Trade Frictions and Agricultural Productivity: Theory and Evidence from Peru (Job Market Paper)

Sebastian Sotelo*†

This version: November 22, 2013

Abstract

How much do trade frictions contribute to developing countries' low and disparate agricultural productivity? To answer this question, I develop a quantitative model of agricultural specialization, trade, and productivity. In the model, farmers in a small open economy grow many crops in plots of varying qualities. Trade frictions within and across national borders reduce productivity through two channels: they impede specialization based on comparative advantage and increase the price of imported intermediate inputs. The model's key parameters determine land heterogeneity, barriers to trade, and consumption substitutability. To estimate them, I construct a detailed data set on Peruvian agriculture with information about crop prices, yields, land allocations, freight rates, and household expenditures. Using these estimates, I assess the model's performance at the baseline and produce counterfactuals in which isolated regions experience improved market access to Lima, Peru's largest urban market. Moving an isolated region from the 90th to the 50th percentile of the distribution of transport costs to Lima improves that region's Total Factor Productivity by almost 10.0 percent: 8.7 percent through improved specialization and 1.2 percent by allowing for cheaper intermediate input use.

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE OR DISTRIBUTE WITHOUT THE AUTHOR'S PERMISSION

*E-mail address: sotelo@uchicago.edu. Webpage: http://home.uchicago.edu/∼/sotelo/

†I am indebted to Sam Kortum, Nancy Stokey, and Thomas Chaney for their advice and encouragement. I would also like to thank Fernando Alvarez, Costas Arkolakis, David Atkin, Saki Bigio, Lorenzo Caliendo, Kerem Cosar, Sara Heller, Erik Hurst, Gita Khun Jush, Joaquin Lopez, Sara Moreira, Ralph Ossa, Alan Sanchez, Danny Tannenbaum, Adriaan Ten Kate, and Mike Waugh, as well as participants of various working groups at the University of Chicago and Yale for helpful comments and suggestions at different stages of this project. Edgar Salgado shared his knowledge about the data used in this paper and Maria Alejandra Zegarra provided excellent research assistance. Part of this paper was written while I was visiting the Economcis 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 and Ministerio de Agricultura for sharing their data with me. All opinions and remaining errors are my own.

1 Introduction

Farmers in developing countries face major barriers to trade in domestic and international markets due to poor infrastructure, adverse geography, and the spatial dispersion characteristic of rural populations. At the same time, agricultural productivity varies enormously across space. Differences in agricultural productivity account for sizable productivity gaps between rich and poor countries.[1] And within developing countries, where agriculture is vital to large fractions of the workforce, this contrast is often mirrored in regional disparities. Comparing regions in Peru, the focus of this paper, I calculate that the the 90-10 quantile ratio of Total Factor Productivity is about 8.[2]

Economic performance and market access are often linked together by researchers and policy makers. The World Bank, for example, considers difficult market access one of the reasons behind the uneven agricultural performance across developing countries (World Bank, 2008, p. 54). 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" (See Mesquita Moreira, Blyde, Volpe, and Molina (2013), p. 3.)

My goal is to measure how trade frictions translate into low and disparate agricultural productivity in developing countries. I extend recent advances in the trade literature to develop a quantitative theory that relates productivity in agriculture to trade and specialization. In the theory, farmers can grow many crops in land plots of different qualities. They can also trade their crops and purchase intermediate inputs, at a cost, within and across national borders. When access to markets is costly, farmers pay high prices for their purchases and collect low prices for their sales. Productivity is then low because farmers find it difficult to specialize according to comparative advantage and because they cut back their use of intermediate inputs.

To quantify the theory, I construct a detailed data set consisting of 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 region, as well as within-region heterogeneity. Second, to estimate within-country trade frictions, I combine data on within-country freight rates with data on geography and the quality of the transportation network. Finally, I use household consumption data to estimate the elasticity of substitution among crops in demand. Bringing the theory to data I find that, in a typical isolated region in Peru, improving access to Lima, the country's largest market, improves total factor productivity by almost to 10 percent: 8.7 percent through improved specialization and 1.2 percent by allowing for cheaper intermediate input use.

Peru is an ideal setting for this study. It is a middle income country where a few large, urban markets are often the destination of traded agricultural produce. Eighty six percent of the national

[1] Gollin, Lagakos, and Waugh (2011); Restuccia, Yang, and Zhu (2008)

[2] Based on my quantitative model, I use Revenue TFP (TFPR) in the sense used by Hsieh and Klenow (2009), who find that the 90-10 inter-quantile ratio of TFPR for manufacturing plants is about 4 in India and China. Agricultural value-added per worker, a measure directly observable in the data at a coarser level of aggregation, points in the same direction. The gap between the most productive region and the least productive one, according to that measure, is about 12.

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.[3] Geography also gives substantial scope for specialization because weather and land quality change drastically within the country. Finally, about 25 percent of the labor force is still employed in agriculture, similar to other developing countries.[4]

In addition to these substantive results for Peru, this paper also makes three methodological contributions. First, it presents a theory that connects tightly with data on land allocations and productivity. While trade models make predictions for both trade shares and factor allocations, the latter are often ignored–either because the model being used only delivers simple expressions for trade shares, or because the allocation of factors across activities is difficult to observe. My approach allows me to focus attention on this other side of trade models and explore its implications for productivity.

Second, I obtain a simple estimating equation for the elasticity of land allocation to the price changes induced by trade. 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. The equation also allows for a careful exploration of measurement error in data sets commonly used in this line of research.

Third, I bring the model to data by estimating its parameters and then comparing simulations to data. This approach contrasts with a current alternative in the literature, which sidesteps the need for estimation by assuming that the model fits the data perfectly, and then exploits the structure of the model to evaluate policy changes. By exploring the sources of error in the model, I am able to learn about its strengths and weaknesses, as well as gain a more nuanced understanding of my policy counterfactuals.

Contributions Relative to the Literature

A recent literature, exemplified by Gollin, Parente, and Rogerson (2007) has highlighted the role of agriculture in understanding the low productivity of developing countries. Restuccia, Yang, and Zhu (2008) have quantified how barriers to the use of modern intermediate inputs and barriers to mobility of labor can generate the productivity gaps observed in the data.[5] I contribute to this literature by showing that the quality of transportation in developing countries is a friction that

[3]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).

[4]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).

[5]Restuccia, Yang, and Zhu (2008) have also shown that there is large dispersion in agricultural labor productivity across countries. They estimate that the GDP 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. Gollin, Lagakos, and Waugh (2011) conclude that, despite measurement and data quality problems, poor countries are disproportionately unproductive in agriculture.

leads to a suboptimal use of modern inputs, and that alleviating this friction can lead to a more productive allocation of land and labor.

Land heterogeneity is a key element in my model. Like Costinot, Donaldson, and Smith (2012) and Fajgelbaum and Redding (2013), I treat unobserved land heterogeneity as draws from a Type II extreme-value distribution. My focus, however, is quite different from those two papers. The first, looks at the role of international trade in mitigating the effects of global warming, and the second at the link between international trade and structural transformation. My goal is to understand how market access –transportation costs– affect agricultural TFP. Hence, my model of technology separates intermediate inputs from land, which allows me to separately measure the role of specialization and input use in accounting for productivity dispersion. Further, I solve the model without resorting to the assumption that goods are differentiated by region; assuming such differentiation is less attractive when within-country trade is a prominent feature of the data. Last, I briefly discuss what restrictions the Fréchet distribution imposes on the observed land and revenue shares of each crop, and how to verify that the data does not deviate systematically from them.[6]

This paper also contributes to a literature that applies quantitative trade models to questions in developing economics. An example is Donaldson (2010), whose path-breaking work establishes the causal effect of transportation infrastructure on welfare, showing in the process how to analyze agriculture trade data through the lens of the Eaton and Kortum (2002) model. My work differs on the microeconomic foundation of the model, which delivers different empirical implications. In Donaldson (2010), there is a continuum of varieties of each crop, and the land is heterogeneous in its productivity to grow each of these varieties. His model thus makes predictions about trade shares for each crop and for welfare, the key parameter being the heterogeneity of land productivity across the varieties of each crop. In contrast, I assume that there is a single variety of each crop, but there is a continuum of land plots whose quality to grow each crop is heterogeneous in each region. Hence, the model emphasizes the allocation of land across crops, and yields implications for productivity. These are precisely the variables I observe, as opposed to within-country trade shares, which are often hard to come by in developing countries. More broadly, In my model the key parameter is the elasticity of land allocation to prices, which is linked to land quality heterogeneity. In practice, my choices seem to reduce the impact of infrastructure on productivity.

I also draw from Costinot and Donaldson (2012) and Costinot and Donaldson (2011) whose innovative work advanced the study of agricultural productivity combining Ricardian trade models with data on productivity from the Global Agro-Ecological Zones (GAEZ) database (see IIASA/FAO (2012)). In particular, Costinot and Donaldson (2011) study the productivity gains due to market integration experienced by the United States in the period 1880-2002. My work complements theirs in two aspects. First, I show how to relate land allocation shares to the information in the GAEZ data set through a new, simple estimating equation. This allows me to apply econometric techniques to deal with potential measurement error in the GAEZ data. Second, I study the consequences of improving market access in an equilibrium with endogenous prices, where all outcomes are functions

[6] In Sotelo (2013) I study more systematically the deviations from the predictions implied by Fréchet heterogeneity and draw some implications for modeling.

of preferences, technology and market structure.

To calculate the effect of market access on factor productivity, I rely on model-based policy counterfactuals. My strategy for doing so contrasts with a currently popular alternative in quantitative trade research. In the wake of Eaton and Kortum (2002) and Dekle, Eaton, and Kortum (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, with the exception of the elasticity of trade flows with respect to trade barriers.[7] I follow the opposite path: I start by estimating the technology and preferences by comparing a set of moments of the model to disaggregated data and then use those estimates to simulate the equilibrium. Naturally, the simulated equilibrium does not replicate the data exactly. The discrepancy between predictions and observations gives me an opportunity to learn about the model's strengths and weaknesses, which are hidden in the more standard approach.

Finally, I also extend the results of a literature that studies agricultural productivity in the context of models of international trade. Tombe (2012) 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. By narrowing the focus to a single country, in this paper I obtain more direct evidence of the mechanisms proposed in those papers. In particular, after directly estimating within-country transportation costs, I measure their impact on resource allocation and productivity.

The rest of the paper is as follows. In Section 2, I present a simple equilibrium framework and the two main assumptions that allow me to connect it to data. In Section 3, I summarize how the model guides the data analysis in three key propositions, and I construct a measure of productivity consistent with the theory. In Sections 4 and 5, I present the data and connect it to the model. In Section 6, I assess the model's strengths and weakness in a simulation of the equilibrium. Finally, in Section 7, I elicit the effect of market access on productivity by computing counterfactuals. All derivations and proofs are contained in the Appendix.

2 Land Allocation and Trade in a Small, Open Economy

To study the link between trade frictions and agricultural productivity, I modify a standard 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. Markets are perfectly competitive, but trade across regions and with the rest of the world is costly. The quality of land to grow different crops varies across, and also within regions; such heterogeneity is a source of comparative-advantage. Trade frictions impede specialization and hence diminish productivity.

[7]For example, see Caliendo and Parro (2012), Parro (2013), Ossa (2011).

To produce, farmers combine land with labor and an imported intermediate input. Regions farther away from major ports use less of the intermediate input because its price is relatively high, which also diminishes productivity.

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 equations that can be used for estimation.

2.1 Environment

2.1.1 Geography

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.

2.1.2 Commodities

There are $k = 1, \ldots, K$ 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.

2.1.3 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 she rents its 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.

2.1.4 Preferences

I only specify the preferences for consumers at home. The consumer in region i spends a fraction b of income on an agricultural aggregate, $C_{i,A}$, and the rest on manufactured goods, $C_{i,M}$:[8]

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

[8] Because Engel's Law holds in the data this simplifying assumption will miss some key aspects of the data. I will improve on it in future versions of the paper. The application of the AIDS of Deaton and Muellbauer (1980) to household expenditure data has proven successful. Moreover, there is a recent literature that explains how non-homotheticity reconciles trade models with observations on international trade. See Fieler (2011).

The agricultural aggregate is

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

where $\sigma>0$ is the elasticity of substitution across agricultural goods. For each region i the weights satisfy $a_{i,k}\geq 0$, so some goods may not be consumed there. I normalize the weights to add up to one: $\sum_{k=1}^{K}a_{i,k}=1$. Also, for each crop k, there is some region i where $a_{i,k}>0$.

2.1.5 Endowments

The household in region i supplies labor inelastically to agriculture, $L_{i,A}$, and to manufacturing, $L_{i,M}$.[9] 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 (measure) of land in the region is $H_i=\int_{\Omega_i}\mathrm{d}\omega$.

2.1.6 Agricultural Technology

To take the model to data requires making assumptions about the functional form of the production function, as well as the distribution of land quality within regions. I make the following two assumptions for tractability, but note that the workings of the rest of the model and the definition of equilibrium are independent of these specific details.

Assumption 1. *Crops are produced with a constant returns to scale technology that combines labor, the imported intermediate and land. The suitability of plot ω in region i for producing crop k is captured by an efficiency parameter $\Lambda_{i,k}(\omega)\geq 0$,*

$$q_{i,k}(\omega)=\phi_{i,k}(\omega)\left(l_{i,k}(\omega)\right)^{\alpha}\left(x_{i,k}(\omega)\right)^{\beta}\left(\Lambda_{i,k}(\omega)\right)^{\gamma} \tag{3}$$

where $q_{i,k}(\omega)$ is the output of crop k, $l_{i,k}(\omega)$ and $x_{i,k}(\omega)$ are labor and intermediates per unit of land, and $\phi_{i,k}(\omega)$ is the share of plot ω allocated to k. The cost shares α, β and γ are the same for all crops k, with $\alpha+\beta+\gamma=1$.

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, $\left(\Lambda_{i,k}(\omega)\right)_{k\in\mathbb{K}_i}$, is distributed as a set of i.i.d Fréchet random variables with parameters $(\tilde{\gamma}A_{i,k},\theta)$,*

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

[9] This assumption is justified by 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, Gollin, Lagakos, and Waugh (2011); Restuccia, Yang, and Zhu (2008); Tombe (2012). Lagakos and Waugh (2013) provide an explanation based on the selection of workers into agriculture, in which a large, unproductive agricultural workforce is an equilibrium outcome.

I normalize $\tilde{\gamma} = \left[\Gamma\left(1 - \frac{1}{\theta}\right)\right]^{-1}$*. Let* $\mathbb{K}_i$ *denote the set of crops that can be grown in region* i*. Then* $\Lambda_{i,k}(\omega) = 0$*, for each plot* ω*, if and only if* $k \notin \mathbb{K}_i$*.*

In this probabilistic representation of within-region land quality, θ is an inverse measure of land heterogeneity. Note also the important difference between $\Lambda_{i,k}(\omega)$ and $A_{i,k}$. The random variable $\Lambda_{i,k}(\omega)$ captures how good plot ω is to grow crop k. In contrast, the parameter $A_{i,k}$, shared by all plots ω in region i, is directly linked 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 all plots in the region is high, although between plots the actual land quality varies according to θ.

2.1.7 Manufacturing technology

Manufacturing uses only labor, with productivity T_i:

$$y_{i,M} = T_i l_{i,M}.$$

2.1.8 Rest of the World

Any region i in Home can trade crops with Foreign at fixed prices. Each region i also imports the intermediate input from Foreign, which is the only producer.[10]

2.2 Markets

2.2.1 Markets and prices

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, $p_{i,M}$ the price of the manufactured good, $p_{i,k}$ the price of crop k, for all k, and let $r_i(\omega)$ denote the rental rate of plot ω.

In Foreign, the price of crop k is $p_{F,k}$, for all k, and the price of the intermediate input there is ρ_F.

2.2.2 Trade costs

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

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., $d_{ni,k} \leq d_{nj,k} \times d_{ji,k}$.

[10]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 (FAOSTAT).

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

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

2.3.2 Producers

The representative manufacturing firm in region i hires labor and produces output. The representative agricultural producer 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 intermediates to use in each plot. 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,

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

where total output of crop k is

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

for all k, and

$$\sum_{k=1}^{K} \phi_{i,k}(\omega) \leq 1,$$

for all $\omega \in \Omega_i$.

2.3.3 Trader decisions

Define $\mathcal{W} = \{1, \ldots, I, F\}$ to be the set of all regions in the economy. Let $z_{ni,k}$ denote the shipments of crop k from region i to region $n \in \mathcal{W}$. The trader in i chooses $\{z_{ni,k} \geq 0\}$ to maximize profits

$$\max \sum_{n \in \mathcal{W}} \sum_{k=1}^{K} z_{ni,k} \left(\frac{p_{n,k}}{d_{ni,k}} - p_{i,k} \right).$$

Without loss of generality, I assume that the trader chooses $z_{ii,k} = 0$, for all crops.

2.4 Remarks

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 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.

Last, I have made the stringent assumption that preferences are homothetic. While this has some downsides that I discuss below, it has the advantage of allowing me to attribute all income to a single representative consumer.

2.5 Competitive Equilibrium

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

Definition 1. A competitive equilibrium in this small economy with costly trade consists of

(a) prices $p_{i,k}$ for all crops k and $p_{i,M}$ 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 $\omega \in \Omega_i$;

(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) shipments of each good $z_{ni,k}$, for regions $n \in \mathcal{W}$, for each crop k and for the manufactured good $z_{ni,M}$;

for each region $i = 1, \ldots, I$, 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 $n, i \in \mathcal{W}$, for all crops k and the manufactured good M; and the intermediate input prices are

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

for all regions i in Home;

5. local markets clear for labor, land, crops and manufactured goods:

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

6. 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 international price of intermediate inputs to one, $\rho_F = 1$. The reason for choosing an international price is that it is not affected by what happens within Home, because we assumed that it is a small economy. Hence the unit of account will remain unchanged across equilibria where policy parameters change, which is helpful for my application to Peruvian data.

3 Quantitative Implications of the Model

3.1 Farmers' Choices and the Empirical Implications of the Model

To connect the model to data on land shares and yields across crops, I start by describing the optimal behavior of the representative agricultural producer –here called a 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 all plots $\omega \in \Omega_i$, 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. Note that, in doing so, we treat each plot as a separate factor of production, since $r_i(\omega)$ 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, in terms of the numéraire, is:

$$c_{i,k}(\omega) = \frac{\bar{c}_i(\omega)}{(\Lambda_{i,k}(\omega))^{\gamma}},$$

where we define

$$\bar{c}_i(\omega) \equiv \alpha^{-\alpha}\beta^{-\beta}\gamma^{-\gamma} w_{i,A}^{\alpha}\rho_i^{\beta}(r_i(\omega))^{\gamma}.$$

The function $\bar{c}_i(\omega)$, which involves only the input prices, is the same for all crops. The unit cost function $c_{i,k}(\omega)$ adjusts this value for the quality of plot ω in producing crop k.

Because of Assumption 2, typically only one crop maximizes profits in a given plot, although in equilibrium each plot earns the farmer zero profits. Then it is optimal to specialize each plot ω completely in a single crop k –an event I denote $\omega \in \Omega_{i,k}$. This insight allows us to characterize the crop choices for the whole region i in a probabilistic way. To see how, note that when plot ω is specialized in the production of crop k, total output in that plot, as a function of input prices, is

$$q_{i,k}(\omega; w_{i,A}, \rho_i, r_i(\omega)) = \frac{r_i(\omega)}{\gamma c_{i,k}(\omega)},$$

which we obtain by setting $\phi_{i,k} = 1$ and noting that γ is the cost share of land. It follows that total profits in plot $\omega \in \Omega_{i,k}$ are

$$[p_{i,k} - c_{i,k}(\omega)] \times q_{i,k}(\omega; w_{i,A}, \rho_i, r_i(\omega)) = \frac{r_i(\omega)}{\gamma}\left[p_{i,k}\frac{(\Lambda_{i,k}(\omega))^{\gamma}}{\bar{c}_i(\omega)} - 1\right].$$

This expression shows that, for plot ω, all the variation in profits across crops is only due to differences in output prices, $p_{i,k}$, and land quality, $\Lambda_{i,k}(\omega)$. Input prices have no effect on relative profitability.

Let $\eta_{i,k}$ denote $\mathbb{P}[\omega \in \Omega_{i,k}]$; then

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

The result is summarized in the following proposition.

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

$$\eta_{i,k} = \frac{\left(p_{i,k}^{\frac{1}{\gamma}} A_{i,k}\right)^{\theta}}{\Phi_i^{\theta}}, \tag{7}$$

where

$$\Phi_i = \left(\sum_{k' \in \mathbb{K}_i} \left(p_{i,k'}^{\frac{1}{\gamma}} A_{i,k'} \right)^{\theta} \right)^{\frac{1}{\theta}}. \tag{8}$$

Equation (7) implies that the relative land allocation between any two crops k and k' depends only on the price $p_{i,k}$ and the land quality $A_{i,k}$ of those two crops. Note that in (7) $p_{i,k}$ is raised to the power of $1/\gamma$, while $A_{i,k}$ does not, which reflects that the cost share of land is γ in production. The prices and land qualities for all other crops are captured in, Φ_i^{θ}, the normalizing term defined in equation (8).[11] This implication is particularly useful when estimating θ, because it allows us to work only with a subset of crops.

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 $\frac{\theta}{\gamma}$ percent. To interpret this elasticity, recall that θ is an inverse measure of land quality heterogeneity. When θ is larger, 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, the output elasticity of land in the production function. A smaller value for γ means that land is less important compared with other inputs, and a given increase in $p_{i,k}$ produces a larger shift in land use.

Although farmers seek to maximize profits, due to perfect competition they will earn zero profits in equilibrium. The rental rate of land in each plot, $r_i(\omega)$, adjusts to ensure that this is so. We can impose the zero profit condition, $p_{i,k} = c_{i,k}(\omega)$, to solve for $r_{i,k}(\omega)$, the value of the rental rate that would prevail in plot ω, if $\omega \in \Omega_{i,k}$:

$$r_{i,k}(\omega) \quad = \quad \gamma \kappa_y p_{i,k}^{\frac{1}{\gamma}} w_{i,A}^{-\frac{\alpha}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} \Lambda_{i,k}(\omega),$$

where we define $\kappa_y = \alpha^{\frac{\alpha}{\gamma}} \beta^{\frac{\beta}{\gamma}}$. As a consequence of profit maximization,

$$r_i(\omega) = \max_{k'} \left\{ r_{i,k'}(\omega) \right\},$$

and $r_i(\omega) = r_{i,k}(\omega)$ happens with probability $\eta_{i,k}$.

While I do not observe rental rates directly in the data, I do observe the land yield, land allocations, and revenue per unit of land across crops in all regions. To characterize yields and revenues in the model –the objects of interest–, we must first calculate the optimal use of labor and intermediate inputs, relative to land, when $\omega \in \Omega_{i,k}$:

$$x_{i,k}(\omega) \quad = \quad \left(\frac{\beta p_{i,k}}{\rho_i} \left(\frac{\alpha \rho_i}{\beta w_{i,A}} \right)^{\alpha} \right)^{\frac{1}{\gamma}} \Lambda_{i,k}(\omega) \tag{9}$$

[11] This implication is similar to the independence of irrelevant alternatives assumption in consumer theory. See, for example, Train (2003).

and

$$l_{i,k}(\omega)=\left(\frac{\alpha p_{i,k}}{w_{i,A}}\left(\frac{\beta w_i}{\alpha\rho_{i,A}}\right)^{\beta}\right)^{\frac{1}{\gamma}}\Lambda_{i,k}(\omega). \tag{10}$$

Given prices, the demand for inputs other than land increases when land has better quality $\Lambda_{i,k}(\omega)$. Equations (9) and (10) relate input use only to prices –taken as given by the farmer– and model primitives.

We are now in a position to compute two measures of productivity for plot $\omega\in\Omega_{i,k}$: physical land yield, $y_{i,k}(\omega)$, and revenue per unit of land, $\psi_{i,k}(\omega)=p_{i,k}y_{i,k}(\omega)$. The optimal physical yield is given by $q_{i,k}(\omega)/\phi_{i,k}(\omega)$, from equation (3), evaluated at the optimal input demands given by (9) and (10):

$$y_{i,k}(\omega) \quad = \quad \kappa_y\Lambda_{i,k}(\omega)\left(\frac{p_{i,k}}{w_{i,A}}\right)^{\frac{\alpha}{\gamma}}\left(\frac{p_{i,k}}{\rho_i}\right)^{\frac{\beta}{\gamma}}, \tag{11}$$

where the constant κ_y is as defined above. This expression highlights that land yield is an endogenous object that reflects the choices of farmers. [12] Multiplying the physical yield by the crop price gives revenue per unit of land:

$$\psi_{i,k}(\omega) \quad = \quad \kappa_y\Lambda_{i,k}(\omega)\,p_{i,k}^{\frac{1}{\gamma}}w_{i,A}^{-\frac{\alpha}{\gamma}}\rho_i^{-\frac{\beta}{\gamma}}. \tag{12}$$

Equations (11) and (12) show that, taking prices as given, land yields and revenue per unit of land are proportional to land quality.

Proposition 2 below formalizes the idea that $y_{i,k}(\omega)$ and $\psi_{i,k}(\omega)$ inherit the properties of land quality, conditional on $\omega\in\Omega_{i,k}$. A takeaway of this proposition is that we cannot infer the value of average land quality, $A_{i,k}$, just by looking at data on physical yields or revenue per unit of land. Instead, these data can only inform us about aggregate land productivity in a region.

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 $\left(\tilde{\gamma}\kappa_y p_{i,k}^{-1}w_{i,A}^{-\frac{\alpha}{\gamma}}\rho_i^{-\frac{\beta}{\gamma}}\Phi_i,\theta\right)$.*

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 $\left(\tilde{\gamma}\kappa_y w_{i,A}^{-\frac{\alpha}{\gamma}}\rho_i^{-\frac{\beta}{\gamma}}\Phi_i,\theta\right)$.

Note that no moment of the distribution of either conditional yield or conditional revenue per unit of land will be informative about the relative values of crop-specific land quality. Particularly, the expected yield and expected revenue per unit of land

$$\mathbb{E}\left[y_{i,k}(\omega)|\omega\in\Omega_{i,k}\right]=\kappa_y\left(\frac{w_{i,A}}{p_{i,k}}\right)^{-\frac{\alpha}{\gamma}}\left(\frac{\rho_i}{p_{i,k}}\right)^{-\frac{\beta}{\gamma}}\left(\frac{\Phi_i^{\gamma}}{p_{i,k}}\right)^{\frac{1}{\gamma}} \tag{13}$$

[12] Kelly (2006) discusses how a high price of intermediates relative to the price of final output reduces the demand for fertilizer in Africa.

and

$$\mathbb{E}\left[\psi_{i,k}\left(\omega\right)|\omega\in\Omega_{i,k}\right]=\kappa_{y}w_{i,A}^{-\frac{\alpha}{\gamma}}\rho_{i}^{-\frac{\beta}{\gamma}}\Phi_{i}, \tag{14}$$

are uninformative about the relative values of $A_{i,k}$ across crops. The first moments of both distributions are especially important: they are the objects in the model that correspond to the observations on yield and revenues.

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 $\frac{\theta}{\gamma}$, while the land allocated to the rest of the crops decreases. Equation (13) 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 remains unchanged. 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.

It is important to realize 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 would increase its productivity in alternative uses. Assumptions 1 and 2 just put constraints on the exact amounts by which productivity changes across alternative uses.

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.[13] The productivity distributions for all crops then

[13] 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

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 note that, in equilibrium, for a given value of Φ_i, 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. A low wage $w_{i,A}$ has an analogous effect through the increased use of labor.[14]

In light of this discussion, the content of Proposition 3 is implied by Propositions 1 and 2. I present it separately because it provides a simple way of empirically evaluating the adequacy of the model. Let $\pi_{i,k}$ be the revenue share of crop k in region i's total revenue, defined as

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

Then we summarize the relationship between land and revenue shares with the following proposition:

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

$$\pi_{i,k} = \eta_{i,k}.$$

Proposition 3 is a direct consequence of Assumptions #1 and #2, about technology and the probabilistic structure of the model. The model predicts that, within a region i, the share of land allocated to crop k is equal to its share in the total agricultural revenue generated in that region. 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. Looking for systematic deviations from this prediction will be helpful in assessing how well the Fréchet structure in Assumption 2 fits the data.

To understand this result, suppose that the relative price of crop k in region i increases. According to the land allocation equation (7), the land share $\eta_{i,k}$ increases with an elasticity of θ/γ (holding Φ_i fixed.) The revenue share will also tend to increase, because the amount of land used to grow crop k determines its total output. But what happens to the average revenue per unit of land? As shown in equation (14), the average revenue per unit of land is always equalized across

(2010) exploits this object in his welfare calculations.

[14] 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 B:

$$\mathbb{E}\left[r_{i,k}(\omega) \,|\omega \in \Omega_{i,k}\right] \quad = \quad \gamma \kappa_y w_{i,A}^{-\frac{\alpha}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} \Phi_i,$$

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

$$r_i = \gamma \kappa_y w_{i,A}^{-\frac{\alpha}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} \Phi_i.$$

Note that 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, $\mathbb{E}\left[\psi_{i,k}(\omega) \,|\omega \in \Omega_{i,k}\right]$, is equal to a constant, γ. However, I emphasize that the restriction that γ is the same for all crops is not they key driver of Proposition 2: complete flexibility in this regard would only allow for crop-specific scale parameters, and still the data would be uninformative about unobserved land quality.

crops. This equalization ensures that the proportional increase in average revenue per unit of land is identical for all crops, 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.

3.1.1 Relation to Other Quantitative Trade Models

The previous set of propositions has a close parallel in two well-known results in the Eaton and Kortum (2002) framework. That model 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 (analogous to Proposition 3 above). The reason is that, in that model, all differences in productivity across suppliers translate into differences in the fraction of goods sold in a destination (Proposition 1). This makes the distribution of unit costs of goods actually sold in region n identical across suppliers (Proposition 2).

In contrast to the trade context, however, in my application both terms in Proposition 3 have empirical counterparts. The reason is that the allocation of land has natural units of measurement (hectares, for example), while it is less clear how to measure a fraction of goods in a continuum.

3.2 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 study two important objects: (i) aggregate labor demand, and (ii) the aggregate value of production. This leads to a discussion of the economic relation of market access and productivity. Finally, I obtain the household's expenditure on each good.

3.2.1 Aggregate Labor Demand

In region i, aggregate labor demand for the production of crop k is

$$L_{i,k} = \kappa_l \Phi_i w_{i,A}^{-\frac{1-\beta}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} \eta_{i,k} H_i,$$

where $\kappa_l = \alpha^{\frac{1-\beta}{\gamma}} \beta^{\frac{\beta}{\gamma}}$. Since the land shares sum to one, the total demand for labor is:

$$L_{i,A} \quad = \quad \kappa_l \Phi_i w_{i,A}^{-\frac{1-\beta}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} H_i. \tag{15}$$

3.2.2 Agricultural Wages and the Total Value of Production

Because we take the stock of agricultural labor as fixed, we can use (15) to express the equilibrium wage as a function of variables that are exogenous to farmers:

$$w_{i,A} = \kappa_l^{\frac{\gamma}{1-\beta}} \Phi_i^{\frac{\gamma}{1-\beta}} \rho_i^{-\frac{\beta}{1-\beta}} \left(\frac{H_i}{L_{i,A}} \right)^{\frac{\gamma}{1-\beta}}. \tag{16}$$

Now we are in a position to calculate the total value of production of crop k in region i,

$$\begin{aligned} V_{i,k} &= p_{i,k}\mathbb{E}\left[y_{i,k}\left(\omega\right)|\omega \in \Omega_{i,k}\right]\eta_{i,k}H_i \\ &= \beta^{\frac{\beta}{1-\beta}}\rho_i^{-\frac{\beta}{1-\beta}}p_{i,k}^{\frac{\theta}{\gamma}}A_{i,k}^{\theta}\Phi_i^{\frac{\gamma}{1-\beta}-\theta}H_i^{\frac{\gamma}{1-\beta}}L_{i,A}^{\frac{\alpha}{1-\beta}}, \end{aligned} \tag{17}$$

where the second line uses (13) to substitute for $\mathbb{E}\left[y_{i,k}\left(\omega\right)|\omega \in \Omega_{i,k}\right]$ and (16) to substitute for $w_{i,A}$. Summing across crops we obtain the total value of agricultural production in region i

$$V_i = \beta^{\frac{\beta}{1-\beta}}\rho_i^{-\frac{\beta}{1-\beta}}\Phi_i^{\frac{\gamma}{1-\beta}}H_i^{\frac{\gamma}{1-\beta}}L_{i,A}^{\frac{\alpha}{1-\beta}}. \tag{18}$$

Equation (18) is the familiar revenue function. It relates the total revenue generated by region i to the prices that are exogenous to the farmer and to the total stock of factors of production. It is a representation of aggregate supply in region i.

To build intuition, consider for the moment the case where the imported intermediate plays no role, so that $\beta = 0$ and $\alpha = 1-\gamma$. Then equation (18) simplifies to

$$V_i = \Phi_i^{\gamma}H_i^{\gamma}L_{i,A}^{1-\gamma}. \tag{19}$$

With Φ_i^{γ} constant, the total value of agricultural production has constant returns to scale in labor and land. The average productivity in agriculture depends on the distribution of land quality and crop prices, both of which affect the exact allocation of land across crops. Equation (19) shows exactly the sense in which Φ_i is a measure of TFPR. For a given stock of land and labor, higher crop prices or a better allocation of land according to comparative advantage increase the total value of production. I turn to an analysis of Φ_i in what follows.

3.2.3 The Economic Relation Between Market Access and Productivity

In the more general expression (18), with $\beta > 0$, the term $\rho_i^{-\frac{\beta}{1-\beta}}\Phi_i^{\frac{\gamma}{1-\beta}}$ can be thought of as the TFPR of Land and Labor. It shows that in location i, agricultural productivity is higher because Φ_i is higher or because the price of intermediates, ρ_i, is lower.

In the model, variation in ρ_i is entirely driven by transportation costs: imported intermediates will be more costly in remote places. This is the first channel through which transport costs lower productivity. The elasticity of TFPR with respect to the price of the intermediate input is $-\beta/\left(1-\beta\right)$, which is higher the larger the cost share of intermediates.

The second channel is related to the farmers' production and consumption choices. Poor market access in this model is equivalent to high transportation costs to and from region i. High transport costs increase the prices of the crops farmers purchase, and decrease the price of the crops they sell. Both effects are summarized in the value of Φ_i. Because producers will tend to sell the goods for 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 Φ_i.

In Appendix H, I discuss a land-only model to clarify the link between trade and comparative advantage through the lens of the model. In such a model, for an autarkic region, the elasticity of the relative price of two crops, $\frac{p_k}{p_{k'}}$, 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 unrelated to land quality A_k. A reduction in the absolute value of the correlation between $p_{i,k}$ and $A_{i,k}$ increases the magnitude of Φ.

I emphasize, however, that Φ_i does not exclusively measure the the effect of specialization due to comparative advantage. Rather, it reflects any factor that increases the productivity of land, and is not explicitly modeled. Thus, if the quality of land in a region doubles –keeping prices constant–, then Φ_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 Φ_i across regions. We return to the impact of trade frictions in Section 7.

3.2.4 Expenditure on each good

Finally, consider regional demand for each good. The solution to the representative consumer's problem yields region n's aggregate expenditure on crop k:

$$E_{n,k} = ba_k \left(\frac{p_{n,k}}{P_n}\right)^{-(\sigma-1)} E_n. \tag{20}$$

where

$$P_n = \left(\sum_{k=1}^{K} a_k p_{n,k}^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \tag{21}$$

is the price index for the agricultural aggregate $C_{n,A}$ in region n. Expenditure on the manufactured good is $E_{n,M} = (1-b)\,E_n$.

3.3 Computation of the Equilibrium

To compute the equilibrium, I apply an iterative algorithm.[15] There are two prevailing ways of computing equilibria in trade models, both of which my strategy differs from. The first approach has been popular for many years in the literature of Computable General Equilibrium (CGE) models. A key element in the CGE approach is the assumption –due to Armington (1969)– that goods are differentiated by country of origin. Although not the original motivation for this assumption, its use sidesteps the need to determine what producer is a destination's cheapest supplier, because every destination buys from every source. [16] The second, more recent approach, is due to Alvarez and Lucas (2007) and is becoming popular to solve quantitative trade models based on Eaton and Kortum (2002). Despite their richer microeconomic underpinnings, Arkolakis, Costinot, and Rodriguez-Clare (2012) have shown that these newer models have a similar general equilibrium

[15]The algorithm is described in detail in Appendix I

[16]See Shoven and Whalley (1992), p. 81, for a discussion of the role of the Armington assumption in CGE trade modeling.

structure to that of the earlier CGE models. Hence, as expected, they also allow for computation based on direct application of CGE methods.

My strategy, in contrast, needs to deal with the problem of finding the cheapest exporter for each importing region, because in the equilibrium of my model some pairs of regions may not trade at all.[17] The algorithm starts with a guess for the crop and manufactured-good prices, and adjusts them according to a tâtonnement process. At each step, the algorithm calculates each region $i's$ desired net imports or exports of each good. Then, it attempts to satisfy all the desired net trades of each region, in a way that is feasible according on traders' profit maximization. The difference between desired and feasible net trades plays the role of excess demand, and guides the update of prices for the next iteration.[18]

4 Data

In this section I discuss the main data sets I have put together. In matching the model to data, I use the administrative division of Peru to define the regions in the model. As of 2012, Peru is hierarchically divided into 24 departments, 194 provinces and 1838 districts. Each region i in Home in the model corresponds to a district in Peru. I use information on consecutive cross-sections, the exact years depending on the sample and source. The size of each district is given by the total amount of land used in agriculture; the quantiles of this size distribution are $2km^2$, $6.5km^2$ and $17.9km^2$.

I also select $K = 51$ crops that the Ministry of Agriculture uses when calculating countrywide statistics and for which the data quality is better. These crops account for between 76 and 79 percent of the total gross value of agricultural production between the years 2008 and 2011. See Table 1 for a summary of all data sources and samples.

National Statistics on Agriculture

These data are collected by the Peruvian Ministry of Agriculture (MINAG). For each district i, crop k, and year t, the data set contains information on farm-gate prices, $p_{i,k,t}$, physical yields, $y_{i,k,t}$, and land use, $\eta_{i,k,t}H_{i,k,t}$. I average each variable at the district and crop level, and interpret these averages as the objects $p_{i,k}$, $y_{i,k}$ and $\eta_{i,k}H_i$ in the model.

I describe two samples, which I use for different purposes depending on their relative strengths. Table 1 summarizes the characteristics of each sample

[17] Although I do not pursue this outcome in this paper, my approach has the desirable implication of being able to generate zeros in the bilateral trade matrix. The absence of trade linkages between pairs of countries is an established fact in the data that quantitative trade models often cannot generate (see Eaton, Kortum, and Sotelo (2012)).

[18] A current shortcoming of the algorithm is that I have no proof of global convergence, although the algorithm has converged in practice. As discussed in Scarf and Hansen (1973), and Mas-Colell, Whinston, and Green (1995) p. 611, some theory results exist for price adjustment mechanisms when preferences display the gross substitution property, which my model satisfies.

Wide sample

This is a cross-section of every district in Peru that produces agricultural goods. It is a balanced panel containing the years 2008-2012. I use this sample to estimate underlying land quality for each region and crop, $A_{i,k}$, which is useful to simulate the model. Descriptive statistics for this sample are shown in Tables 11 and 12 in the Appendix.

Long sample

This is a sample of the districts contained in four out of 24 departments.[19] The advantage of this sample is that it includes the years 1997-2011, although the panel is unbalanced. I use it to get more precise estimates of the long-run equilibrium values of $\eta_{i,k}$ and $p_{i,k}$, which are necessary to improve the estimation of θ, the land heterogeneity parameter.

Global Agro-Ecological Zones (GAEZ)

I describe the data set briefly –Costinot and Donaldson (2011) provide a more detailed discussion of it. The goal of the GAEZ project is to assess the agricultural potential for land cells in a fine partition of the World. FAO and IIASA have developed a methodology to estimate the potential land yield (see IIASA/FAO (2012)). That is, they estimate the land yield that would prevail if all land in a cell is entirely devoted to growing a crop. This method transforms information on land types, water resources and weather conditions into potential yields, through a model of agricultural production. Importantly, actual statistics on agricultural production are not inputs into the model. Hence, the database contains truly independent measures of potential agricultural productivity.

To access the data on potential land yields, the user must make a choice about management conditions: low, medium and high level of inputs. In estimation, I use both low and medium input levels, and show that the results are robust to the choice of management conditions.

Geography and the Transportation Network

To construct the road network of Peru, I use the georeferenced data set put together by the Peruvian Ministry of Transportation (MTC). The system of roads in Peru is classified in three subsystems: National, Departmental, Rural. For each road in each subsystem, the data set contains its exact location and its quality (dirt, graded and paved roads).

Freight Rates

I use a sample of freight rates between 45 pairs of districts, averaged over the years 2010-2013, where at least one of the districts in the pair belongs to the department of La Libertad. [20] Most of the freight rates are expressed in terms of local currency per unit of weight.

[19] The departments are Arequipa, Huánuco, La Libertad, Puno. They account for 22 percent of the total value of production in 2008.

[20] The scope of the data is restricted this way because the source is the Direccion Regional de Agricultura de La Libertad.

Although this is a relatively small sample it is an informative one. For each pair, I observe the freight rates for several goods, for each month in the year, so I have some confidence in the precision of the estimates. Also, there is substantial geographic variation within La Libertad: some districts are in the coast while others in the mountains; in the sample, some pairs are connected by paved highways and others by low-quality roads. Finally, La Libertad is an important agricultural producer and accounts for 8 percent of the total value of agricultural production in Peru.

National Household Surveys (ENAHO)

This survey is conducted by the Instituto Nacional de Estadistica e Informatica (INEI); it is Peru's main tool for learning about living standards. Every quarter, INEI samples households and applies a survey about income, expenditures, living conditions, etc. In this paper, I focus on the following two modules:

Food Consumption Module

INEI applies this module to all households. Respondents give a detailed account of their expenditures on food consumption during the previous fifteen days. Categories are narrow and can be partially matched to the government statistics database. This allows me to examine the consumption patterns of households in terms of those crops.

5 The Connection between the Model and Agricultural Data

In this section, I first provide a brief discussion of the empirical content of Proposition 3, which is the model prediction that land and revenue shares are equalized within regions. I show that even if that prediction does not hold exactly in the data, we can proceed using the model as a good approximation to reality.

Next, I explain the estimation of the model's parameters, which I obtain by comparing selected moments in the model with informative micro 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 (2010): 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, I predict freight rates for all possible origin-destination pairs in the country. Third, I combine expenditure household data with my previous estimates of land quality, $A_{i,k}$, to estimate the elasticity of substitution between crops in demand.

Based on these estimates, in the last part of this Section I calculate TFPR for each district in the country, and show its substantial dispersion. I defer the simulations of the equilibrium to

Section 7, where I assess the overall fit of the model, and carry out counterfactuals to dissect the effect of market access on productivity.

5.1 The model of land heterogeneity is consistent with the data

Figure 2 shows the relation between $\eta_{i,k}$ and $\pi_{i,k}$, in a log scale. The data show a clear positive relation between the two shares, although there are obvious departures from the model's prediction. In Table 2, I regress $\pi_{i,k}$ on $\eta_{i,k}$, without an intercept, as implied by the theory. The slope estimate is 1.017 (0.001) and $R^2 = 0.97$. The coefficient is close to 1, but one can easily reject the null hypothesis that the coefficient is different from 1. At this point, however, the motivation for such a regression is unclear, and so is the source of uncertainty in the estimates. I instead interpret these statistics as evidence that the model does not do extreme violence to the data and, although it does ignore part of the data variation, we can proceed with some confidence.[21]

5.2 Estimation of α and β

My goal is to calibrate α, β and γ to national income shares. Currently, I use data on technical requirements for growing a set of crops and interpret them through the lens of the model.[22] Ignoring the variation across crops, I focus on the averages and estimate $\hat{\alpha} = 0.5$, $\hat{\beta} = 0.3$ and $\hat{\gamma} = 1 - \hat{\alpha} - \hat{\beta} = 0.2$. Importantly, these estimates are in line with others in the literature. As a check, I compare my estimates to those in Hayami and Ruttan (1985), –later quoted in Restuccia, Yang, and Zhu (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. Importantly, the land share is almost the same.[23]

5.3 Estimation of θ using National Statistics and GAEZ data

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

$$\eta_{i,k} = \frac{\left(p_{i,k}^{\frac{1}{\gamma}} A_{i,k}\right)^{\theta}}{\Phi_i^{\theta}} \tag{22}$$

Now we tie this expression to the data on unconditional yields produced by GAEZ, following a similar logic to that of Costinot, Donaldson, and Smith (2012). 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

[21] In Sotelo (2013) I explore the deviations from the prediction in more detail. Also see Appendix G for this same exercise in the shares' natural scale.

[22] The source of data is Direccion Regional de Arequipa, a department that accounts for approximately 6% of the total value of agricultural production.

[23] The relative values of α and γ are also in line with those used in the GTAP 8 database. See Hertel, Tsigas and Narayanan (2013).

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 (11)

$$\tilde{y}_{i,k} = \kappa_y A_{i,k} p_{i,k}^{\frac{(\alpha+\beta)}{\gamma}} w_{i,A}^{-\frac{\alpha}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}}.$$

I assume that this object corresponds to the measures produced by the GAEZ project, 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.[24] Then we relate each observation in the GAEZ data to objects in the theory, particularly $A_{i,k}$:

$$\tilde{y}_{i,k}^G = \kappa_y A_{i,k} \left(p_k^G\right)^{\frac{(\alpha+\beta)}{\gamma}} \left(w_A^G\right)^{-\frac{\alpha}{\gamma}} \left(\rho^G\right)^{-\frac{\beta}{\gamma}} e^{u_{i,k}}$$

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 (22) to substitute for $A_{i,k}$ in the expression for $\tilde{y}_{i,k}^G$, we obtain an expression that relates GAEZ yield measures to observed land allocations and prices:

$$\log\left(p_{i,k}^{\frac{1}{\gamma}} \tilde{y}_{i,k}^G\right) = \frac{1}{\theta}\log\eta_{i,k} + k^G + \iota_k + \iota_i + u_{i,k}. \tag{23}$$

where, ι_k is a dummy that absorbs $\left(p_k^G\right)^{-\frac{\alpha+\beta}{\gamma}}$, ι_i absorbs $\log\Phi_i$, and k is a constant that absorbs

$$\log\kappa_y + log\left[\left(w_A^G\right)^{-\frac{\alpha}{\gamma}}\left(\rho^G\right)^{-\frac{\beta}{\gamma}}\right].$$

I construct the left-hand side of equation (23) imposing the restriction that the coefficient on log-prices is $\frac{1}{\gamma}$. The reason is that I want to focus attention on 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 (22)– we would predict that $A_{i,k}$ is relatively large, too. By the same logic we would 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 (23), I take a particular stand 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

[24]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 have a model for the spatial variation of those prices.

show zero potential productivity in the GAEZ data set. Another reason for measurement error is the way I process the GAEZ data and assign it to each Peruvian district, as explained below.[25]

Note that the unobserved heterogeneity $\Lambda_{i,k}(\omega)$ does not, by itself, generate an error term. In the model, there is a continuum of plots in each region, which means that the model's predictions about endogenous variables, like land allocation, must hold exactly. Hence, land-quality heterogeneity cannot serve as motivation for the error, unless one abandons the assumption that there is a continuum of plots.

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 its price and land quality, $p_{i,k}$ and $A_{i,k}$, together with a region-wide shifter motivated by Φ_i. We do not need to take into account the prices and land qualities of all other crops.[26]

Sample and Results

To estimate equation (23) 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 long-run, I focus on the long sample of National Statistics which averages more than ten years, and contains information for four departments. Averaging a longer time series reduces the variation induced by weather shocks, pests, etc.

To construct the $\tilde{y}^G_{i,k}$ data at the district level, I overlay the administrative division of Peru on top of the GAEZ grid. To deal with the fact that the boundaries of both divisions do not coincide, I further partition the GAEZ grid according to Peru's map. It often happens one cell gets assigned to more than one district; also, on many occasions, this procedure assigns pieces from more than one cell to a single district, in which case I assign the maximum cell value to that district.[27] The crops included in the regression are those that are both observed in national statistics and in the GAEZ dataset.

Table 3 shows the results of estimating equation (23). The coefficient on land allocation is 0.483, which implies an estimate $\hat{\theta}= 2.06$. Figure 3 shows the variation that identifies $\hat{\theta}$: it relates $\log\left(p_{i,k}^{\frac{1}{\gamma}}\tilde{y}^G_{i,k}\right)$ to $\log\eta_{i,k}$ after removing district and crop fixed effects. The coefficient is precisely

[25] This motivation for the error is related, but conceptually different from, a common treatment of errors in the quantitative trade literature. Here we assume that the theoretical object of interest is measured with noise, and hence we try to predict it with observables. In contrast, 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.

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

[27] To match the GAEZ grid to the districts, I use the actual administrative division of Peru, from which I cannot observe what fraction of the land is actual agricultural land. The quartiles of the administrative district-size distribution are $93.8km^2$, $210.6km^2$ and $497.7km^2$. Due to the way it is projected, the cell size in the GAEZ grid is approximately $86.km^2$ at the Equator, but it grows larger at higher latitudes. The fact that the total agricultural land is usually much smaller than the total amount of land in a district justifies using the maximum GAEZ value for each district that contains more than one GAEZ cell.

estimated. This value of $\hat{\theta}$ implies a large elasticity of land allocation with respect to prices: $\theta/\gamma \approx \frac{2}{0.2} = 10$. To understand this magnitude, consider an exogenous change in the relative price of two crops –as could happen, for example, with a change in international prices. If the price of potatoes relative to tomatoes increases by one percent, then the allocation of land to potatoes relative to tomatoes increases by about 10 percent.[28]

Note that a key assumption to obtain estimating equation (23) 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. In Appendix G, I discuss how to treat the estimating equation in a way similar to an errors-in-variables model. I discuss the possibility of bounds identification for dealing with measurement error in and $\eta_{i,k}$, exploiting the method proposed by Leamer (1978).[29] [30]

5.4 Absolute and Relative Values of Land Quality Parameters, $A_{i,k}$

To estimate the $A_{i,k}$ parameters I rely heavily on the model structure. As I have discussed in detail in Section 3, Assumption #2 imposes strong restrictions on what data are informative about land quality. 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. Recall, that relative values of $A_{i,k}$ are not directly observable, except for the limiting case where $\theta \to \infty$, in which the observed land yield of crop k fully reveals $A_{i,k}$.

My approach, which extracts the model parameters using data on the endogenous variables, is an alternative to the use of external measures of productivity.[31] Costinot, Donaldson, and Smith

[28] Santos Silva and Tenreyro (2006) have argued that to estimate constant elasticity models by taking logarithms and running a linear regression, one must make more stringent assumptions about the error than would be necessary if one used, for example, a Poisson Pseudo Maximum-Likelihood estimator. I will explore alternative estimators in future versions of the paper.

[29] Chapter 7, p. 239; see also Cameron and Trivedi (2005), pp. 906-908, and the references there.

[30] 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. Together with the assumption that the potential productivity does not vary much over time –so we can subsume it in a time-invariant crop fixed effect– we can use directly the land allocation equation

$$\log \eta_{i,kt} = \theta \log p_{i,kt} + a_{i,k} + b_{i,t} + u_{i,kt},$$

where $a_{i,k} = \log A_{i,k}$ and $b_{i,t} = \log \left[\left(\sum_k \left(p_{i,kt}^{\frac{1}{\gamma}} A_{i,k} \right)^{\theta} \right)^{\frac{1}{\theta}} \right]$.

[31] The idea of combining the model structure with data on endogenous variables to estimate the model's primitives has antecedents in the trade literature. My approach is a variation on Eaton and Kortum (2002), who recover the productivity parameters in a Ricardian trade model from the fixed effects in a gravity equation. Waugh (2010) discusses the correct econometric specification of the gravity equation, 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, Donaldson, and Komunjer (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, Hurst, Jones, and Klenow (2013) back out the frictions to the allocation of labor across occupations using the structure that a Fréchet distribution for unobserved talent imposes.

(2012) and Costinot and Donaldson (2011), 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. The main shortcoming of this method 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.

I start by relating the aggregate regional data on output and factor endowments in region i to that region's aggregate **level** of productivity. Under the assumption that labor is immobile between sectors and districts, we can use equation (18), and solve for Φ_i as a function of observable variables:

$$\Phi_i = \kappa_y^{-\frac{1-\beta}{\gamma}} \kappa_l^{\frac{\alpha}{\gamma}} H_i^{-\frac{1-\beta}{\gamma}} \left(\frac{H_i}{L_i}\right)^{\frac{\alpha}{\gamma}} V_i^{\frac{1-\beta}{\gamma}} \rho_i^{\frac{\beta}{\gamma}}. \tag{24}$$

Equation (24) strips from the total value of production, V_i, the contribution of endowments, H_i, L_i, and the intermediate input (through its price ρ_i). This residual is informative about the aggregate level of productivity, as shown in the distribution of yields and revenues in Proposition 2. A higher value of V_i relative to inputs will lead us to infer a higher land quality for all crops in i, because it means that region i produces more value for a given factor use.[32]

Just like aggregate output and endowments are informative of a common component of land quality for all crops in region i, data on prices and land allocations are informative about the relative land qualities within that region. Using equation (7) to solve for $A_{i,k}$ we obtain:

$$A_{i,k} = p_{i,k}^{-\frac{1}{\gamma}} \eta_{i,k}^{\frac{1}{\theta}} \Phi_i, \tag{25}$$

which we can take to data because $p_{i,k}$ and $\eta_{i,k}$ are measured directly, and equation (24) tells us how to measure Φ_i with the regional aggregates. It is clear that these estimates, $\hat{A}_{i,k}$, are independent of the unit of account in the data, since equation (25) is homogeneous of degree zero in prices.

Let us take a moment to interpret this equation. As already said, the residual $\hat{\Phi}_i$ 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 a large price $p_{i,k}$, which also tends to generate

[32] Recall that equation (14) relates the expected yield in crop k to Φ_i:

$$\mathbb{E}\left[\psi_{i,k}(\omega)\,|\omega \in \Omega_{i,k}\right] = \kappa_y \Phi_i w_i^{-\frac{\alpha}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}}.$$

In principle, solving for Φ_i from this equation, we obtain one estimate of Φ_i, for each $k \in \mathbb{K}_i$, conditional on wages and the price of intermediates:

$$\hat{\Phi}_i\left(\hat{\psi}_{i,k}\right) = \frac{\hat{\psi}_{i,k}}{\kappa_y w_i^{-\frac{\alpha}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}}}.$$

But note also that the theory predicts no variation in $\mathbb{E}\left[\psi_{i,k}(\omega)\,|\omega \in \Omega_{i,k}\right]$ across crops. The within-region variation in the expected yield is the same as the variation around the model's prediction in Figure (2). The approach I follow in the paper amounts to constructing an average of all the $\hat{\Phi}_i\left(\hat{\psi}_{i,k}\right)$, where each crop is weighted by the fraction of land allocated to it.

a large land allocation to crop k.

There is an alternative interpretation of equation (25) that will help understand the results of the simulations in Sections 6 and 7. The estimate of $A_{i,k}$ combines information on prices and land allocations. In this estimation, how important is the variation in land allocations relative to the variation in prices? Variation in land allocations is more important the larger $1/\theta$. Given γ, a lower value of θ (high heterogeneity) gives less importance to land allocations in the estimation of $\hat{A}_{i,k}$. Thus, with lower values of θ the model will simulate land allocations that are closer to the data.

Last, note that 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. Even if the model is correct, we are ignoring the sampling variation and hope for an unbiased estimate of $A_{i,k}$, at best.

Sample and Results

My primary goal in obtaining $\hat{A}_{i,k}$ is coverage, so I sacrifice precision in the estimation to be able to obtain an estimate for every region in the country. Hence, I use the wide sample of national statistics, which contains repeated cross-sections from 2008-2011 and covers the whole country. Data on $p_{i,k}$, and $\eta_{i,k}$ are averages across time. I use the corresponding data on land yields, $y_{i,k}$, to construct the total value of production in region i, $V_i = \sum_{k \in \mathbb{K}_i} p_{i,k} y_{i,k} \eta_{i,k} H_i$. The data on agricultural labor comes from the 2007 population census. I delay the description of the intermediate input price to later in this section.

Table 5 shows, for each crop k, the summary statistics of the estimates of $A_{i,k}$. The estimates vary substantially between crops, reflecting the fact that the price of a ton of output also varies much between crops. For example, the average price of a kilogram of coffee in the data is 6.45 LCU, while that of cassava is 0.46 LCU; accordingly, $\bar{A}_{i,coffee}$=2.63 and $\bar{A}_{i,cassava}$=$2,088,482$.

5.5 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.

5.5.1 Projection of Freight Rates on Road and Geography Data

I follow Donaldson (2010) 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[33], and (iii) the shape, length and quality of the road network, which I group into high quality (paved), medium quality (graded) and low quality (unpaved).

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 100 km. apart.[34]

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:

$$\mathbb{E}\left[\log f_{ni} | \text{geography, roads}\right] = \beta_0 + \beta_{\text{distance}} \times \log\left(\text{effective distance}_{ni}\left(\lambda\right)\right). \tag{26}$$

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

$$\text{effective distance}_{ni}\left(\lambda\right) = \sum_{q \in \{\text{high, med, low}\}} \sum_{\text{edge} \in E_{q,ni}(\lambda)} \left[\exp\left(\lambda_s s_{\text{edge}}\right) \cdot \left(\lambda_q \text{distance}_{\text{edge}}\right)\right]. \tag{27}$$

In equation (27), $E_{q,ni}\left(\lambda\right)$ is the set of edges of quality q that form part of the optimal route between i and n; λ_q is the cost of traversing a kilometer of road of quality q, and λ_s is the effect of traversing an edge with slope s_{edge}. Without loss, we normalize the weight of high-quality distance, λ_{high}, to one.[35]

Figure 4 compares the fit of three versions of the model: (i) a model that constrains $\lambda_q = 1$ and $\lambda_s = 0$, (ii) a model that constrains only $\lambda_s = 0$, and (iii) a completely unconstrained model. It is clear that taking into account the quality and geography of the road substantially improves the estimation. Table 6 shows the point estimates of each model.[36] I focus attention on the completely unrestricted model (column 3), which is the one I use to predict freight rates out of sample. I find that $\hat{\lambda}_{\text{med}} = 3.09$, $\hat{\lambda}_{\text{low}} = 9.416$, $\hat{\lambda}_s = 15.7$ and $\hat{\beta}_{\text{distance}} = 0.857$. To interpret these values, suppose that the route between two regions n and i is completely paved and flat. If that same route were unpaved, and the slope were at the 90th percentile in the sample (0.023), the freight cost would increase by a factor of $\left(\exp\left(\hat{\lambda}_s \cdot 0.023\right)\hat{\lambda}_{\text{low}}\right)^{\hat{\beta}_{\text{distance}}} = 9.31$. With these estimates at hand, we predict freight rates for the whole country based on its geography and the quality of its road

[33] The size of each cell in the grid is 30 arc-seconds by 30 arc-seconds.

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

[35] Expression (27) 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 relative cost,λ. In the extreme, if $\lambda_{\text{med}} = \lambda_{\text{low}} = 1$, the algorithm minimizes the simple distance between two points. As λ_{med} and λ_{low} grow, the algorithm gives priority to high-quality edges.

[36] The standard errors remain to be computed.

network.

5.5.2 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, $\bar{p}_k$:

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

This calculation 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.[37]

For any region i in Home, the cost of trading with Foreign 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 $o(i)$ is the port closest to i. Thus, I compute for each good:

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

Results

Table 7 summarizes the distribution of $\hat{d}_{ni,k}$, by crop k, pulling together across all pairs of regions. The most salient feature of the table is that my estimates of $d_{ni,k}$ are low, which has two possible explanations. First, as suggested by Figure 4, the model of transportation is unable to replicate transportation costs at the higher end of the sample. Second, 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.[38] When we evaluate the model at the baseline equilibrium, it will become clear that these measures of transportation costs generate too much specialization and too high intermediate input use.

5.6 The Price of Intermediate Inputs

The cost of intermediate inputs in region i, ρ_i, has two components: the price of the input bundle at the port and the cost of transporting it to region i. To calculate the price of the good at the port, I construct a bundle of fertilizers and average their price. This is akin to assuming that, to obtain a unit of intermediate, farmers combine all available fertilizers in fixed, equal proportions, as with a Leontieff production function. In Table 8 I show the unit FOB price and the import quantity of each fertilizer I include in the input bundle. I take the simple average in the bottom row to be

[37] See Hummels and Skiba (2004)

[38] For example, Chaney (2011) has recently explored the implications of networks in trade. Allen (2012) has shown that information costs are an important part of total trade costs.

ρ_F.[39] Then I multiply the cost of the bundle of the input by the each region's cost of trading with Foreign, $\hat{d}_{iF,x}$

$$\rho_i = \hat{d}_{iF,x}\rho_F.$$

5.7 Estimation of Domestic Demand Parameters

I set $\sigma = 3$, with the notion that substitution in consumption is relatively easy between similar crops (e.g., grains) but not between different crops (e.g., vegetables versus grains). I am currently working on estimating σ by combining household expenditure and unit-value (consumption price) data, with the estimates $\hat{A}_{i,k}$, which the model suggests should be good instruments for unit prices. [40] Estimation of the other parameters is described in the Appendix.

5.8 Assessing the Sources of Productivity

At this point we have estimated the transportation costs, $\hat{f}_{ni}$, which allow us to calculate the cost of the intermediate input in each region, ρ_i. We have also measured the productivity component Φ_i implied by the aggregate regional data. We are in a position to assess the contribution of each to agricultural productivity.

The first column in Table 4 shows the estimates of TFPR. The variation in TFPR is large: the difference between the 90th and the 10th percentile of the distribution is a factor of 8.

According to the Table (third column), the variation in $\rho_i^{-\frac{\beta}{1-\beta}}$ is relatively small, and hence accounts for a small part of the variation in TFPR. Consider the 90th and 10th percentiles in the distribution of access to intermediate inputs. The ratio of the effect of this channel is $\frac{0.98}{0.93} = 1.05$, which means that moving from the 90th to the 10th percentile increases value productivity by 5 percent.

The effect of Φ_i is much larger (middle column), but we must bear in mind two caveats. First, I obtain the component as a residual: all the variation in labor and land productivity that cannot be accounted for by variation in ρ_i is attributed to Φ_i. Second, as discussed above, Φ_i captures more than the effect of specialization based on comparative-advantage, as better land quality on average will translate into a higher value of Φ_i, regardless of trading possibilities. Therefore, to elicit the effect of trade policy on productivity, we work with counterfactuals in Section 7.

[39] The data on costs of production described in Section 5.2 contains information regarding the cost share of different types of intermediates. For the goods for which I have data on production costs, the cost share of fertilizer is a large fraction of the cost share of intermediates. FAOSTAT data for the years 2008 and 2009 show: (i) the fertilizers included in the Table account for more than 95 percent of total imports of fertilizer, (ii) imported fertilizer is more than 99 percent of consumption, (iii) exports are about 3 percent of consumption (all measured by weight). Taken together, this evidence suggests that the assumption that intermediate inputs are imported from abroad is not too far from reality.

[40] A value of $\sigma = 3$ 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. My preliminary results, which also suggest that $\hat{\sigma} = -1.2$, point in the same direction.

6 Simulation and Evaluation of the Model

The first task is to simulate the model and compare its predictions to the data.[41] We identify the features of the data that the model is able to replicate and we study why the model cannot capture other aspects.

6.1 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. The comparisons I show next are only for those observations. This is important because, to the extent that the model predicts a higher price in an non-producing region –which we would expect in reality–, we will not be able to observe such price variation in the following figures. All we can examine is whether the model can replicate producer prices, which I think is a tougher test of the model's performance.

The top panel of Figure 5 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 allowed them to capture the crop-specific variation in prices. For example, if on average a ton of potatoes is cheaper than a ton of asparagus, then on average $\hat{A}_{i,\text{potatoes}}$ will be larger than $\hat{A}_{i,\text{asparragus}}$. Table 5 already suggests the outcome, as it shows large differences in $\hat{A}_{i,k}$ between crops. Note, however, that the model predicts too high prices relative to the data.

Consider now the bottom panel of Figure 5, which compares land allocation $\eta_{i,k}$ in the data and model (in a log scale). The top panel shows clearly that, at the baseline parametrization, the model predicts extreme patterns of specialization relative to the data. This is shown by the spread of the predictions, which is much larger than in the data.

Generally, the model will not reproduce exactly the endogenous variables, even when we estimated many of the parameters of the model by looking precisely at those data, in particular producer prices. The key is that prices are equilibrium outcomes, too, and neither preferences, technology nor geography are exact representations of reality.

The model does not match the data perfectly for at least four reasons reasons. First, in Section 5, I estimated transport costs to be quite modest. The model did not do a good job of fitting the larger freight-rate observations. More importantly, as has been argued in the literature, freight-rates are just a small part of the story when it comes to trade frictions. Hence, in the simulation, the model allows every region in Peru to reap almost entirely the international or urban price for the cash crops they sell there; through the factor markets, this raises the prices of all other crops. Low transportation costs also induce too much specialization, which increases productivity.

[41] Calculating the equilibrium with more than 1800 districts and 51 crops is a computationally daunting task. In this version of the paper, therefore, by aggregating the Peruvian data at the province level, I reduce the number of regions to 194. The number of crops remains the same.

Second, recall that the elasticity of land allocation with respect to prices is relatively large, $\theta/\gamma \approx \frac{2}{0.2} = 10$. Such a large magnitude explains why the model predicts extreme patterns of specialization we observed before. Recall that unbiased estimation of θ by means of estimating equation (23) requires that $\eta_{i,k}$ be measured without error. If not, attenuation bias will yield too high an estimate of θ. If we simulate the model with too little heterogeneity (high θ) and too low transportation costs, the model will deliver too much specialization and too high producer prices. This is the rationale for the alternative calibration.

Third, I estimate the scale productivity parameters, $A_{i,k}$, by taking the land allocation equation as correct and then calculating the values of $A_{i,k}$ to reflect both $p_{i,k}$ and $\eta_{i,k}$, with relative importance determined by θ. The land allocation equation is only true under the extreme assumptions that (i) the land productivity is, in fact, distributed like a set of i.i.d. Fréchet random variables and (ii) the technology is Cobb-Douglas and constant across crops. As has been discussed before, these assumptions have empirical implications that are not perfectly reflected in the data and at this stage we pay the price of tractability.[42]

Finally, I assume that farmers are entirely driven by expected profit maximization, which is at odds with reality, as discussed in the development literature. The primary concern is that cropping choices also reflect farmers' intention to diversify risk, so the model predicts too much specialization.

6.2 The Rationale for an Alternative Parametrization

The performance of the model, in terms of its predictions, can be improved. The model predicts too much specialization and too high equilibrium prices (in terms of the numéraire). A large part of the performance of the model can be traced back to the $\hat{\theta}$ estimate, which implies that land allocation is too sensitive to changes in relative prices, and to the low estimates of $\hat{d}_{ni,k}$, which allow for too large trade integration both within Peru and with the rest of the world. In what follows, I compare the baseline parametrization with an alternative one. In the alternative parametrization, first, I set $\hat{\theta}' = 1.25$, lower than the baseline $\hat{\theta} = 2.06$, and which goes in line with the possibility of attenuation bias in the estimation of equation (23). Second, I increase trade frictions according to

$$\hat{d}'_{ni,k} = 1 + \delta\left(\hat{d}_{ni,k} - 1\right),$$

which means that I increase the trade frictions preserving the order in that distribution, and set $\delta = 7$. This value of δ sets the average level of simulated prices to the one in the data. The value of $\hat{\theta}'$ aims at improving the fit the data on land shares.

With this in mind, consider the top panel of Figure 6, which compares farm-gate price data with simulated prices under the alternative parametrization. The model does now better at reproducing the level of prices relative to the numéraire. Also, there is a small increase in the correlation between

[42] Choosing parameters to fit bilateral trade flows perfectly is the approach followed by Dekle, Eaton, and Kortum (2008) and, more recently, Caliendo and Parro (2012), Parro (2013), Ossa (2011). This approach brings other benefits that I cannot exploit, chief among them, the ability to compute counterfactuals without estimating all the primitives in the model.

data and simulation within crops, as shown in Table 9. In that table, as a measure of fit, we regress the price data on the model simulation using crop fixed effects, because we want to remove the variation that is due to the fact that some crops are simply more expensive than others on average. The fact that in Column (1) the estimate is positive, but larger than one (1.77) means that the baseline model generates the right correlation, but too little price variation relative to the data. Column (3) shows that the alternative parametrization the slope estimate is closer to, but smaller than, 1.

In contrast, Panel B shows that increasing trade frictions and increasing the heterogeneity improves the prediction about land allocations substantially: there is less specialization and a tighter relation between model and data. This improvement is quantified in Table 10, which pools between and within-crop variation. Regressing data on land shares on the model's prediction, we see that both models generate the right correlations, although the alternative parametrization improves on the baseline.

Being assured that the model gets the main correlations right –even if it could be improved– we now turn to a counterfactual to evaluate the effect of market access on productivity.

7 The Effect of Improving Market Access

In this section I ask: What are the effects of improving the access of an isolated region to a large market? In a more realistic counterfactual, which I will pursue in a future version of this paper, 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. The feasibility of the conterfactual will be then clear, and the exercise will be more in line with the spirit of the paper, which emphasizes the general equilibrium interactions of small productive units.

To gain a quick sense of the magnitudes involved, I select a small and isolated region and evaluate the effect of improving its connection to Lima, the largest urban market in the country. I proceed in two steps. First, in the distribution of freight rates of shipping to Lima, $\hat{f}_{Lima,i}$, I identify the region that corresponds to the 90th percentile of that distribution. Denote that region by i_1. Second, to capture an improvement in the transportation network, I reduce the freight rate of trading to and from Lima, $\hat{f}_{Lima,i_1}$ and $\hat{f}_{i_1,Lima}$, to the 50th percentile of the distribution of $\hat{f}_{Lima,i}$.

I will compare the estimated productivity gain from in both parametrizations. As will become clear, the alternative parametrization, gives more room for productivity gains.

Productivity Gains Under the Baseline Parametrization

Productivity is improved by almost 10 percent. Let us start by decomposing the effect on value productivity of land and labor. Recall that in this model, the equilibrium TFPR of land and labor is proportional to

$$\rho_i^{-\frac{\beta}{1-\beta}} \Phi_i^{\frac{\gamma}{1-\beta}},$$

as shown in equation (18). The first term, $\rho_i^{-\frac{\beta}{1-\beta}}$, captures the effect of access to intermediate inputs (the first channel in the theory). This comes about because, in this particular example, Lima becomes the closest port –in the sense of lowest cost– and so the price of the intermediate input decreases from 1.182 to 1.149 in terms of the numéraire, ρ_F. The productivity increase due to this channel is approximately 0.43 for every percent point decrease in the price of the intermediate:

$$\begin{aligned} &\approx && -\frac{\beta}{1-\beta} \times \text{percent change in } \rho_i \\ &= && 1.2\% \end{aligned}$$

The second channel comes from better land use, due to improved specialization. The percent change on value productivity is $\frac{\gamma}{1-\beta}$ for each percentage point increase in Φ_i:

$$\begin{aligned} &\approx && \frac{\gamma}{1-\beta} \times \text{ percent change in } \Phi_i \\ &= && 8.7\% \end{aligned}$$

The total increase in value productivity is almost 10 percent.

Productivity Gains Under the Alternative Parametrization

Under the alternative, when we set $\hat{\theta}' = 1.25$ and $\delta = 7$, the total gain in TFPR is about 35 percent. In this scenario, 4.4 percent is gained by having access to cheaper intermediate inputs –the price drops from 2.27 to 2.04–, while 30.7 percent is gained by a better land allocation.

Why do the gains triple when moving from the baseline to the alternative parametrization? Most of the increase in the productivity gains come from the change in the Φ_i term. As discussed before, when trade costs are high as in this parametrization, there is a negative relationship between prices and land quality. Such relationship arises because the market for each crop tends to clear locally. Lowering the trade costs that the region faces to trade with a large market severs the link between that region's local prices and its land quality, inducing farmers in that region to allocate more land to those crops in which they are better or for which they can now reap a higher price.

8 Conclusion

In this paper, I estimate the effect of market access on agricultural productivity by tightly 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 central message of the paper is that barriers to market access have a negative effect on farmers' productivity, especially by preventing them from allocating land to its most valuable use. Take the distribution of costs of trading with Lima, Peru's largest urban market. According to my baseline estimation, moving a remote region from the 90th to the 50th percentile of that distribution increases Revenue TFP by about 10 percent. Under an alternative parametrization – which improves the model's fit of the data – the productivity increase is approximately 35 percent. Hence, the isolation of low-productivity regions suggests that trade frictions help account for within-country dispersion in productivity.

A broader contribution of this paper is to develop an applied equilibrium framework to study the interaction between urban and rural economies that are not completely integrated. In future work, I will extend this tool and apply it to other important questions at the intersection of trade, regional, and development economics. My approach, for example, is well suited to study the spatial effect of a shock to world food prices, as trade frictions isolate some areas from world-market volatility and prevent stronger supply responses from those farmers. Moreover, since Engel's Law is a prominent feature in consumption data, augmenting the model in this paper to incorporate non-homothetic preferences will shed light on how income inequality shapes the urban-rural exchange within countries.

Finally, in this paper I have uncovered two puzzling features of the data. First, I documented that to make sense of observed prices and land allocations, we need large barriers to the movement of goods across space. But these barriers are an order of magnitude larger than those implied by freight rate data, which raises the immediate question of what additional frictions they capture. Second, I estimated large geographic dispersion in productivity within Peru. Why 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 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(2).

Allen, T. (2012): "Information Frictions in Trade," 2012 Meeting Papers 125, Society for Economic Dynamics.

Alvarez, F., and R. J. Lucas (2007): "General equilibrium analysis of the Eaton-Kortum model of international trade," *Journal of Monetary Economics*, 54(6), 1726–1768.

Anderson, J. E., and Y. V. Yotov (2010): "The Changing Incidence of Geography," *American Economic Review*, 100(5), 2157–86.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012): "New Trade Models, Same Old Gains?," *American Economic Review*, 102(1), 94–130.

Armington, P. S. (1969): "A Theory of Demand for Products Distinguished by Place of Production," *IMF Staff Papers*, 16(1), 159–178.

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(4), 666–72.

Caliendo, L., and F. Parro (2012): "Estimates of the Trade and Welfare Effects of NAFTA," NBER Working Papers 18508, National Bureau of Economic Research, Inc.

Cameron, A., and P. Trivedi (2005): *Microeconometrics: Methods and Applications*. Cambridge University Press.

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

Costinot, A., and D. Donaldson (2011): "How Large are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-2002," .

——— (2012): "Ricardo's Theory of Comparative Advantage: Old Idea, New Evidence," *American Economic Review*, 102(3), 453–58.

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(2), 581–608.

Costinot, A., D. Donaldson, and C. Smith (2012): "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from a 9 Million-Field Partition of the Earth," .

Deaton, A. S., and J. Muellbauer (1980): "An Almost Ideal Demand System," *American Economic Review*, 70(3), 312–26.

Dekle, R., J. Eaton, and S. Kortum (2008): "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, 55(3), 511–540.

Donaldson, D. (2010): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," Working Paper 16487, National Bureau of Economic Research.

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

Fajgelbaum, P., and S. Redding (2013): "External Integration and Internal Development: Evidence from Argentina, 1870-1914," .

Fieler, A. C. (2011): "Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation," *Econometrica*, 79(4), 1069–1101.

Gollin, D., D. Lagakos, and M. E. Waugh (2011): "The Agricultural Productivity Gap in Developing Countries," Working Papers 11-14, New York University, Leonard N. Stern School of Business, Department of Economics.

Gollin, D., S. L. Parente, and R. Rogerson (2007): "The food problem and the evolution of international income levels," *Journal of Monetary Economics*, 54(4), 1230–1255.

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.

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.

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

Hummels, D., and A. Skiba (2004): "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture," *Journal of Political Economy*, 112(6), 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(2), 948–80.

Leamer, E. (1978): *Specification searches: ad hoc inference with nonexperimental data*, Wiley series in probability and mathematical statistics. Wiley.

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.

Mas-Colell, A., M. Whinston, and J. Green (1995): *Microeconomic Theory*, Oxford student edition. Oxford University Press.

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. (2011): "Trade Wars and Trade Talks with Data," Working Paper 17347, National Bureau of Economic Research.

Parro, F. (2013): "Capital-Skill Complementarity and the Skill Premium in a Quantitative Model of Trade," *American Economic Journal: Macroeconomics*, 5(2), 72–117.

Ramondo, N., and A. Rodriguez-Clare (2013): "Trade, Multinational Production, and the Gains from Openness," *Journal of Political Economy*, 121(2), 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(2), 234–250.

Santos Silva, J. M. C., and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658.

Scarf, H., and T. Hansen (1973): *The Computation of Economic Equilibria*, Cowles Foundation Monograph Series. Yale University Press.

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

Tombe, T. (2012): "The Missing Food Problem," Working Papers tt0060, Wilfrid Laurier University, Department of Economics.

Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

Waugh, M. E. (2010): "International Trade and Income Differences," *American Economic Review*, 100(5), 2093–2124.

9 Tables and Figures

Table 1: Description of Data Sets

Name	Source	Unit of Observation	Contents	Time	Coverage
National Statistics on Agriculture	Ministry of Agriculture of Peru (MINAG)	(Crop, District) pairs	-Land use -Physical land yields -Price	-Wide Sample: 2008-2011 -Long Sample: 1999-2011	-Wide Sample: 24 departments which produce agricultural goods -Long Sample: 4 selected departments.
Global Agro-Ecological Zones (GAEZ)	FAO-IIASA	5 arc-minute x 5 arc-minute grid of the country, by crop	Potential land yield, if all the land in the pixel is used in that crop	Average for 1961- 1990	All of Peru
National Household Survey (ENAHO)	Instituto Nacional de Estadistica e Informatica (INEI)	Household	- Agricultural Producer Module - Food Consumption Module	2008-2011	Nationally Representative Samples
Transport Network	Ministry of Transport and Communications of Peru (MTC)	Each road in the national system of roads (National, Departmental, Rural)	Exact location, shape, length and road quality of each road	As of 2012	All of Peru
Altitude	Shuttle Radar Topography Mission	3 arc-second by 3 arc-second grid of the country	Altitude of each cell	-	All of Peru
Freight Rates	Direccion Regional de Agricultura de La Libertad	Sample of 45 pairs of districts where at least one district belongs to La Libertad	Cost in LCU of shipping different goods, per corresponding unit of measure	Years 2011-2013	At least one of the trading districts belongs to the department of La Libertad
Price of fertilizer	Direccion Regional de Agricultura de La Libertad	(Fertilizer type, Province) pairs	Price in LCU per kilogram	2008-2011	- All provinces in La Libertad. - Main fertilizers
Employment	National Population and Housing Census	1838 Districts in Peru	Number of people working in agriculture as main or secondary activity	2007	All of Peru
Technical Requirements	Direccion Regional de Agricultura de Arequipa	(Crop, Province) pairs	Complete list of costs and activities to grow a crop	-	- Sample of crops - Sample of Provinces in Arequipa
Manufacturing GDP	INEI	Department	Total GDP in LCU	Years 2008-2011	All of Peru

Table 2: An Empirical Assessment of Proposition #3: Equalization of Land and Revenue Share

	(1)
	$\log \pi_{ik}$
$\log \eta_{ik}$	1.017^{***}
	(0.00119)
Observations	22717
R^2	0.970
Adjusted R^2	0.970

Standard errors in parentheses

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

Table 3: Estimation of Inverse Heterogeneity θ

	(1)
	$\log \tilde{y}_{ik} p_{ik}^{1/\gamma}$
$\log \eta_{ik}$	0.483^{***}
	(0.0656)
Crop FE	Yes
Observations	640
R^2	0.851
Adjusted R^2	0.801

Standard errors in parentheses

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

Table 4: Summary Statistics of the Estimates of Value Productivity and its Components

	Value Productivity	$\Phi^{\frac{\gamma}{1-\beta}}$	$\rho^{\frac{-\beta}{1-\beta}}$
count	1751.00	1751.00	1751.00
mean	3.51	6.13	0.96
std	3.46	5.96	0.02
min	0.03	0.05	0.81
10.0%	0.93	1.63	0.93
50%	2.43	4.27	0.96
90%	7.37	12.98	0.98
max	43.54	73.31	1.00

Value Productivity is the model-based measure $\beta^{\frac{\beta}{1-\beta}} \Phi_i^{\frac{\gamma}{1-\beta}} \rho_i^{-\frac{\beta}{1-\beta}}$.

Table 5: Summary Statistics of the Estimates of Land Quality, $\hat{A}_{i,k}$

	count	mean	std	min	10.0%	50%	90%	max
achiote	83.00	92.11	245.39	0.02	0.16	6.76	254.22	1655.83
apple	417.00	33212.91	219530.80	0.13	8.95	310.80	32188.76	2775583.97
asparragus	77.00	12516.62	71009.09	0.48	9.34	203.70	5193.63	594706.70
avocado	580.00	25054.63	230033.49	0.00	2.03	143.04	14062.67	5022709.61
banana	553.00	867464.82	3684655.11	0.52	674.45	55407.80	1400240.02	44298477.87
barley grain	1112.00	4467.06	20078.43	0.07	11.62	333.26	7606.31	375235.69
cacao	240.00	5.43	23.52	0.00	0.02	0.38	7.47	231.09
canahua	87.00	154.94	605.56	0.00	0.25	9.07	253.02	4407.37
cassava	605.00	2135658.75	24609250.39	0.21	194.36	18127.43	620113.50	419824888.51
chocho	425.00	84.56	465.81	0.00	0.12	4.27	118.21	7081.56
coconut	176.00	1653012.74	10982079.93	28.61	1109.94	30113.91	1168967.72	136411196.36
coffee	329.00	2.73	8.71	0.00	0.01	0.34	5.23	100.52
cotton branch	190.00	463.63	1639.91	0.02	3.54	84.37	629.01	18265.62
dry bean	983.00	2740.42	57181.16	0.00	0.23	7.57	344.86	1304334.94
dry faba bean	1007.00	535.95	3717.85	0.01	1.12	24.47	632.26	96761.76
dry pallar	86.00	550.98	2207.90	0.08	0.39	32.64	683.89	18966.21
dry pea	927.00	321.81	2882.22	0.00	0.74	12.89	248.94	59290.32
dry zarandaja	56.00	863.43	1502.04	0.01	13.15	371.31	2115.04	8424.54
garbanzo beans	40.00	32.79	48.96	0.02	0.15	5.10	89.71	212.73
garlic	319.00	919.28	8078.27	0.00	0.01	1.19	150.24	103901.98
grape	184.00	9444.55	31220.54	0.01	3.27	544.37	21304.34	321182.84
green bean	186.00	4312.90	12440.97	0.05	9.78	468.38	11345.38	132596.08
green faba bean	733.00	17299.04	101986.56	0.17	25.89	589.80	24406.42	1974391.61
green pea	787.00	2183.41	10271.46	0.01	4.46	114.47	2858.53	139932.64
key lime	471.00	210570.76	2132120.87	0.21	27.69	1222.64	123494.76	40151483.30
kiwicha amaranth	131.00	19.80	107.16	0.00	0.04	0.54	13.37	920.22
lentil	98.00	53.66	167.95	0.02	0.16	2.72	105.94	1023.32
maize (amilaceo)	1205.00	1286.94	11056.93	0.00	1.65	33.78	1185.46	270781.79
maize (choclo)	827.00	107170.22	983639.15	0.33	41.21	1567.24	65751.07	24149886.19
maize (yellow hard)	784.00	81366.18	387685.02	0.08	105.33	4892.28	143491.95	7920207.66
mango	321.00	411525.82	2676759.70	0.08	279.02	10261.34	393930.52	41661778.95
mashua	513.00	81788.90	333932.25	0.67	97.03	2772.08	170315.85	4544737.25
oca	846.00	57667.81	324632.41	1.04	40.99	1521.64	84452.90	7100334.29
oil palm	17.00	1418494.54	2551171.95	623.03	25987.13	170687.66	3784236.12	9358622.37
olive	53.00	4697.60	14570.01	0.00	1.80	232.55	8686.91	103458.71
onion	466.00	561702.65	3506826.69	0.02	55.29	3342.55	376010.74	51318013.32
orange	558.00	6671274.44	53859829.70	0.14	143.70	4638.83	485633.62	623610746.69
papaya	346.00	356338.48	2284115.16	1.69	100.51	6665.46	215326.96	32116991.94
pecan	33.00	1.08	3.60	0.00	0.00	0.12	1.27	20.13
pineapple	204.00	133462.08	429556.72	4.90	308.15	7633.06	225257.06	3155568.68
potato	1357.00	153164.54	1002180.63	0.01	101.87	3950.06	187529.28	24443177.41
quinoa	673.00	94.74	2005.73	0.00	0.04	0.75	22.11	51975.60
rice	367.00	96492.78	414723.48	1.30	83.34	7099.09	179183.44	6052610.84
sorghum	5.00	231563.69	330597.69	139.30	13638.26	46569.40	587960.54	785356.58
soy	84.00	249.73	738.53	0.05	3.28	26.69	391.07	4657.77
sweet potato	369.00	1138676.55	5675634.16	0.03	136.30	11647.73	1470353.26	73746544.01
tangerine	173.00	139871.53	600850.21	17.49	483.90	8819.21	302566.47	5467275.46
tea	4.00	60439.21	56517.66	158.52	12086.52	53558.65	114296.34	134481.00
tomato	390.00	208494.54	1744028.92	0.06	43.96	2969.91	105382.66	23873396.81
ulluco	1027.00	9827.06	41248.40	0.24	19.00	700.22	16308.55	568679.84
wheat	1119.00	9865.68	119939.03	0.02	3.32	98.54	2536.53	3383171.19

Table 6: Estimates of the Transportation Model

	$\lambda_q = 1, \lambda_s = 0$	$\lambda_s = 0$	Unconstrained
Effective Distance, β_{distance}	0.272	0.897	0.857
Relative Cost of Low Quality Road, λ_{med}	1.000	4.672	3.09
Relative Cost of Low Quality Road, λ_{low}	1.000	12.843	9.416
Effect of Slope, λ_s	0.000	0.000	15.700
Intercept, β_0	-3.465	-7.8665	-7.569
Correlation$\left(f_{ni}, \hat{f}_{ni}\right)$	0.319	0.569	0.581
Number of observations	1345	1345	1345

Standard Errors remain to be computed.

Table 7: Summary Statistics of the Estimates of Iceberg Trade Costs, $\hat{d}_{ni,k}$

	mean	std	10.0%	50%	90%	max
achiote	1.15	0.08	1.07	1.14	1.23	1.60
apple	1.38	0.19	1.17	1.35	1.58	2.53
asparragus	1.14	0.07	1.06	1.13	1.22	1.57
avocado	1.27	0.14	1.12	1.25	1.42	2.09
banana	1.74	0.38	1.33	1.69	2.14	3.98
barley grain	1.35	0.18	1.16	1.33	1.55	2.43
cacao	1.07	0.04	1.03	1.07	1.11	1.28
canahua	1.18	0.09	1.08	1.17	1.28	1.72
cassava	1.54	0.28	1.24	1.51	1.84	3.21
chocho	1.15	0.08	1.07	1.14	1.23	1.61
coconut	1.84	0.43	1.37	1.79	2.30	4.39
coffee	1.07	0.03	1.03	1.06	1.10	1.27
cotton branch	1.14	0.07	1.06	1.13	1.22	1.57
dry bean	1.14	0.07	1.06	1.13	1.22	1.57
dry faba bean	1.22	0.11	1.10	1.21	1.34	1.90
dry pallar	1.12	0.06	1.05	1.12	1.19	1.50
dry pea	1.19	0.10	1.09	1.18	1.30	1.78
dry zarandaja	1.21	0.11	1.09	1.20	1.32	1.84
garbanzo beans	1.11	0.06	1.05	1.10	1.17	1.44
garlic	1.11	0.05	1.05	1.10	1.17	1.43
grape	1.22	0.11	1.10	1.20	1.33	1.87
green bean	1.30	0.15	1.13	1.28	1.47	2.22
green faba bean	1.45	0.23	1.20	1.42	1.69	2.81
green pea	1.29	0.15	1.13	1.27	1.45	2.18
key lime	1.47	0.24	1.21	1.44	1.73	2.91
kiwicha amaranth	1.12	0.06	1.05	1.11	1.19	1.48
lentil	1.15	0.07	1.07	1.14	1.23	1.59
maize (amilaceo)	1.20	0.10	1.09	1.19	1.32	1.83
maize (choclo)	1.45	0.23	1.20	1.42	1.70	2.82
maize (yellow hard)	1.43	0.22	1.19	1.41	1.67	2.75
mango	1.55	0.28	1.25	1.52	1.85	3.23
mashua	1.63	0.32	1.28	1.59	1.97	3.55
oca	1.56	0.28	1.25	1.52	1.86	3.25
oil palm	2.13	0.58	1.50	2.06	2.74	5.56
olive	1.16	0.08	1.07	1.15	1.24	1.64
onion	1.47	0.24	1.21	1.44	1.73	2.91
orange	1.60	0.31	1.27	1.56	1.92	3.42
papaya	1.56	0.29	1.25	1.53	1.87	3.28
pecan	1.04	0.02	1.02	1.03	1.05	1.14
pineapple	1.61	0.31	1.27	1.57	1.95	3.47
potato	1.53	0.27	1.24	1.50	1.82	3.14
quinoa	1.12	0.06	1.05	1.11	1.18	1.48
rice	1.38	0.19	1.17	1.35	1.58	2.52
sorghum	1.51	0.26	1.23	1.48	1.79	3.06
soy	1.19	0.10	1.09	1.18	1.30	1.78
sweet potato	1.60	0.31	1.27	1.56	1.93	3.42
tangerine	1.55	0.28	1.25	1.52	1.86	3.25
tea	1.52	0.26	1.23	1.49	1.80	3.09
tomato	1.43	0.22	1.19	1.41	1.67	2.76
ulluco	1.45	0.23	1.20	1.42	1.69	2.81
wheat	1.28	0.14	1.12	1.26	1.43	2.11

Table 8: The Intermediate Input Bundle

Input Name	Unit Price (LCU / Kg)	Import Quantity (Tons)
Potassium chloride	1.529	52742.80
Diammonium phosphate (DAP)	2.144	87398.30
Ammonium nitrate	1.158	117545.50
Ammonium sulfate	0.705	112678.80
Magnesium and potassium sulphate	1.009	16075.60
Potassium sulphate	2.916	36281.90
Superphosphate	1.599	1647.00
Urea	1.349	277113.70
Average	1.551	

Source: Ministry of Agriculture of Peru, Monthly Bulletin. Year 2008

Table 9: Within-Crop Regression of Price Data on Model Simulation

	(1) Baseline	(2) High Trade Barrier	(3) High Heterogeneity and Trade Barriers
model	1.770	0.733	0.669
	(0.0411)	(0.0143)	(0.0145)
R2 within	0.334	0.415	0.365
R2 between	0.724	0.817	0.755
R2 overall	0.692	0.750	0.690
N	3739	3739	3739

Standard errors in parentheses

Table 10: Pooled Regression of Land Allocation Data on Model Simulation

	(1) Baseline	(2) High Trade Barrier	(3) High Heterogeneity and Trade Barriers
etamodel	0.180	0.229	0.352
	(0.0104)	(0.0120)	(0.0123)
R2	0.0744	0.0895	0.180
N	3739	3739	3739

Standard errors in parentheses

Figure 1: Map of Peru

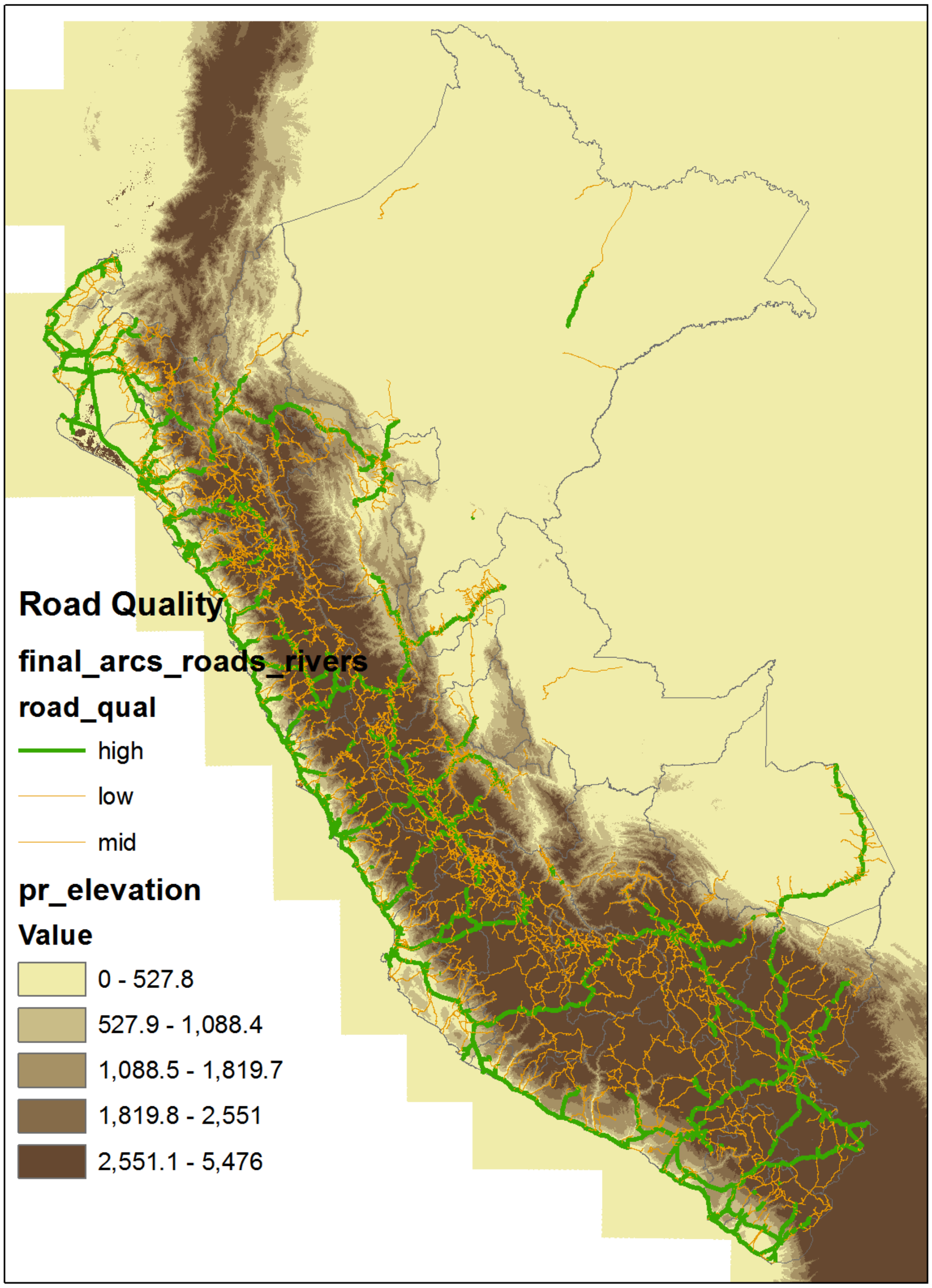


Figure 2: Comparison of Land Shares, $\eta_{i,k}$, and Revenue Shares, $\pi_{i,k}$

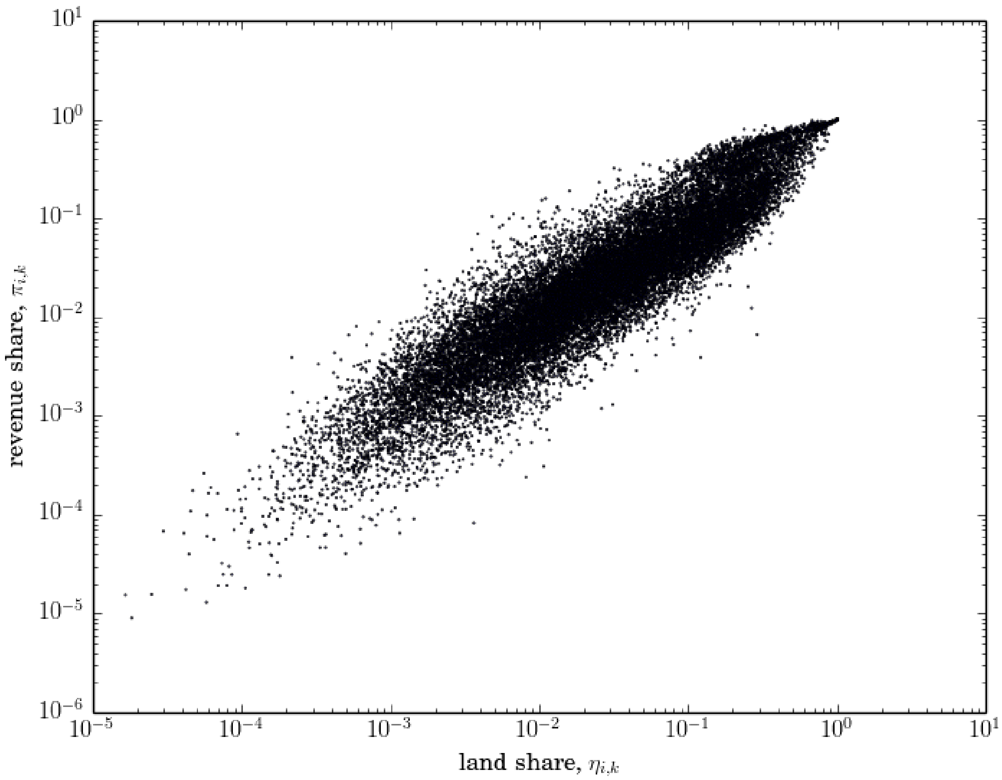


Figure 3: Estimation of θ

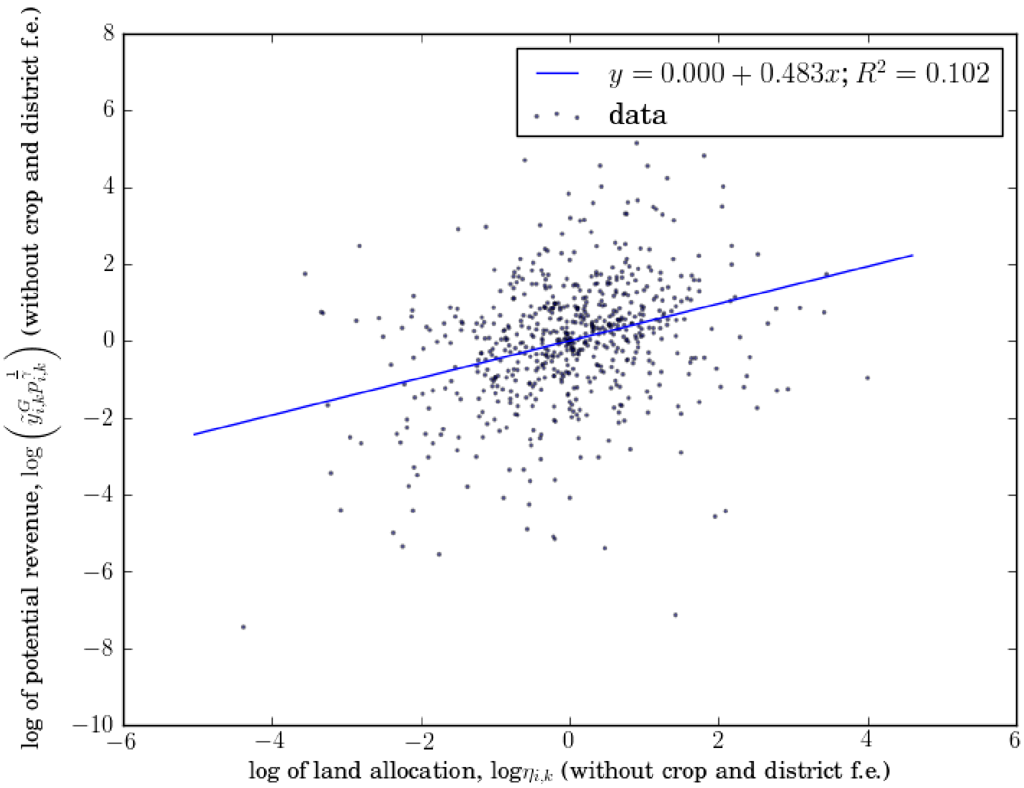


Figure 4: Comparison of Freight Rate Predictions (Constrained and Unconstrained Transportation Model)

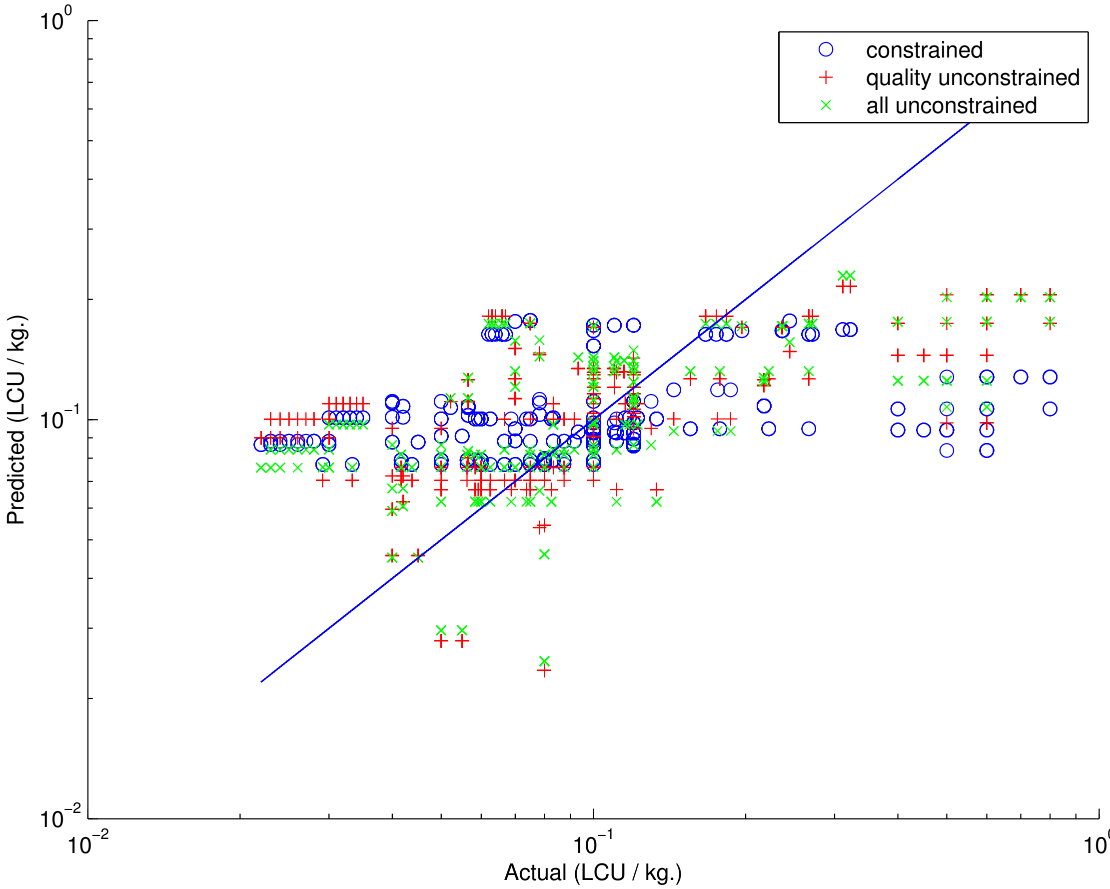
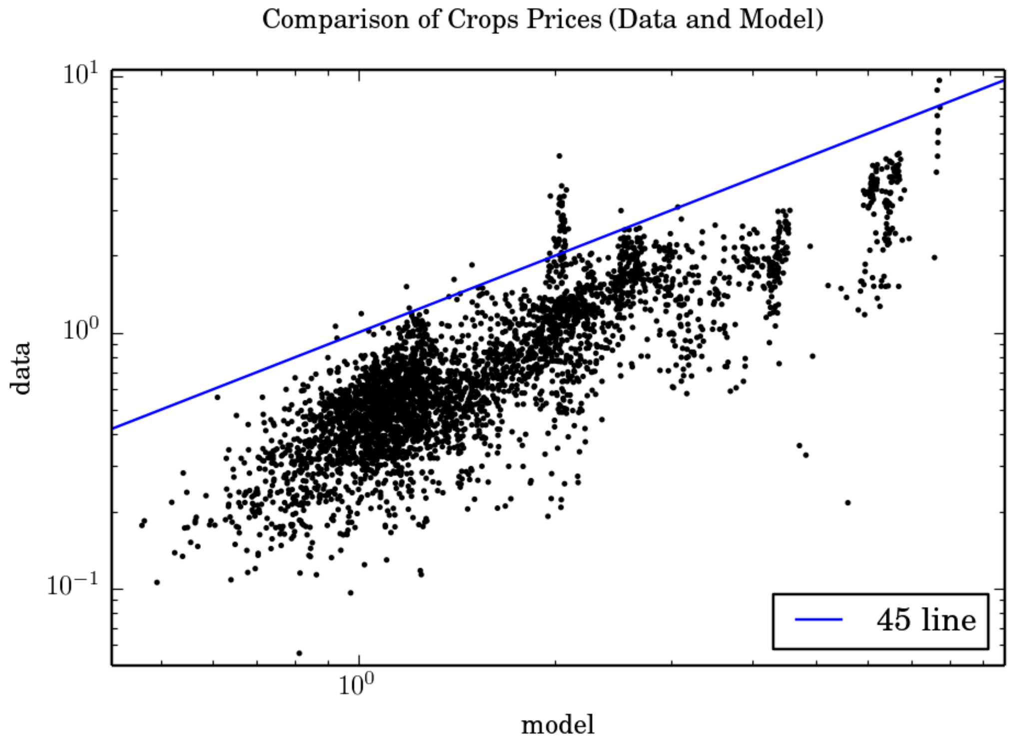


Figure 5: Baseline Parametrization: Fitting Price and Land Allocation Data in the



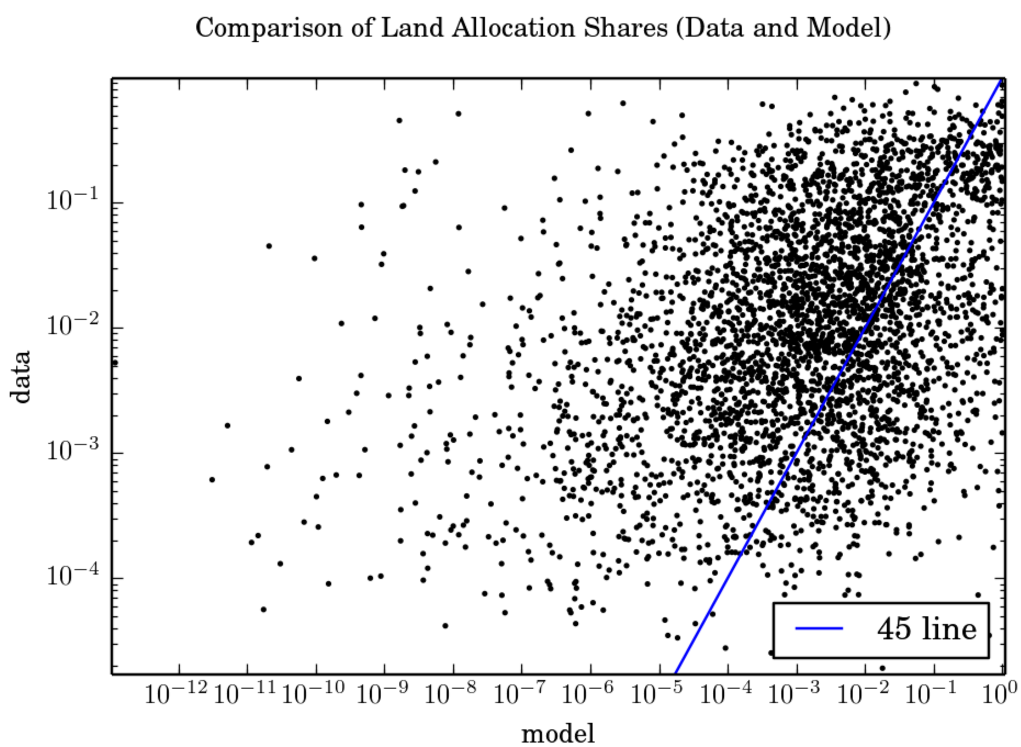


Figure 6: Alternative Parametrization: Fitting Price and Land Allocation Data