RESEARCH SEMINAR IN INTERNATIONAL ECONOMICS

Gerald R. Ford School of Public Policy
The University of Michigan
Ann Arbor, Michigan 48109-3091

Discussion Paper No. 641

Domestic Trade Frictions and Agriculture

Sebastian Sotelo
University of Michigan

May 9, 2015

Domestic Trade Frictions and Agriculture

Sebastian Sotelo*†

This version: May 9, 2015

Abstract

Trade costs are a major barrier to efficient farming in developing countries. I study land use patterns and input demand in Peru, a country where goods are traded at a high cost, both domestically and with the rest of the world. I then quantify the equilibrium effect of paving existing roads on productivity and real incomes. To do so, I develop a model of agricultural specialization and trade, and quantify it using a new dataset on Peruvian agriculture, which includes disaggregated information on crop prices, yields and land allocations. While typically raising productivity, paving roads on a large scale creates both winners and losers, depending on whether prices are set in domestic markets, or whether workers are net food buyers. In the simulations, an average farmer gains 14% in productivity and 5% in welfare.

Keywords: assignment models, trade costs, equilibrium, agriculture, productivity

---

*Department of Economics, University of Michigan, Ann Arbor. E-mail address: ssotelo@umich.edu. Webpage: http://www-personal.umich.edu/∼ssotelo/

†I am indebted to Sam Kortum, Nancy Stokey, and Thomas Chaney for their advice and encouragement. I also thank Fernando Alvarez, Costas Arkolakis, David Atkin, Saki Bigio, Lorenzo Caliendo, Kerem Cosar, Javier Cravino, Jonathan Eaton, Jeremy Fox, Sara Heller, Erik Hurst, Gita Khun Jush, Joaquin Lopez, Sara Moreira, Ralph Ossa, Edgar Salgado, Danny Tannenbaum, Adriaan Ten Kate, and Mike Waugh, as well as participants in seminars at UChicago, Yale, the Minneapolis Fed, the St. Louis Fed, Michigan, Penn State, the Fed Board, Drexel, Toronto, Brown for helpful comments and suggestions at different stages of this project. Special thanks to Alan Sanchez and Ben Faber for detailed discussions. Maria Alejandra Zegarra provided excellent research assistance. Part of this paper was written while I was visiting the Economics Department at Yale University. Funding from the Sjaastad Research Fellowship Fund is gratefully acknowledged. Finally, I would like to thank Ministerio de Transportes y Comunicaciones de Peru and Ministerio de Agricultura y Riego de Peru for sharing their data with me. All opinions and remaining errors are my own. This paper circulated previously as "Trade Frictions and Agricultural Productivity: Theory and Evidence from Peru."

1 Introduction

In developing countries, a large majority of the poor live in rural areas. Their livelihoods are often tied to unproductive agriculture, limited by major barriers to trade such as poor infrastructure, adverse geography, and the spatial dispersion that is characteristic of rural populations. Not surprisingly, researchers and policy makers perceive the costs of domestic and international trade as a drag on incomes and productivity.[1]

An assessment of policies that reduce trade costs, however, requires an understanding of how farmers and consumers react to improved trading opportunities, as well as how their choices interact in the aggregate. On the one hand, policies that reduce trade costs can increase allocative efficiency and welfare by unlocking the forces of comparative advantage and the use of modern inputs. On the other hand, such policies also affect the equilibrium prices of crops, especially of those that are traded only domestically. Because these crops usually constitute an important part of the diet of subsistence farmers and net food buyers, improved exchange opportunities that induce shifts in local supply also affect welfare through their effect on consumption prices.[2]

In this paper, I provide a framework to measure the consequences of policies aimed at reducing trade costs. I develop a model that relates agricultural incomes and productivity to trade and specialization, and use it to calculate the equilibrium effect of large-scale infrastructure policies, like paving existing roads. In the model, farmers can grow various crops in land plots of varying quality. They can also trade their crops and purchase intermediate inputs, at a cost, in local, urban, and international markets. The model is a hybrid between a small open economy, which takes international prices as given, and a closed economy, where prices are determined by regional trade within the country. Differences in land quality are a source of comparative advantage, and generate trade across regions and with the rest of the world. In

[1] According to the World Bank's World Development Report, as of 2002, 75% of the world poor were rural dwellers (The World Bank (2007)). The same Report relates developing countries' agricultural performance to within-country variation in access to markets and land quality (p. 54.) Likewise, a recent Inter-American Development Bank report reflects on how transport costs limit overall exporting activity: "high domestic transport costs can push exports to concentrate in just a few areas [...], while squeezing gains or simply locking out of trade large swaths of the country" (Mesquita Moreira et al. (2013), p. 3.)

[2] See The World Bank (2007) p. 109

equilibrium, farming income and productivity are deeply tied to market access and comparative advantage.

To quantify the theory, I construct a detailed data set combining several sources of data on Peruvian agriculture. First, I use government statistics on land allocation, production, and prices to estimate crop-specific land quality across regions, as well as within-region heterogeneity. Second, to estimate within-country trade frictions, I project freight rates on a complete dataset of the geography and the quality of the transportation network of Peru. Finally, I use disaggregated household consumption data to estimate the elasticity of substitution across crops in demand.

Once quantified, the model allows me to calculate the effect of trade costs on productivity through relative prices, land allocation and intermediate input use.[3] In the model, when access to markets is costly, farmers pay high prices for their purchases and collect low prices for their sales. Hence, farmers have less incentive to specialize according to comparative advantage and also cut back their use of intermediate inputs. I find that a country-wide policy that paves existing major roads increases a multi-crop index of TFP in agriculture in almost every region. In the average region, for example, TFP (measured in units of intermediate inputs at the port) increases by about 14%. The overall productivity effect in a region depends on the farmers' crop choices, and their ability to reallocate land across crops in response to price changes.

I then move on to assess the welfare effects of a counterfactual policy that paves roads. For a farmer, better roads increase the price he fetches for his products, by improving his own access to domestic and international markets. But better roads also improve the access of farmers in other regions, increasing the supply of crops to domestic markets and thus decreasing their price. The total effect on the farmer's welfare depends on the increase in the price of consumption relative to income. I find that, as a result of this policy, a farmer in an average region gains 5% in terms of welfare. Nevertheless the policy can generate winners and losers. A farmer in the first quartile of the distribution of welfare changes loses 5%, while one in the third quartile gains 13%. Rural dwellers employed in the non-agricultural sector sometimes see their welfare decrease as a result of the policy, especially in areas that are remote at the baseline. But the dispersion in the welfare changes for these workers is small,

[3] A recent literature suggests there exists large differences in agricultural productivity across space, partly caused by differences in input use. Restuccia et al. (2008) estimate that value added per worker in agriculture in the top 5 percent richest countries in the world is 78 times larger than that of the bottom 5 percent.

and they usually gain from better roads.

Peru is an ideal setting for this study because its geography is diverse, and its agriculture shares features with developed and developing countries alike. It is a middle income country where a few large urban markets are often the destination of traded agricultural produce, but where some well-connected regions produce for export markets. Eighty six percent of the national highway system is unpaved, yet dirt roads coexist with modern highways. Geography also plays a major role in shaping trade patterns: the country is divided in two by the Andes, with rainforests to the east and deserts and fertile valleys to the west. Transport and geography in Peru produce large variation in access to markets, as shipping crops even between relatively close locations can be very costly.[4] Geography is also the basis for specialization based on comparative advantage, because weather and land quality vary drastically within the country.[5] And while large farms on the coast often employ modern techniques, isolated Andean and jungle regions still use traditional farming methods. Finally, about 25 percent of the labor force is employed in agriculture, much like in other developing countries.[6] Figure 1 summarizes the large variation in Peru's geography and transportation quality.

In addition to producing substantive results for Peru, this paper also makes two methodological contributions. First, it presents a theory of general equilibrium with production that connects tightly with data on land allocations and productivity, enabling the estimation of the model based solely on agricultural statistics and aggregate trade statistics, which are collected by many countries. This approach is especially useful for studying economies where much trade occurs within borders, because it sidesteps the need to use domestic trade data, often a limiting factor. Second, I obtain a simple estimating equation for the elasticity of land allocation with respect to relative prices. The estimating equation captures a basic economic intuition inherent to models where factors of production are heterogeneous: as more land is allocated to a crop, the average productivity of the land used to grow that crop decreases, with an elasticity directly related to the heterogeneity of land.

[4]For example, in 2013 a 209 kilometer (130 mile) trip from the district of Uchumarca to the district of Chachapoyas doubles the price of a kilogram of potatoes, due to freight rates alone (source Regional Direction of Agriculture, La Libertad).

[5]As noted by Escobal and Torero (2005), Peru contains 84 different climate zones.

[6]The share of labor in agriculture in developing countries ranges from 64 percent in Sub-Saharan Africa to 22 percent in Eastern Europe and Latin America (See World Bank, 2008, p.27-28).

My treatment of land as a heterogeneous factor is closely related to Costinot et al. (2015), who look at the role of international trade in mitigating the effects of global warming, and Fajgelbaum and Redding (2014), who study the link between international trade and structural transformation. Relative to them, I introduce two methodological differences. First, my model allows me to trace separately the effect of domestic trade costs on productivity through specialization and input use. Second, I close the model allowing for domestic trade in a set of homogeneous crops, in the presence of trade costs, which is a key feature of the environment I study.

In quantifying the extent of land heterogeneity, I also draw from Costinot and Donaldson (2014) and Costinot and Donaldson (2012) who combine Ricardian trade models with data on productivity from the Global Agro-Ecological Zones (GAEZ) database (IIASA/FAO (2012)). In particular, Costinot and Donaldson (2014) study the gains due to market integration experienced by the United States in the period 1880-1997. My work complements theirs by showing how to estimate the relationship between land allocation shares and potential agricultural yields in the GAEZ data set, delivering a key parameter of the model. Moreover, I provide a model for the joint determination of prices and domestic trade flows in equilibrium.

This paper also presents an alternative framework for questions related to trade in developing economics. It complements that of Donaldson (2015), whose work establishes the causal effect of transportation infrastructure on welfare, and shows how to integrate agriculture trade data to the Ricardian model of Eaton and Kortum (2002). But unlike the canonical Ricardian model, my model naturally yields predictions for land shares across crops, allowing me to make contact with land use data, which is more widely available than data on within-country trade. I am also able to decompose the effect of trade and inputs on productivity.

My work also complements two related papers that explore the interactions of trade and agricultural productivity. Tombe (2015) finds that high import barriers together with barriers to labor movement help account for poor countries' low food imports, even when their relative agricultural productivity is low compared to rich countries. Adamopoulos (2011) argues that in a two sector model, low transport productivity can distort the allocation of resources within and between sectors, leading to low productivity. In this paper, I focus on domestic trade costs and, after estimating them directly, I measure their impact on factor allocation, productivity, and incomes.

More broadly, this paper also speaks to the literature on the role of agriculture in understanding the low productivity of developing countries. Gollin et al. (2013) show that poor countries allocate larger fractions of labor to agriculture, an unproductive activity relative to other sectors.[7] Restuccia et al. (2008) document that poor countries are disproportionately unproductive in agriculture and quantify how barriers to the use of modern intermediate inputs and to the mobility of labor can account for the productivity differences observed in the data.[8] I contribute to this literature by showing that transportation technology in developing countries is a constraint that limits the use of modern inputs, and that improving this technology can lead to a more productive allocation of land and labor.

2 A Model of Specialization, Input Use, and Trade

To study the link between trade frictions, agricultural productivity, and welfare, I develop a model of factor allocation and trade based on comparative advantage. In the model, the Home country consists of many regions that differ in terms of their population and land endowment. Within a region, the quality of land to grow different crops varies across plots. In equilibrium, plots are allocated according to comparative advantage, as to maximize the region's revenue. But, on average, some regions are relatively better suited than others for growing particular crops. This source of comparative advantage fuels trade between regions and with the rest of the world, and drives patterns of specialization across regions. Thus, land heterogeneity provides two sources of comparative advantage for the allocation of land.

In contrast to land, each crop is homogeneous. To grow crops, farmers combine land with labor and an imported intermediate input. Markets are perfectly competitive, but trade across regions and with the rest of the world is costly. Trade costs impede specialization and hence diminish productivity. Regions farther away from major ports use less of the intermediate input because its price is relatively high, which also diminishes productivity.

The assumption that land is heterogeneous reflects that, in reality, the suitability

[7] They find that after correcting for measurement and data quality problems poor countries have systematically larger "agricultural productivity gaps", suggestive of misallocation of labor across sectors.

[8] Several papers offer explanations for these observed gaps, among them self-selection and workers' comparative advantage (Lagakos and Waugh (2013)), policy barriers to efficient farm size (Adamopoulos and Restuccia (2014)), etc.

of a location to grow a crop depends on the quality of the soil, altitude, weather, etc. To make contact with observed land allocations, I introduce assumptions on technology and the distribution of crop-specific land quality that ensure that land allocation adjusts smoothly with changes in crop prices and average land quality. With these assumptions, the model delivers simple, estimable equations for land allocation and revenue shares across crops.

2.1 Environment

2.1.1 Geography and Commodities

I divide the world into Home –the focus of attention– and Foreign. Home consists of regions indexed $i = 1, \ldots, I$. I denote Foreign by $i = F$. When a region is treated as an importer, I use the index n.

There are $k = 1, \ldots, K$ homogeneous agricultural goods (crops, for short). The rest of goods for consumption are summarized in a "manufactured" good, denoted by M. There is also an intermediate input x, used in agricultural production, which is imported from Foreign.

2.1.2 Agents

In each region i, there are three agents: a representative consumer, a representative producer and a representative trader. The representative consumer owns land and supplies labor. The consumer trades in local markets, where he rents his factor inputs and purchases consumption goods. The representative producer also trades in local markets, where he hires labor, rents land, and sells the output he produces. The trader in i purchases goods in i's local market, ships them to other regions in Home and sells them there. The trader can also buy and sell goods for trade between region i and Foreign.[9]

2.1.3 Preferences and Endowments

For this application, we only need to specify the preferences for consumers at Home. The consumer in region i spends a fraction b of income on an agricultural aggregate,

[9] Because this model does not produce analytical expression for trade flows, trading technologies play an essential role in the definition of the equilibrium. Traders are not a standard modeling choice, but they are quite useful to explain the equilibrium use of these technologies.

$C_{i,A}$, and the rest on manufactured goods, $C_{i,M}$:[10]

$$U_i = C_{i,A}^b C_{i,M}^{1-b}. \tag{1}$$

The agricultural aggregate is

$$C_{i,A} = \left(\sum_{k=1}^{K} a_k^{\frac{1}{\sigma}} C_{i,k}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where $\sigma > 0$ is the elasticity of substitution across agricultural goods. I normalize the weights to add up to one: $\sum_{k=1}^{K} a_k = 1$, with $a_k > 0$.

The household in region i supplies labor inelastically to agriculture, $L_{i,A}$, and to manufacturing, $L_{i,M}$.[11] The household also supplies inelastically its endowment of heterogeneous land, which consists of a continuum of plots. I denote the set of plots by Ω_i, and all plots, indexed by ω, have size one. The total amount of land in the region is $H_i = \int_{\Omega_i} \mathrm{d}\omega$.

2.1.4 Technology

I introduce two assumptions about the production function and the distribution of land quality to take the model to data. The workings of the rest of the model, however, and the definition of equilibrium are independent of these specific details.

Assumption 1. *The technology to grow crop k exhibits constant returns to scale. It combines labor, the intermediate input, and land. The suitability of plot ω in region*

[10] While this assumption is stringent, it simplifies the analysis, allowing me to attribute all income to a single representative consumer. Because Engel's Law holds in the data this simplifying assumption will miss expenditure changes induced by large changes in income. An alternative is to let b vary with level of income of a region. There is a recent literature that explains how non-homotheticity helps reconcile trade models with observations on international trade. See Fieler (2011), Markusen (2013), and Fajgelbaum and Khandelwal (2014). Moreover, Atkin (2013) shows how local abundance shapes preferences and the benefits of trade.

[11] This polar assumption captures in a simple way the findings in recent research. To rationalize the data on labor allocation across agriculture and non-agriculture requires large barriers to the movement of labor between sectors, given productivity data. See, for example, Restuccia et al. (2008); Tombe (2015); Gollin et al. (2013). Lagakos and Waugh (2013) provide an alternative explanation based on the selection of workers into agriculture, in which a large, unproductive agricultural workforce is an equilibrium outcome.

i for producing crop k is captured by an efficiency shifter $\Lambda_{i,k}(\omega) \geq 0$,

$$q_{i,k}(\omega) = (l_{i,k}(\omega))^{\alpha_k} (x_{i,k}(\omega))^{\beta_k} (\phi_{i,k}(\omega) \Lambda_{i,k}(\omega))^{\gamma_k} \tag{3}$$

where $q_{i,k}(\omega)$ is the output of crop k, $l_{i,k}(\omega)$ and $x_{i,k}(\omega)$ are labor and intermediate inputs, and $\phi_{i,k}(\omega)$ is the share of plot ω allocated to k. The cost shares α_k, β_k and γ_k vary across crops k, but $\alpha_k + \beta_k + \gamma_k = 1$, $\forall k$.

The following assumption ensures that we obtain a structural equation for the allocation of land across crops,

Assumption 2. *The vector of land qualities for producing crops in region i, plot ω, $(\Lambda_{i,k}(\omega))_k$, is distributed as a set of i.i.d Fréchet random variables with parameters $(\tilde{\gamma} A_{i,k}, \theta)$,*

$$\mathbb{P}[\Lambda_{i,k}(\omega) \leq \Lambda] = e^{-\tilde{\gamma}^{\theta} A_{i,k}^{\theta} \Lambda^{-\theta}}.$$

I normalize $\tilde{\gamma} = \left[\Gamma\left(1 - \frac{1}{\theta}\right)\right]^{-1}$. In a region where growing crop k is impossible, $\Lambda_{i,k}(\omega) = 0$ in each plot ω.

In this probabilistic representation, the parameter $A_{i,k}$, shared by all plots ω in region i, relates to the average land quality for growing crop k in that region. Thus, a high value of $A_{i,k}$ means that the land quality of every plot in the region is high for crop k. Within region i, between-plot dispersion in land quality decreases with increases in θ, an inverse measure of land heterogeneity.[12]

Manufacturing uses only labor, $l_{i,M}$, to produce a homogeneous output: $y_{i,M} = T_i l_{i,M}$, where T_i is a labor productivity coefficient.

2.2 Markets

2.2.1 Trade costs

Trade in agricultural goods is costly. I formalize this notion by introducing iceberg trade costs: for a unit of crop k to arrive from i to n, $d_{ni,k} \geq 1$ units must be shipped. I normalize $d_{nn,k} = 1$, all n, k, and $d_{ni,k} > 1$, all $n \neq i$, all k. I also assume that costs are symmetric, so $d_{ni,k} = d_{in,k}$, and I impose the triangle inequality, i.e.,

[12] Allowing for certain types of correlation across plots, within a region, does not change the results, and only requires redefining some variables. See for example Eaton and Kortum (2002), footnote 14.

$d_{ni,k} \leq d_{nj,k} \times d_{ji,k}$. The tractability of the model, however, does not hinge on the iceberg formulation, which I adopt to avoid spelling out the details of transportation sector.

Labor is immobile across regions and sectors. Manufactured goods are costlessly traded within Home, but cannot be traded between Home and Foreign.[13]

2.2.2 Prices in Domestic and International Markets

Each region at Home has local markets for land, labor, the imported agricultural intermediate and consumption goods. In region i, let $w_{i,A}$ and $w_{i,M}$ be the wages for agricultural and manufacturing labor, ρ_i the price of the intermediate input, $p_{i,k}$ the price of crop k, for all k, and let $r_i(\omega)$ denote the rental rate of plot ω. The price of the manufactured good, p_M, is the same across regions.

Any region i in Home can trade crops with Foreign at fixed prices, $p_{F,k}$. Foreign is also the only producer of the intermediate input, which costs ρ_F there.[14]

2.3 Consumer, producer and trader decisions

2.3.1 Consumers

The representative household inelastically supplies land and both kinds of labor. It uses all of its income to purchase consumption goods. The consumer's problem is, therefore, to maximize (1) subject to the budget constraint

$$\sum_{k=1}^{K} p_{i,k} C_{i,k} + p_{i,M} C_{i,M} = E_i, \tag{4}$$

where expenditure E_i is equal to the household's income from all sources

$$E_i = w_{i,A} L_{i,A} + w_{i,M} L_{i,M} + \int_{\Omega_i} r_i(\omega)\, \mathrm{d}\omega,$$

[13] The focus of the paper is on two trade relationships: farmers and domestic urban markets, and farmers and foreign markets. Therefore, we simplify the decisions of non-agricultural producers by shutting down trade with ROW, assuming that the counterfactuals will not change the propensity to trade non-agricultural products with the rest of the world. Moreover, I assume that manufacturing is relatively less affected by trade costs due to spoiling, bruising, etc.

[14] The assumption that Foreign is the only producer of intermediate inputs is a good representation of reality. In the case of Peru, between 2008 and 2011, nearly 100 percent of the fertilizer used in production was imported, according to FAOSTAT.

assuming that in equilibrium each plot ω is fully rented.

2.3.2 Producers

The representative farmer in region i rents land, hires labor and buys the imported intermediate input. He decides how to allocate plots of land across crops, and how much labor and intermediate inputs to use in each plot. Formally, the producer's problem is to choose $\{\phi_{i,k}(\omega), l_{i,k}(\omega), x_{i,k}(\omega), \omega \in \Omega_i, \text{all } k\}$, to maximize profits,

$$\max\left\{\sum_{k=1}^{K} p_{i,k}q_{i,k} - \int_{\Omega_i}\sum_{k=1}^{K}\left[w_{i,A}l_{i,k}(\omega) + \rho_i x_{i,k}(\omega) + r_i(\omega)\,\phi_{i,k}(\omega)\right]\mathrm{d}\omega\right\}, \tag{5}$$

where total output of crop k is

$$q_{i,k} = \int_{\Omega_i}\left[(l_{i,k}(\omega))^{\alpha_k}(x_{i,k}(\omega))^{\beta_k}(\phi_{i,k}(\omega)\,\Lambda_{i,k}(\omega))^{\gamma_k}\right]\mathrm{d}\omega$$

for all k. The representative manufacturing firm in region i also hires labor as to maximize profits.

2.3.3 Trader decisions

The traders in i have access to technologies for exporting to n, as well as for bilateral trade between i and ROW.[15] Therefore, their problem is to maximize the profits from domestic trade

$$\max_{\{z_{ni,k}\}_{n,k}}\sum_{n=1}^{I} z_{ni,k}\left(p_{n,k} - d_{ni,k}p_{i,k}\right)$$

as well as from international trade

$$\max_{\{z_{Fi,k},z_{iF,k}\}_k}\sum_k\left[z_{Fi,k}\left(p_{F,k} - d_{Fi,k}p_{i,k}\right) + z_{iF,k}\left(p_{i,k} - d_{iF,k}p_{F,k}\right)\right] + z_{iF,x}\left(\rho_i - d_{iF,x}\rho_F\right),$$

where $z_{ni,k}$ are domestic trade flows of good k to n from i, $z_{Fi,k}$ and $z_{iF,k}$ are international trade flows, all expressed in units of good k at the destination, and $z_{iF,x}$ are imports of intermediate inputs, expressed in units of intermediates at region i.

[15] In equilibrium, traders earn zero profits, so they do not affect demand. Their location is thus not important.

2.4 Remarks

2.4.1 Returns to scale and profits

All agricultural technologies have constant returns to scale at the plot level, and all factors are paid their marginal products, so all producers earn zero profits. The rental rate for each plot of land adjusts to ensure that this is so, absorbing the difference between total revenue and the total cost of labor and intermediate inputs. But note that land quality varies across plots, and as more land is allocated to a crop, the average quality of land used in that crop decreases. Hence, at the regional level, an increase in the amount of labor, intermediate inputs, and land allocated to the production of a crop does not increase its output in the same proportion.

The trading technology also displays constant returns to scale, and prices across regions must adjust to eliminate arbitrage opportunities. Thus, traders also earn zero profits. Producers of manufactured goods also make zero profits, for the same reason.

2.4.2 The nature of regional trade

In this model, regions trade for two reasons: productivity differences and relative factor abundance. On the one hand, as in a Ricardian model, if region i is relatively better at growing crop k, as captured by a relatively higher $A_{i,k}$ term, it will tend to produce and export that crop. This is clearest in the limiting case of $\theta \to \infty$ and $\gamma_k = \gamma \ \forall k$, which brings us to a Ricardian world with many goods and three factors. On the other hand, if for example region i is relatively abundant in land, it will tend to specialize in goods that use land intensively (high γ_k). In fact, the limiting case of $\theta \to \infty$ and $A_{i,k} = A_i \ \forall i, k$ is similar to a Heckscher-Ohlin model with many goods and three factors. On top of these forces that produce regional trade, within-region heterogeneity adds curvature to the production possibility frontier of each region, controlling how land allocations change with changes in relative prices.

2.5 Competitive Equilibrium

Regions in Home take the prices in Foreign as given, and these prices remains unchanged regardless of how much is imported or exported.

Definition 1. A competitive equilibrium consists of, for each region $i = 1, \ldots, I$: (a) prices $p_{i,k}$ for all crops k and p_M for manufactured goods; (b) wage rates $w_{i,M}$

and $w_{i,A}$, input prices ρ_i, and rental rates for land $r_i(\omega)$, for each plot $\omega \in \Omega_i$; (c) final goods consumption $C_{i,M}$ and $C_{i,k}$ for all crops k; (d) labor input $l_{i,M}$ and output $y_{i,M}$ for the manufactured good; (e) inputs $\{\phi_{i,k}(\omega), l_{i,k}(\omega), x_{i,k}(\omega), \omega \in \Omega_i\}$, and outputs $q_{i,k}$ for all crops $k = 1, \ldots, K$; (f) trade flows: (f.1) domestic $z_{ni,k}$, for all regions $n = 1, \ldots, I$ and crops $k = 1, \ldots, K$, (f.2) international $z_{Fi,k}$ and $z_{iF,k}$ for all crops $k = 1, \ldots, K$ (f.3) international $z_{iF,x}$ of the intermediate input X

such that,

1. the quantities in (c) solve the consumer's problem, given income and prices;

2. the inputs and outputs in (d) solve the manufactured goods producer's problem, given prices;

3. the inputs and outputs in (e) solve the agricultural producer's problem, given prices;

4. the agricultural goods prices in (a) are consistent with profit-maximization by traders

$$p_{n,k} \leq d_{ni,k} p_{i,k}$$

with equality if $z_{ni,k} > 0$, for all regions $n, i \in \{1, \ldots, I, F\}$, for all crops k; the intermediate input prices are

$$\rho_i = d_{iF,x} \rho_F$$

for all regions i in Home, and the law of one price holds for the manufactured good;

5. In each region, local markets clear for labor, land, and crops:

$$\begin{aligned}
L_{i,A} &= \sum_{k=1}^{K} \int_{\Omega_i} l_{i,k}(\omega)\, \mathrm{d}\omega \\
L_{i,M} &= l_{i,M} \\
1 &= \sum_{k=1}^{K} \phi_{i,k}(\omega), \quad \text{all } \omega \in \Omega_i \\
C_{i,k} &= q_{i,k} - \sum_{n \in \mathcal{W}} d_{ni,k} z_{ni,k} + \sum_{i' \in \mathcal{W}} z_{ii',k}, \quad \text{all } k = 1, \ldots, K \\
z_{iF,x} &= \sum_{k=1}^{K} \int_{\Omega_i} x_{i,k}(\omega)\, d\omega
\end{aligned}$$

6. The domestic market for manufactured goods clears:

$$\sum_{i=1}^{I} C_{i,M} = \sum_{i=1}^{I} y_{i,M}$$

7. trade with Foreign is balanced: the value of exports is equal to the value of imports

$$\sum_{k=1}^{K} p_{F,k} \sum_{i=1}^{I} \frac{z_{Fi,k}}{d_{Fi,k}} = \sum_{k=1}^{K} p_{F,k} \sum_{n=1}^{I} d_{nF,k} z_{nF,k} + \rho_F \sum_{n=1}^{I} d_{nF,x} \sum_{k=1}^{K} \int_{\Omega_n} \phi_{n,k}(\omega)\, x_{n,k}(\omega)\, \mathrm{d}\omega.$$

To complete the description of the equilibrium, we must make a choice of a numeraire. In what follows, I normalize the price of the manufactured good to one, $p_M = 1$. With this choice, the prices of the crops have a natural terms-of-trade interpretation, since they speak to the urban-rural exchange. A counterfactual change in the numeraire will reveal how infrastructure policy affects the terms of that exchange.

3 Quantitative Implications of the Model

3.1 Expenditure on each good

The solution to the representative consumer's problem is standard. Region n spends a share $ba_k \left(p_{n,k}/P_n\right)^{-(\sigma-1)}$ of income in crop k. The price of the agricultural bundle is $P_n = \left(\sum_{k=1}^{K} a_k p_{n,k}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$, and the cost of living in n is $P_n^b p_{n,M}^{1-b}$.

3.2 How to relate farmers' choices to the data

To connect the model to data on land shares, income shares and yields across crops, I start by describing the optimal behavior of the representative farmer at the plot level. This behavior is naturally represented in a probabilistic way exploiting Assumptions 1 and 2. The three propositions at the end of this section condense the model's empirical predictions, taking as given the equilibrium prices and returns to factors.

The farmer in region i seeks to maximize profits over each plot $\omega \in \Omega_i$, as shown in expression (5). As in standard trade theory, it is quite useful to work with unit cost functions to describe the farmer's choices. In doing so, we treat each plot as a separate factor, since the rental rate $r_i(\omega)$ is plot specific.

For the Cobb-Douglas production function in (3), the unit cost function, which measures the cost of producing a unit of crop k in plot ω, is:

$$c_{i,k}(\omega) = \frac{\bar{c}_k w_{i,A}^{\alpha_k} \rho_i^{\beta_k} (r_i(\omega))^{\gamma_k}}{(\Lambda_{i,k}(\omega))^{\gamma_k}},$$

where we define $\bar{c}_k \equiv \alpha_k^{-\alpha_k} \beta_k^{-\beta_k} \gamma_k^{-\gamma_k}$.[16] Profit maximization

$$\max_{q_{i,k}(\omega)} \{(p_{i,k} - c_{i,k}(\omega)) q_{i,k}(\omega)\}$$

pins down a relation between the rental rate in ω, $r(\omega)$, and the price of crop k. Let $\omega \in \Omega_{i,k}$ denote the event that ω is used to grow k. Then profit maximization dictates that conditional on $\omega \in \Omega_{i,k}$

$$p_{i,k} = c_{i,k}(\omega),$$

and in that case the equilibrium rental rate is

$$r_{i,k}(\omega) = \lambda_{i,k} \Lambda_{i,k}(\omega)$$

where we define $\lambda_{i,k} \equiv \bar{c}_k^{-1/\gamma_k} p_{i,k}^{1/\gamma_k} w_{i,A}^{-\alpha_k/\gamma_k} \rho_i^{-\beta_k/\gamma_k}$. We can interpret $\lambda_{i,k}$ as a benefit-cost index of growing k in ω, ignoring land.

The farmer behaves competitively, which ensures that all the difference between revenues and the input costs is transferred to the landowners. Therefore, a competitive farmer will choose crops such that the rental rate is the maximum that can be attained in that plot,

$$r_i(\omega) = \max_k \{\lambda_k \Lambda_{i,k}(\omega)\}.$$

Because of Assumption 2, typically only one crop maximizes profits for plot ω (although in equilibrium each plot earns the farmer zero profits.) With constant returns to scale, this implies that it is optimal to specialize each plot completely in a single crop k. Those plots where specialization is incomplete are measure zero.[17] In what follows, we use these results to characterize crop choices for a region in a probabilistic

[16] Proofs to this and other claims are contained in the Appendix.

[17] This is the same argument as one would make if there were a mass of competitive farmers bidding to rent land from a land owner. The appendix shows that the optimal behavior would be the same if we wrote the problem in such a way that farmers try to maximize the total payoff to land owners over each plot, with plots of some constant size different from one. The appendix also shows that a representation where the land owner and the farmer are the same agent yields the same behavior.

way.

Let us denote the probability that k is rent-maximizing by $\eta_{i,k}$,

$$\eta_{i,k} = \mathbb{P}\left[k = \arg\max_{l} \lambda_l \Lambda_{i,l}(\omega)\right]. \tag{6}$$

Proposition 1 derives optimal land shares for region i.

Proposition 1. *Profit maximization, together with Assumptions 1 and 2, implies that the fraction of land allocated to crop k is*

$$\eta_{i,k} = \frac{(\lambda_{i,k} A_{i,k})^{\theta}}{\Phi_i^{\theta}}, \tag{7}$$

where

$$\Phi_i = \left(\sum_{l=1}^{K} (\lambda_{i,l} A_{i,l})^{\theta}\right)^{\frac{1}{\theta}}. \tag{8}$$

Equation (7) implies that the relative land allocation between any two crops k and k' depends on input prices, which are common across crops, as well as the price $p_{i,k}$ and the land quality $A_{i,k}$ of those two crops. The prices and land qualities for all other crops are captured in, Φ_i^{θ}, the normalizing term defined in equation (8).

Equation (7) also gives the elasticity of land allocation with respect to prices. Ignoring its effect on Φ_i, a one percent increase in $p_{i,k}$ increases crop k's share of land by $\frac{\theta}{\gamma_k}$ percent. To interpret this elasticity, recall that θ is an inverse measure of land quality heterogeneity. When θ is large, land is more homogeneous, and a given increase in $p_{i,k}$ produces a larger shift in the land use pattern. The elasticity is also inversely proportional to γ_k, the output elasticity of land in the production function. A smaller value for γ_k means that land is less important compared with other inputs, and so the elasticity of marginal cost with respect to ω's rental rate is also lower. Since land rents absorb profits, a lower value of γ_k transforms a given increase in $p_{i,k}$ into a larger rental rate when growing crop k, and hence induces a larger shift in land use.

While I do not observe rental rates directly in the data, I do observe the land yield and revenue per unit of land across crops in all regions. We can compute these measures using the technology from equation (3), evaluated at the optimal input use.

The optimal physical yield is given by:

$$y_{i,k}(\omega) = \frac{\bar{c}_k^{-\frac{1}{\gamma_k}}}{\gamma_k}\left(\frac{p_{i,k}}{w_{i,A}}\right)^{\frac{\alpha_k}{\gamma_k}}\left(\frac{p_{i,k}}{\rho_i}\right)^{\frac{\beta_k}{\gamma_k}}\Lambda_{i,k}(\omega), \tag{9}$$

where the constant $\bar{c}_k$ is as defined above. Multiplying the physical yield by the crop price gives revenue per unit of land:

$$\psi_{i,k}(\omega) = \frac{\lambda_{i,k}\Lambda_{i,k}(\omega)}{\gamma_k}. \tag{10}$$

Equations (9) and (10) show that, taking prices as given, land yields and revenue per unit of land are endogenous objects that reflect the choices of farmers.[18] They are both proportional to land quality in ω – a reflection of how the demand for inputs other than land increases when land has better quality $\Lambda_{i,k}(\omega)$.

Proposition 2 below formalizes the idea that $y_{i,k}(\omega)$ and $\psi_{i,k}(\omega)$ inherit the properties of land quality, conditional on $\omega \in \Omega_{i,k}$. While it appears from equations (9) and (10) that we can infer average land quality, $A_{i,k}$, just by looking at data on physical yields or revenue per unit of land, a takeaway of this proposition is that we cannot. Instead, these data can only inform us about aggregate land quality in a region because the average quality of the land used to grow a crop is inversely related to the amount of land allocated to that crop.

Proposition 2. *A) The physical land yield of crop k, conditional on $\omega \in \Omega_{i,k}$, denoted by $y_{i,k}(\omega)|\omega \in \Omega_{i,k}$, is distributed like a Fréchet r.v, with parameters $\left(\tilde{\gamma}\gamma_k^{-1}p_{i,k}^{-1}\Phi_i, \theta\right)$.*

B) The revenue per unit of land for crop k, conditional on $\omega \in \Omega_{i,k}$, denoted by $\psi_{i,k}(\omega)|\omega \in \Omega_{i,k}$, is distributed like a Fréchet r.v., with parameters $\left(\tilde{\gamma}\gamma_k^{-1}\Phi_i, \theta\right)$.

The immediate implication is that expected yield,

$$\mathbb{E}\left[y_{i,k}(\omega)|\omega \in \Omega_{i,k}\right] = \frac{\Phi_i}{\gamma_k p_{i,k}}, \tag{11}$$

and the expected revenue per unit of land,

$$\mathbb{E}\left[\psi_{i,k}(\omega)|\omega \in \Omega_{i,k}\right] = \frac{\Phi_i}{\gamma_k}, \tag{12}$$

[18] Kelly (2006) discusses how a high price of intermediates relative to the price of final output reduces the demand for fertilizer in Africa. The World Development Report (2008) argues that transaction costs are one of the causes for low seed and fertilizer use in Sub-Saharan Africa (p. 12).

are uninformative about the relative values of $A_{i,k}$ across crops. But expected yields and revenues are precisely the objects in the model that correspond to the observations on yield and revenues. Also note that, given prices, average yields and revenues per unit of land are decreasing in the cost share of land, γ_k, reflecting that land is optimally combined with less of the other inputs.

Propositions 1 and 2 summarize how each region will adjust to differences in relative prices and relative land qualities. To illustrate, let us focus on what happens with average yields and revenues per unit of land, when the relative price of some particular crop $\hat{k}$ increases – as would be the case if it engaged in trade with a region where that crop is more expensive. Proposition 1 tells us that in region i the amount of land allocated to crop $\hat{k}$ increases, with an elasticity of $\frac{\theta}{\gamma_k}$, while the land allocated to the rest of the crops decreases. Equation (11) then guides us through the changes in physical yields. An increase in the price of crop $\hat{k}$ reduces the relative price of labor and intermediate inputs, and increases their use in production. This force, which pushes towards higher productivity, is more than offset by a decrease in average land quality: as more land is used to produce crop $\hat{k}$, the corresponding average land quality must decrease. By the same reasoning, the average land quality used in each of i's other crops must increase, while for those crops the use of labor and intermediates is not affected directly by the price change. Part A of Proposition 2 describes these changes precisely. The increase in $p_{i,\hat{k}}$ increases the aggregate productivity parameter Φ_i, thus improving the distribution of yields for all crops but $\hat{k}$. Crop $\hat{k}$'s yield actually falls, as $\frac{\Phi_i}{p_{i,\hat{k}}}$, which summarizes the effect of an increase in $p_{i,\hat{k}}$, decreases.

Having established the change in the physical yields of each crop, it is straightforward to understand the change in the revenue per unit of land. For all crops whose price did not increase, the proportional increase in revenue per unit of land is identical to the proportional increase in physical yields. For crop $\hat{k}$, the increase in the price $p_{i,\hat{k}}$, though partly countered by the decrease in crop $\hat{k}$'s physical yield, increases the revenue per unit of land. Assumptions 1 and 2 ensure that this increase is identical to that of the rest of the crops, as part B of Proposition 2 indicates.

The two propositions together show that all within-region variation in relative prices and relative land qualities only translates into observable variation in relative land allocations across crops, not into observable variation in either measure of productivity. Formally, the $A_{i,k}$ terms are included in Φ_i, which summarizes the effect of land quality on the productivity of land. Everything else constant, when region i

is more productive for some crop, or when the price that the crop commands in i increases, so does this measure of productivity.[19] The productivity distributions for all crops then shift to the right. This means that, although one might hope that observed land yields and revenues per unit of land would be informative about unobserved land quality, the model imposes the strong restriction that they are not.

Finally, as shown in the Appendix, the average rental rate of land in region i, $\mathbb{E}\left[r_i\left(\omega\right)\right]$ is proportional to Φ_i. Thus, given crop prices, land commands a higher rental rate when labor and intermediates are cheaper. When the intermediate input price ρ_i is low, farmers increase its use in production, thus increasing the output per unit of land; they also shift land use towards crops that are more intensive in the use of intermediates. A low wage $w_{i,A}$ has an analogous effect through the increased use of labor.[20]

In light of this discussion, the content of Proposition 3 is implied by Propositions 1 and 2. This result is important, however, because it provides a basis for identifying γ_k by comparing data on land shares and data on revenue shares within regions. Let $\pi_{i,k}$ be the revenue share of crop k in region i's total revenue, defined as

$$\pi_{i,k} = \frac{p_{i,k} q_{i,k}}{\sum_{k' \in \mathbb{K}_i} p_{i,k'} q_{i,k'}}.$$

Proposition 3. *Within a region, the land share and the revenue share that crop k*

[19]The variable $\Phi_i^{\frac{1}{\theta}}$ in this model is akin to the expression $\sum_i T_i \left(w_i d_{ni}\right)^{\theta}$ in Eaton and Kortum (2002), which summarizes a destination country's access to the world technology, given the cost of labor and trade costs. Donaldson (2015) exploits this object in his welfare calculations.

[20]Just like in any model of optimal resource allocation, in this model the return to land across crops is equalized at the margin, across alternative uses. But the model implies more: the average return to land is also equalized across crops. Formally, as shown in Appendix A, the rental rate of land, conditional on growing crop k, is drawn from a Fréchet distribution with parameters $(\tilde{\gamma}\Phi_i, \theta)$. The conditional mean is then:

$$\mathbb{E}\left[r_{i,k}\left(\omega\right)|\omega \in \Omega_{i,k}\right] \quad \propto \quad \Phi_i,$$

which does not depend on k. Because this expression is independent of k, it follows that the average rent is also

$$r_i \propto \Phi_i.$$

The assumption about the shape of the production function is not innocuous: for each crop k the ratio of the expected return to land, r_i, and the expected revenue per unit of land, $\mathbb{E}\left[\psi_{i,k}\left(\omega\right)|\omega \in \Omega_{i,k}\right]$, is equal to a constant, γ_k.

commands are equalized, up to a crop-specific constant

$$\pi_{i,k} = \frac{\gamma_k^{-1}\eta_{i,k}}{\sum \gamma_l^{-1}\eta_{i,l}}. \tag{13}$$

This means that a crop will have a high revenue share when (i) it has a relatively high land share, and (ii) it has a relatively low cost share of land. In reaction to, say, a change in prices, equation (12) ensures that the proportional increase in average revenue per unit of land is identical for all crops, up to γ_k, and captured by the increase in Φ_i. The change in the revenue share of crop k, therefore, is entirely driven by the change in its land allocation.[21] This is an outcome that holds at any vector of prices –in particular the equilibrium vector of prices– and is derived only from optimal farmer behavior.

Note that the farmer's economic behavior described by these propositions would be predicted by any model of optimal use of heterogeneous factors. Increasing the amount of land allocated to a given use will always decrease its productivity and will increase its productivity in alternative uses. Assumptions 1 and 2 simply put constraints on the exact amounts by which average productivity changes across alternative uses.

3.3 Aggregation, Market Access and Productivity

To close the model in general equilibrium, we must first aggregate the farmer's and the consumer's choices at the regional level, given prices. To that end, I start by studying supply and revenue in each region, which are remarkably tractable and smooth in the price of outputs. I then move on to study aggregate input demands. Finally, in the context of a discussion of the economic relation between market access and productivity, I contrast this model with a simplified model where where technologies do not differ across crops.

[21] Propositions 1, 2, and 3 have parallels in two well-known results in the Eaton and Kortum (2002) model, which predicts that the probability that a region i is region n's cheapest supplier for some good is equal to the share of region n's total expenditure on region i's goods (Proposition 3). In that model, differences in productivity across suppliers translate into differences in the fraction of goods sold in a destination (Proposition 1), keeping the distribution of unit costs of goods actually sold in region n identical across suppliers (Proposition 2). In my application both terms in Proposition 3 have empirical counterparts because the allocation of land can be measured.

3.3.1 Regional Production

Revenues from crop k can be calculated as

$$V_{i,k} \;\; = \;\; \gamma_k^{-1} \left(\lambda_{i,k} A_{i,k}\right)^{\theta} \Phi_i^{1-\theta} H_i.$$

where, recall, $\lambda_{i,k} = \bar{c}_k^{-1/\gamma_k} p_{i,k}^{1/\gamma_k} w_{i,A}^{-\alpha_k/\gamma_k} \rho_i^{-\beta_k/\gamma_k}$ measures the profitability of crop k in region i.[22] Keeping the statistic Φ_i fixed, the revenue from crop k is increasing in the price of the crop, with a constant partial elasticity of θ/γ_k, and decreasing in the price of labor and intermediates, with constant elasticities, too. We can further aggregate the revenue across crops for this region, to obtain the total value of production in agriculture:

$$V_i \;\; = \;\; \frac{\Phi_i H_i}{\bar{\gamma}_i}, \tag{14}$$

where $\bar{\gamma}_i = \sum_k \gamma_k p_{i,k} q_{i,k} / \sum_l p_{i,l} q_{i,l}$ is the revenue-weighted cost share of land in region i.

3.3.2 Regional Demand for Labor and Intermediate Inputs

The aggregate demand for labor in region i, coming from crop k is

$$l_{i,k} \;\; = \;\; \frac{\alpha_k}{\gamma_k w_{i,A}} \Phi_i \eta_{i,k} H_i$$

and, similarly, aggregate regional demand for intermediate inputs is

$$x_{i,k} = \frac{\beta_k}{\gamma_k \rho_i} \Phi_i \eta_{i,k} H_i$$

Aggreggating across crops, within region i, delivers the aggregate regional demand for these inputs

$$l_{i,A} \;\; = \;\; \frac{\Phi_i H_i}{w_{i,A}} \sum \frac{\alpha_k}{\gamma_k} \eta_{i,k} \tag{15}$$

[22] It follows that the total output of crop k in region i (measured there) is given by $q_{i,k} = \gamma_k^{-1} p_{i,k}^{-1} \left(\lambda_{i,k} A_{i,k}\right)^{\theta} \Phi_i^{1-\theta} H_i$.

and

$$x_i = \frac{\Phi_i H_i}{\rho_i} \sum_k \frac{\beta_k}{\gamma_k} \eta_{i,k}. \tag{16}$$

With crop-specific technologies, it is not possible find closed-form solutions for aggregate input demands. But it turns out we can still write these demands in a way that resembles the usual input demands associated with Cobb-Douglas technology. In fact, we can write $l_{i,A} = \bar{\alpha}_i V_i / w_{i,A}$ and $x_i = \bar{\beta}_i V_i / \rho_i$, where $\bar{\alpha}_i$, and $\bar{\beta}_i$ (the average labor and input shares) are endogenous objects that depend on the primitive cost shares of each crop and their relative uses.[23]

Because technologies are crop-specific, changes in input prices have two effects. For example, if the wage increases, the amount of land allocated to relatively labor intensive crops will go down. That effect is additional to the decrease induced by an input mix that is less labor intensive, for all crops.

3.3.3 The Economic Relation Between Trade Costs and Productivity

To gain a clear understanding of how trade costs affect allocations and productivity, I discuss a stripped down version of the model. Suppose that $\gamma_k = \gamma$, for all crops. Then Propositions 1 trough 3 above simplify quite a bit. The key distinction is that, since all crops have the same input shares, changes in the factor rewards do not affect the allocation of land across crops. Therefore, the land allocation is independent of factor rewards, and depends only on relative average land qualities and relative output prices.[24]

With this simplification, we can calculate the equilibrium revenue function in terms of endowments, crop prices, and the price of intermediate inputs:

$$V_i = \kappa_V \quad \rho_i^{-\frac{\beta}{1-\beta}} \tilde{\Phi}_i^{\frac{\gamma}{1-\beta}} H_i^{\frac{\gamma}{1-\beta}} L_{i,A}^{\frac{\alpha}{1-\beta}}. \tag{17}$$

where $\tilde{\Phi}_i^{\theta} = \sum_k \left(p_{i,k}^{1/\gamma} A_{i,k} \right)^{\theta}$. Equation (17) is the familiar revenue function. It relates the total revenue generated by region i to prices that are exogenous to the farmer and to the total stock of factors of production.

[23] Specifically, $\bar{\alpha}_i = \sum_k \alpha_k p_{i,k} q_{i,k} / \sum_l p_{i,l} q_{i,l}$ and $\bar{\beta}_i = \sum_k \beta_k p_{i,k} q_{i,k} / \sum_l p_{i,l} q_{i,l}$

[24] In this case, relative land shares are given by $\eta_{i,k}/\eta_{i,l} = \left(p_{i,k}^{\frac{1}{\gamma}} A_{i,k} \right)^{\theta} / \left(p_{i,k}^{\frac{1}{\gamma}} A_{i,k} \right)^{\theta}$. Moreover, land shares and revenue shares are equalized across crops $\pi_{i,k} = \eta_{i,k}$.

In this context, where a region produces many crops, we may measure physical productivity in each crop directly by looking at yields. But to study aggregate productivity at the regional level requires a method for aggregating consistently across crops. The multi-crop index V_i offers just such a measure. In Section 7, where I take this index to data, I express V_i in terms of units of intermediate inputs at the port, or V_i/ρ_F. This choice of units is appropriate for productivity, since it measures revenue in quantities whose value does not change in counterfactual exercises.

Equation (17) shows exactly the sense in which the coefficient $\rho_i^{-\frac{\beta}{1-\beta}}\tilde{\Phi}_i^{\frac{\gamma}{1-\beta}}$ is a measure of productivity, or TFP. Keeping the coefficient constant, the total revenue of agricultural production has constant returns to scale in land and labor. Equation (17) also shows that in location i, agricultural productivity is higher because $\tilde{\Phi}_i$ is higher (capturing, in part, better land allocations) or because the relative price of intermediates, ρ_i, is lower.[25]

In the model, variation in ρ_i is entirely driven by transportation costs: imported intermediates will be more expensive in remote places. This is the first channel through which transport costs lower productivity. The elasticity of TFP with respect to the price of the intermediate input, keeping all other prices constant, is $-\beta/(1-\beta)$, which is higher the larger the cost share of intermediates. As shown before, however, input use depends on the price of the intermediate relative to the price of output. In the exchange between the farmer and the rest of the world, trade costs increase this relative price twice: once when the farmer ships his output to the closest port and once when he brings the intermediate input back to the farm.[26]

The second channel is related to the farmers' production and consumption choices. High transport costs increase the prices of the crops that farmers purchase, and decrease the price of the crops they sell. Both effects are summarized in the value of $\tilde{\Phi}_i$. Because producers will tend to sell the goods in which they have a comparative advantage and buy those in which they do not, high transport costs will induce a negative correlation between $p_{i,k}$ and $A_{i,k}$ across k, thus lowering $\tilde{\Phi}_i$.[27]

I emphasize, however, that $\tilde{\Phi}_i$ does not exclusively measure the effect of specializa-

[25] Note that subtraction of intermediate input costs leaves a constant proportion of revenue, $(1-\beta)V_i$, so the TFP coefficient is the same.

[26] Consider the use of intermediate inputs relative to total output in region i, crop k, in the case when region i exports crop k to Foreign and obtains inputs in return. The model predicts $x_{i,k}/q_{i,k} = \beta_k p_{i,k}/\rho_i = \beta_k \frac{p_{F,k}}{\rho_F} d_{Fi,k} d_{iF,x}$. Insofar as modern intermediates increase productivity, trade costs will decrease measured productivity.

[27] In a land-only model, for an autarkic region, the elasticity of the relative price of two crops,

tion due to comparative advantage. Rather, it also reflects other factors that increase the productivity of land, but are not explicitly modeled. Thus, if the quality of land in a region doubles – keeping prices constant –, then $\tilde{\Phi}_i$ will also double, regardless of that region's access to markets. The education of the workforce, for example, or the presence of increasing returns to scale at the farm level can generate differences in $\tilde{\Phi}_i$ across regions. We return to the quantitative impact of trade frictions in Section 7.

4 Data

This section gives a brief discussion of the dataset I have assembled.[28] The first task is to match the regions and crops in the model to the data. A region i in the model corresponds to one of 194 provinces according to Peru's 2012 administrative division. A crop k is one of the top 20 crops by value of production between the years 2008 and 2011.

The main data sources I exploit are:

1. National Statistics on Agriculture: Collected by the Ministry of Agriculture of Peru at a finely disaggregated geographic level. It contains comparable data on prices, physical land yields, and land use, corresponding to $p_{i,k}$, $y_{i,k}$ and $\eta_{i,k}H_i$ in the model.

2. Global Agro-Ecological Zones (GAEZ): Estimates of the land yield that would prevail if all land in a given cell (or pixel) is entirely devoted to growing a given crop (see IIASA/FAO (2012)). Costinot and Donaldson (2014) provide a detailed discussion of these data.

3. Geography and Transportation: Georeferenced data from Peruvian Ministry of Transportation (MTC), which indicates each road's location, length, and quality (dirt, graded, or paved). Peru's road system is hyerarchically divided in three levels: National, Departmental and Neighborhood roads.

$\frac{p_k}{p_{k'}}$, with respect to their relative land qualities, $\frac{A_k}{A_{k'}}$, is $-\dfrac{\theta}{\theta+\sigma-1}$. In contrast, if a small region is integrated with the rest of the economy, then the relative price of crop k is not related endogenously to land quality A_k. Weakening the negative correlation between $p_{i,k}$ and $A_{i,k}$ that prevails in autarky increases the magnitude of Φ_i.

[28]The Appendix contains a full discussion of the data, as well as summary statistics.

4. Freight rates: A sample of freight rates between 46 pairs of districts, averaged over the years 2010-2013.[29]

5. National Household Surveys (ENAHO): A living standards survey, collected yearly by the Instituto Nacional de Estadistica e Informatica (INEI). It constains disaggregated information on expenditures, consumption quantities and unit values.

5 Connecting the Model and Agricultural Data

In this section, I explain the estimation of the model's parameters, which I obtain by comparing selected moments in the model with disaggregated data. The estimation consists of three main parts. First, I estimate the heterogeneity parameter θ by fitting the model's land allocation equation, using exogenous yield estimates from from the GAEZ project. Once I have obtained an estimate of θ, I calculate the levels of land quality, $A_{i,k}$, relying solely on Peruvian national statistics data. Second, I estimate a statistical model of transportation costs, following the approach in Donaldson (2015): I project freight rates (for a sample district pairs) on road and geography data, and estimate the cost of traversing roads of different qualities, and with different slopes. Using these estimates, and the fact that road and geography are observable for the whole country, I predict freight rates for all possible origin-destination pairs. Third, I combine expenditure household data with my previous estimates of freight rates, and observations on international crop prices, to estimate the elasticity of substitution between crops in demand.[30]

5.1 Estimation of Factor Cost Shares: γ_k, α_k and β_k

To estimate the cost shares of land, I exploit systematic differences between crop revenue shares, $\pi_{i,k}$, and land shares, $\eta_{i,k}$, within a region i. Proposition 3 shows that crops whose revenue share is systematically higher than their land share also have lower land cost shares, γ_k. Because Proposition 3 only informs us about how to

[29]At least one of the districts in the pair belongs to the department of La Libertad. The scope of the data is restricted this way because the source is the Direccion Regional de Agricultura de La Libertad. To construct the observation for each origin-destination pair, I average an unbalanced panel of monthly data. The original database contains about a thousand observations.

[30]Estimation and calibration of the other parameters is described in the Appendix.

retrieve the relative values of the land cost shares, we need additional information and a normalization to pin down their levels. Based on Dias Avila and Evenson (2010), 0.22 is appropriate for a country-wide measure of land cost shares.

Figure 2 shows the results.[31] There are sizable differences in the coefficient γ_k, as one would expect if some crops were, in fact, more land intensive than others. Fruit crops, for example, appear to have lower land intensities than grains. For simplicity, I constrain the ratio of labor and intermediate input shares to be constant across crops and equal to the average $\bar{\alpha}/\bar{\beta} = 2.5$.[32]

5.2 Estimation of θ using National Statistics and GAEZ data

The farmer's land-allocation decision is the basis for the estimation of θ. Recall that the fraction of land allocated to crop k is described by equation (7), which I repeat here

$$\eta_{i,k} = \frac{\left(\lambda_{i,k}^{\theta} A_{i,k}^{\theta}\right)^{\theta}}{\Phi_i^{\theta}} \tag{18}$$

Now we tie this expression to the data on unconditional yields produced by GAEZ, following a similar logic to that of Costinot et al. (2015). Using the model, we calculate the physical land yield given prices, but unconditional on $\omega \in \Omega_{i,k}$. That is, we calculate the yield that would be obtained using labor and land optimally, but allocating all land in region i to the production of crop k. I denote it by $\tilde{y}_{i,k}$, omitting the dependence on prices. To obtain it, we calculate the unconditional expectation of $y_{i,k}(\omega)$ in (9)

$$\tilde{y}_{i,k} \quad = \quad \gamma_k^{-1} \bar{c}_k^{-\frac{1}{\gamma_k}} w_{i,A}^{-\frac{\alpha_k}{\gamma_k}} \rho_i^{-\frac{\beta_k}{\gamma_k}} p_{i,k}^{\frac{1-\gamma_k}{\gamma_k}} A_{i,k}.$$

I assume that this object corresponds to the measures produced by the GAEZ project,

[31] In practice, this means that I impose a normalization for the revenue-weighted cost share for the whole country $\left(\sum_k \sum_i \gamma_k p_{i,k} q_{i,k}\right) / \left(\sum_k \sum_i p_{i,k} q_{i,k}\right) = 0.22$. The appendix gives further detail on the estimation. The F-statistic associated with the crop fixed effects in the regression that identifies γ_k is 213.81, largely above the cutoff at standard levels of significance. The interpretation is that there are systematic differences between land shares and revenue shares across crops, within regions.

[32] These numbers come from and Dias Avila and Evenson (2010), Table A.3a, attributing "Mechanization" and "Animal Power" to intermediate inputs, which yields $\bar{\alpha} = 0.56$ and $\bar{\beta} = 0.22$. This is somewhat different from the estimates in Hayami and Ruttan (1985), –later quoted in Restuccia et al. (2008)–, who estimate for that, for a sample of countries, the labor cost share is 0.42, the intermediate input share is 0.4 and the land share is 0.18.

although it is not an object that we would observe in equilibrium. To connect it to the data, I assume that there exist prices p_k^G, w_A^G, ρ^G that rationalize the technological assumptions used by IIASA and FAO to construct the GAEZ dataset.[33] Then we relate each observation in the GAEZ data to objects in the theory, in particular $A_{i,k}$:

$$\tilde{y}_{i,k}^G = \gamma_k^{-1} \bar{c}_k^{-\frac{1}{\gamma_k}} \left(w_A^G\right)^{-\frac{\alpha_k}{\gamma_k}} \left(\rho^G\right)^{-\frac{\beta_k}{\gamma_k}} \left(p_k^G\right)^{\frac{1-\gamma_k}{\gamma_k}} A_{i,k} e^{u_{i,k}} \tag{19}$$

where $\tilde{y}_{i,k}^G$ is the GAEZ measure of unconditional yields, and $e^{u_{i,k}}$ is a term that captures the possibility that the true $A_{i,k}$ is measured with error. Using (18) to substitute for $A_{i,k}$ in the expression for $\tilde{y}_{i,k}^G$, we obtain a relation between GAEZ yield measures and observed land allocations and prices:

$$\log\left(p_{i,k}^{\frac{1}{\gamma_k}} \tilde{y}_{i,k}^G\right) = \frac{1}{\theta} \log \eta_{i,k} + \iota_k + \iota_i + \delta_i \frac{1-\gamma_k}{\gamma_k} + u_{i,k}. \tag{20}$$

where, ι_k and ι_i are dumies that absorb components that do not vary simultaneously at the region-crop level.[34] I construct the left-hand side of equation (20) imposing the restriction that the coefficient on log-prices is $\frac{1}{\gamma_k}$. The reason is that I want to estimate as precisely as possible the coefficient of $\log \eta_{i,k}$, which is the only coefficient informative of θ in the regression.

What is the economic interpretation of this estimating equation? Suppose we observe that in region i a large fraction of land is allocated to crop k. Because farmers optimally allocate more land to crops for which land is better suited –equation (18)– we would predict that $A_{i,k}$ is relatively large, too. Using equation (19), we would then predict a large GAEZ estimate of potential productivity. But the farmers can also choose to allocate a large fraction of land to a crop when its price is high; that is why the dependent variable in the estimating equation "values" the GAEZ productivities at the equilibrium prices.

In estimating equation (20), I also take a particular stance on what is the source of error $u_{i,k}$. There is reason to believe that the GAEZ data are a noisy measure of potential productivity $\tilde{y}_{i,k}$. For example, there are regions that actually grow a crop in the national statistics, which nonetheless show zero potential productivity in the

[33]Note that I assume that the prices that rationalize the GAEZ data are independent of i. I take the stance that, although the GAEZ data set models input use as a function of input prices relative to output prices, they do not take into account the spatial variation of those relative prices.

[34]The Appendix explains in detail what variables these fixed effects absorb.

GAEZ data set. Another reason for measurement error is that the match between the GAEZ data and the Peruvian regions is not perfect.[35]

Before discussing the results, note that the model allows us to estimate θ based only on a sub-sample of goods: assuming that the Fréchet draws are i.i.d. allows us to write optimal land allocation to crop k only as a function of land quality, $A_{i,k}$, input prices and the price of crop k, together with a region-wide shifter. We do not need to take into account the prices and land qualities of all other crops.[36]

To estimate equation (20) we need data on crop prices $p_{i,k}$, land allocations $\eta_{i,k}$ and GAEZ potential productivity measures $\tilde{y}^{G}_{i,k}$. Data on $p_{i,k}$ and $\eta_{i,k}$ come from the Peruvian Ministry of Agriculture. Since this model is best thought of as a description of the medium run, I focus on a long sample of National Statistics which averages more than ten years, and contains information for four departments, at the district level.[37]

Table 1 shows the results of estimating equation (20): the coefficient on land allocation is 0.596, which implies an estimate $\hat{\theta}$= 1.68. Figure 3 shows the variation that identifies $\hat{\theta}$; it relates $\log\left(p_{i,k}^{\frac{1}{\gamma_k}}\tilde{y}^{G}_{i,k}\right)$ to $\log\eta_{i,k}$ after removing the other regressors in equation 20. The coefficient is precisely estimated; its value implies a large elasticity of land allocation with respect to prices: $\theta/\bar{\gamma}\approx\frac{1.68}{0.22}=7.63$. That is, after an exogenous one percent change in the relative price of a crop, the land share for a typical crop would increase by about 7.63 percent.[38]

[35]Since there is a continuum of plots in each region, land heterogeneity $\Lambda_{i,k}(\omega)$ does not, by itself, generate an error term. Therefore, here we assume that the theoretical object of interest is measured with noise, and hence we try to predict it with observables. The typical approach in trade assumes that the proxies for trade costs in a trade-flow equation (e.g. distance between countries) are measured correctly, but do not capture all variation in trade costs. See Head and Mayer (2013) for a detailed exposition.

[36]In a study of trade and multinational production, Ramondo and Rodriguez-Clare (2013) extend the unobserved heterogeneity approach to a multivariate Fréchet distribution.

[37]By averaging the data for each crop k and region i, I eliminate the time variation, so we cannot use it to infer θ from changes in behavior through time. Such variation, however, especially at the yearly frequency, is subject to short run fluctuations like weather shocks. Since equation (20) is essentially a supply equation – which is part of a system – short run fluctuations may raise concerns of endogeneity of $\eta_{i,k}$. Averaging over a long time series reduces these concerns.

[38]A key assumption to obtain estimating equation (20) is that $\eta_{i,k}$ is observed without error. If that is not the case, then the estimate of $1/\theta$ is subject to attenuation bias, which means that the estimate of θ is too large. Another possibility, which I leave for the future, is to exploit time-series variation to study the farmers' land allocation changes in reaction to exogenous variation in prices.

5.3 Estimation of Transportation Costs

The goal of this section is to produce an estimate of the iceberg trade costs between any two pairs of regions in Peru, and for each good in the data set. The first step is to estimate a statistical model of transport costs: I project a sample of within-country freight rates on data about the quality and geography of the road that connects each origin-destination pair in the sample. Because data on geography and road quality are available for the whole country, I then use this estimated model to predict the freight rates for all possible origin-destination pairs in Peru. The second step is to transform predicted freight rates –measured in local currency per kilogram– into iceberg trade costs by comparing them to the average farm-gate price of each crop in the data.

5.3.1 Projection of Freight Rates on Road and Geography Data

I follow Donaldson (2015) and represent the transportation network with a graph. To form the graph, I combine GIS data on (i) the exact location of the capital of each district i, (ii) a fine grid of altitude, and (iii) the shape, length and quality of the road network.

Each region i in the model corresponds to a node in the graph. The rest of the nodes represent the connections between segments in the road network. For example, if a highway splits in two, my procedure places a node at the point where the split occurs. Two nodes are connected if at least one of the two following conditions is met: (i) there is a segment of road of any quality that connects them or (ii) they are two district capitals at most 50 km. apart.[39]

I use the sample of freight rates to estimate a statistical transport cost model, which will give estimates of the relative costs of traversing roads of different qualities and with different slopes. Let f_{ni} be the observed freight rate of shipping a kilogram of goods from region i to region n. I estimate the following equation by non-linear least squares:

$$\mathbb{E}\left[\log f_{ni} | \text{geography, roads}\right] = \beta_0 + \beta_{\text{distance}} \times \log\left(\text{effective distance}_{ni}\left(\lambda\right)\right). \tag{21}$$

where $\beta_{distance}$ translates effective distance into freight rates. For a given choice of the parameter vector λ, "effective distance$_{ni}$" is the lowest-cost path between regions

[39]If there is no road, I use the straight-line distance and assign low quality to the connection.

n and i, calculated according to Dijkstra's algorithm, which minimizes the following weighted sum of distances:

$$\text{effective distance}_{ni}(\lambda; Q) = \min_{R_{ni}} \sum_{q \in Q} \sum_{\text{edge} \in E_{q,ni}(\lambda)} \left[h(\lambda_s s_{\text{edge}}) \cdot (\lambda_q \text{distance}_{\text{edge}}) \right]. \quad (22)$$

In equation (22), the effective distance between i and n is the route R_{ni} over the network, composed of the sets of edges of quality $q \in Q$, $E_{q,ni}(\lambda)$. The cost of traversing a kilometer of road of quality q is λ_q and λ_s is the effect of traversing an edge with slope s_{edge}, captured through the function $h(\cdot)$. Without loss, we normalize the weight of high-quality distance, λ_{high}, to one.[40]

In practice, I set $Q = \{\text{high, low}\}$, where only paved roads are high quality, and $h(x) = (1 + x)$. Table 2 compares two versions of the model: (i) a model that constrains $\lambda_q = 1$ and $\lambda_s = 0$, and (ii) a model that constrains only $\lambda_s = 0$. It is clear that taking into account the quality of the road substantially improves the estimation.[41] Therefore, I focus attention on the specifcation that gives a role to infrastructure, where I find that $\hat{\lambda}_{\text{low}} = 24$, and $\hat{\beta}_{\text{distance}} = 0.8$. To interpret these values, suppose that the route between two regions n and i is completely paved If that same route were unpaved, the freight cost would increase by a factor of $\left(\hat{\lambda}_{\text{low}}\right)^{\hat{\beta}_{\text{distance}}} = 24^{0.8} = 12.7$. With these estimates at hand, we can predict freight rates for the whole country.[42]

5.3.2 Transformation of Freight Rates into Iceberg Costs

Let $\hat{f}_{ni}$ be the predicted freight rate between n and i. To transform $\hat{f}_{ni}$ into an iceberg cost $\hat{d}_{ni,k}$, we divide $\hat{f}_{ni}$ by crop k's average producer price, $\bar{p}_k$, and correct

[40] Expression (22) emphasizes that the optimal road depends on the actual value of λ. The reason is that, given λ, Dijkstra's algorithm chooses among alternative ways of reaching n from i, over the network, and these choices may change with the cost vector λ. In the extreme, if $\lambda_{\text{q}} = 1 \; \forall q \in Q$, the algorithm minimizes the simple distance between two points. As λ_q increases, for $q \neq$ high, the algorithm gives priority to high-quality edges.

[41] I do not report the results of a specification with $\lambda_s \geq 0$. The reason is that for the sample at hand, little is gained by allowing $\lambda_s \geq 0$, since the standard errors on that coefficient are too large to draw any inference. Part of the problem is that the sample of freight rates is small, and does not allow to separate the effect of road quality and slope, which one would expect to be correlated.

[42] Limao and Venables (1999) find empirical evidence for the role of infrastructure as a determinant of trade costs. Donaldson (2015) estimates that transporting goods on dirt roads increases transport costs by a factor of 7.9 relative to railroads. My estimates are larger, which possibly reflects that infrastructure plays a larger role in Peru, as well as the fact that I have a small sample of freight rates (as evidenced by the large standard errors).

for a crop-independent factor δ:

$$\hat{d}_{ni,k} = 1 + \delta \frac{\hat{f}_{ni}}{\bar{p}_k}.$$

The reason for estimating a factor $\delta \geq 1$ is that, as has been discussed at large in the trade literature, actual transportation costs are only a small fraction of the total costs needed to rationalize trade flows relative to a frictionless benchmark.[43] Without it the model is unable to reproduce the variation of crop prices within the country, as equilibrium prices deviate little with respect to international prices.

For any region i in Home, the cost of trading with Foreign has two components. The first is captured by the cost of trading with the closest international port. To find the closest international port, I select the three main sea ports by value and find the closest to region i according to the predicted freight rate $\hat{f}_{o(i)i}$, where $o\left(i\right)$ is the port closest to i. The second component, denoted by τ below, is a barrier that prevents the traders from realizing the full price of goods at the port. For exporters, this component may capture that goods require additional packaging to be shipped abroad. For importers, it may reflect other costs associated with getting the goods out of the port and into the roads. Thus, I compute for each good, including the intermediate input:

$$\hat{d}_{iF,k} = \left(1 + \tau\right)\left(1 + \frac{\delta \hat{f}_{o(i)i}}{\bar{p}_k}\right).$$

The calculation of iceberg costs delivers a unitless estimate, as $\bar{p}_k$ and $\hat{f}_{ni}$ are measured in units of currency per weight. Note that the freight rate is constant across crops for a given origin-destination pair, and therefore the iceberg trade cost is inversely related to the observed price of the crop. This captures the fact that goods with higher value to weight are more likely to be traded.[44]

I choose δ and τ to help the model replicate, as best as possible, the variation in farm-gate prices observed in the data. In a coarse grid search, values of $\delta = 2.5$ and $\tau = 0.5$ maximize the correlation between the equilibrium farm-gate prices produced by the model, and those observed in the data. Table 3 summarizes the distribution of $\hat{d}_{ni,k}$, by crop k, pulling together across all pairs of domestic regions. It shows that

[43]For example, Chaney (2011) has recently explored the implications of networks in trade. Allen (2014) has shown that information costs are an important part of total trade costs.

[44]See Hummels and Skiba (2004)

there is substantial variation across regions and crops (with averages not unlike those reported, in a different context, by Anderson and van Wincoop (2004))

5.4 Land Quality Parameters, $A_{i,k}$

To estimate the $A_{i,k}$ parameters I rely heavily on the model structure. Assumption 2 imposes that the only way to learn about the relative values of the parameters $A_{i,k}$ is by comparing land allocations across crops, within a region. In contrast, data on revenue per unit of land and physical yields are informative about the common component of all $A_{i,k}$ within a region.

My approach, which extracts the model parameters using data on the endogenous variables, is an alternative to the use of external measures of productivity.[45] Costinot et al. (2015) and Costinot and Donaldson (2014), for example, use directly the potential quality measures produced by the GAEZ project. This method has the benefit that the productivity measures are independent of the model, insofar as the researcher only needs to choose how to interpret the productivity data. Its main shortcoming is that, although constructed with extreme care, the GAEZ measures are an imperfect measure of actual land quality. For my application, there is an additional complication: GAEZ does not estimate potential productivity data for some goods that are important in my database.

My procedure follows two steps. First, I construct the value of the productivity index Φ_i at the baseline, using equation (14):

$$\Phi_i = \frac{\bar{\gamma}_i V_i}{H_i}, \tag{23}$$

which uses the aggregation properties of the model to infer region i's aggregate land productivity from data on its land share of income, total value of production, and land endowment. In the second step, I combine this common component of land

[45]The idea of combining the model structure with data on endogenous variables to estimate the model's primitives has many antecedents in the trade literature. Waugh (2010) discusses the correct econometric specification of the gravity equation in Eaton and Kortum (2002), and Levchenko and Zhang (2011) refine this technique in a many-sector, many-factor model. Using a different model, Anderson and Yotov (2010) show also how to use the model structure to recover model parameters. See Head and Mayer (2013) for an evaluation of recent approaches to the estimation of gravity equations in trade. Closer to my approach, Costinot et al. (2012) exploit the Eaton-Kortum model to relate bilateral trade flows to the level of productivity in the source country. In a growth and talent allocation context, Hsieh et al. (2013) back out the frictions to the allocation of labor across occupations using the structure that a Fréchet distribution for unobserved talent imposes.

quality for all crops, with data on prices and land allocations across crops. Using equation (7) to solve for $A_{i,k}$ we obtain:[46]

$$A_{i,k} = \eta_{i,k}^{\frac{1}{\theta}} \frac{\Phi_i}{\lambda_{i,k}}. \tag{24}$$

We can take these expressions to data because $p_{i,k}$, $\eta_{i,k}$, V_i and $L_{i,A}$ are measured directly, and equation (23) tells us how to measure Φ_i with the regional aggregates.

Let us take a moment to interpret this equation. The statistic $\hat{\Phi}_i$ shifts all estimates $\hat{A}_{i,k}$ proportionally, based on how much output is produced in i, compared to its endowments. A large value of $\eta_{i,k}$ requires a higher land quality for crop k, relative to the other crops, to rationalize it. But we must also net out the effect of the profitability of growing that crop in i, $\lambda_{i,k}$, which also tends to generate a large land allocation to crop k.[47]

5.5 Estimation of Domestic Demand Elasticity

To estimate the elasticity of substitution across crops, σ, I bring in a new dataset with detailed information of household expenditures and the prices they pay.[48] The data come form a living-standard survey called called Encuesta Nacional de Hogares (ENAHO). The survey samples, for each year, many households for many regions i. I treat each household as randomly sampled from the model, and match its consumption to the goods k used in the simulation.

Consistent with the model of demand, I estimate β^{ENAHO} in

$$\log\left(\text{share}_{i,k,t,h}^{ENAHO}\right) = \iota_k + \iota_h + \iota_t + \beta^{ENAHO} \log v_{i,k,t,h}^{ENAHO} + \epsilon_{i,k,t,h}^{ENAHO}$$

where $\text{share}_{i,k,t,h}^{ENAHO}$ is the expenditure share and $v_{i,k,t,h}$ is the unit value (expenditure divided by quantity) of crop k for household h at time t in region i. In this regression, the coefficient on the unit values, β^{ENAHO}, is an estimate of $-(\sigma-1)$ after controlling for fixed effects for household, crop and time. The error $\epsilon_{i,k,t,h}^{ENAHO}$ reflects

[46]The appendix explains how to construct the values of $\lambda_{i,k}$ at the baseline.

[47]The estimation of $A_{i,k}$ is not free of error; the observations for $p_{i,k}$, $\eta_{i,k}$, and the aggregate variables used to infer Φ_i are themselves estimates, just like the values of θ and γ_k. Even if the model is correct, we are ignoring the sampling variation and hope for an unbiased estimate of $A_{i,k}$, at best.

[48]Appendix G discusses the estimation of expenditure shares α_k.

that expenditure in each category and physical quantities are measured with error. Moreover, since unit values are often not measured directly, but rather as the ratio of expenditure and quantity, these measurement errors create a bias concern.[49]

For this reason I implement a simple IV strategy based on the assumption that international prices and transportation costs are orthogonal to the error $\epsilon^{ENAHO}_{i,k,t,h}$. In particular, I instrument $\log v_{i,k,t,h}$ with

$$Z_{i,k} = \begin{cases} \log\left(p_{F,k} + f_{i,k}\right), & \text{if } \eta_{i,k} = 0 \\ \log\left(p_{F,k} - f_{i,k}\right), & \text{if } \eta_{i,k} > 0 \end{cases}$$

This instrument captures a simple idea. On the one hand, if $\eta_{i,k} = 0$, the region cannot produce crop k so, unless its importing costs are too high, the supply of the crop in question will be affected by the price of delivering the crop from abroad, approximated by, $p_{F,k} + f_{i,k}$. On the other hand, if $\eta_{i,k} > 0$, region i produces some amount of crop k, and, provided trade costs are not too high, will export it, so the price will be close to $p_{F,k} - f_{i,k}$. Combining the information on trade costs and international prices is crucial to generate enough variation in the data such that household and crop fixed can be included in the estimation. Results are similar when using region, instead of household, fixed effects.

Table 4 displays the estimate, $\hat{\beta}^{ENAHO}$ using OLS and an IV strategy based on the instrument just discussed. The observations that I can match between ENAHO and the National Statistics on agriculture account for 8.3% of total household expenditure in the survey. The results are consistent with a story where there observed expenditure shares and unit values are correlated, and where the instruments are orthogonal to the errors. In the first stage (results not shown), the coefficient of the instrument $Z_{i,k}$ is .307 (.004). Based on the second stage, I set $\hat{\sigma} = 2.64$.[50]

6 Baseline Simulation

In this section, I discuss how to simulate the model and compare its predictions to the data. The key numerical challenge in solving the model is to find the cheapest

[49] See Deaton (1997)

[50] A value of $\sigma = 2.6$ seems to be on the higher end of plausible values, as compared, for example, with Behrman and Deolalikar (1989), who estimate 1.2 for the elasticity of substitution between broad food groups, at low levels of income.

exporter for each importing region, because in the equilibrium some pairs of regions may not trade at all.[51] Hence, the simplest representation of the equilibrium includes prices, trade flows and a set of complementarity constraints arising from the traders' problem.[52]

The rest of the section discusses the model's fit at the baseline. As shown below, the model does a good job of capturing variation in the price and land allocation data.

6.1 A Simplified Representation of Equilibrium

To explain the basic structure of the model and to discuss how to solve it, I now show that the model's simplest representation corresponds to a standard general equilibrium model with linear production technologies. The first step is to redefine the commodity space, which will now "stack" the crops in each region, the manufactured good, and the agricultural labor in each region. Thus, for example, the vector of prices is

$$\mathbf{p} = \left[\begin{array}{ccccccccccccccc} p_{1,1} & \dots & p_{1,K} & \dots & p_{I,1} & \dots & p_{I,K} & p_M & w_{1,A} & \dots & w_{I,A} & \rho_1 & \dots & \rho_I \end{array} \right]^T .$$

Likewise, one can define the corresponding vectors for market demands, $\mathbf{C}$, supplies, $\mathbf{Q}$, and endowments, $\mathbf{L}$, of each commodity. The vector of excess demands is then $\mathbf{Z} = \mathbf{C} - (\mathbf{Q} + \mathbf{L})$. All these vectors contains functions of $\mathbf{p}$. Note that land plots ω have effectively disappeared from the commodity space in this formulation. The reason is that the probabilistic model of heterogeneity allows us to embed optimal land allocations in the expressions for crop supplies contained in $\mathbf{Q}$.

Next, trading possibilities are captured in a matrix of linear technologies. Techniques for trading with Foreign produce crops using other crops. For example, if

[51]See the appendix for more details. The absence of trade linkages between pairs of countries is an established fact in the data that quantitative trade models often cannot generate (see Helpman et al. (2008) and Eaton et al. (2012)).

[52]This approach differs from the standard in quantitative trade models, which usually exploit the the assumption – due to Armington – that goods are differentiated by country of origin, or boil down the model to one of pure exchange in factors of production (Alvarez and Lucas (2007).) See Shoven and Whalley (1992), p. 81, for a discussion of the role of the Armington assumption in CGE trade modeling. Arkolakis et al. (2012) have shown that, despite their richer microeconomic underpinnings, the newer models have a similar general equilibrium structure to that of the earlier CGE models.

region i wants to import good l by exporting good k, then one unit of l delivered in i, requires $d_{Fi,k}d_{iF,l}p_{F,l}/p_{F,k}$ units of crop k. Domestic trade simply transforms crop k in region i into crop k in region n, with unit requirement $d_{ni,k}$. Denote the matrix of such techniques by $\mathbf{B}$, which is not a function of $\mathbf{p}$.

It becomes clear that both prices $\mathbf{p}$ and trade flows $\mathbf{s}$ need to be part of the equilibrium definition. In fact, an equilibrium is now a set of prices $\mathbf{p}$ and activity levels $\mathbf{s}$ such that (i) Markets clear $\mathbf{Z}(\mathbf{p}) = \mathbf{Bs}$ and (ii) Traders make non-negative profits $\mathbf{pB} \leq 0$, with profits equal to zero if the activity level is strictly positive. Since alternative trading techniques can "produce" the same good, one should expect that many of those techniques go unused. A zero in the vector $\mathbf{s}$ corresponds to a zero in the bilateral trade matrix, so often regions will not trade.

Solving the model requires finding $\mathbf{p}$ and $\mathbf{s}$ such that (i) and (ii) above hold. Walras Law implies that one market clearing condition can be dropped to choose a normalization. In practice, computation of the solution is not trivial and the choice of normalization affects the speed of convergence. Choosing $p_M = 1$ seems to allow the solver to find the solution quickest.[53] Note also that, with condition (ii), it is not obvious how to use excess demands to produce a mapping, should one decide to construct an iterative algorithm from scratch.

6.2 Fitting Farm-Gate Prices and Land Allocations

I start by comparing equilibrium crop prices, $p_{i,k}$, in the model and in the data. Since the national statistics database focuses on producers, we only observe farm-gate prices for the crops that are being produced in a given region. This is important because, to the extent that the model predicts a higher price in a non-producing region –which we would expect in reality–, we will not be able to use such price variation in the following comparisons. All we can examine is whether the model can replicate producer prices, which is a tougher test of the model's performance.

The left panel of Figure 4 shows the relation between price data and predictions, combining between-region and between-crop variation (in a log scale). There is a clear, positive relation between model and data. We would expect the model to be able to capture this relation because, when we estimated the $A_{i,k}$ parameters, we

[53] In this version of the paper, I use KNITRO as a solver. Solving the model for a given set of parameters usually takes between 30 minutes and 1 hour. Alternative normalizations, like $\sum_i \mathbf{p}_i = 1$, take much longer.

allowed them to capture the crop-specific variation in prices. International prices, however, also exert an effect. A within-crop regression of log prices in the data on log prices in the model – which measures the spatial dispersion of prices, ignoring the fact that some crops are pricier on average – shows a coefficent of .93 and an R^2 of .36.

The right panel of Figure 4 compares land allocation $\eta_{i,k}$ in the data and model (in a log scale). The relation is noisier, but the predicted land shares cluster clearly around the 45 degree line, especially for larger land shares. A pooled regression of log shares in the data on log shares in the model gives a coefficient of .28 and an R^2 of .18.

There are a few reasons why the model does not fit the data perfectly. For one thing, the elasticity of land allocation with respect to prices is relatively large, $\theta/\bar{\gamma} \approx \frac{1.68}{0.22} = 7.63$; for another, many parameters, like $d_{ni,k}$ and $A_{i,k}$, are just estimates, subject to sampling variation. Finally, actual cropping choices may also reflect the farmers' intention to diversify risk, so the model may predict too much specialization.

6.3 Non-analytic gravity

Although even simple versions of this heterogenous-land model do not produce an analytical gravity equation, the model does generate a numerical one, at least as it relates trade flows and trade barriers. Unfortunately, I do not observe domestic trade flows in the data, which makes it impossible to use the empirical gravity relation to test the adequacy of the theory. But given that gravity is one of the most successful empirical relations in all of economics, one would expect that this relation applies also in this environment.[54]

A standard log regression of the model's predicted trade flows, $p_{n,k}z_{ni,k}$, on the effective distance measure that was an input in the solution of the model, yields a coefficient of -0.9.[55] Moreover, the elasticity of trade flows with respect to trade barriers is equal to -1.94. I both cases, the relationship is approximately log-linear with substantial dispersion around the regression mean. In this context, such dispersion

[54] Current trade workhorses such as Eaton and Kortum (2002), quantitative versions of Melitz (2003), like Chaney (2008), or the Armington version of the model in this paper, due to Costinot et al. (2015) do produce analytical gravity equations. For an up to date discussion of theory and empirics of gravity equations, see Head and Mayer (2013)

[55] After controlling for origin, destination and crop fixed effects. The value is a bit below -1, around which many estimates seem to cluster. See Head and Mayer (2013).

around the regression line is a sign that gravity is not a perfect description of the data, instead of an indication that that $z_{ni,k}$ or $d_{ni,k}$ are measured incompletely or with error.[56]

7 The Effect of Improving Market Access

What are the effects of large-scale market access policies in this environment? I compute a counterfactual equilibrium where I simulate the paving of all departmental highways (recall, second in the hierarchy). The extent of market access improvement is governed by the estimates of the transport cost model in Section 5. Those estimates show what is the reduction in freight rates that occur, for example, if a dirt road is paved. These policy counterfactuals are also a way of showing the model at work.[57]

The policy requires paving approximately 33,000 km. of roads. At the baseline, approximately 11,000 km. are dirt roads, while the rest is graded. This policy has a pervasive but disparate impact on the cost of trading. Pooling across all crops, for example, there is a median reduction of 9.4% in iceberg costs. The distribution of trade cost reductions is skewed to the right, with a mean of 10.5% and a standard deviation of 4.08%. The effect is asymmetric and trade costs that, at the baseline, were relatively low are essentially unaffected because they were generated by traversing high quality highways to begin with.

7.1 Productivity

The effect on productivity depends on the units in which we express the multi-crop revenue function. As discussed in Section 3, for thinking about productivity, it is convenient to choose a unit of account whose value does not change as a result of the

[56] These results are in line with Deardorff (1998), who constructs special versions of a neoclassical model of trade that deliver empirical gravity relations. Provided that patterns of specialization in production are stable, bilateral trade flows are analytically determined by his model, and the elasticity of trade flows with respect to trade barriers is equal to $-(\sigma-1)$. Simulations not included in the paper confirm that, in my model, trade flows decrease when trade barriers increase, with a strength directly related to σ.

[57] In the wake of Eaton and Kortum (2002) and Dekle et al. (2008), much recent work has used models whose equilibrium predictions can be matched exactly to data on trade shares, and has then exploited the model's analytical properties to evaluate policy counterfactuals. A main benefit of this strategy is that it circumvents the need to estimate most of the underlying model parameters. For example, see Caliendo and Parro (2015), Parro (2013), Ossa (2014). The strategy is not applicable here.

policy. Therefore, I express revenue in terms of units of the intermediate input at the port.

To build intuition, I start by analyzing the multi-crop index of productivity defined in equation (17), which measures the amount of intermediate inputs at at the port, that can be produced with one unit of the land and labor bundle $H_i^{\frac{\gamma}{1-\beta}} L_{i,A}^{\frac{\alpha}{1-\beta}}$. A first order approximation is useful to understand the effect of the policy and why one needs to simulate the model. To a first order, the log change in $\rho_i^{-\frac{\beta}{1-\beta}} \tilde{\Phi}_i^{\frac{\gamma}{1-\beta}}$ is

$$-\frac{\beta}{1-\beta} d \log \rho_i + \frac{1}{1-\beta} \sum_k \eta_{i,k} d \log p_{i,k}.$$

This formula is useful in special cases. Suppose a region exports all its output to the rest of the world, while it imports the intermediate input, with iceberg costs independent of the good. In that case, the approximation will yield an increase equal to the absolute value of $(1+\beta)/(1-\beta)\, d \log d$, in the same order of magnitude of the change in trade costs. This is just the envelope theorem at work: to a first order, changes in prices do not change factor allocations nor revenue shares, which are used to weigh the price changes across crops (see Costinot and Vogel (2014)).

But recall that the model is substantially richer. Each region exports and imports some goods, and equilibrium prices reflect more than just the prices at international markets. Crop-specific technologies make it impossible to compute a simple index like in equation (17). Finally, for some regions the changes in trade costs are large enough that a first order approximation will miss the effect of land reallocation.[58]

Figure 5 presents the distribution of productivity gains, which we can measure as $(V_i^{\text{counterfactual}} / V_i^{\text{baseline}} - 1) \times 100$. Since the amound of land and labor supplied to the sector does not change, the measure reflects change in the efficiency with which factors are transformed into revenue. This policy has large effects and the vast majority of regions experience productivity increases. The median region's productivity increases by 9%, whereas a region in the 90th percentile sees its productivity increase

[58] One can show that in the model with homogeneous technologies, the second order approximation to the percent change in $\tilde{\Phi}_i$ is

$$\frac{1}{\gamma} d\boldsymbol{\eta}^T d \log \mathbf{p} + \theta \frac{1}{2\gamma^2} d \log \mathbf{p}^T M(\boldsymbol{\eta})\, d \log \mathbf{p} + \mathbf{K}(\boldsymbol{\eta}, \gamma).$$

This expression shows that, conditional on the baseline land shares $\boldsymbol{\eta}$, the second order effect of the change in prices is indexed by θ. That is, if there is less heterogeneity, farmers can take advantage of the changes in prices by reallocating land towards pricier crops.

by as much as 33%.[59]

7.2 Welfare

But in calculating the benefits of this policy, one must also look for the changes in real income across regions and types of workers.[60] At the baseline, the inter-quartile range of log real income is 1.58. Figure 6 shows that, in terms of welfare, this policy generates winners and losers among farmers. As one would expect, welfare gains for farmers tend to increase with how much their trade costs decline. The farmers hurt by the policy are mostly those who were initially well connected.

The overall welfare effect for farmers reflects two forces. On the one hand, the policy increases the supply of food relative to manufactured goods everywhere, thus decreasing food's relative price, P_i/p_M. On the other hand, the relative prices of crops within the food category also adjust. Consider how the increased supply of domestic crops to urban markets affects farmers differently, depending on how connected they were at the baseline. If the farmer was initially well connected, he was fetching higher prices in urban markets, and therefore the policy decreases the relative price of his output. In contrast, for a farmer in a remote region, the decrease in trade costs pushes up the price he collects. If a farmer can specialize in cash crops, however, the shift in domestic supply is not as important, as he can export his output. How much each farmer benefits from changing crop prices depends on the strength of comparative advantage, governed by θ.

To gain a more concrete understanding of the effects of the policy, let us compare the regions at the 10th, 25th, 75th and 90th percentiles of the distribution of farmer welfare changes induced by the policy. Figure 7 plots, for each one of these regions, the change in land allocation and change in crop prices (now relative to the numeraire, the manufactured good), as a function of the land allocation at the baseline. The region at the 10th percentile (8% welfare loss) experiences drops in all crop prices, which is the bulk of the net effect. It was quite specialized to begin with, and has no room to maneuver. In contrast, the region at the 25th percentile (5% welfare loss) is able to repurpose its land to crops whose price increases. The experience of the 90th

[59]Note that these measurements are different from those reported, for example, in Restuccia et al. (2008). The FAO statistics used by them evaluate quantities across countries at a fixed set of prices.

[60]Since preferences are homothetic in the model, we can study separately the changes in welfare associated with farming and non-farming labor.

percentile and 75th percentile regions (21% and 13% welfare gains) illustrate the joint effect of better prices and improved allocation.

Although the model is simple in terms of the division of labor between agriculture and non-agriculture, it captures something essential: in rural areas, there are large fractions of the population that are net food buyers. A simple way of assessing the impact of the policy on net food buyers is to calculate the change in welfare of the non-farming population. Although the changes are relatively small, Figure 6 shows that in places where the price of food increases due to better access to markets, net food buyers can be harmed.[61] Not surprisingly, it is precisely the workers in the regions that saw the larger improvements in market access the ones that are harmed the most.

8 Conclusion

I estimate the effect of domestic trade costs on agricultural productivity and incomes by connecting data on Peruvian agriculture with a model of trade and specialization. The model makes predictions about land allocations across crops within a region and gives precise indications of how to interpret other geographically disaggregated data on crop prices and yields.

The message of the paper is that barriers to market access have a negative effect on farmers' productivity. Large-scale infrastructure policy, however, can generate winners and losers if crops are substitutable and there are barriers to the movements of factors, in spite of the possibility of reallocating land according to comparative advantage.

The framework I have presented can be further developed to answer other questions at the intersection of development and trade. For example, since Engel's Law is a prominent feature in consumption data, augmenting the model in this paper to incorporate non-homothetic preferences can help shed light on how income inequality shapes the urban-rural exchange. Also, I have estimated large geographic dispersion in productivity and welfare within Peru. How can this productivity dispersion persist in the long run, especially if people can move within a country? This paper suggests that there is a payoff to frameworks that jointly explain barriers to trade and barriers

[61] The quartiles of the distribution of welfare gains for manufacturing workers are 1.3%, 2.1%, and 2.4%

to labor mobility from less to more productive farming pursuits.

References

Adamopoulos, T. (2011): "Transportation Costs, Agricultural Productivity, And Cross-Country Income Differences," *International Economic Review*, 52.

Adamopoulos, T. and D. Restuccia (2014): "The Size Distribution of Farms and International Productivity Differences," *American Economic Review*, 104, 1667–97.

Allen, T. (2014): "Information Frictions in Trade," *Econometrica*, 82, 2041–2083.

Alvarez, F. and R. J. Lucas (2007): "General equilibrium analysis of the Eaton-Kortum model of international trade," *Journal of Monetary Economics*, 54, 1726–1768.

Anderson, J. E. and E. van Wincoop (2004): "Trade Costs," *Journal of Economic Literature*, 42, 691–751.

Anderson, J. E. and Y. V. Yotov (2010): "The Changing Incidence of Geography," *American Economic Review*, 100, 2157–86.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012): "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.

Atkin, D. (2013): "Trade, Tastes, and Nutrition in India," *American Economic Review*, 103, 1629–63.

Behrman, J. R. and A. B. Deolalikar (1989): "Is Variety the Spice of Life? Implications for Calorie Intake," *The Review of Economics and Statistics*, 71, 666–72.

Caliendo, L. and F. Parro (2015): "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies*, Forthcoming.

Chaney, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98, 1707–21.

——— (2011): "The Network Structure of International Trade," NBER Working Papers 16753, National Bureau of Economic Research, Inc.

Costinot, A. and D. Donaldson (2012): "Ricardo's Theory of Comparative Advantage: Old Idea, New Evidence," *American Economic Review, Papers and Proceedings*, 102, 453–58.

——— (2014): "How Large are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-1997," .

Costinot, A., D. Donaldson, and I. Komunjer (2012): "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies*, 79, 581–608.

Costinot, A., D. Donaldson, and C. Smith (2015): "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World," *Journal of Political Economy*, Forthcoming.

Costinot, A. and J. Vogel (2014): "Beyond Ricardo: Assignment Models in International Trade," Working Paper 20585, National Bureau of Economic Research.

Deardorff, A. (1998): "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in *The Regionalization of the World Economy*, National Bureau of Economic Research, Inc, NBER Chapters, 7–32.

Deaton, A. (1997): *The analysis of household surveys: a microeconometric approach to development policy*, The World Bank.

Dekle, R., J. Eaton, and S. Kortum (2008): "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, 55, 511–540.

Dias Avila, A. F. and R. E. Evenson (2010): *Total Factor Productivity Growth in Agriculture: The Role of Technological Capital*, Elsevier, vol. 4 of *Handbook of Agricultural Economics*, chap. 72, 3769–3822.

Donaldson, D. (2015): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, Forthcoming.

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

Eaton, J., S. S. Kortum, and S. Sotelo (2012): "International Trade: Linking Micro and Macro," NBER Working Papers 17864, National Bureau of Economic Research, Inc.

Escobal, J. and M. Torero (2005): "Measuring the Impact of Asset Complementarities: The Case of Rural Peru," *Cuadernos de Economia*, 42, 137–164.

Fajgelbaum, P. and S. Redding (2014): "External Integration and Internal Development: Evidence from Argentina, 1870-1914," NBER Working Papers 20217, National Bureau of Economic Research, Inc.

Fajgelbaum, P. D. and A. K. Khandelwal (2014): "Measuring the Unequal Gains from Trade," Working Paper 20331, National Bureau of Economic Research.

Fieler, A. C. (2011): "Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation," *Econometrica*, 79, 1069–1101.

Gollin, D., D. Lagakos, and M. E. Waugh (2013): "The Agricultural Productivity Gap," *The Quarterly Journal of Economics*.

Hayami, Y. and V. Ruttan (1985): *Agricultural development: an international perspective*, Johns Hopkins studies in development, Johns Hopkins University Press.

Head, K. and T. Mayer (2013): "Gravity Equations: Workhorse, Toolkit, and Cookbook," CEPR Discussion Papers 9322, C.E.P.R. Discussion Papers.

Helpman, E., M. Melitz, and Y. Rubinstein (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 123, 441–487.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2013): "The Allocation of Talent and U.S. Economic Growth," NBER Working Papers 18693, National Bureau of Economic Research, Inc.

Hummels, D. and A. Skiba (2004): "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture," *Journal of Political Economy*, 112, 1384–1402.

IIASA/FAO (2012): *Global Agro-ecological Zones (GAEZ v3.0)*, IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Lagakos, D. and M. E. Waugh (2013): "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, 103, 948–80.

Levchenko, A. A. and J. Zhang (2011): "The Evolution of Comparative Advantage: Measurement and Welfare Implications," NBER Working Papers 16806, National Bureau of Economic Research, Inc.

Limao, N. and A. J. Venables (1999): "Infrastructure, geographical disadvantage, and transport costs," Policy Research Working Paper Series 2257, The World Bank.

Markusen, J. R. (2013): "Putting per-capita income back into trade theory," *Journal of International Economics*, 90, 255–265.

Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

Mesquita Moreira, M., J. Blyde, C. Volpe, and D. Molina, eds. (2013): *Too far to export: domestic transport cost and regional export disparities in Latin America and the Caribbean*, Inter-American Development Bank.

Ossa, R. (2014): "Trade Wars and Trade Talks with Data," *American Economic Review*, 104, 4104–46.

Parro, F. (2013): "Capital-Skill Complementarity and the Skill Premium in a Quantitative Model of Trade," *American Economic Journal: Macroeconomics*, 5, 72–117.

Ramondo, N. and A. Rodriguez-Clare (2013): "Trade, Multinational Production, and the Gains from Openness," *Journal of Political Economy*, 121, 273 – 322.

Restuccia, D., D. T. Yang, and X. Zhu (2008): "Agriculture and aggregate productivity: A quantitative cross-country analysis," *Journal of Monetary Economics*, 55, 234–250.

Shoven, J. and J. Whalley (1992): *Applying General Equilibrium*, Cambridge University Press.

The World Bank (2007): *World Development Report 2008 : Agriculture for Development*, The World Bank.

Tombe, T. (2015): "The Missing Food Problem," *American Economic Journal: Macroeconomics*, 7, 1–33.

Waugh, M. E. (2010): "International Trade and Income Differences," *American Economic Review*, 100, 2093–2124.

9 Tables and Figures

Table 1: Estimation of Inverse Heterogeneity θ

	(1)
	$\log \tilde{y}_{ik} p_{ik}^{1/\gamma_k}$
$\log \eta_{ik}$	0.596^{***}
	(0.0969)
Crop FE	Yes
Region FE	Yes
Region x $\frac{1-\gamma_k}{\gamma_k}$	Yes
Observations	574
Adjusted R^2	0.826

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Estimates of the Transportation Model

	Constrained Model	Road Quality Model
log effective distance β_{dist}	0.209	0.831
	(0.110)	(0.211)
high quality λ_{hi}	1.000	1.000
	—	—
low quality λ_{lo}	1.000	23.919
	—	(16.503)
Intercept β_0	-3.139	-7.812
	(0.507)	(1.456)
N	46	46
R-squared	0.499	0.786

Bootstrapped standard errors in parentheses

Table 3: Summary Statistics of the Estimates of Iceberg Trade Costs, $\hat{d}_{ni,k}$

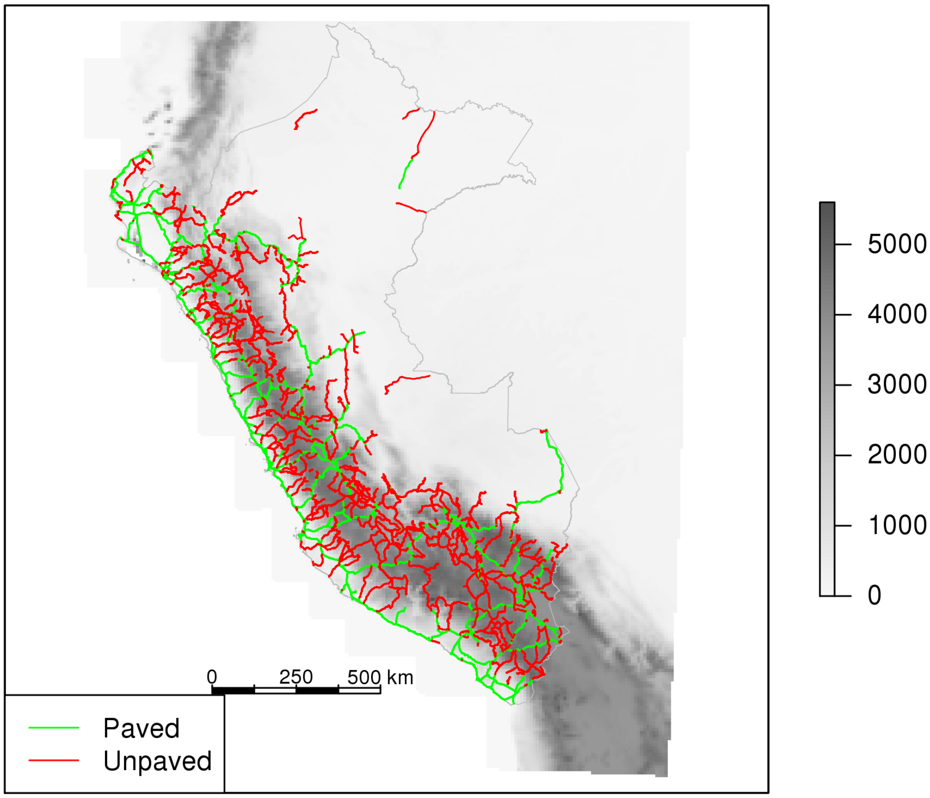
	mean	std	10%	25%	50%	75%	90%	max
asparagus	1.36	0.23	1.17	1.23	1.31	1.40	1.66	2.05
avocado	1.68	0.42	1.31	1.43	1.57	1.74	2.23	2.94
banana	2.71	1.08	1.79	2.09	2.45	2.88	4.13	5.94
barley grain	1.90	0.57	1.41	1.57	1.76	1.99	2.65	3.60
cacao	1.18	0.12	1.08	1.12	1.16	1.20	1.34	1.53
cassava	2.34	0.84	1.61	1.85	2.13	2.47	3.45	4.85
coffee	1.17	0.11	1.08	1.11	1.15	1.19	1.32	1.50
cotton branch	1.37	0.23	1.17	1.24	1.31	1.41	1.68	2.07
dry bean	1.36	0.22	1.16	1.23	1.30	1.39	1.65	2.03
grape	1.53	0.33	1.24	1.34	1.45	1.58	1.97	2.53
key lime	2.13	0.71	1.52	1.72	1.96	2.24	3.07	4.26
maize (amilaceo)	1.53	0.33	1.24	1.34	1.45	1.58	1.97	2.52
maize (choclo)	2.13	0.71	1.52	1.71	1.95	2.24	3.06	4.24
maize (yellow hard)	2.07	0.67	1.49	1.68	1.90	2.18	2.96	4.08
onion	2.19	0.75	1.55	1.76	2.00	2.31	3.17	4.42
orange	2.36	0.86	1.62	1.86	2.15	2.49	3.49	4.92
potato	2.35	0.85	1.62	1.85	2.14	2.48	3.46	4.87
rice	1.88	0.55	1.40	1.56	1.74	1.96	2.60	3.52
tangerine	2.33	0.83	1.61	1.84	2.12	2.46	3.42	4.82
wheat	1.71	0.45	1.33	1.45	1.60	1.78	2.30	3.05

Table 4: Instrumental Variable Estimation of Elasticity of Substitution, σ

	OLS	IV
log unit value	0.298***	-1.648***
	(0.00402)	(0.0418)
Observations	362287	357731
R^2	0.573	0.225
Adjusted R^2	0.446	0.006

Standard errors in parentheses
Regression includes Province, Crop and Year FE
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1: Peru: Geography and Roads



Note: Peru's road system is divided in three levels: National, Departmental and Neighborhood roads. The map plots only the first two, to avoid clutter.

Figure 2: Cost Shares of Land across Crops γ_k

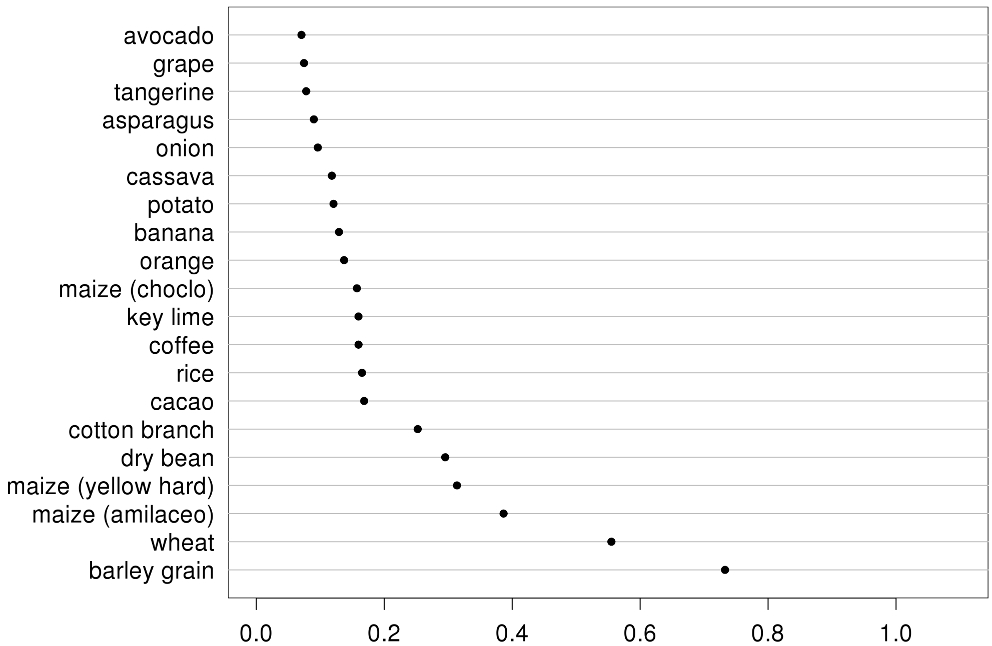


Figure 3: Identification of Inverse Heterogeneity θ

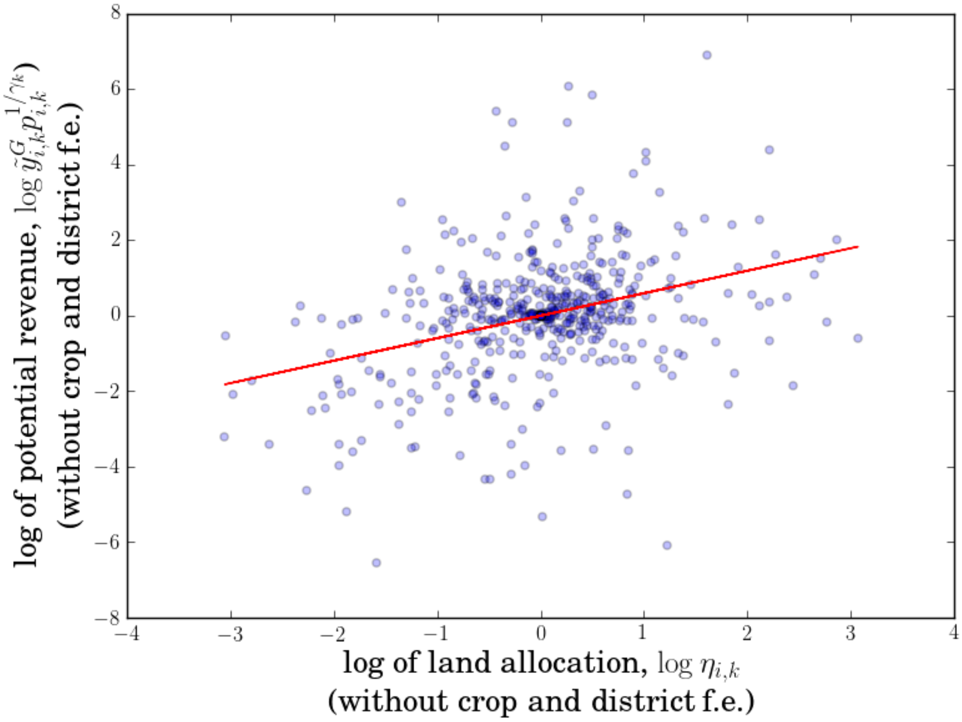
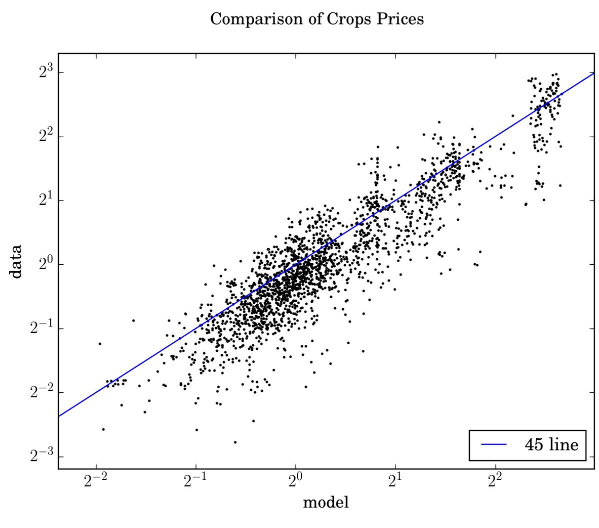


Figure 4: Fitting Price and Land Allocation Data at the Baseline



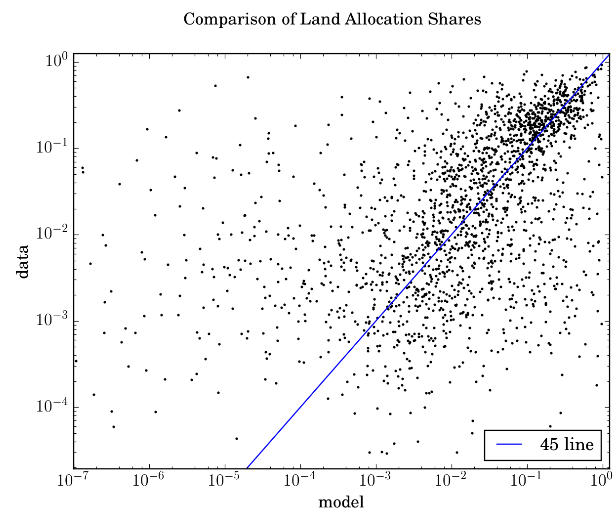
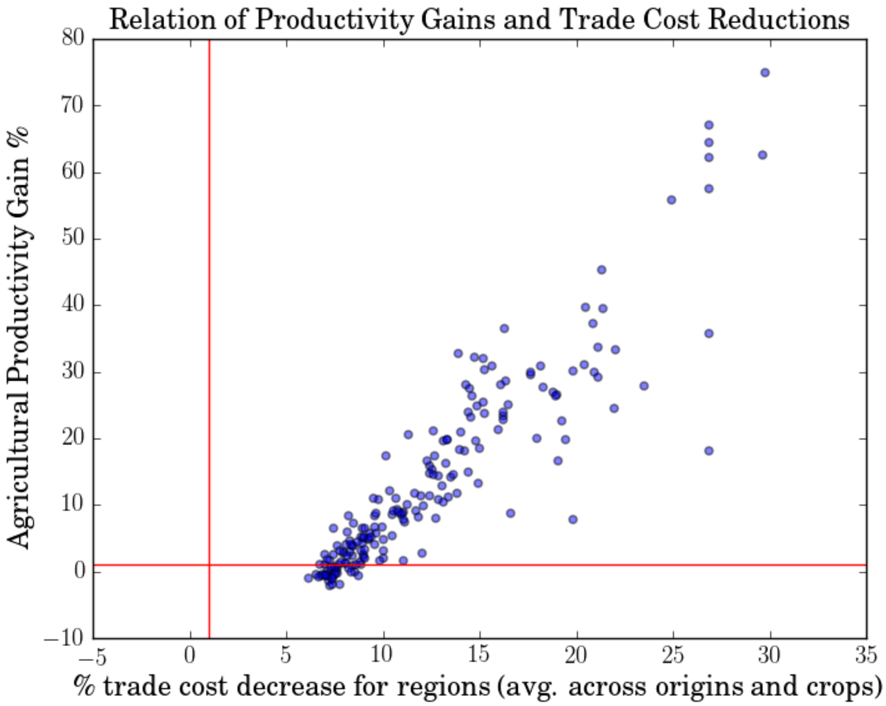


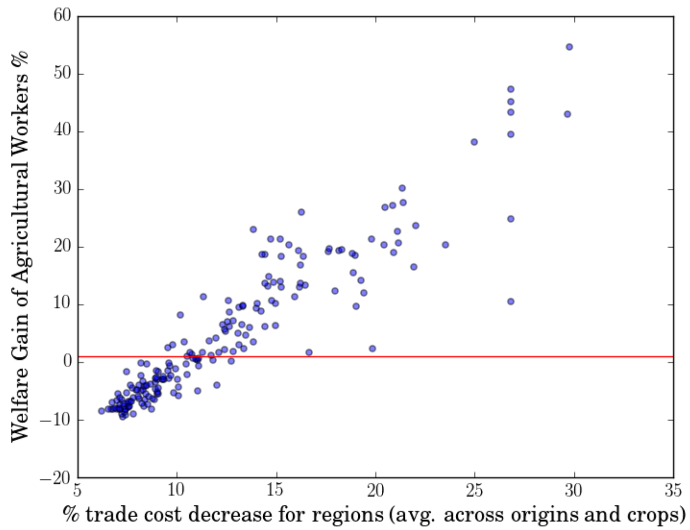
Figure 5: Counterfactual Change in Agricultural Productivity: Departmental roads paved



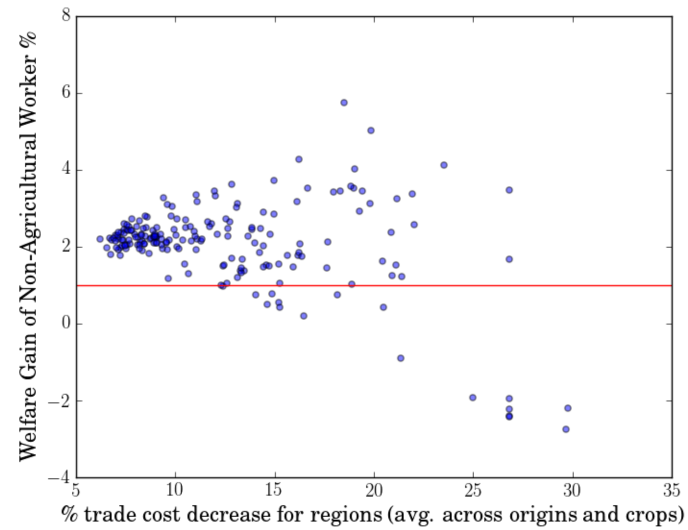
Note: The Y axis plots the percent change in V_i, in terms of intermediates at the output. At constant endowments, this measures changes in productivity.

Figure 6: Counterfactual Change in Worker Welfare: Departmental roads paved

(a) Agricultural Workers



(b) Non-agricultural Workers



Note: In the left panel, the Y axis plots the percent change in $\frac{(1-\bar{\beta}_i)V_i/L_{i,A}}{P_i^b p_M^{1-b}}$. In the right panel, The Y axis plots the percent change in $\frac{w_{i,M}}{P_i^b p_M^{1-b}}$

Figure 7: Experiences with Paving Roads