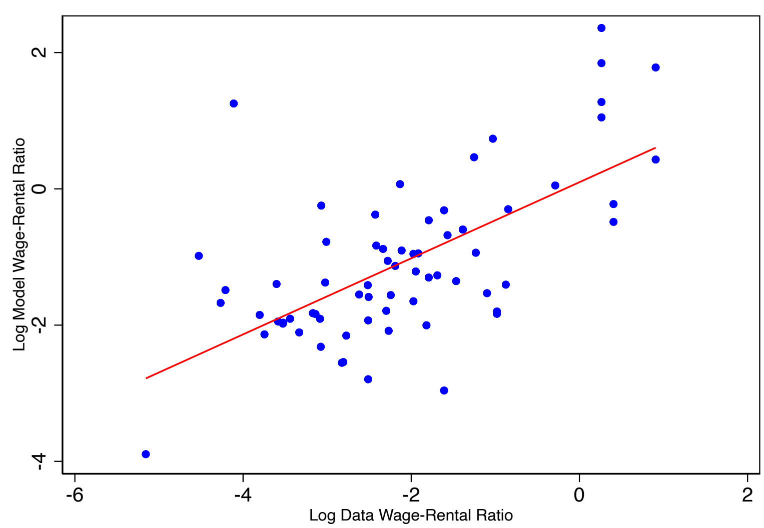
[4]The only year for which comprehensive data on land values are reported is 1895. In contrast, only the distribution of agricultural establishments across a number of discrete land value bins is reported in 1914.

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.[5]

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 overall district income of 0.45, which is close to the aggregate labor share of 0.48 in 1913 reported in Frankema (2010). In Figure A.6, we demonstrate a strong, positive and statistically significant correlation of 0.64 between the model's predictions and the observed wage-rental ratio. 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 as reported in Section 4.2 in the paper.

In Figure A.8, we display a map of the model's predictions for the wage-rental ratio ($\omega(\ell)$) across the full sample of Argentinian districts in 1914. Darker brown shading denotes higher values for the wage-rental ratio. Consistent with the predictions of spatial Balassa-Samuelson effect, we find low wage-rental ratios in districts close to the top-four ports surrounding Buenos Aires, and in districts with good railroad connections to these top-four ports. In contrast, as we move further inland towards more remote districts, with worse railroad connections to these top-four ports, we find progressively higher wage-rental ratios.

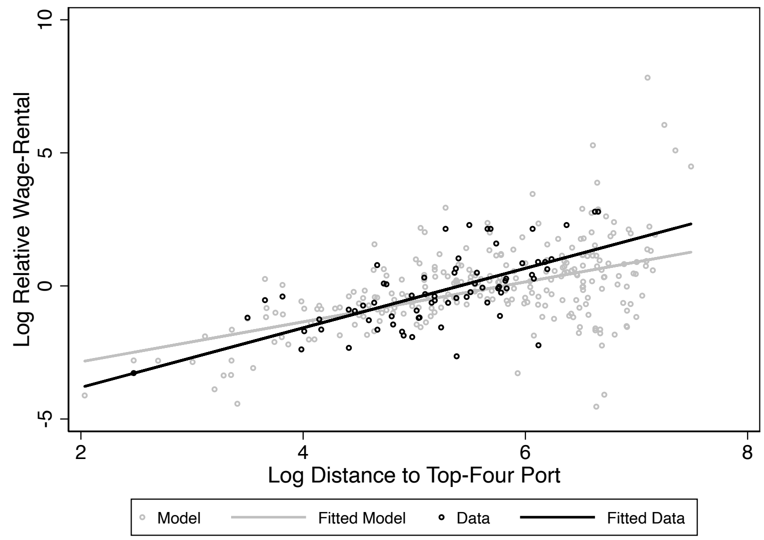
Figure A.6: Log Wage-Rental Ratio ($\omega(\ell)$) in the Model for the Estimated Production Cost Share Parameters (α_A, α_N) and in the Data



Note: Observations are a cross-section of districts in years around 1895 for which we observe both the wages of agricultural laborers and the value of land per hectare; vertical axis shows the model's predictions for the log wage-rental ratio based on observed population density and the agricultural employment share and our estimated production cost share parameters (α_A, α_N) from equation (A.68); horizontal axis shows the log wage-rental ratio in the data, measured as the ratio of the wages of agricultural laborers to land value per hectare, assuming a constant proportional relationship between land rents and land values; production cost share parameters (α_A, α_N) chosen to minimize the sum of squared deviations between the model predictions and observed data for the log wage-rental ratio.

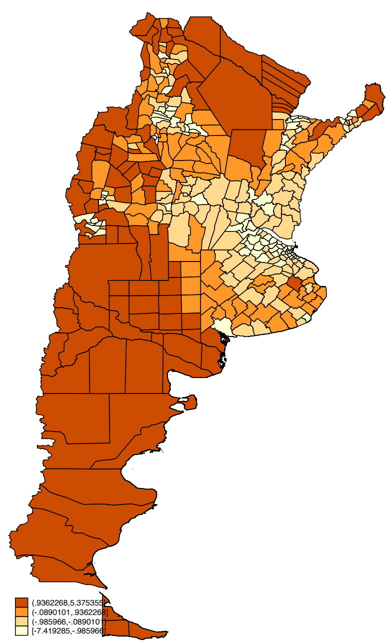
[5]This log specification allows for the proportional relationship between land rents and land values per hectare (captured by an additive constant in logs) and for proportional changes in the overall price level between the two different years of 1895 and 1913 for which the land value and wage data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

Figure A.7: Log Wage-Rental Ratio ($\omega(\ell)$) in the Model for the Estimated Parameters (α_A, α_N) and in the Data Against Distance from the Closest Top-Four Port (Spatial Balassa-Samuelson Effect)



Note: Gray dots show the model's predictions for the log wage-rental ratio for all districts in our model sample in 1895; black dots show the log wage-rental ratio in the data, measured as as the ratio of the wages of agricultural laborers to land value per hectare, assuming a constant proportional relationship between land rents and land values, for districts for which we observe both of these variables; vertical axis shows the log wage-rental ratio in the model and data; horizontal axis shows log geographical (Great Circle) distance to the nearest top-four port (Buenos Aires, Rosario, La Plata and Bahia Blanca); gray line shows the linear fit for the model's predictions; black line shows the linear fit for the observed data.

Figure A.8: Log Wage-Rental Ratio ($\omega(\ell)$) in the Model in 1914 for the Estimated Production Cost Share Parameters (α_A, α_N) (Spatial Balassa-Samuelson Effect)



Note: Map shows the log wage-rental ratio ($\omega(\ell)$) in the model for the estimated production cost share parameters (α_A, α_N) across the full sample of Argentinian districts in 1914; log wage-rental ratio normalized to have a mean of zero.

A.4.4 Demand Parameters (Step 2)

As discussed in Section 6.2 of the paper, 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 this 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) in the paper implies the following relationship between the agricultural employment share ($\nu_A(\ell)$) and the relative tradeables price index ($E_T(\ell)/E(\ell)$):

$$\ln s_T(\ell) = \ln\left[\frac{(1-\alpha_N)\,\nu_A(\ell)}{(1-\alpha_A)+(\alpha_A-\alpha_N)\,\nu_A(\ell)}\right] = \kappa_0 + \kappa_1 \ln\left(\frac{E_T(\ell)}{E(\ell)}\right) + \ln h_T(\ell), \tag{A.69}$$

where $\kappa_0 = \ln(\beta_T)$ and $\kappa_1 = (1-\sigma)$; 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 the model's prediction for the share of tradeables ($s_T(\ell)$) on the left-hand side and observed data on the relative tradeables price index ($E_T(\ell)/E(\ell)$) on the right-hand side. We measure both the tradeables price index ($E_T(\ell)$) and the overall price index ($E(\ell)$) by combining our 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, as discussed in Section 3 of the paper. We construct the two 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 in further detail in Section A.6.7 of this 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 Table 3 in the paper, we report the results of estimating equation (A.69) using ordinary least squares (OLS) and instrumental variables (IV). Consistent with the model's assumptions, we estimate an elasticity of substitution between traded and non-traded goods of less than one in both specifications. In our IV estimates, we find an elasticity of substitution of $\sigma = 0.49$, and a weight of tradeables in consumer expenditure of $\beta_T = 0.77$.

In Figure A.9, we illustrate the fit of the regression (40) by scattering the model's prediction for the share of tradeables ($s_T(\ell)$) against the data on the relative tradeables price index ($E_T(\ell)/E(\ell)$) for the districts for which these data are available. There are several reasons why the model's predictions and the observed data can differ, including the fact that our measure of the relative tradeables price index in the data is necessarily imperfect, with a relatively small number of traded goods for which we observe price data.[6] Nevertheless, we find a strong relationship between the model's predictions and the data.

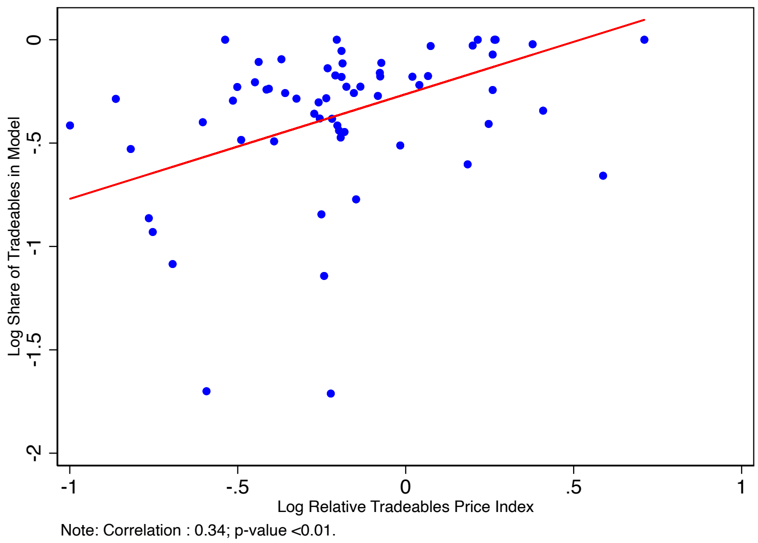
In Figure A.10, we use the estimated parameters (β_T, σ) from the regression (40) to compute the model's prediction for the relative tradeables price index ($[(1/\beta_T)\,s_T(\ell)]^{1/(1-\sigma)}$), and display this prediction against the data on the relative tradeables price index ($E_T(\ell)/E(\ell)$). We find 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 reported in Section 4.2 above.

In Figure A.11, we display a map of the model's predictions for the relative tradeables price index ($E_T(\ell)/E(\ell)$) across the full sample of Argentinian districts in 1914. Darker brown shading denotes higher values for the relative tradeables price index. Consistent with the predictions of spatial Balassa-Samuelson effect, we find low relative trade-

[6] We observe prices for the following expenditure shares categories and traded goods: Meat (beef, lamb, pork), Bread, Other Foods (cooking oil, Bremen rice, Tucumán sugar, Brazilian coffee, milk, Tobacco, Potatoes, Wine, Corn Flour, Wheat flour, Beans, Noodles, Crushed Corn, Check Peas, Corn, Cow fat, Herbs, and Salt); Other Household Expenses (brooms, soap, starch, kerosene and phosphorus).

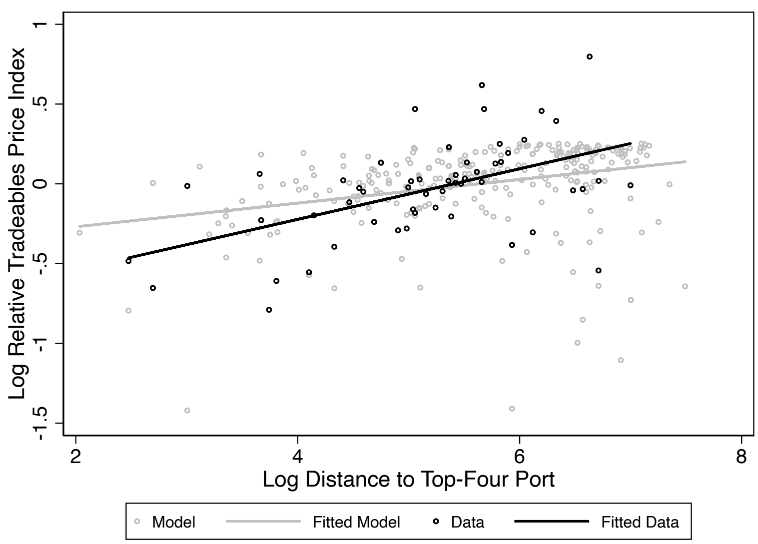
ables price indexes in districts close to the top-four ports surrounding Buenos Aires, and in districts with good railroad connections to these top-four ports. In contrast, as we move further inland towards more remote districts with worse railroad connections to these top-four ports, we find progressively higher relative tradeables price indexes.

Figure A.9: Model's Prediction for the Share of Tradeables ($s_T(\ell)$) in Expenditure for the Estimated Parameters (α_A, α_N, σ, β_T) and the Relative Tradeables Price Index ($E_T(\ell)/E(\ell)$) in the Data



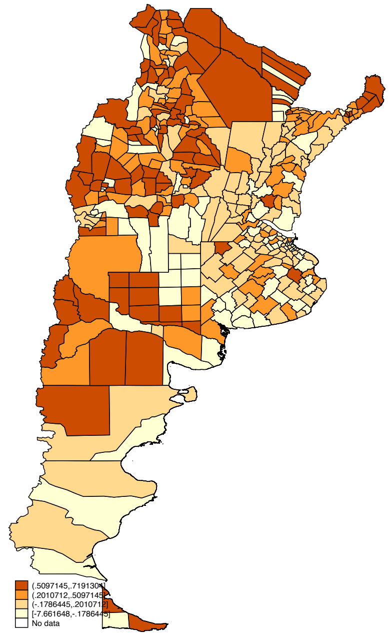
Note: Observations are a cross-section of 63 districts in years around 1895 for which we can compute both the tradeables price index ($E_T(\ell)$) and the overall price index ($E(\ell)$); vertical axis shows the model's prediction for the log share of tradeables in expenditure ($s_T(\ell)$) from the left-hand side of equation (A.69); horizontal axis shows the log relative tradeables price index in the data ($E_T(\ell)/E(\ell)$); correlation reports the correlation between the model's prediction and the fitted values from the regression specification (A.69).

Figure A.10: Relative Tradeables Price Index ($E_T(\ell)/E(\ell)$) in the Model for the Estimated Parameters (α_A, α_N, σ, β_T) and in the Data Against Distance from the Closest Top-Four Port (Spatial Balassa-Samuelson Effect)



Note: Gray dots show the model's predictions for the log relative tradeables price index ($E_T(\ell)/E(\ell)$) for all districts in our model sample in 1895; black dots show the log relative tradeables price index ($E_T(\ell)/E(\ell)$) in the data for the 63 districts for which data are available; the log relative tradeables price index ($E_T(\ell)/E(\ell)$) in the data is measured using our aggregate household expenditure shares, district prices of traded goods and district land value per hectare as a measure of housing costs; vertical axis shows the log relative tradeables price index ($E_T(\ell)/E(\ell)$); horizontal axis shows low geographical (Great Circle) distance to the nearest top-four port (Buenos Aires, Rosario, La Plata and Bahia Blanca); gray line shows the linear fit for the model's predictions; black line shows the linear fit for the observed data.

Figure A.11: Log Relative Tradeables Price Index ($E_T(\ell)/E(\ell)$) in the Model in 1914 for the Estimated Parameters (α_A, α_N, σ, β_T) (Spatial Balassa-Samuelson Effect)



Note: Map shows the log relative tradeables price index ($E_T(\ell)/E(\ell)$) in the model for the estimated parameters (α_A, α_N, σ, β_T) across the full sample of Argentinian districts in 1914; log relative tradeable price index normalized to have a mean of zero.

A.4.5 Land Shares Estimation (Step 6)

In this section of the online appendix, we report additional empirical results for the land shares estimation in Section 6.6 of the paper. In Subsection A.4.5.1, we provide further evidence on the evolution of aggregate agricultural land shares for each of the disaggregated goods over time.

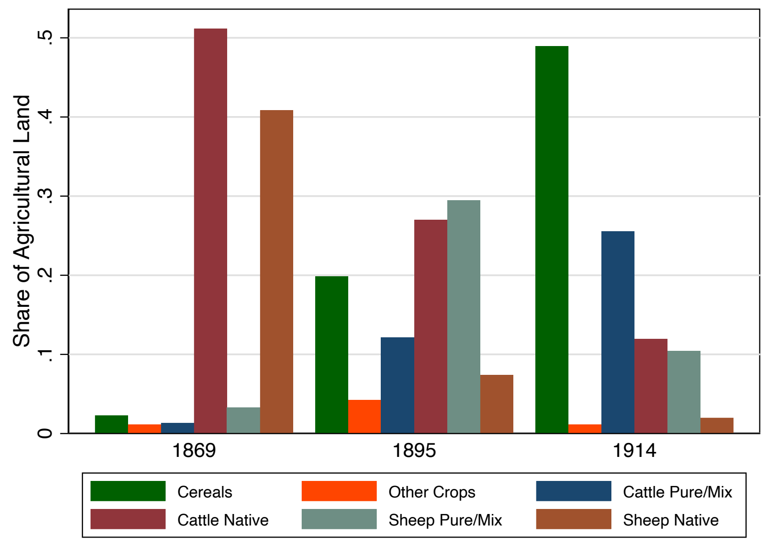
In Subsection A.4.5.2, we report a first overidentification check, in which we compare our estimated changes in relative technology-adjusted prices at Argentina's trade hub ($\widehat{\mu}_{g\chi}$) to changes in relative transatlantic freight rates. In Subsection A.4.5.3, we present a second overidentification check, in which we compare our model's predictions for district-level agricultural land shares for the disaggregated goods in 1895 with observed data on these district-level agricultural land shares in that year.

In Subsection A.4.5.4, we introduce a third overidentification check, in which we compare our model's predictions for the quantity of cereals produced in each district in 1895 and 1914 with the observed data on railroad shipments of cereals from each district in those years. In Subsection A.4.5.5, we include a fourth overidentification check, in which we compare our model's predictions for the revenue from cereals production in each district in 1914 with the observed data on the value of cereals machinery used in each district in that year.

A.4.5.1 Aggregate Agricultural Land Shares 1869, 1895 and 1914

In Figure A.12, we display the observed aggregate agricultural land shares for the disaggregated goods for Argentina as a whole over time. One of the most striking features of the figure is the transformation of economic activity within the agricultural sector over time, consistent with the evidence of the large-scale change in export composition in Section 2 of the paper. In 1869, Argentinian agriculture was was dominated by the extensive farming of native cattle and sheep on ranches ("estancias"), primarily for hides, skins, bones, fat, tallow and other animal products. Over the course of our sample period, native cattle are replaced by pure/mixed-breed cattle, which yielded higher quantity and quality of meat. Similarly, native sheep are displaced by pure/mixed-breed sheep, which produced higher quantity and quality of wool. Nevertheless, perhaps the most dramatic change is the expansion of cereals cultivation, which rises from a negligible share to around half of agricultural land area.

Figure A.12: Aggregate Agricultural Land Shares in Argentina in 1869, 1895 and 1914



Note: Aggregate shares of agricultural land in Argentina allocated to each of the disaggregated agricultural goods; agricultural land area for cereals and other crops is from reported cultivated area; agricultural land area for cattle and sheep is from the reported numbers of each type of cattle and sheep and estimates of animals per hectare for each type of animal, as discussed further in Section A.6.8 of this online appendix.

A.4.5.2 Overidentification Check Using Transatlantic Freight Rates

In a first overidentification check, we compare our estimated changes in relative technology-adjusted prices at Argentina's trade hub ($\widehat{\mu}_{g\chi} = \mu_{g\chi}/\mu_{gt}$) from Section 6.6 of the paper to separate data on changes in relative transatlantic freight rates over the periods 1869-1914 and 1895-1914.

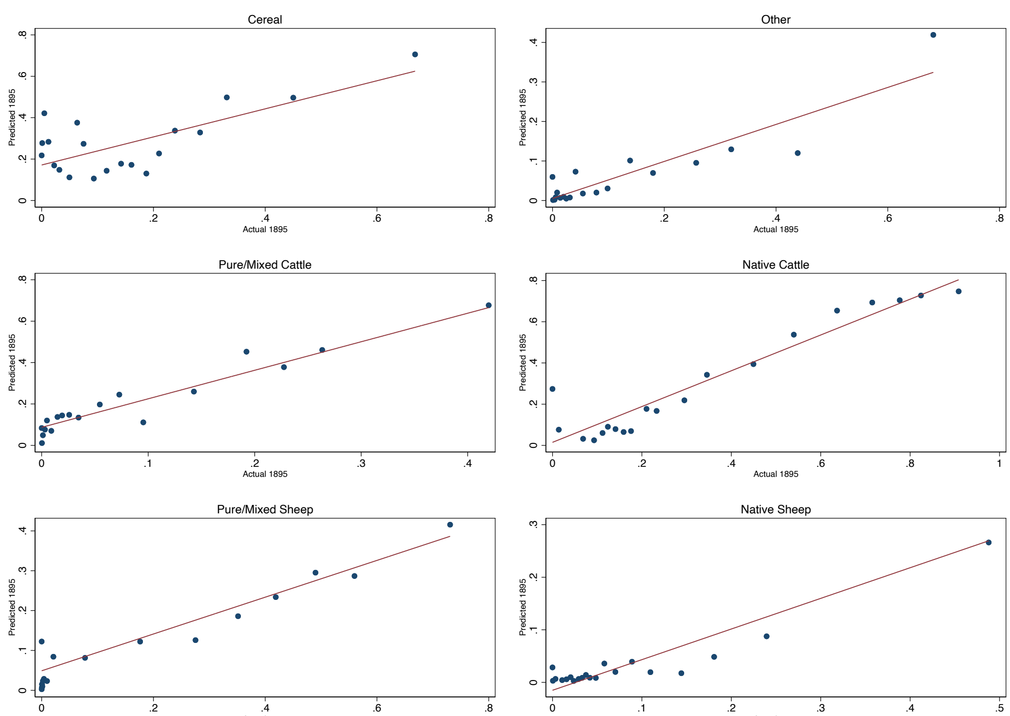
For each of our disaggregated agricultural goods, we assign the following transatlantic freight rates from Tena-Junguito and Willebald (2013): (i) Cereals (Wheat); (ii) Other Crops (Corn); (iii) Pure/Mixed-breed Cattle (Beef); (iv) Native-breed Cattle (Hides); (v) Pure/Mixed-breed Sheep (Wool); (vi) Native-breed Sheep (Hides). We estimate our changes in technology-adjusted prices ($\widehat{\mu}_{g\chi}$) relative to the numeraire of Native-breed cattle (a value of one). Therefore, we compute changes in transatlantic freights from 1869-1914 relative to the excluded category of hides, which is used for both native-breed cattle and native-breed sheep. Therefore, we can compare our estimated changes in relative technology-adjusted prices ($\widehat{\mu}_{g\chi}$) to changes in relative transatlantic freights for the remaining four categories of Cereals, Other Crops, Pure/Mixed-breed Cattle and Pure/Mixed-breed Sheep.

We can make this comparison for a relatively small number of agricultural goods. Furthermore, our estimated intercepts ($\widehat{\mu}_{g\chi}$) capture common changes in relative production technology, whereas the transatlantic freight rates focus on changes in international transport costs. Nevertheless, we find the expected negative correlation, with lower relative transatlantic freight rates implying higher relative prices for these exported goods. Across the four goods, we find correlations of -0.44 from 1869-1914 and -0.81 from 1895-1914.

A.4.5.3 Overidentification Check Using 1895 District-Level Agricultural Land Shares

In a second overidentification check, we compare our model's predictions for district-level agricultural land shares in 1895 with the observed data on these agricultural land shares in that year. We use the model's predictions from equation (49) in Section 6.6 of the paper, which uses (i) our cross-section estimates for 1914 of the impact of travel time to the nearest top-four port on agricultural specialization from equation (48) in the paper; (ii) the observed changes in the railroad network from 1895-1914 ($\widehat{\tau}_{\chi}(\ell, \ell^{*}) = \tau_{\chi}(\ell, \ell^{*}) / \tau_{t}(\ell, \ell^{*})$); and (iii) our estimated intercepts for common changes in relative technology-adjusted prices for each disaggregated good at Argentina's trade hub from 1895-1914 ($\widehat{\mu}_{g\chi} = \mu_{g\chi}/\mu_{gt}$). In Figure A.13, we display Binscatters of the model's predictions against the actual data, where the blue dots correspond to quantiles of the distribution of agricultural land shares, and the red line shows the linear fit between the predicted and actual agricultural land shares. We find a strong positive and statistically significant correlation between our model's predictions and the observed data, with an average correlation across the disaggregated agricultural goods of 0.64. Therefore, we find that our approach of predicting historical patterns of agricultural specialization using the land shares in our baseline year of 1914, observed changes in the transport network, and estimated changes in the intercept capturing technology-adjusted prices at Argentina's trade hub has predictive power for the observed agricultural land shares in 1895.

Figure A.13: Predicted and Actual Agricultural Land Shares 1895



Note: Binscatters of predicted agricultural land shares in 1895 against actual agricultural land shares in that year for each disaggregated good; vertical axis shows predicted agricultural land share; horizontal axis shows actual agricultural land share; blue dots show twenty quantiles of the distribution of agricultural land shares; red line shows the linear fit between the predicted and actual values; predicted agricultural land shares from equation (49) in Section 6.6 of the paper, which uses (i) our cross-section estimates for 1914 of the impact of travel time to the nearest top-four port (Buenos Aires, Rosario, La Plata and Bahia Blanca) on agricultural specialization from equation (48) in Section 6.6 of the paper; (ii) the observed changes in the railroad network from 1895-1914 ($\widehat{\tau}_{\chi}(\ell, \ell^{*})$); and (iii) our estimated intercepts for common changes in relative technology-adjusted prices for each disaggregated good at Argentina's trade hub from 1895-1914 ($\widehat{\mu}_{g\chi}$).

A.4.5.4 Overidentification Check Using District-Level Railroad Shipments

In a third overidentification check, we compare our model's predictions for the quantity of cereals produced in each district in 1895 and 1914 with separate data on railroad shipments of cereals in those years, which are not used in any part of the estimation of the model's parameters.

We begin by deriving the model's predictions for the quantity of each disaggregated agricultural good produced. Revenue for each agricultural good ($R_{gt}(\ell)$) is equal to total agricultural revenue ($R_{At}(\ell)$) times that good's share of total agricultural revenue ($\xi_{gt}(\ell)$):

$$R_{gt}(\ell) = \xi_{gt}(\ell) R_{At}(\ell). \tag{A.70}$$

Under our assumption of a Cobb-Douglas production technology, the share of each agricultural good in revenue is equal to its share in agricultural land ($l_{gt}(\ell)$), and zero profits implies that total agricultural revenue is equal to total agricultural income. Using these two properties, we can rewrite revenue for each agricultural good in equation (A.70) in terms of the agricultural land share ($l_{gt}(\ell)$), the wage ($w_t(\ell)$), agricultural employment ($N_{At}(\ell)$), land rents ($r_t(\ell)$) and agricultural land area ($L_{At}(\ell)$):

$$R_{gt}(\ell) = l_{gt}(\ell)\left[w_t(\ell) N_{At}(\ell) + r_t(\ell) L_{At}(\ell)\right]. \tag{A.71}$$

Using the definition of the wage-rental ratio ($\omega_t(\ell) \equiv w_t(\ell)/r_t(\ell)$), we can further re-write this expression for agricultural revenue for each good as:

$$R_{gt}(\ell) = l_{gt}(\ell) r_t(\ell)\left[\omega_t(\ell) N_{At}(\ell) + L_{At}(\ell)\right]. \tag{A.72}$$

Assuming a constant proportional relationship between land rents ($r_t(\ell)$) and land values ($v_t(\ell)$), such that $r_t(\ell) = \varsigma_t v_t(\ell)$, agricultural revenue in equation (A.72) can be further re-written as:

$$R_{gt}(\ell) = \varsigma_t l_{gt}(\ell) v_t(\ell)\left[\omega_t(\ell) N_{At}(\ell) + L_{At}(\ell)\right], \tag{A.73}$$

where agricultural land area is given by:

$$\frac{L_{At}(\ell)}{L_{At}(\ell) + L_{Nt}(\ell)} = \frac{\frac{\alpha^A}{1-\alpha^A} N_{At}(\ell)}{\frac{\alpha^A}{1-\alpha^A} N_{At}(\ell) + \frac{\alpha^N}{1-\alpha^N} N_{Nt}(\ell)}. \tag{A.74}$$

Noting that revenue for each agricultural good ($R_{gt}(\ell)$) equals price ($P_{gt}(\ell)$) times quantity ($Q_{gt}(\ell)$), equation (A.73) implies the following expression for the quantity of each agricultural good produced:

$$Q_{gt}(\ell) = \frac{\varsigma_t l_{gt}(\ell) v_t(\ell)}{P_{gt}(\ell)}\left[\omega_t(\ell) N_{At}(\ell) + L_{At}(\ell)\right]. \tag{A.75}$$

As in our land shares estimation in Section 6.6 of the paper, we assume that agricultural prices ($P_{gt}(\ell)$) depend on the international price at Argentina's trade hub (P^*_{gt}) and a constant elasticity function of travel time ($\tau_t(\ell, \ell^*)$):

$$P_{gt}(\ell) = P^*_{gt} \tau_t(\ell, \ell^*)^{\phi_g}, \tag{A.76}$$

where we estimated the coefficient ϕ_g in our land shares estimation in Section 6.6 of the paper. Using equation (A.76) to substitute for the price of each agricultural good in equation (A.75), we obtain the following expression for the quantity of each agricultural good produced:

$$Q_{gt}(\ell) = \frac{\varsigma_t l_{gt}(\ell) v_t(\ell)}{P^*_{gt} \tau_t(\ell, \ell^*)^{\phi_g}}\left[\omega_t(\ell) N_{At}(\ell) + L_{At}(\ell)\right]. \tag{A.77}$$

Assuming that local consumption of each agricultural good is small relative to local production, which is consistent with the export-orientation of agricultural activity in Argentina during our sample period, total shipments of each agricultural good equal the total quantity of the good produced. Additionally, assuming that all shipments of each agricultural good are sent by rail, railroad shipments of each agricultural good ($\mathbb{Q}_{gt}^{\text{Rail}}(\ell)$) equal the total quantity of the good produced: $\mathbb{Q}_{gt}^{\text{Rail}}(\ell) = Q_{gt}(\ell)$. Using these assumptions and equation (A.77), we obtain the following prediction of the model for railroad shipments of each agricultural good:

$$\ln \mathbb{Q}_{gt}^{\text{Rail}}(\ell) = \iota_{gt} + \ln \left\{ \frac{l_{gt}(\ell)\, v_t(\ell)}{\tau_t(\ell, \ell^*)^{\phi_g}} \left[\omega_t(\ell)\, N_{At}(\ell) + L_{At}(\ell)\right] \right\} + h_{gt}(\ell), \tag{A.78}$$

where the constant (ι_{gt}) captures the price of this good at Argentina's trade hub (P_{gt}^*), the constant proportional relationship between land rents and land values (ς_t), and the choice of units in which railroad shipments are measured; the error term ($h_{gt}(\ell)$) captures measurement error in railroad shipments, and allows for random departures from the assumptions that local consumption is small and all shipments are sent by rail.

On the left-hand side of equation (A.78), we observe the quantity of railroad shipments ($\mathbb{Q}_{gt}^{\text{Rail}}(\ell)$) of each agricultural good in 1895 and 1914 in the data. On the right-hand side of equation (A.78), we observe agricultural land shares ($l_{gt}(\ell)$) for each agricultural good in both of these years; we observe land values ($v_t(\ell)$) in 1895, which we use for both 1895 and 1914; we have already estimated the constant elasticity function of travel time for each agricultural good ($\tau_t(\ell, \ell^*)^{\phi_g}$) in our land shares estimation in Section 6.6 of the paper; we recover the wage-rental ratio ($\omega_t(\ell)$) from our model inversion in Section 6.5 of the paper; we observe agricultural employment ($N_{At}(\ell)$); and we can recover agricultural land area ($L_{At}(\ell)$) from observed agricultural employment ($N_{At}(\ell)$) and non-traded employment ($N_{Nt}(\ell)$) using equation (A.74).

We find strong, positive and statistically significant relationships between the data on railroad shipments and the model's predictions for each agricultural good, but we focus on cereals for the following two reasons. First, from the historical literature, cereals were mainly shipped by rail rather than by other modes of transport, whereas cattle in particular were sometimes transported by foot. Second, cereals are a relatively homogeneous category, and railroad shipments of each cereal (Wheat, Linseed, Corn (Maize), Barley, Oats, Bran, Rice and Other Cereals) are reported in consistent units of tons. In contrast, for the other agricultural goods, quantities are typically not reported in consistent units (e.g. numbers of cattle and tons of cattle hides), which makes it hard to consistently aggregate these different types of shipments within goods in the absence of prices in the railroad shipments data.

In Table A.3, we report the results of estimating equation (A.78) for cereals for the subset of the Argentinian districts in our model sample for which we have positive values for cereals' railroad shipments and production. As shown in Columns (1) and (2), we find a positive and statistically significant relationship between the data and the model's predictions in both 1895 and 1914.[7] We find estimated coefficients below one, which is consistent with the idea that some cereals production could be used for local consumption or could be shipped using other modes of transport. In Columns (3) and (4), we show that these estimated relationships are robust to controlling for geographical location within Argentina and agroclimatic conditions using latitude and longitude. We continue to find strong, positive and statistically significant relationships, with estimated coefficients that remain within the 95 percent confidence intervals of those reported in Columns (1) and (2).

[7]Despite the caveats about the problems of aggregating shipments using different units within the other agricultural goods, we find positive and statistically significant relationships for all of the other goods, with the following estimated coefficients (standard errors) in 1914: other crops 0.371 (0.070), pure/mixed and native cattle 0.293 (0.077), and pure/mixed and native sheep 0.432 (0.072).

Table A.3: Rail Shipments of Cereals in the Data Versus Quantity of Cereals Produced in the Model

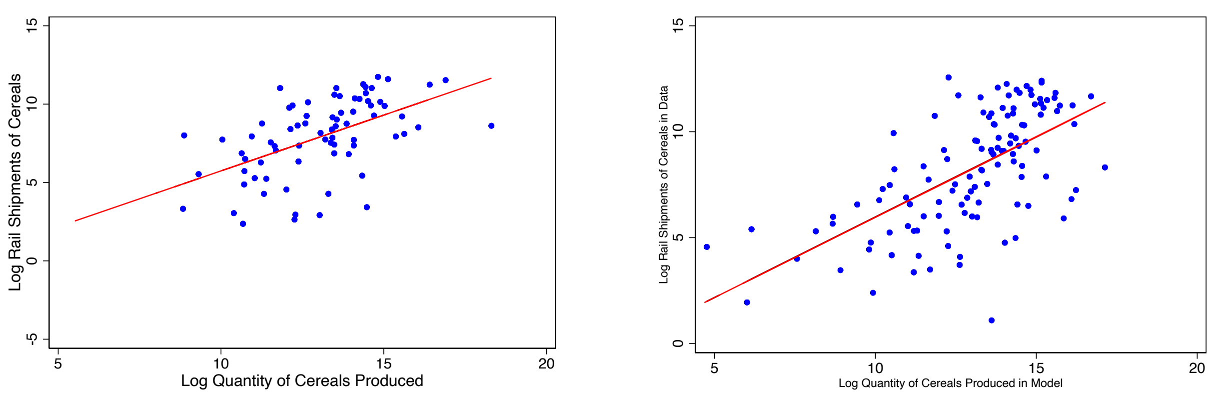
	(1)	(2)	(3)	(4)
	Rail Shipments of Cereals in the Data	Rail Shipments of Cereals in the Data	Rail Shipments of Cereals in the Data	Rail Shipments of Cereals in the Data
	1895	1914	1895	1914
Quantity of Cereals Produced in the Model	0.712***	0.759***	0.710***	0.704***
	(0.130)	(0.080)	(0.136)	(0.090)
Latitude and Longitude	–	–	Yes	Yes
Estimation	OLS	OLS	OLS	OLS
Observations	75	121	75	121
R-squared	0.286	0.392	0.431	0.422

Notes: Observations in each year are the subset of Argentinian districts in our model sample with positive values for cereals' rail shipments and production. Rail shipments in the data is the quantity of cereals (Wheat, Linseed, Corn (Maize), Barley, Oats, Bran, Rice and Other Cereals) in tons loaded at railroad stations in each district. Quantity of cereals produced in the model is the model's prediction for the quantity of cereals produced in each district from equation (A.78) in this section of the online appendix. Latitude and longitude are the latitude and longitude of the centroid of each district. 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 Figure A.14, we provide further evidence on the fit of the model throughout the observed distribution in the data. We scatter cereals' railroad shipments in the data against cereals' production in the model for 1895 and 1914, and include the linear fit between the two variables. As apparent from the two panels of the figure, we find an approximately log linear relationship between the observed data and the model's predictions in both years. Although there are several reasons why the observed data need not exactly equal the model's predictions, including local consumption and shipments using other modes of transport, these empirical results demonstrate that the model has predictive power for separate data on railroad shipments that were not used in the estimation of the model's parameters.

Figure A.14: Rail Shipments of Cereals in the Data Versus Quantity of Cereals Produced in the Model

(a) 1895 (b) 1914



Note: the vertical axis shows the data on the quantity of cereals (in tons) shipped from railroad stations in each district; the horizontal axis shows the model's prediction for the quantity of cereals produced in each district from equation (A.78) in this section of the online appendix; each dot corresponds to the observation for a district in the relevant year; the red line shows the linear fit between the two variables.

A.4.5.5 Overidentification Check Using District-Level Machinery Use

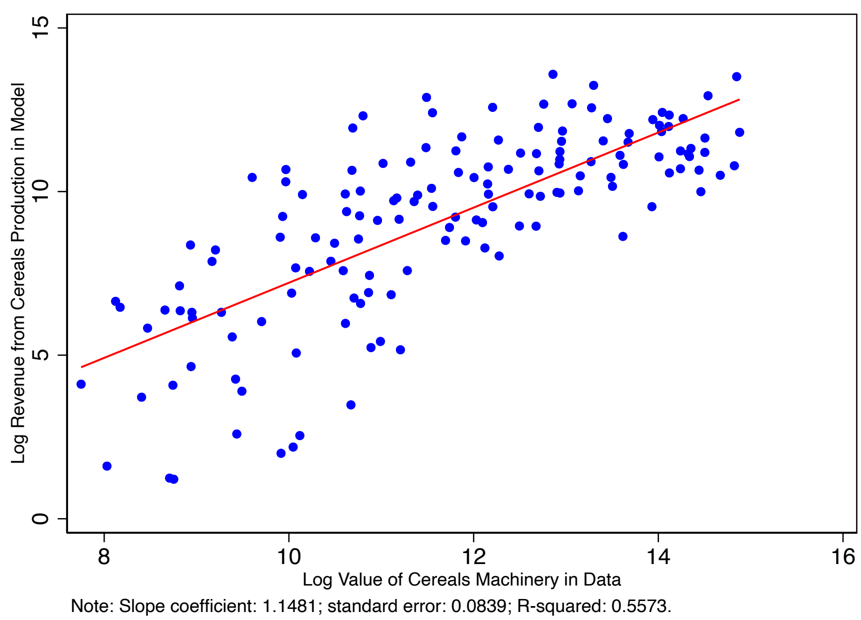
In a fourth overidentification check, we compare our model's predictions for the revenue from cereals production in each district in 1914 with separate data on the value of machinery used for cereals production in that year. From equation (A.73) from our overidentification check using railroad shipments in the previous subsection, revenue from cereals production in the model ($R_{gt}(\ell)$) can be expressed in terms of the observed cereals agricultural land share ($l_{gt}(\ell)$), the observed value of land per hectare ($v_t(\ell)$), the wage-rental ratio from our model inversion ($\omega_t(\ell)$), observed employment in agriculture ($N_{At}(\ell)$), and total agricultural land area ($L_{At}(\ell)$) as:

$$R_{gt}(\ell) = \varsigma_t l_{gt}(\ell) v_t(\ell) [\omega_t(\ell) N_{At}(\ell) + L_{At}(\ell)], \tag{A.79}$$

where ς_t captures our assumed constant proportional relationship between land rents and land values.

As discussed in Section 3 of the paper, we observe data on the number of agricultural machines in both 1895 and 1914 and on the value of agricultural machines in 1914. We focus on the year 1914 for which values are reported and for which a more disaggregated classification of agricultural machinery is provided. We construct the total value of machines used for cereals production by aggregating the following categories of machinery: (i) Baler (used to compress cut and raked crop into compact bales that are easy to handle, transport, and store); (ii) Breakers (used to break up the ground for cultivation); (iii) Combines (used to harvest grains); (iv) Gleaner (a self-propelling combine used to harvest grains); (v) Mowers (used to cut grasses and grains); (vi) Ploughs (used for loosening or turning the soil before sowing seed or planting); (vii) Rakes (used to collect cut grasses and grains); (viii) Rollers (used for flattening land or breaking up large clumps of soil); (ix) Seeders (used for sowing seeds); (x) Shellers (used for processing Corn (Maize)); and (xi) Threshers (used for removing the seeds from the stalks and husks of grain).

Figure A.15: Revenue from Cereals Production in the Model Versus Value of Machinery Used for Cereals Production in the Data



Note: the vertical axis shows the log revenue from cereals production in the model in each district in 1914 from equation (A.79) in this section of the online appendix; horizontal axis shows the log value of cereals machinery in the data in each district in 1914; cereals machinery aggregates types of agricultural machinery used for the cultivation and harvesting of cereals, as discussed in the text above; each dot corresponds to the observation for a district; the red line shows the linear fit between the two variables.

In Figure A.15, we display the model's prediction for revenue from cereals production in each district in 1914 against the data on the value of cereals machinery in 1914. We find a positive and statistically significant relationship between the model's predictions and the observed data. We observe an approximately log linear relationship between the two variables, suggesting that the value of cereals machinery used scales approximately proportionately with the value of cereals production. Although there are many idiosyncratic factors that could affect the relationship between the value of cereals production and the value of cereals machinery used in individual districts that are not captured by the model, these empirical results provide further evidence that the model has predictive power for separate data not used in the estimation of its parameters.

A.5 Counterfactuals

In this section of the online appendix, we provide further information on the counterfactuals that are discussed in Section 7 of the paper. We evaluate the impact of external integration (changes in transatlantic freight rates) and internal integration (the construction of the railroad network) on macroeconomic aggregates and the spatial distribution of economic activity within Argentina.

We undertake our counterfactuals starting from the observed equilibrium in the data in our baseline year of $t = 1914$, for which we have district-level data on agricultural land shares ($l_{gt}(\ell)$) for each of the disaggregated 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 the year 1869. We focus on the set of districts in our model sample for which we have data in both 1869 and 1914 and can compute adjusted agricultural productivity for both years. As in the main text of the paper, we use a hat above a variable to denote a relative change between year $t = 1914$ and year $\chi < t$, such that $\widehat{x}_\chi = x_\chi / x_t$.

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 in the paper and this section of the online appendix, we treat the international terms of trade ($\left\{P^*_{gt}\right\}_{g=1}^{G}$, P^*_{Mt}) as exogenous, which implicitly assumes that Argentina is a small open economy. In Section A.5.2 of this online appendix, we report a robustness exercise in which we allow for endogenous changes in the international terms of trade.

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 $\{n_t(\ell), \nu_{At}(\ell), \omega_t(\ell), E_{Tt}(\ell), E_t(\ell)\}$.

In Section A.5.1, we report the system of equations that we use to solve for a counterfactual equilibrium. We show that we can determine the unique 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. In Section A.5.2, we report a robustness exercise for our counterfactuals, in which we allow for endogenous changes in the international terms of trade in response to changes in economic activity within Argentina. In Section A.5.3, we report an additional robustness exercise, in which we allow for agglomeration forces in both the traded and non-traded sectors. In Section A.5.3, we report a further robustness exercise, in which we generalize our baseline specification with CES preferences from Section 5 of the paper to allow for non-homothetic CES preferences.

A.5.1 System of Counterfactual Equations

We now show that we can solve for a counterfactual equilibrium in year $\chi < t$ by solving the system of general equilibrium equations in the model for the wage-rental ratio $(\omega_\chi(\ell))$ that equates the demand and supply for labor in each location, given our assumed values for the four sufficient statistics $\{\widetilde{z}_{A\chi}(\ell), z_{N\chi}(\ell), u_\chi^{RW}, N_\chi^W\}$.

We start with an initial guess for the wage-rental ratio $(\omega_\chi(\ell))$ and then update this guess until labor demand equals labor supply in each location. First, for each value of the wage-rental ratio $(\omega_\chi(\ell))$, we compute labor demand in each location $(N_\chi^D(\ell))$. In particular, collecting together the common component of utility (equation (A.80) below), the relative price of tradeables (equation (A.81) below), the agricultural employment share (equation (A.82) below), population density (equation (A.83) below), and the relation between population and population density (equation (A.84) below), we can determine labor demand $(N_\chi^D(\ell))$:

$$u_\chi(\ell) = \left[\beta_T \left(\widetilde{z}_{A\chi}(\ell)\,\omega_\chi(\ell)^{\alpha_A}\right)^{\sigma-1} + (1-\beta_T)\left(z_{N\chi}(\ell)\omega_\chi(\ell)^{\alpha_N}\right)^{\sigma-1}\right]^{\frac{1}{\sigma-1}}, \tag{A.80}$$

$$\frac{E_{T\chi}(\ell)}{E_\chi(\ell)} = \frac{u_\chi^*}{\widetilde{z}_{A\chi}(\ell)\omega_\chi(\ell)^{\alpha_A}}. \tag{A.81}$$

$$\nu_{A\chi}(\ell) = \frac{(1-\alpha_A)\,\beta_T\left(\frac{E_{T\chi}(\ell)}{E_\chi(\ell)}\right)^{1-\sigma}}{1-\left(\alpha_N + (\alpha_A-\alpha_N)\,\beta_T\left(\frac{E_{T\chi}(\ell)}{E_\chi(\ell)}\right)^{1-\sigma}\right)}, \tag{A.82}$$

$$n_\chi(\ell) = \left(\frac{1}{\alpha_N + (\alpha_A-\alpha_N)\,\beta_T\left(\frac{E_{T\chi}(\ell)}{E_\chi(\ell)}\right)^{1-\sigma}} - 1\right)\frac{1}{\omega_\chi(\ell)}, \tag{A.83}$$

$$N_\chi^D(\ell) = n_\chi(\ell)\,L(\ell). \tag{A.84}$$

Second, for each value of the wage-rental ratio $(\omega_\chi(\ell))$, we compute labor supply in each location $(N_\chi^S(\ell))$. In particular, collecting together the common component of utility (equation (A.80) below), expected utility within Argentina (equation (A.85) below), international population mobility (equation (A.86) below) and population mobility within Argentina (equation (A.87) below), we can determine labor supply $(N_\chi^S(\ell))$:

$$u_\chi^* = \left[\sum_{\ell\in\mathcal{L}} u_\chi(\ell)^\varepsilon\right]^{\frac{1}{\varepsilon}}, \tag{A.85}$$

$$N_\chi^S(u_t^*) = \frac{(u_\chi^*)^{\varepsilon^{INT}}}{(u_\chi^{RW})^{\varepsilon^{INT}} + (u_\chi^*)^{\varepsilon^{INT}}} N_\chi^W, \tag{A.86}$$

$$N_\chi^S(\ell) = \left(\frac{u_\chi(\ell)}{u_\chi^*}\right)^\varepsilon N_\chi^S(u_\chi^*). \tag{A.87}$$

Third, for each guess for the wage-rental ratio $(\omega_\chi(\ell))$, we compare labor demand $(N_\chi^D(\ell))$ from equation (A.84) to labor supply $(N_\chi^S(\ell))$ from equation (A.87). If these two variables are equal to one another, we have found the equilibrium wage-rental ratio $(\omega_\chi(\ell))$. If not, we update our guess for the wage-rental ratio $(\omega_\chi(\ell))$, based on the relative values of labor demand $(N_\chi^D(\ell))$ and labor supply $(N_\chi^S(\ell))$. As we have shown that there exists a unique equilibrium in the model in Proposition 1 in the paper, there is a unique value for the wage-rental ratio $(\omega_\chi(\ell))$ at which labor demand $(N_\chi^D(\ell))$ equals labor supply $(N_\chi^S(\ell))$.

A.5.2 Counterfactuals with an Endogenous Terms of Trade

In this subsection of the online appendix, we report a robustness exercise for our counterfactuals in Section 7 of the paper, in which we relax the assumption that Argentina is a small open economy, and allow for endogenous changes in the international terms of trade. We choose the price of the manufacturing good on world markets as our numeraire ($\widehat{P}^{W}_{M\chi} = 1$). Therefore, the price of the manufacturing good at Argentina's trade hub ($P^{*}_{M\chi}$) is fully determined by transatlantic freight rates ($\tau^{*}_{M\chi}$), where lower transatlantic freight rates imply lower import prices for the manufacturing good ($\widehat{P}^{*}_{M\chi} = \widehat{\tau}^{*}_{M\chi}$). The price of each disaggregated agricultural good at Argentina's trade hub ($P^{*}_{g\chi}$) depends on both the endogenous price of that good on world markets ($P^{W}_{g\chi}$) and transatlantic freight rates ($\tau^{*}_{g\chi}$), where lower transatlantic freight rates imply higher export prices for each disaggregated agricultural good for each value of the price of the good on world markets ($\widehat{P}^{*}_{g\chi} = \left(\widehat{\tau}^{*}_{g\chi}\right)^{-1} \widehat{P}^{W}_{g\chi}$).

The key new feature in this extension is that the effects of changes in transatlantic freight rates or the construction of the railroad network on agricultural productivity ($z_{At}(\ell)$) and the tradeables price index ($E_{Tt}(\ell)$), and hence on adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$), are now modified by the endogenous response in the prices of the disaggregated agricultural goods on world markets ($\left\{\widehat{P}^{W}_{g\chi}\right\}_{g=1}^{G}$). In Subsection A.5.2.1, we generalize our baseline specification from Section 7 of the paper to allow for these endogenous changes in the international terms of trade. In Subsection A.5.2.2, we report our counterfactual results incorporating these endogenous changes in the terms of trade. We compute our counterfactuals for trade elasticities ranging from 3 to 5, which includes the central values used in the existing empirical international trade literature.

A.5.2.1 Introducing an Endogenous Terms of Trade

We begin by generalizing our baseline specification in Section 7 of the paper to introduce an endogenous terms of trade. First, we show how we use the structure of the model to solve for actual and counterfactual exports of each disaggregated good. Second, we specify a constant elasticity import demand curve for the rest of the world, which we use to solve for the endogenous change in international prices for each disaggregated agricultural good as a function of the counterfactual change in exports of that disaggregated good. Third, we incorporate these endogenous responses in international prices into the system of equations that we used to solve for a counterfactual equilibrium in Section A.5.1 of this online appendix.

Exports of each Disaggregated Agricultural Good We start by using the structure of the model to solve for exports for each disaggregated good. Recall that from equation (42) in Section 6.4 of the paper that we can solve for real income in each location as follows:

$$Y_t(\ell) = \frac{y_t(\ell) L(\ell)}{E_t(\ell)} = \frac{u_t^*}{(N_t)^{\frac{1}{\epsilon}}} \frac{(1-\alpha_A) + (\alpha_A - \alpha_N)\nu_{At}(\ell)}{(1-\alpha_A)(1-\alpha_N)} L(\ell)^{\frac{\epsilon+1}{\epsilon}} n_t(\ell)^{\frac{\epsilon+1}{\epsilon}}, \tag{A.88}$$

where $y_t(\ell)$ is income per unit of land; $L(\ell)$ is land area; $E_t(\ell)$ is the tradeables price index; u_t^* is expected utility; N_t is total population; $\nu_{At}(\ell)$ is the agricultural employment share; and $n_t(\ell)$ is population density.

Using expenditure and cost minimization, aggregate agricultural exports ($x_{At}(\ell) L(\ell)$) in each location are a multiple of income in that location ($y_t(\ell) L(\ell)$) that depends on the agricultural employment share ($\nu_{At}(\ell)$):

$$X_{At}(\ell) = x_{At}(\ell) L(\ell) = \frac{(1-\gamma_A)(1-\alpha_N)\nu_{At}(\ell)}{(1-\alpha_A) + (\alpha_A - \alpha_N)\nu_{At}(\ell)} y_t(\ell) L(\ell), \tag{A.89}$$

where $x_{At}(\ell)$ is the aggregate value of agricultural exports per unit of land and γ_A is the aggregate share of all the disaggregated agricultural goods in consumer expenditure.

From equation (24) in Section 5.8 of the paper, exports of each disaggregated good in each location ($x_{gt}(\ell) L(\ell)$) can be expressed in terms of the share of agricultural land area for that good ($l_{gt}(\ell)$) and the aggregate value of agricultural exports ($x_{At}(\ell) L(\ell)$) in that location:

$$X_{gt}(\ell) = x_{gt}(\ell) L(\ell) = \frac{l_{gt}(\ell) - \gamma_g}{1-\gamma_A} x_{At}(\ell) L(\ell) \approx \frac{l_{gt}(\ell)}{1-\gamma_A} x_{At}(\ell) L(\ell), \tag{A.90}$$

where the final term uses small local consumption ($\gamma_g \approx 0$).

Using equations (A.88), (A.89) and (A.90), and summing across locations, we can solve for aggregate real exports of each disaggregated good in purchasing power parity terms as:

$$X_{gt} = \sum_{\ell \in \mathcal{L}} \frac{x_{gt}(\ell) L(\ell)}{E_t(\ell)} \approx \frac{u_t^*}{(N_t)^{\frac{1}{\epsilon}}} \sum_{\ell \in \mathcal{L}} \left[\frac{\frac{(1-\gamma_A)(1-\alpha_N)\nu_{At}(\ell)}{(1-\alpha_A)+(\alpha_A-\alpha_N)\nu_{At}(\ell)} \times \frac{l_{gt}(\ell)}{1-\gamma_A} \times}{\left[\frac{(1-\alpha_A)+(\alpha_A-\alpha_N)\nu_{At}(\ell)}{(1-\alpha_A)(1-\alpha_N)} L(\ell)^{\frac{\epsilon+1}{\epsilon}} n_t(\ell)^{\frac{\epsilon+1}{\epsilon}}\right]} \right]. \tag{A.91}$$

Constant Elasticity Import Demand Given these real exports for each disaggregated good, we assume a constant elasticity import demand curve for the rest of the world. Using this assumption, the change in the international price of each disaggregated agricultural good ($\widehat{P}_{gt}^W$) is the following constant elasticity function of the change in real exports of that good ($\widehat{X}_{gt}$):

$$\widehat{P}_{gt}^W = \widehat{X}_{gt}^{1/\vartheta}, \tag{A.92}$$

where ϑ is the elasticity of the rest of the world's import demand with respect to the price of the good.

Counterfactual Equilibrium with an Endogenous Terms of Trade Finally, we incorporate these endogenous responses in the international terms of trade into our system of equations for a counterfactual equilibrium. Recall that we undertake our counterfactuals starting from our baseline year of $t = 1914$ back to year $\chi < t$. The changes in agricultural productivity ($\widehat{z}_{A\chi}(\ell)$) and the tradeables price index ($\widehat{E}_{T\chi}(\ell)$), and hence in adjusted agricultural productivity ($\widehat{\widetilde{z}}_{A\chi}(\ell)$), back to this earlier year $\chi < t$, are now endogenously determined as part of the equilibrium, because of these endogenous responses in the international terms of trade. In our counterfactual for external integration, the change in agricultural productivity ($\widehat{z}_{A\chi}^{\text{External}}(\ell)$) is now:

$$\widehat{z}_{A\chi}^{\text{External}}(\ell) = \left[\sum_{g=1}^{G} l_{gt}(\ell) \left(\widehat{P}_{g\chi}^W\right)^{\theta} \left(\widehat{\tau}_{g\chi}^*\right)^{-\theta} \right]^{\frac{1}{\theta}}, \tag{A.93}$$

and the change in the tradeables price index ($\widehat{E}_{T\chi}^{\text{External}}(\ell)$) is:

$$\widehat{E}_{T\chi}^{\text{External}}(\ell) = \left(\widehat{\tau}_{M\chi}^*\right)^{1-\gamma_A} \prod_{g=1}^{G} \left(\widehat{P}_{g\chi}^W \left(\widehat{\tau}_{g\chi}^*\right)^{-1} \right)^{\gamma_g}, \qquad \text{where} \qquad \sum_{g=1}^{G} \gamma_g = \gamma_A, \tag{A.94}$$

where recall that a fall in transatlantic freight rates implies a rise in export prices ($\left(\widehat{\tau}_{g\chi}^*\right)^{-1}$) and a fall in import prices ($\widehat{\tau}_{M\chi}^*$); and we choose the international price of the manufacturing good as our numeraire ($\widehat{P}_{M\chi}^W = 1$).

Equations (A.93) and (A.94) are generalization of equations (56) and (57) in Section 7 of the paper to incorporate the endogenous response in the international terms of trade ($\widehat{P}_{g\chi}^W$). The corresponding change in adjusted agricultural productivity ($\widehat{\widetilde{z}}_{A\chi}^{\text{External}}(\ell)$) is given by:

$$\widehat{\widetilde{z}}_{A\chi}^{\text{External}}(\ell) = \widehat{z}_{A\chi}^{\text{External}}(\ell) / \widehat{E}_{T\chi}^{\text{External}}(\ell). \tag{A.95}$$

We solve for the counterfactual equilibrium with an endogenous terms of trade using the following shooting algorithm. First, we guess a change in the international prices of each disaggregated agricultural good ($\left\{\widehat{P}_{g\chi}^{W}\right\}_{g=1}^{G}$). Second, we compute the change in adjusted agricultural productivity ($\widehat{z}_{A\chi}^{\text{External}}(\ell)$), given these assumed international price changes $\left\{\widehat{P}_{g\chi}^{W}\right\}_{g=1}^{G}$. Third, we solve for the counterfactual equilibrium wage-rental ratio ($\omega_{\chi}(\ell)$) using the system of equations in Section A.5.1 above, given these assumed international price changes $\left\{\widehat{P}_{g\chi}^{*}\right\}_{g=1}^{G}$. Fourth, we compute the change in real exports for each disaggregated good ($\widehat{X}_{g\chi}$) and the implied change in international prices $\left\{\widehat{P}_{g\chi}^{*}\right\}_{g=1}^{G}$ using equations (A.88)-(A.92). If the implied change in the international price for each disaggregated agricultural good equals our guess, we have found an equilibrium. If not, we update our guess using a weighted average of the implied change in international prices and our previous guess.

A.5.2.2 Counterfactual Results with an Endogenous Terms of Trade

In Table A.4, we report the results of our counterfactuals for external integration, internal integration, and both external and internal integration with an endogenous terms of trade. The top, middle and bottom panels of the table report results for real GDP, total population and expected utility, respectively. In Column (1), we reproduce the results with an exogenous terms of trade from our baseline specification in Table 7 from Section 7 of the paper. We use the same numbering of rows as in Table 7 in the paper: Rows (4), (5) and (6) correspond to counterfactuals for external integration, internal integration, and both external and internal integration, respectively. In Columns (2), (3) and (4) we report results with an endogenous terms of trade for central values of trade elasticities from the existing empirical international trade literature. For simplicity, we assume the same trade elasticity for all of the disaggregated agricultural goods, and report results for trade elasticities ranging from 3 to 5.

Introducing endogenous changes in the terms of trade dampens the impact of external integration on real GDP, population and expected utility. The reason is that our baseline specification implies complete passthrough of changes in transatlantic freight rates to export and import prices. In contrast, this extension to incorporate an endogenous terms of trade gives rise to incomplete passthrough. An increase in transatlantic freight rates that reduces export prices in Argentina, and hence diminishes export values from Argentina, is partially offset by an improvement in the terms of trade that raises export prices in Argentina. While we obtain smaller predicted impacts of external integration in this extension with an endogenous terms of trade, the change in magnitude is small for conventional values for trade elasticities, such that we obtain similar quantitative results as in our baseline specification. For example, for a trade elasticity of 4 in Column (3), we find that reversing external integration from 1869-1914 reduces real GDP, population and expected utility to 83.6, 87.3 and 93.5 percent of 1914 values, respectively (Row (4) of each panel). By comparison, in our baseline specification in Column (1), we find reductions in real GDP, population and expected utility to 82.3, 86.2 and 92.9 percent of 1914 values, respectively (Row (4) of each panel).

Introducing endogenous changes in the terms of trade has more subtle effects on the impact of internal integration, because changes in the terms of trade and the construction of the railroad network have heterogeneous effects across locations, as determined by initial patterns of specialization. Depending on the correlation of the price changes from the construction of the railroad and those from the endogenous terms of trade, the impact of the construction of the railroad network with an endogenous terms of trade can be either larger or smaller than for a small open economy. Again, for central values for trade elasticities from the existing empirical literature, we find that allowing for an endogenous terms of term has only modest effects on our quantitative conclusions. For example, for a trade elasticity of 4 in Column (3), we find that removing railroad lines constructed from 1869-1914 reduces real GDP, population and expected utility to 89.7, 89.2 and 94.5 percent of 1914 values, respectively (Row (5) of each panel). By comparison, in our baseline specification in Column (1), we find reductions in real GDP, population and expected utility to 87.2, 90.6 and 95.2 percent of 1914 values, respectively (Row (5) of each panel).

When we consider both external and internal integration (Row (6) of each panel), we again find the same pattern of results, with an endogenous terms of trade implying only modest adjustments to the predicted changes in real GDP, population and expected utility. Taking the results of this section together, we find that our quantitative conclusions for the impact on external and internal integration on real GDP, total population and expected utility are robust to allowing for endogenous changes in the terms of trade using central values for trade elasticities from the existing empirical international trade literature.

Table A.4: Counterfactual Predictions for Real GDP, Total Population and Expected Utility in Argentina (Endogenous Terms of Trade)

	Panel (A)	Real GDP 1869/1914			
		(1) Baseline Specification	(2) TOT with Trade Elasticity 3	(3) TOT with Trade Elasticity 4	(4) TOT with Trade Elasticity 5
	External and Internal Integration				
(4)	Transatlantic freights back to 1869	0.823	0.840	0.836	0.833
(5)	Railroad network back to 1869	0.872	0.899	0.897	0.894
(6)	Transatlantic freights and railroad back to 1869	0.720	0.748	0.742	0.738
	Panel (B)	Total Population 1869/1914			
		(1) Baseline Specification	(2) TOT with Trade Elasticity 3	(3) TOT with Trade Elasticity 4	(4) TOT with Trade Elasticity 5
	External and Internal Integration				
(4)	Transatlantic freights back to 1869	0.862	0.876	0.873	0.870
(5)	Railroad network back to 1869	0.906	0.895	0.892	0.889
(6)	Transatlantic freights and railroad back to 1869	0.782	0.805	0.800	0.796
	Panel (C)	Expected Utility 1869/1914			
		(1) Baseline Specification	(2) TOT with Trade Elasticity 3	(3) TOT with Trade Elasticity 4	(4) TOT with Trade Elasticity 5
	External and Internal Integration				
(4)	Transatlantic freights back to 1869	0.929	0.936	0.935	0.933
(5)	Railroad network back to 1869	0.952	0.946	0.945	0.943
(6)	Transatlantic freights and railroad back to 1869	0.885	0.898	0.895	0.893

Notes: Table reports counterfactual values in 1869 divided by actual values in 1914; Panel (A) reports results for Real GDP; Panel (B) reports results for population; Panel (C) reports results for expected utility; Column (1) reproduces the results from our baseline specification in Table 7 in Section 7 of the paper, which treats Argentina as a small open economy; Row (4) changes adjusted agricultural productivity ($\tilde{z}_{At}(\ell)$) by changes in transatlantic freight rates back to 1869 using equations (56) and (57) in the paper; Row (5) changes adjusted agricultural productivity ($\tilde{z}_{At}(\ell)$) by removing the railroad network back to 1869 using equations (58) and (59) in the paper; Row (6) changes adjusted agricultural productivity ($\tilde{z}_{At}(\ell)$) 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 paper; Columns (2), (3) and (4) allow for endogenous changes in the international terms of trade, as discussed in this section of the online appendix, and assuming trade elasticities of 3, 4 and 5, respectively.

A.5.3 Counterfactuals with Agglomeration Forces

In this subsection of the online appendix, we report another robustness exercise for our counterfactuals in Section 7 of the paper, in which we allow for agglomeration forces in the agricultural and non-traded sectors. We adopt the standard neoclassical formulation of these agglomeration forces as external economies of scale, as in Ciccone and Hall (1996) and Allen and Arkolakis (2014). In particular, we assume that productivity for each disaggregated agricultural good ($T_{gt}(\ell)$) and non-traded productivity ($z_{Nt}(\ell)$) are constant elasticity functions of aggregate employment in the agricultural sector ($N_{At}(\ell)$) and non-traded sector ($N_{Nt}(\ell)$) respectively:

$$T_{gt}(\ell) = \overline{T}_{gt}(\ell) N_{At}(\ell)^{\theta\eta_A}, \qquad \eta_A > 0, \tag{A.96}$$

$$z_{Nt}(\ell) = \overline{z}_{Nt}(\ell) N_{Nt}(\ell)^{\eta_N}, \qquad \eta_N > 0, \tag{A.97}$$

where $\overline{T}_{gt}(\ell)$ and $\overline{z}_{Nt}(\ell)$ are exogenous components of productivity for each agricultural good and the non-traded sector, respectively; η_A and η_N control the strength of agglomeration forces; and we scale agglomeration forces for each individual agricultural good by θ to ensure that the elasticity of aggregate agricultural productivity ($z_{At}(\ell)$) with respect to employment is equal to η_A. In particular, as the external economies of scale for each disaggregated agricultural good depend on overall employment in the agricultural sector, we can write aggregate agricultural productivity ($z_{At}(\ell)$) as:

$$z_{At}(\ell) = \overline{z}_{At}(\ell) N_{At}(\ell)^{\eta_A}, \tag{A.98}$$

$$\overline{z}_{At}(\ell) = \Gamma\left(\frac{\alpha_A\theta - 1}{\alpha_A\theta}\right)\left[\sum_{g=1}^{G}\overline{T}_{gt}(\ell) P_{gt}(\ell)^{\theta}\right]^{\frac{1}{\theta}}, \tag{A.99}$$

and adjusted agricultural productivity can be written as:

$$\widetilde{z}_{At}(\ell) = N_{At}(\ell)^{\eta_A}\,\widetilde{\overline{z}}_{At}(\ell), \tag{A.100}$$

$$\widetilde{\overline{z}}_{At}(\ell) = \frac{\overline{z}_{At}(\ell)}{E_{Tt}(\ell)}. \tag{A.101}$$

Given these assumptions on agglomeration forces, the system of equations for a counterfactual equilibrium remains the same as in Subsection A.5.1 of this online appendix, except that equation (A.80) for the common component of utility and equation (A.81) for the relative tradeables price index must be modified to take account of these agglomeration forces. Starting at the observed equilibrium in the data in our baseline year of $t = 1914$, these two equations for a counterfactual year $\chi < t$ become:

$$u_\chi(\ell) = \left[\beta_T\left(N_{A\chi}(\ell)^{\eta_A}\,\widetilde{\overline{z}}_{A\chi}(\ell)\,\omega_\chi(\ell)^{\alpha_A}\right)^{\sigma-1} + (1-\beta_T)\left(N_{N\chi}(\ell)^{\eta_N}\,\overline{z}_{N\chi}(\ell)\omega_\chi(\ell)^{\alpha_N}\right)^{\sigma-1}\right]^{\frac{1}{\sigma-1}}, \tag{A.102}$$

$$\frac{E_{T\chi}(\ell)}{E_\chi(\ell)} = \frac{u_\chi^*}{N_{A\chi}(\ell)^{\eta_A}\,\widetilde{\overline{z}}_{A\chi}(\ell)\omega_\chi(\ell)^{\alpha_A}}. \tag{A.103}$$

We solve for a counterfactual equilibrium using a similar procedure as in our baseline specification in Section A.5.1 of this online appendix, modified to take into account that aggregate productivity in the agricultural and non-traded sectors is now endogenous to employment levels in each sector. In particular, we use the following shooting algorithm. First, we guess a relative change in employment in each sector and location ($\widehat{N}_{A\chi}(\ell)$, $\widehat{N}_{N\chi}(\ell)$) and hence the endogenous change in productivity in each sector and location from agglomeration forces ($\widehat{N}_{A\chi}^{\eta_A}(\ell)$, $\widehat{N}_{N\chi}^{\eta_N}(\ell)$). Second, given this assumed change in productivity ($\widehat{N}_{A\chi}^{\eta_A}(\ell)$, $\widehat{N}_{N\chi}^{\eta_N}(\ell)$), we compute labor demand in each location ($N_\chi^D(\ell)$) and labor supply in each location ($N_\chi^S(\ell)$) for each value of the wage-rental ratio ($\omega_\chi(\ell)$). Third, given this assumed change in productivity ($\widehat{N}_{A\chi}^{\eta_A}(\ell)$, $\widehat{N}_{N\chi}^{\eta_N}(\ell)$), we find the wage-rental ratio ($\omega_\chi(\ell)$) at which labor demand

Table A.5: Counterfactual Predictions for Real GDP, Total Population and Expected Utility in Argentina (with Agglomeration Forces)

	Panel (A): Baseline Specification	Real GDP 1869/1914	Total Population 1869/1914	Expected Utility 1869/1914
	External and Internal Integration			
(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
	Panel (B): Agglomeration Forces	Real GDP 1869/1914	Total Population 1869/1914	Expected Utility 1869/1914
	External and Internal Integration			
(4)	Transatlantic freights back to 1869	0.810	0.851	0.923
(5)	Railroad network back to 1869	0.862	0.898	0.948
(6)	Transatlantic freights and railroad back to 1869	0.701	0.766	0.876

Notes: Table reports counterfactual values in 1869 divided by actual values in 1914; Top panel reproduces the results from our baseline specification in Table 7 in Section 7 of the paper; Bottom panel reports a robustness exercise allow for agglomeration forces in each sector, with an elasticity of productivity in each sector with respect to aggregate employment in that sector of 0.10; Row (4) changes adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) by changes in transatlantic freight rates back to 1869 using equations (56) and (57) in the paper; Row (5) changes adjusted agricultural productivity ($\widetilde{z}_{At}(\ell)$) by removing the railroad network back to 1869 using equations (58) and (59) in the paper; 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 paper.

($N_{\chi}^{D}(\ell)$) equals labor supply ($N_{\chi}^{S}(\ell)$) in each location. Fourth, we compare the implied relative changes in employment in each sector at this wage-rental ratio to the assumed values in our guess. If these implied relative employment changes in each sector are equal to the assumed values in our guess, we have found an equilibrium. If not, we update our guess using a weighted average of the implied relative changes in employment and our previous guess.

In the urban economics literature, the conventional range of estimates for the elasticity of productivity with respect to changes in employment density for non-agricultural activity is from 3-8 percent, as reviewed in Rosenthal and Strange (2004). In the existing empirical literature on the agricultural sector, there is much less consensus about the existence and magnitude of agglomeration economies, although Holmes and Lee (2012) and Kantor and Whalley (2019) find evidence of externalities in agricultural settings. As a check on the robustness of our results to the introduction of agglomeration economies, we assume elasticities of productivity with respect to employment of 0.10 in both the agricultural and non-traded sectors, just above the conventional range of estimates in urban economics. For these parameter values, we find that the system of equations for a counterfactual equilibrium converges to the same equilibrium regardless of our assumed initial values for the endogenous variables of the model.

In Table A.5, we report the results of our counterfactuals incorporating agglomeration forces. We again use the same numbering of rows as in Table 7 in the paper: Rows (4), (5) and (6) correspond to counterfactuals for external integration, internal integration, and both external and internal integration, respectively. The top panel reproduces results from our baseline specification from Table 7 in Section 7 of the paper. The bottom panel reports results for our extension incorporating agglomeration economies with an elasticity of productivity with respect to employment in each sector of 0.10. Comparing the top and bottom panels, we find that incorporating agglomeration forces magnifies the impact of external and internal integration on real GDP, population and expected utility.

This pattern of results is intuitive, as can be seen by considering our counterfactual for internal integration. Since the construction of the railroad network raises expected utility in Argentina, it induces a population inflow as a result of international labor mobility, which increases employment in each sector, and hence raises productivity through agglomeration forces. However, the changes in total population from the construction of the railroad network are relatively modest (around 10 percent in our baseline specification in Table 7 in the paper). Furthermore, although the construction of the railroad network redistributes employment across locations, this increases productivity because of greater agglomeration forces in locations where employment increases, but decreases productivity because of smaller agglomeration forces in locations where employment decreases.[8] As a result, for empirically-reasonable values for the agglomeration parameters, we find that these magnification effects are relatively modest. For example, for external integration in Row (4), we find reductions in real GDP, population and expected utility to 81.0, 85.1 and 92.3 percent of 1914 values, respectively, using our assumed elasticity of productivity with respect to employment in each sector of 0.10. By comparison, we find reductions in real GDP, population and expected utility to 82.3, 86.2 and 92.9 percent of 1914 values, respectively, in our baseline specification with exogenous productivity in the top panel.

Taking the results of this section together, we find that our quantitative conclusions for the impact on external and internal integration on real GDP, total population and expected utility are robust to allowing for agglomeration forces in each sector for standard values for agglomeration parameters from the existing empirical literature.

A.5.4 Counterfactuals with Non-Homothetic Preferences

In this subsection of the online appendix, we report a final robustness exercise for our main counterfactuals in Section 7 of the paper, in which we now allow for non-homothetic preferences as described in Section A.3.1. We calibrate the non-homothetic model and then implement the internal- and external-integration counterfactuals corresponding to rows (3) to (6) of Table 7, i.e.: taking estimated agricultural productivity back to 1869, bringing transatlantic freights back to 1869, bringing the railroad network back to 1869, and bringing both transatlantic freights and the railroad network back to 1869, respectively.

To implement these counterfactuals, for the parameters that are common to both the homothetic and the non-homothetic model (α_A, α_N, σ, β_T, ε, and ε^{INT}), we use our estimates from Section 6.3 of the paper. In addition, consistent with the discussion in Appendix A.1 of Comin, Lashkari and Mestieri (2021), the local expenditure shares in our model are defined only up to a normalization of one of the non-homothetic parameters.[9] We normalize $\zeta_T = 1$ and set ζ_N consistent with the value of $\varepsilon_N \equiv \frac{\zeta_N - \sigma}{1-\sigma} = 1.65$ reported in column (1) of Table 1 of Comin, Lashkari and Mestieri (2021). As their estimate is implemented relative to a normalized value of 1 for manufacturing while our tradeable sector also includes food (which has a lower income elasticity than manufacturing), we also implement counterfactuals using a value of the non-homothetic parameter for the non-tradeable sector that is twice as large as the benchmark.

Given these parameters, we recover adjusted agricultural productivity and non-traded productivity following the

[8] For further analysis of these offsetting effects of enhanced and diminished agglomeration economies from the reallocation of economic activity, see for example Busso, Gregory and Kline (2013).

[9] Letting $\varepsilon_i \equiv \frac{\zeta_i - \sigma}{1-\sigma}$ and combining conditions (A.56) and (A.57) with (A.54) and (A.55), the expenditure shares of workers and landowners can be expressed as follows as a function of the ratio $\frac{\varepsilon_T}{\varepsilon_N}$:

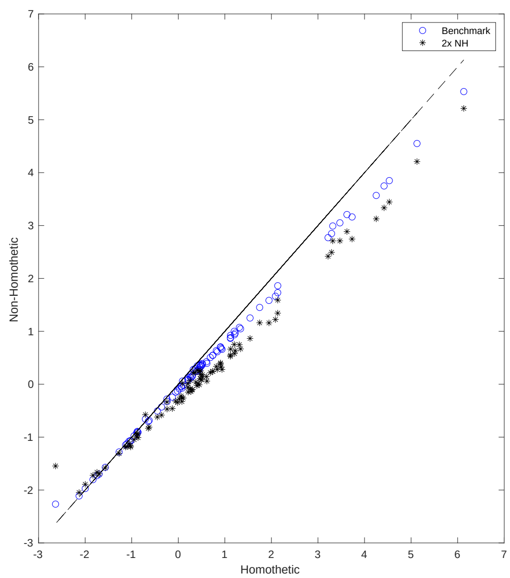
$$s_T^W(\ell)\left(1 - s_T^W(\ell)\right)^{-\frac{\varepsilon_T}{\varepsilon_N}} = \beta_T (1-\beta_T)^{-\frac{\varepsilon_T}{\varepsilon_N}} \left(Z_A(\ell) Z_N(\ell)^{-\frac{\varepsilon_T}{\varepsilon_N}}\right)^{\sigma-1} \omega(\ell)^{(\sigma-1)\left(\alpha_A - \alpha_N \frac{\varepsilon_T}{\varepsilon_N}\right)}, \tag{A.104}$$

$$s_T^L(\ell)\left(1 - s_T^L(\ell)\right)^{-\frac{\varepsilon_T}{\varepsilon_N}} = \beta_T (1-\beta_T)^{-\frac{\varepsilon_T}{\varepsilon_N}} \left(Z_A(\ell) Z_N(\ell)^{-\frac{\varepsilon_T}{\varepsilon_N}}\right)^{\sigma-1} \omega(\ell)^{(\sigma-1)\left((\alpha_A - 1) - (\alpha_N - 1) \frac{\varepsilon_T}{\varepsilon_N}\right)}. \tag{A.105}$$

steps described in Section A.3.1.2. Figure A.16 compares adjusted agricultural productivity, $\tilde{z}_A(\ell)$, between the benchmark parametrization of the homothetic model and parameterizations with non-homothetic preferences. The black line is the 45 degree line, that would correspond to no differences between the location fundamentals between homothetic and non-homothetic models. The ranking of agricultural productivities is preserved. The differences across models are small and manifest themselves more strongly for larger locations, in which the non-homothetic model requires a smaller adjusted agricultural productivity than the homothetic model to rationalize the observed data.

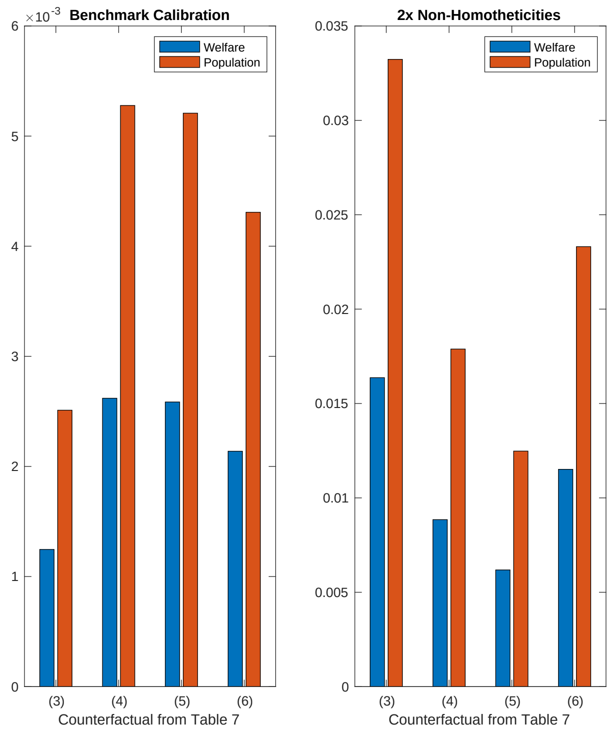
We implement counterfactuals with the non-homothetic model using the equilibrium described in Section A.3.1.1. Figure A.17 shows the percent difference between the outcomes of our main counterfactuals in rows (3) to (6) of Table 7 in the paper for the homothetic and non-homothetic models. In the left panel, we find that the absolute percentage differences in welfare and population changes between the homothetic model and the benchmark parametrization of the non-homothetic model are very small (no greater than 0.6 percent for each variable and counterfactual). The right panel shows that these differences are magnified when we increase non-homotheticities to twice their benchmark value, but that still the differences are small (less than 3.5 percent for each variable and counterfactual).

Figure A.16: Calibrated Adjusted Agricultural Productivities, Homothetic versus Non-Homothetic Case



Note: the vertical axis shows the log of the adjusted agricultural productivities calibrated for the year 1914 in the benchmark parametrization with non-homothetic preferences and in a parametrization where non-tradeables are twice as income elastic as in the benchmark. The horizontal axis shows the log of the adjusted agricultural productivities calibrated for the year 1914 in the benchmark calibration with homothetic preferences from the paper.

Figure A.17: Population and Welfare Effects in the Non-Homothetic Case Relative to the Homothetic Case



Note: the left panel shows the percent difference between the welfare and population changes in the homothetic and benchmark non-homothetic models. Each counterfactual corresponds to the corresponding row in Table 7 in the paper. The right panel shows these outcomes for the model with twice as large a parameter value for non-homotheticities in non-tradeables compared to the benchmark.

We conclude that our main counterfactual results as well as the model-implied differences in adjusted agricultural productivities are robust to including non-homothetic preferences in the analysis. In our main analysis in the paper, we focus on our baseline homothetic specification, because of its parsimony and predictive power in terms of both within-sample fit and our overidentification checks.

A.5.5 Comparison with Railroad Construction Costs

In this section of the appendix, we report additional results for the comparison with railroad construction costs discussed in Section 7 of the paper. 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 two counterfactuals corresponds to the economic impact of the railroad network starting from 1869 levels of external integration. We begin by discussing the counterfactual impact of the construction of the railroad network on real GDP and land income. We next report the results of these two sets of counterfactuals.

We evaluate the impact of the construction of the railroad network on real GDP using equation (42) in the paper. We compute its impact on real land income analogously using the structure of the model. From observed population $(N_t(\ell))$ and land area $(L(\ell))$, and our solutions for the wage-rental ratio $(\omega_t(\ell))$, the share of land in each location's income is:

$$\xi_t(\ell) = \frac{L(\ell)}{\omega_t(\ell) N_t(\ell) + L(\ell)}. \tag{A.106}$$

Combining this land share ($\xi_t(\ell)$) with overall real income ($y_t(\ell) L(\ell) / E_t(\ell)$) in each location, and summing across locations, we thus obtain the following expression for aggregate real land income (Y_t^L):

$$Y_t^L = \sum_{\ell \in \mathcal{L}} \xi_t(\ell) \frac{y_t(\ell) L(\ell)}{E_t(\ell)} = \frac{u_t^*}{N_t^{\frac{1}{\epsilon}}} \sum_{\ell \in \mathcal{L}} \frac{[(1-\alpha_A) + (\alpha_A - \alpha_N) \nu_{At}(\ell)] L(\ell)^{\frac{2\epsilon+1}{\epsilon}} n_t(\ell)^{\frac{\epsilon+1}{\epsilon}}}{(1-\alpha_A)(1-\alpha_N)[\omega_t(\ell) N_t(\ell) + L(\ell)]}. \tag{A.107}$$

Our assumption of a Cobb-Douglas production technology implies the same percentage change in land income and overall income within a given sector and location. However, both the size of this percentage change and the share of land in income ($\xi_t(\ell)$) differ across locations, because of compositional differences in the relative importance of the agricultural and non-traded sectors. As a result, the percentage change in aggregate real land income can differ from the percentage change in real GDP. In particular, aggregate real land income is a weighted sum across locations using $\xi_t(\ell)$ as weights in equation (A.107), whereas aggregate real GDP is an unweighted sum across locations in equation (42).

In Column (1) of Table A.6, we report the model's counterfactual predictions for the economic impact of the removal of the railroad network starting from 1914 levels of external integration at the observed equilibrium in the data in 1914. As shown in Rows (1) and (2), we find that real GDP and land income fall by 249 and 127 million pesos in 1914 prices, respectively. These figures correspond to percentage reductions in real GDP and land income of 12.8 and 12.7 percent, respectively, with the decline in real land income corresponding to 6.5 percent of real GDP. As shown in Rows (3)-(6), these flow reductions correspond to a decline in the net present value of real GDP from 4,976-8,293 million pesos and of real land income from 2,537-4,228 million pesos, depending on whether we assume a 5 or 3 percent discount rate. As reported in Row (7), total construction costs as measured by the capital issued by all railroad lines were 1,308 million pesos in 1914 prices. Therefore, we find ratios of changes in the net present value of income and land payments to construction costs that are substantially greater than one using either discount rate, as summarized in Rows (8)-(11). A relevant caveat is that these comparisons of net present values and construction costs do not correspond to a full cost-benefit analysis, because for example we abstract from railroad operating costs, revenues and environmental externalities. Nevertheless, these findings suggest that the large-scale investments in the construction of Argentina's 19th-century railroad network can be rationalized in terms of their effects on the net present value of economic activity.

In Column (2), we report the model's counterfactual predictions for the economic impact of the removal of the railroad network starting from 1869 levels of external integration. As shown in Rows (1) and (2), we find that real GDP and land income fall by 200 and 102 million pesos in 1914 prices, respectively. These figures correspond to percentage reductions in real GDP and land income of 10.26 and 10.23 percent, respectively, with the decline in real land income corresponding to 5.24 percent of real GDP. As shown in Rows (3)-(6), these flow reductions correspond to a decline in the net present value of real GDP from 3,993-6,655 million pesos and of real land income from 2,039-3,398 million pesos, depending on whether we assume a 5 or 3 percent discount rate. Comparing with construction costs in Row (7), we again find ratios of the reduction in the net present value of real GDP and land income to construction costs that are greater than one, as summarized in Rows (8)-(11). With a 5 percent discount rate, the net present value of the increase in land income starts to become closer to the value of construction costs.

Comparing Rows (8)-(11) in Columns (1) and (2), we find that the ratios of net present values to construction costs are substantially larger when we start from levels of external integration in 1914 than from those in 1869. This pattern of results 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 integration, relative to the fixed costs of the construction of the railroad network.

Table A.6: Counterfactual Predictions for Removing the Railroad Network from 1869-1914 and Construction Costs

	(1)	(2)
	Starting from 1914 External Integration	Starting from 1869 External Integration
Economic Impact		
(1) GDP	248.79	199.66
(2) Land Income	126.85	101.95
(3) NPV GDP (3%)	8292.93	6655.43
(4) NPV GDP (5%)	4975.76	3993.26
(5) NPV Land Income (3%)	4228.39	3398.22
(6) NPV Land Income (5%)	2537.04	2038.93
Construction Costs		
(7) Total Construction Costs	1308.00	1308.00
Ratio Economic Impact to Construction Costs		
(8) NPV GDP (3%) / Construction Cost	6.34	5.09
(9) NPV GDP (5%) / Construction Cost	3.80	3.05
(10) NPV Land Income (3%) / Construction Cost	3.23	2.60
(11) NPV Land Income (5%) / Construction Cost	1.94	1.56

Notes: Values are reported in 1914 millions of pesos; Column (1) reports counterfactuals for the reversing the construction of the railroad network starting from the observed equilibrium in the data in 1914 (starting from 1914 levels of external integration) and going back to 1869; Column (2) reports the difference between a counterfactual for reversing external integration and a counterfactual for reversing both external integration and the construction of the railroad network, which corresponds to a counterfactual for reversing the construction of the railroad network starting from 1869 levels of external integration; real GDP computed using equation (42) in the paper; real land income computed using equation (A.107) in this online appendix; NPV denotes the net present value assuming an infinite lifetime and either a 3 or 5 percent discount rate; construction costs are based on the total capital issued by all railroad lines, as reported in reported in Direccion Nacional de Ferrocarriles (1895, 1914), and discussed in further detail in Section A.6 of this online appendix.

A.6 Data Appendix

A.6.1 District Boundaries

The unit of analysis is the *partido* or *departamento* (which we refer to as "district" from now on). These districts correspond to the first administrative division within a province (the name *partido* is only used in the province of Buenos Aires, whilst in the other provinces the name *departamento* prevails).

The actual administrative division used in this paper corresponds to the one reported in the 1895 population census, when there were 23 provinces or national territories and 386 districts. The boundaries correspond to those drawn by Cacopardo (1967) with reference to that year. The same publication includes maps corresponding to the administrative divisions in place in 1869 and 1914 and a concordance table which links the districts listed in the different census years. Both sets of information – the maps and the table – were used to assign the data contained in the 1869 and 1914 census to constant spatial units based on the districts reported in the 1895 census. The process of reassigning the data is discussed for each variable in turn in the remaining subsections of this appendix.

A.6.2 Urban Population

Urban population for each district is reported in the population censuses of 1869, 1895 and 1914 (República Argentina 1869, 1895, 1914). We constructed the 1869 and 1914 urban populations of districts within 1895 boundaries using the following procedure:

- When a 1869 or 1914 district was entirely contained within a given 1895 district, the urban population for 1869 and 1914 was entirely assigned to that 1895 district.
- When a 1869 or 1914 district was split among several 1895 districts, the location of the main urban center at each year was established using secondary sources (Google Earth) and it was determined to which 1895 district it belonged. All the urban population reported for the 1869 and 1914 districts was assigned to the 1895 district where the main urban center was located.

A.6.3 Rural Population

Rural population for each district is reported in the population censuses of 1869, 1895 and 1914 (República Argentina 1869, 1895, 1914). We constructed the 1869 and 1914 rural populations of districts within 1895 boundaries using the following procedure:

- When a 1869 or 1914 district was entirely contained within a given 1895 district, the rural population for 1869 and 1914 was entirely assigned to that 1895 district.
- When a 1914 district was split among several 1895 districts, an overlap of the 1895 administrative map with the 1914 administrative map was constructed using GIS software, and it was determined which portion of the 1914 districts corresponded to the 1895 districts. Under the assumption of a uniform density of the rural population within each 1914 district, the 1914 rural population was assigned to the relevant 1895 districts.
- When a 1869 district was split among several 1895 districts, an equivalent method to the one used for the 1914/1895 match was used. Since no 1869 district ceded territory to more than one new 1895 district, it was possible to estimate the portions of each 1869 district that corresponded to a given 1895 district simply by comparing the land areas reported in 1869 and 1895. Under the assumption of a uniform density of the rural population within each 1869 district, the 1869 rural population was assigned to the relevant 1895 districts.

A.6.4 Total Population

Total population for each district is the sum of the urban and rural populations, as defined above.

A.6.5 Land Rents

Data on the value of land in pesos per hectare are reported for each district in the statistical yearbook for 1895 (Dirección General de Estadística 1895).

A.6.6 Wages of Agricultural Laborers

Data on the daily wages ("peones por dia") of agricultural laborers are reported in the agricultural statistics for 1913 (República Argentina 1913). Data are reported for the year of 1913 for a number of districts in the provinces of Buenos Aires, Catamarca, Córdoba, Corrientes, Entre Ríos, La Rioja, Mendoza, San Luis, Santa Fé, Santiago del Estero, San Juan and Tucumán.

A.6.7 Tradeables and Overall Consumption Price Index

We construct a price index for the overall cost of living ($E(\ell)$) and the cost of tradeables ($E_T(\ell)$) using data on (i) aggregate household expenditure shares for components of the consumer price index from Bunge (1918) and (ii) prices of individual traded goods that are reported for towns and cities in Argentina in Alsina (1905) and land values from the statistical yearbook for 1895 (Dirección General de Estadística 1895).

(i) Expenditure Shares. We use the following expenditure shares from a random sample of 377 households in Buenos Aires in 1913-14 that are reported in Bunge (1918):

- Beef (7.2 percent)
- Lamb (0.9 percent)
- Pork (0.9 percent)
- Bread (6 percent)
- Other Foods (15 percent)
- Other Household Expenses (50 percent)
- Housing (20 percent)

(ii) Prices We use the data on the local prices of traded goods in towns and cities in Argentinian districts from Alsina (1905) and the data on land values in each district from the statistical yearbook for 1895 (Dirección General de Estadística 1895). We allocate these prices to the expenditure share categories in Bunge (1910) as follows:

- Price of beef by district from Alsina (1905) (7.2 percent of expenditure)
- Price of lamb by district Alsina (1905) (0.9 percent of expenditure)
- Price of pork by district from Alsina (1905) (0.9 percent of expenditure)
- Price of Bread by district from Alsina (1905) (6 percent of expenditure)
- Other Foods (15 percent of expenditure)
  - Prices by district from Alsina (1905) for cooking oil, Bremen rice, Tucumán sugar, Brazilian coffee, milk, tobacco, potatoes, wine, corn flour, wheat flour, beans, noodles, crushed corn, chick peas, corn, cow fat, herbs, and salt

- Other Household Expenses (50 percent)
  - Price by district from Alsina (1905) for brooms, soap, starch, kerosene and phosphorus
- Housing (20 percent)
  - Value of land per hectare by district from the statistical yearbook for 1895 (Dirección General de Estadística 1895).

Consumer Price Index ($E(\ell)$). We follow a similar methodology as the United States Bureau of Labor Statistics (BLS) to compute the overall consumer price index ($E(\ell)$). First, for the lower tier of individual goods within expenditure share categories (e.g. milk within Other Foods) for which we only observe prices in each district and do not observe expenditure shares, we aggregate prices across these individual goods using a Jevons Index. Second, for the upper tier of expenditure share categories, for which we either observe prices in each district or have constructed prices in each district using the Jevons Index, we aggregate prices across these expenditure share categories using a Cobb-Douglas Index, which weights the prices for these categories by their expenditure shares. We use the same expenditure share weights for all districts from the household survey data in Bunge (1918). We normalize the price of each disaggregated good by its geometric mean across districts before aggregating prices, to ensure that our price index is not sensitive to differences in units of measurement for the individual goods.

Tradeables Price Index ($E_T(\ell)$). We construct the tradeables price index ($E_T(\ell)$) in the same way as the overall consumer price index, except that we exclude the non-traded good of housing, and we express the expenditure share for each traded good as a share of all expenditure on traded goods (excluding housing).

A.6.8 Disaggregated Agricultural Goods

The 1895 and 1914 population censuses (República Argentina 1895, 1914) report cultivated area in each district for a number disaggregated crops. We distinguish between the main new export crops of Cereals (which together account for around one half of aggregate export value in 1914) and Other Crops as follows:

- Cereals
  - Barley, Linseed, Maize (Corn), Oats, Rice, Rye, Sorghum and Wheat
- Other Crops
  - Beans, Bird Seed, Cassava, Cotton, Hops, Linen, Sugar Cane, Peanuts, Potato, Tobacco, Vegetables, Wine, Yams and other crops

The 1895 and 1914 population censuses (República Argentina 1895, 1914) also report the number of cattle and sheep in each district that are pure breed, mixed breed and native breeds. Whereas native breeds were historically used for hides, skins, bones, fat and tallow, mixed and pure breeds where introduced for chilled and frozen meat and wool (Scobie 1971 and Perren 2017). We therefore distinguish the following categories of animals:

- Pure and mixed-breed cattle

- Native-breed cattle
- Pure and mixed-bread sheep
- Native-breed sheep

We convert numbers of animals into land area following the U.S. department of agriculture recommendations of 1 acre for each head of cattle and 0.2 acres for each head of sheep: https://www.nrcs.usda.gov/. We thus obtain agricultural land area for the following six disaggregated goods: (i) Cereals; (ii) Other Crops; (iii) Pure and mixed-breed cattle; (iv) Native-breed cattle; (v) Pure and mixed-breed sheep; (vi) Native-breed sheep.

A.6.9 Transport Network

We compute bilateral travel times from the centroid of each district in Argentina to the centroid of all other districts and to each of the ports on the coast or navigable rivers. To compute these bilateral travel times, we construct transport networks for each year of our sample (1869, 1895 and 1914), which include the railroad network, the coast and navigable rivers, and land transport.

To construct the railroad network in each year, we start with a shapefile of the modern Argentinian railroad network. We modify this modern shapefile using the historical maps of the Argentinian railroad network in 1869, 1895 and 1914 in Randle (1981). We combine these historical shapefiles of the railroad network with data on the opening/closing and latitude and longitude coordinates of each railroad station in Argentina. The list of stations operating in each year was obtained from *Estadísticas de los ferrocarriles en explotación*, a yearly publication of the *Dirección nacional de ferrocarriles* (National Railroad Direction) for 1895 and 1914 and from historical sources for 1869. The geographical coordinates of stations were determined using Google Earth taking into account changes in station names over time. To determine the opening dates of stations, we used the opening date of the section of the railroad line on which the station is located. This information is available in *Estadísticas de los ferrocarriles en explotación.* Three dates for each section of a railroad line are available: the date when the construction was authorized, the date when the decree opening the section was issued and the actual date when service was started (generally a few months after the issuance of the decree). We used the last date whenever it was available. If this last date was unavailable, the opening decree issuance date was used. We also constructed a shapefile of the coast and main navigable rivers (the Paraná, Plate and Uruguay rivers) in Argentina

Using the resulting transport network for each year of our sample, we compute least-cost paths between the centroids of districts and to each of the ports on the coast or navigable rivers. We assume the following weights for each transport mode based on the estimates for the 19th century in Donaldson (2018): Rail (1); Coast/River (3) and Land (4.5). We abstract from automobiles/trucks, because they were of negligible importance in Argentina in 1914. We assume that agents can only connect to the railroad network at a railroad station.

A.6.10 Railroad Shipments

The railroad shipments data comes from *Estadísticas de los ferrocarriles en explotación.* The data is available on a yearly basis for all years starting in 1895. The records include the cargo loaded by station. With respect to the records for the years 1895 and 1914 used in this paper, there is no data for *Ferrocarril Central Córdoba* (Central Córdoba Railroad) in 1895 and for the *Ferrocarril Midland* (Midland Railroad) in 1914. The classification of products differs somewhat across the railroad lines. To construct a common product classification, we aggregated similar products to obtain the following categories: (i) Cereals (Wheat, Linseed, Corn (Maize), Barley, Oats, Bran, Rice and Other Cereals); (ii) Other Crops; (iii) Pure/mixed and Native-Breed Cattle and Cattle Products; (iv) Pure/Mixed and Native-Breed Sheep and Sheep Products.

A.6.11 Railroad Construction Costs

Our railroad construction cost data come from *Estadística de los Ferrocarriles en Explotación.* For each railroad line and year, this publication reports total capital issued in current price pesos and capital issued per kilometer, which yields implied railroad length. We find a total cumulative capital issued of 1,308 million current price gold pesos in 1914, which compares closely with the separate estimate of 1,267 million current price gold pesos in 1913 reported in U.S Department of Commerce (1926). We find a total implied railroad length of 33,844 kilometers in 1914, which compares closely to the separate estimate of 33,510 kilometers reported in Tornqvist (1919) for that year. The small difference between these numbers could reflect for example the different treatment (inclusion or exclusion) of narrow-gauge railroads.

Capital issued per kilometer varies across the railroad lines, depending among other things on the topography of the terrain traversed by the railroad line. We use the overall reported capital issued per kilometer of 38,661 current price pesos in 1914 as our estimate of average construction cost, which corresponds to a weighted average, in which the construction cost for each railroad line is weighted by the length in kilometers of that railroad line. As an additional check on these figures, we convert construction costs into current price pounds sterling using the reported exchange rates from Denzel (2010), and convert kilometers into miles, to obtain an average construction cost per mile in Argentina in current price pounds sterling of 53,907 in 1914. This figure is comparable to the average construction cost per mile for overground railroads in England and Wales in current price pounds sterling of 60,000 pounds in 1921 from Heblich, Redding and Sturm (2020). This similarity of construction costs is consistent with the two countries having relatively similar levels of gross domestic product (GDP) per capita at this time. Although lower land costs in Argentina are likely to have reduced construction costs relative to those in England and Wales, the more mountainous terrain in parts of Argentina is likely to have increased construction costs relative to those in England and Wales.

A.6.12 Aggregate and Customs trade

International trade data were collected from a number of official publications. For 1870, trade data were obtained from the statistical yearbook *Estadísticas de las aduanas de la República Argentina* published by the *Oficina de estadística general de la dirección de aduanas* (the Statistical Office of the Customs Direction). Both the quantity and value of exports (by destination) and imports (by origin) are reported for each customs (port) in Argentina using a relatively aggregated product classification. For 1895 and 1914, the same variables were obtained from the statistical yearbook of the *Dirección General de Estadística* (Statistics General Direction). The product classification in 1895 and 1914 is more disaggregated than that in 1870. We construct a constant product classification across all three years by concording the product categories for each year and aggregating were necessary. Finally, the geographical coordinates for each customs (port) were determined using Google Earth. The customs were then mapped into our spatial units of analysis, the districts reported in the 1895 census, using GIS software.

A.6.13 Export Prices, Import Prices and Transatlantic Freight Rates

Export and Import Prices We measure import prices, export prices and the terms of trade using data from Francis (2017), which are based on wholesale prices in Argentina, and hence capture the prices received by exporters after transport costs have been incurred. Since transport costs across the Atlantic fell substantially during our sample period, these wholesale prices in Argentina rise more rapidly over time than the prices in destination markets including the transport costs.

Export prices and export value weights are reported for the following disaggregated goods: 1 Dried Hides; 2 Salted Hides; 3 Jerked Beef; 4 Wool; 5 Fat and Tallow; 6 Cattle; 7 Sheep Skins; 8 Wheat; 9 Maize; 10 Wheat Flour; 11 Linseed; 12 Goat Skins; 13 Barley; 14 Chilled Beef; 15 Conserved Beef; 16 Frozen Beef; 17 Bran; 18 Butter; 19 Oats; 20 Sugar; 21 Quebracho Logs; 22 Quebracho Extract; 23 Rye.

A key feature of the data emphasized by our theoretical framework and quantitative analysis is the change in specialization across disaggregated goods within the agricultural sector, with new export goods such as Cereals and Chilled and Frozen Beef emerging over our sample period. In Figure 1 in the paper, we report the aggregate export and import price indexes from 1869-1914 from Francis (2017), which are constructed using a chained geometric Laspeyres index, such that new goods enter the price index when they are present for two consecutive time periods (through the chaining of these two consecutive time periods).

In Figure A.4 in the paper, we report export prices separately for each good from 1869-1914, for the 12 goods with the largest export value shares over our sample period, out of the 23 goods listed above. In this specification, prices appear for each good in the first year for which positive exports are reported. We also present export value shares in this figure for each of the 23 goods for the years 1869, 1895 and 1914.

Transatlantic Freight Rates We use data on transatlantic freight rates from Tena-Junguito and Willebald (2013). Data are reported for seven categories: 1 Beef; 2 Mutton; 3 Hides; 4 Wool; 5 Wheat; 6 Linseed; 7 Corn. We assign these seven categories to our six disaggregated agricultural goods as follows: (i) Cereals (Wheat); (ii) Other Crops (Corn); (iii) Pure/Mixed-breed Cattle (Beef); (iv) Native-breed Cattle (Hides); (v) Pure/Mixed-breed Sheep (Wool); (vi) Native-breed Sheep (Hides). Transatlantic freight rates are estimated as the ratio of the "cost inclusive of freight and insurance (cif)" price in destination markets to the "free on board (fob)" price at the origin. See Tena-Junguito and Willebald (2013) for further details on the data sources and definitions. In Section A.4.1 of this online appendix, we also report data on freight rates for coal from Wales to the River Plate (Río de la Plata) from Angier (1920) and for wheat from South America to London from North (1958).

External Integration and Tradeables Price Index In our counterfactuals for external integration in Section 7 of the paper, we measure the impact of external integration on the tradeables price index using the aggregate household expenditure shares from Bunge (1918) and the transatlantic freight rates from Tena-Junguito and Willebald (2013). In particular, we measure the change in the tradeables price index from external integration from 1869-1914 as:

$$\begin{aligned}\widehat{E}^*_{T\chi} = & \left(\left(\widehat{\tau}^*_{\text{Beef}\chi}\right)^{-1}\right)^{\gamma_{\text{Beef}}} \times \left(\left(\widehat{\tau}^*_{\text{Lamb}\chi}\right)^{-1}\right)^{\gamma_{\text{Lamb}}} \times \left(\left(\widehat{\tau}^*_{\text{Pork}\chi}\right)^{-1}\right)^{\gamma_{\text{Pork}}} \\ & \left(\left(\widehat{\tau}^*_{\text{Bread}\chi}\right)^{-1}\right)^{\gamma_{\text{Bread}}} \times \left(\left(\widehat{\tau}^*_{\text{Other}\chi}\right)^{-1}\right)^{\gamma_{\text{Other}}} \times \left(\widehat{\tau}^*_{\text{Household}\chi}\right)^{\gamma_{\text{Household}}},\end{aligned}$$

where export prices are inversely proportional to transatlantic freight rates, whereas import prices are proportional to transatlantic freight rates; we assume that the agricultural products of Beef, Lamb, Pork, Cereal (used to make bread) and Other Foods are exported and the non-agricultural products of Other Household Expenses are imported; γ_{Beef} is the household share of tradeables expenditure on beef; $\widehat{\tau}^{*}_{\text{Beef}\chi}$ is the change in the transatlantic freight for native-breed cattle used for local consumption; γ_{Lamb} is the household share of tradeables expenditure share on lamb; $\widehat{\tau}^{*}_{\text{Lamb}\chi}$ is the price change for native-breed sheep used for local consumption; γ_{Pork} is the household share of tradeables expenditure on pork; $\widehat{\tau}^{*}_{\text{Pork}\chi}$ is the change in the transatlantic freight rate for native-breed pigs used for local consumption, which we proxy with the change in the transatlantic freight rate for native-breed cattle; γ_{Bread} is the household share of tradeables expenditure on bread; $\widehat{\tau}^{*}_{\text{Bread}\chi}$ is the change in the transatlantic freight rate for bread, which we proxy with the change in the transatlantic freight rate for wheat; γ_{Other} is the household share of tradeables expenditure on other food; $\widehat{\tau}^{*}_{\text{Other}\chi}$ is the change in the transatlantic freight for other food, which we proxy with the change in the transatlantic freight for corn; $\gamma_{\text{Household}}$ is the household share of tradeables expenditure on other household expenses; $\widehat{\tau}^{*}_{\text{Household}\chi}$ is the transatlantic freight rate for other household expenses, which we proxy with the average change in transatlantic freights across our six disaggregated agricultural goods.

A.6.14 Spanish Colonial Sixteenth Century Cities

Table A.7 reports the Spanish colonial 16th-century cities used to construct our instrument based on the maps of Spanish colonial postal routes from Randle (1981).

Table A.7: Spanish Colonial 16th-Century Cities

Province	**Name**	**Year Founded**
Buenos Aires	Buenos Aires	1580
Buenos Aires	Santa María del Buen Ayre	1536
Catamarca	Londres	1558
Catamarca	San Pedro de Mercado de Andalgalá	1582
Chaco	Matará y Guacará	1585
Chaco	Nuestra Señora de la Concepción del Bermejo	1585
Corrientes	Vera en las 7 Corrientes	1588
Córdoba	Alta Gracia	1590
Córdoba	Córdoba	1573
Córdoba	Santa María	-
Córdoba	Santa Rosa de Calamuchita	-
Jujuy	Humahuaca	1596
Jujuy	Nieva	1561
Jujuy	San Francisco de Alava	1575
Jujuy	San Salvador de Jujuy	1593
La Rioja	Todos Santos de La Nueva Rioja	1591
Mendoza	Ciudad de la Resurección de Mendoza	1561
Mendoza	Mendoza	1562
Salta	Córdoba del Calchaquí	1551
Salta	Esteco (Caceres)	1566
Salta	Lerma en el Valle de Salta	1582
Salta	Madrid de Las Juntas	1592
Salta	Metán Viejo	-
Salta	Nuestra Señora de Talavera	1567
Salta	Primera San Clemente de la Nueva Sevilla	1577
Salta	Segunda y Tercera San Clemente de la Nueva Sevilla	1577
Salta	Segundo Barco	1559
San Juan	San Juan de la Frontera	1562
San Luis	San Luis de Loyola	1593
Santa Fe	Corpus Christi	1536
Santa Fe	Nuestra Señora de la Buena Esperanza	1536
Santa Fe	Sancti Spiritus	1527
Santa Fe	Santa Fe (Cayasta)	1573
Santiago del Estero	Santiago del Nuevo Maestrazgo del Estero	1553
Santiago del Estero	Tercer Barco	1552
Tucumán	Amaicha	-
Tucumán	Cañete	1560
Tucumán	El Barco	1550
Tucumán	Quilmes	-
Tucumán	Ranchillos	-
Tucumán	San Miguel de Tucumán	1565

References

Allen, Treb and Costas Arkolakis (2014) “Trade and the Topography of the Spatial Economy,” *Quarterly Journal of Economics*, 129(3), 1085-1140.

Alsina, Juan A. (1905) *El Obrero en la República Argentina*, Buenos Aires: Imprenta, Calle de Mexico.

Angier, E. A. V. (1920) *Fifty Years’ Freights 1869-1919*, London: Fairplay.

Bunge, Alejandro E. (1918) “Costo de la vida en la Argentina, de 1910 a 1917,” *Revista de Economía Argentina*, 1(1), 39-63.

Busso, Matias, Jesse Gregory and Patrick Kline (2013) “Assessing the Incidence and Efficiency of a Prominent Place Based Policy,” *American Economic Review*, 103(2), 897-947.

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

Ciccone, Antonio and Robert Hall (1996) “Productivity and the Density of Economic Activity,” *American Economic Review*, 86(1), 54-70.

Comin, Diego, Danial Lashkari and Martí Mestieri (2021) “Structural Change with Long-run Income and Price Effects,” *Econometrica*, 89(1), 311-374.

Denzel, Markus A. (2010) *Handbook of World Exchange Rates, 1590-1914*, New York: Routledge.

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

Dirección General de Estadística (1914) *Anuario de la Dirección General de Estadística*, Buenos Aires.

Dirección Nacional de Ferrocarriles (1895) *Estadísticas de los Ferrocarriles en Explotación*, Buenos Aires.

Dirección Nacional de Ferrocarriles (1895) *Estadísticas de los Ferrocarriles en Explotación*, Buenos Aires.

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

Duranton, Gilles, Peter Morrow, and Matthew A. Turner (2014) “Roads and Trade: Evidence from the US,” *Review of Economic Studies*, 81(2), 681-724.

Duranton, Gilles and Matthew A. Turner (2012) “Urban Growth and Transportation,” *Review of Economic Studies*, 79(4), 1407-1440.

Francis, Joseph 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, Ewout (2010) “Reconstructing labor income shares in Argentina, Brazil and Mexico, 1870-2000,” *Revista de Historia Economica*, 28(2), 343-374.

Hanoch, Giora (1975) “Production and Demand Models with Direct or Indirect Implicit Additivity,” *Econometrica*, 43 395-419.

Harley, C. K. (1980) “Transportation, World Wheat Trade, and the Kuznets Cycle, 1850-1913,” *Explorations in Economic History*, 17, 218-250.

Harley, C. K. (1988) “Ocean Freight Rates and Productivity, 1740-1913: The Primacy of Mechanical Invention Reaffirmed,” *Journal of Economic History*, 48(4), 851-876.

Heblich, Stephan, Redding, Stephen and Daniel Sturm (2020) “The Making of the Modern Metropolis: Evidence from London,” *Quarterly Journal of Economics*, 135(4), 2059-2133.

Holmes, Thomas J. and Sanghoon Lee (2012) "Economies of Density Versus Natural Advantage: Crop Choice on the Back Forty," *Review of Economics and Statistics*, 94(1), 1-19.

Kantor, Shawn and Alex Whalley (2019) "Research Proximity and Productivity: Long-Term Evidence From Agriculture," *Journal of Political Economy*, 127(2), 819-854.

North, Douglas (1958) "Ocean Freight Rates and Economic Development 1750-1913," *Journal of Economic History*, 18(4), 537-555.

Oficina de Estadística General de la Dirección de Aduanas (1870) *Estadísticas de las Aduanas de la República Argentina*, Buenos Aires.

O'Rourke, K. H. and Williamson, J. G. (1990) *Globalization and History: The Evolution of a 19th-century Atlantic Economy*, Cambridge: MIT Press.

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

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

República Argentina (1869) *Primer Censo de la República Argentina*, Buenos Aires: Superintendente del Censo.

República Argentina (1895) *Segundo Censo de la República Argentina*, Buenos Aires: Comision Directiva.

República Argentina (1913) "Boletin del Departamento Nacional de Trabajo", issue 25, pp. 1050, Buenos Aires.

República Argentina (1914) *Tercer Censo Nacional*, Buenos Aires: Comisión Nacional.

Rosenthal, Stuart S. and William C. Strange (2004) "Evidence on the Nature and Sources of Agglomeration Economics," in (ed.) J. V. Henderson and J. F. Thisse, *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: Elsevier North Holland.

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

Tena-Junguito, Antonio and Henry Willebald (2013) "On the Accuracy of Export Growth in Argentina 1870-1913," *Economic History of Developing Regions*, 28(1), 28-68.

Tornqvist, Ernesto (1919) *The Economic Development of the Argentine Republic in the Last Fifty Years*, Buenos Aires: Ernesto Tornqvist and Company

United States Department of Commerce (1926) *Railways of South American, Part 1: Argentina*, Washington DC: Government Printing Office.