

# Domestic Trade Frictions and Agriculture

Sebastian Sotelo  
University of Michigan\*

March, 2019

## Abstract

I develop a model of agriculture on heterogeneous land to study the relation between trade, productivity, and welfare in Peru, where farmers face high internal and external trade costs. I quantify the model with new data on crop prices, yields, and land allocations. Using the model, I measure the welfare and productivity effects of changes to trade opportunities. A policy of paving roads raises aggregate productivity by 4.9 percent and the median farmer's welfare by 2.7 percent, but increased competition from remote domestic suppliers harms more than 20 percent of farmers. An increase in the international relative price of grains spreads unevenly across regions, benefiting farmers but hurting urban consumers close to ports.

Keywords: assignment models, trade costs, equilibrium, agriculture, productivity  
JEL Codes: F11, O13, Q17, D58

---

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

# 1 Introduction

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

Assessing the welfare and productivity effects of improvements in trade opportunities, such as infrastructure, requires an understanding of how individual farmers and consumers react to these improvements and how their choices interact in the aggregate. On the one hand, policies that reduce trade costs can increase allocative efficiency and welfare by unlocking the forces of comparative advantage and the use of modern inputs. On the other hand, such policies also affect the equilibrium prices of crops, especially those that are traded only domestically, thereby potentially harming producers who face increased competition. Because these crops usually constitute an important part of the diet of food buyers, improved exchange opportunities can also affect welfare through their effect on consumption prices.

In this paper, I develop a framework to measure the consequences of trade-related shocks to the agricultural sector and quantify it using Peruvian data. I start by documenting four facts about agriculture in Peru that guide my modeling choices. First, there is substantial price dispersion across regions, which is indicative of domestic trade costs. Consistent with the presence of domestic and international trade, prices are also higher in urban areas, and the prices of export crops rise with proximity to ports. Second, farmers within a region allocate land to many crops, and third, average revenue per unit of land varies substantially across crops. The last two facts suggest that land quality varies within regions and that land intensity differs across crops. Fourth, the quality of roads varies throughout the country, which produces spatial variation in trade costs.

Informed by these facts, I then develop a quantitative model of specialization and trade. In the model, farmers can grow various crops on land plots of varying quality. They can also trade their crops, at a cost, for intermediate inputs, non-agricultural goods, and other crops in local, urban, and international markets. Differences in land quality across regions and in land intensity across crops generate domestic and international trade, while domestic trade

---

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

costs discourage it. The resulting model is a hybrid between a small open economy, which takes international prices as given, and a closed economy, in which prices are determined by regional trade within the country. The equilibrium features price dispersion and incomplete land specialization across regions, as in the data. Productivity and welfare, moreover, are determined by market access and comparative advantage.

To quantify the theory, I construct a novel data set that combines several sources of data on Peruvian agriculture. I use government statistics on land allocation, production, and prices to estimate crop-specific land quality across and within regions. To estimate within-country trade frictions, I bring in a complete data set of the transportation system of Peru. Finally, I use disaggregated household survey data to estimate the elasticity of substitution across crops in consumption.

I next conduct two sets of counterfactual experiments to understand and quantify how changes in trade opportunities affect welfare and productivity in the context of imperfect regional integration. First, informed by policy plans of the Ministry of Transportation of Peru, I simulate two infrastructure shocks that improve the transportation system and reduce domestic trade costs unevenly across regions. Second, I simulate a shock to international crop prices based on a World Bank scenario for the effects of the Doha trade talks.

There are two key points to understanding regional responses in these counterfactual scenarios. First, the initial production and consumption allocations govern the extent to which changes in prices translate into changes in productivity and welfare; these allocations, in turn, are driven by the interaction of comparative advantage and market access. Second, equilibrium prices respond endogenously and transmit shocks across regions. These price responses depend also on the substitutability of crops in production—which is linked to the degree of land heterogeneity—and in consumption.

To understand, for example, how improved transportation directly impacts farmers' productivity and welfare, note that when access to markets is costly, farmers pay high prices for their purchases and collect low prices for their sales. By reducing the cost of accessing domestic and foreign markets, better roads allow farmers to specialize according to comparative advantage and increase their use of imported intermediate inputs.

The general equilibrium effects of this policy are more subtle. While infrastructure policy increases the price a farmer gets for his products by improving his own access to domestic and international markets, the policy also improves the access of farmers in other regions and increases the supply of crops to domestic markets, thus decreasing their price. Increased competition from remote suppliers therefore leads to reduced prices for farmers who originally had better access to markets.

Using this quantitative model, I measure the aggregate and distributional effects of two

infrastructure policies. I find that a Ministry of Transportation plan to pave major roads increases productivity in agriculture, measured as a multi-crop index of productivity (3.8 percent for the average region and 4.9 percent in aggregate). The median farmer, moreover, experiences a 2.7 percent welfare gain. There is also substantial heterogeneity: The farmer in the 25th percentile of the welfare change distribution gains 0.1 percent and that in the 75th gains 13.2 percent. Rural dwellers employed in the non-agricultural sector mostly benefit as a result of the policy, but their gains are limited in remote areas because consumption prices increase and because their ability to export non-agricultural goods is limited. Along the same lines, a different policy of building new roads in targeted remote areas allows the median farmer to gain 0.02 percent of welfare, while farmers at the 25th percentile of the welfare gains distribution lose 0.1 percent and those at the 75th percentile experience 0.25 percent gains. In both simulated policies, an increased supply of crops not traded internationally drives down the welfare of farmers who were originally well connected to urban markets.<sup>2</sup>

In the second counterfactual exercise, I simulate a shock to international crop prices, based on a World Bank analysis of the potential effects of the Doha trade talks. This exogenous shock mildly increases the international price of cereals and cotton. Unlike in a small open economy, however, international shocks are unevenly spread in regional markets because many Home regions do not trade every crop with the rest of the world. Although this is a relatively small shock, its general equilibrium effects are quantitatively important and lead to the opposite welfare prediction of a simple small open economy model in about 23 percent of regions. Proximity to ports makes it more likely that a region will trade with the rest of the world, and therefore strengthens the transmission of international price shocks to regional markets—including to crops that are only traded domestically. Among non-agricultural workers, those in regions close to ports tend to lose more, because their consumption baskets reflect more closely the price increase of high-consumption cereals.

In addition to producing substantive results for Peru, this paper also makes three methodological contributions. First, the theory connects tightly with data on land allocations and productivity; hence the model can be estimated based solely on agricultural and aggregate trade statistics, which are collected by many countries. This approach is especially useful for studying economies in which trade also occurs within borders, because it sidesteps the need to use domestic trade data, which are typically unavailable.

Second, the model treats each crop as a homogeneous good, instead of using the standard Armington modeling device in which the same good is regionally differentiated. This

---

<sup>2</sup>Sizable welfare reductions are not uncommon, even though both policies reduce trade costs for all regions. The farmer in the 5th percentile of the welfare gains distribution loses 2.4 percent in the first policy and 0.7 percent in the second.

approach allows for a simple analysis of the dissemination of price shocks across regional markets, based only on initial consumption and land shares, and the elasticities of supply and demand. Moreover, when studying trade within a country, this treatment provides an alternative to regional differentiation, which, besides being implausible at high spatial resolutions, forces all regions to trade bilaterally in all goods. In contrast, the model in this paper generates sparse trade patterns—a well-known feature of trade data—using finite trade costs.

Third, I obtain a simple estimating equation for the elasticity of land allocation with respect to relative prices, which is a key quantity that governs adjustments to shocks. The estimating equation captures a basic economic intuition inherent to models in which production factors are heterogeneous, namely, that factors are optimally allocated to their best use. Hence, increasing a crop's land allocation reduces yields because it requires incorporating land less suited to that use. The strength of this effect is directly related to the heterogeneity of land. This parameter also governs all cross-elasticities in production—a practical compromise when the number of parameters that can be estimated is limited by the availability of a few cross sections of data.

Peru is an ideal setting for this study, because its geography is diverse and its agricultural sector resembles that of both developed and developing countries. It is a middle-income country in which a few large urban markets are often the destination for traded agricultural produce, but some well-connected regions produce for export markets. Eighty-six percent of roads are 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 rain forests 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 costly.<sup>3</sup> Geography is also a basis for specialization based on comparative advantage, because weather and land quality vary drastically within the country. And while large farms on the coast often employ modern techniques, isolated Andean and jungle regions still use traditional farming methods. Finally, about 25 percent of Peru's labor force is employed in agriculture, similar to other developing countries.<sup>4</sup>

This paper relates to the literature on equilibrium models of trade in agriculture, such as Costinot, Donaldson, and Smith (2016); Fajgelbaum and Redding (2014); Costinot and Donaldson (2014); and Costinot and Donaldson (2012), who study how trade opportunities

---

<sup>3</sup>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 (Regional Direction of Agriculture, La Libertad).

<sup>4</sup>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, pp. 27-28).

mitigate the consequences of climate change, drive structural transformation, and improve welfare. I contribute to this literature by providing a parsimonious framework to study prices and allocations when regions that produce a finite set of homogeneous goods are partially integrated with each other and the rest of the world. These modeling elements are important in my application to Peru and can provide an apt description of other developing economies. Yet in terms of the general equilibrium structure of the model, they represent a departure from the literature and provide new insights on regional adjustment to trade shocks. Finally, unlike Donaldson (2015), who integrates the Ricardian model of Eaton and Kortum (2002) with Indian regional trade data, the model in this paper yields predictions for land shares across crops; this allows me to connect to land use data, which is more widely available than data on within-country trade.

Previous research has studied the impact of transportation policy in different settings and for different purposes. Faber (2014) shows that China’s National Trunk Highway System, by targeting the connection of large cities, led to a concentration of industrial production in large production locations. Using data from 15 countries in Africa, Storeygard (2016) provides evidence that cities farther away from main city-ports face lower growth in the face of shocks to oil prices, showing that trade costs are a key determinant of economic activity. Fajgelbaum and Redding (2014) quantitatively explore how structural change and the relative rewards to production factors depend on trade shocks and proximity to ports. In this paper, I focus on how infrastructure policy and foreign price shocks affect the agricultural sector, which is key in developing countries. Moreover, I quantitatively assess the effect of these shocks on the distribution of welfare across workers and space.<sup>5</sup>

More broadly, this paper also speaks to the literature that documents the role of agriculture in the low productivity of developing countries (Gollin, Lagakos, and Waugh, 2013; Restuccia, Yang, and Zhu, 2008) and proposes explanations for this finding, such as worker sorting across sectors (Lagakos and Waugh, 2013); policy barriers to efficient farm size (Adamopoulos and Restuccia, 2014); exposure to uninsurable shocks (Donovan, 2018); and trade frictions (Tombe, 2015; Adamopoulos, 2011; Gollin and Rogerson, 2014). This paper shows that transportation technology in developing countries is a constraint that limits the use of modern inputs, and that improving this technology can lead to a more productive allocation of land and labor.

---

<sup>5</sup>Earlier theoretical work has anticipated the distributional within-country effects of trade—which I measure quantitatively—using the Heckscher-Ohlin (Courant and Deardorff, 1992) and Ricardian models (Rauch, 1991). More recently, using a Ricardian model with labor mobility, Cosar and Fajgelbaum (2016) highlight the effect of internal geography on factor rewards, and provide evidence of these effects using Chinese data.

## 2 Data on Agricultural Production, Consumption, and Roads

Throughout the paper, I combine four main data sets that allow me to measure production and consumption in Peru, as well as transport costs. I include in the analysis the largest 20 crops by nationwide value of production (rice, potato, coffee, yellow maize, alfalfa, asparagus, banana, cassava, amilaceo maize, grape, cotton, onion, choclo maize, bean, avocado, wheat, cacao, orange, barley, and tangerine) between 2008 and 2011. I aggregate the data to the 194 provinces in Peru. This section gives a brief discussion of these main data sets (see the Online Appendix for more detail and a discussion of other supporting data sets).

**National Statistics on Agriculture.** This data set is collected by the Ministry of Agriculture of Peru at a finely disaggregated geographic level. It contains comparable data on farm-gate prices, physical land yields, and land use for each region and crop.

**National Household Survey (ENAHO).** This is Peru’s main living standards survey, which is collected yearly by Instituto Nacional de Estadística e Informática (INEI). It contains information on household expenditures, consumption quantities, and unit values, and is disaggregated by regions and commodities.

**Global Agro-Ecological Zones (GAEZ).** The GAEZ project (IIASA/FAO, 2012) estimates the attainable yield—i.e., the average product of land—if all land in a 5 arc-minute cell is used in a particular crop. These crop-specific estimates are obtained by combining information on weather, soil suitability, altitude, etc., and assumptions about management techniques.<sup>6</sup> Costinot and Donaldson (2014) provide a detailed discussion of these data.

**Geography and Transportation.** These Geo-referenced data from the Peruvian Ministry of Transportation (MTC) report each road’s location, length, and quality (paved or unpaved). Peru’s road system is hierarchically divided in three levels: (i) National roads, (ii) Departmental roads, and (iii) Rural roads.<sup>7</sup>

---

<sup>6</sup>For example, “Under a low level of inputs (...), the farming system is largely subsistence based. Production is based on the use of traditional cultivars (...), labor intensive techniques, and no application of nutrients, no use of chemicals for pest and disease control and minimum conservation measures” (IIASA/FAO, 2012, p. 38). Throughout, I use yields attainable with rain-fed agriculture and “low” levels of management, although the results are robust to the choice of “intermediate” levels of management.

<sup>7</sup>National roads (Red Vial Nacional) consist of 3 north-south axes, which connect the northern and southern frontiers of the country, and 20 west-east axes, which link the north-south axes at different latitudes. Departmental roads (Red Vial Departamental) serve as intermediate connectors between National roads and the more local Rural roads (Red Vial Vecinal). The latter connect populated and production centers with

### 3 Four Motivating Facts

In this section, I document four observations about agriculture in Peru and explain how they motivate my modeling approach. First, there are large, systematic differences in crop prices across regions and between urban and rural areas, indicative of domestic barriers to trade. Second, regions tend to grow many crops, which suggests land quality heterogeneity. Third, average revenue per unit of land varies significantly across crops, which suggests that crops use land with different intensities. Fourth, road quality varies throughout the country, creating spatial differences in trade costs.

**There are large spatial differences in farm-gate and consumer prices.** Figures 1(a) and 1(b) give an example of variation in farm-gate prices across regions using coffee, one of Peru’s main exports. Panel (a) displays the large variation in the price of coffee across regions that produce it, while panel (b) shows that farm-gate prices decline with distance to the capital (Lima), which also contains the main seaport.

To test whether prices are systematically different across regions, I use unit values calculated from household surveys, which approximate consumer prices and for which we have many observations per region (Deaton, 1997, Ch. 7). Table 1 shows the results of regressing log unit values on two different sets of dummies. Column 1 shows, for each crop, the F statistic of a joint significance test for province dummies. For each crop, we reject the null that these dummies have no explanatory power, which is indicative of spatial dispersion. Column 2 shows the  $R^2$  of the regression, which indicates that a large fraction of the price variation is spatial. Columns 3 and 4 report the coefficient and standard error of regressing log unit values on a dummy for whether the observation corresponds to an urban area. These regressions show that consumer prices in urban areas are usually higher than in rural areas, ranging from -0.02 to 0.6 log points higher.<sup>8</sup>

These spatial price differences suggest that trade barriers prevent price equalization, and are consistent with urban centers that import food from rural areas. They illustrate that for crops such as coffee, one of the country’s main exports, foreign markets exert an effect on domestic prices. To generate this spatial price dispersion in equilibrium, I write down a model that includes domestic trade frictions and allows for international and domestic trade, and where a non-agricultural sector employs urban workers.

---

Departmental roads. See Ministerio de Transportes y Comunicaciones de Peru (2011b).

<sup>8</sup>An alternative interpretation is that price differences reflect differences in quality. One way to control for quality differences is to include a measure of household income in these regressions, under the assumption that the quality of consumption increases with total income. Online Appendix H.11 shows that spatial and urban-rural differences remain after controlling for household food spending in each region, which suggests quality differences alone cannot rationalize price dispersion.



Table 1: Spatial Dispersion of Unit Values

Crop	Region dummies		Urban dummy	
	F statistic	R-sq	Coefficient	s.e
alfalfa				
asparagus	49.160	0.238	0.336	0.078
avocado	367.949	0.528	0.363	0.008
banana	225.344	0.362	0.123	0.003
barley grain	31709.330	0.671	0.402	0.013
cacao	18.362	0.555	0.302	0.092
cassava	590.747	0.331	0.246	0.004
coffee	102.337	0.560	0.194	0.022
cotton branch				
dry bean	151.542	0.285	0.190	0.005
grape	113848.500	0.293	-0.016	0.006
maize (amilaceo)				
maize (choclo)	200.788	0.428	0.335	0.007
maize (yellow hard)				
onion	98.850	0.110	-0.016	0.003
orange	364.937	0.316	0.128	0.004
potato	259.760	0.352	0.242	0.003
rice	168.931	0.188	0.054	0.002
tangerine	155.203	0.318	0.156	0.005
wheat	231.343	0.663	0.604	0.007

Notes: Each row presents results associated with two regressions of log unit values measured at the household level on different sets of dummies for a given crop. Column 2 presents the F-statistic associated with a set of 194 regional dummies in a regression that also contains year fixed effects. Column 3 presents the corresponding R-squared. Columns 4 and 5 present the coefficient and standard error of regressing unit values on a dummy indicating whether a region is urban. Unit values are not available for crops that are infrequently consumed directly by households. In those cases, the corresponding row is empty. Sample size varies with each crop.

**Regions tend to grow many crops.** In the median region, total arable land is  $82km^2$  and 11 crops are grown (see Figure 1(c)). Table 2 shows that most regions tend to allocate a large amount of land to a few crops and use small amounts of land to grow several others—a pattern of incomplete specialization. For the median region, the most important crop captures a 0.4 share of total land, while the median crop gets a share of 0.03. With many producing regions, this pattern of incomplete specialization suggests a degree of curvature in regional production possibility frontiers, which I model as within-region heterogeneity in land quality to grow different crops.

Table 2: Distribution of Largest, Median, and Smallest Land Shares across Regions

	min	1st quartile	median	3rd quartile	max
largest share	0.205	0.317	0.404	0.538	1.000
median share	0.001	0.015	0.028	0.063	1.000
smallest share	0.000	0.000	0.001	0.003	1.000

Notes: The table shows, across columns, the distribution across regions of the largest land shares, median land shares, and smallest land shares.

**Average revenue per unit of land differs systematically across crops.** Figure 1(d) presents two distributions of residual log-revenues per hectare, across crops and regions. The distribution in the solid line is obtained by removing region fixed effects, and the one in the dashed line is obtained by additionally removing crop fixed effects. Crop-specific effects account for a large fraction of the variation in residual log-revenue per hectare, reducing its variance by more than half, from 0.63 to 0.20.

This fact guides how I model technology. To see why, consider a competitive farmer who combines homogeneous land and labor using identical, constant returns to scale technologies across crops. Profit maximization ensures that land-labor ratios be identical across crops and, consequently, that there be no variation in revenue per unit of land.<sup>9</sup> While I develop a heterogeneous-land model later, this simple intuition carries over under my specific distributional assumptions. Recognizing the role of crop fixed effects in accounting for variation in revenue per unit of land, I allow for different land cost shares across crops. Given prices, these technological differences—together with land heterogeneity across and within regions—will determine the allocation of productive resources as well as the adjustment of prices in equilibrium.<sup>10</sup>

**Road quality varies throughout the country.** Figure 1(e) presents the two main components of the national road system (National and Departmental Highways) as of 2011. The

<sup>9</sup>Given factor rewards, cost minimization equates land-labor ratios, which in turn pin down output per hectare across crops. Marginal costs, which equal price for each crop actually grown by the farmer, will also be identical. Therefore, price  $\times$  yield (i.e., revenue per hectare) will be constant across crops. This reasoning also holds with factor neutral productivity shifters across crops, since their effect on marginal costs and yields offsets when calculating revenue per hectare. An alternative interpretation is that frictions in factor markets cause small household farms to have larger labor to land ratios. The Online Appendix shows that variation in farm size cannot explain away within-region, farm-level variation in revenues per unit of land.

<sup>10</sup>I assume that observed price differences capture all distortions faced by farmers. A complementary view is that farmers face other types of distortions that prevent an efficient allocation of land across crops—such as frictional land markets—which the literature on misallocation in agriculture has studied (see, e.g., Adamopoulos and Restuccia, 2014). Unfortunately, my data set does not allow me to explore these possibilities. When I estimate land intensities in Section 5, however, I rely only on systematic differences between crops, so a crop will have a relatively large land intensity whenever its revenue share is larger than that of other crops on average. Doing so limits the effect of idiosyncratic distortions faced by farmers within regions.

figure shows that locations along the coast are connected through paved roads, as are a few locations in the highlands and the jungle (towards the east). In contrast, most other locations are only served by unpaved roads. According to the Ministry of Transportation, in 2011, 12 percent of existing roads were paved (Ministerio de Transportes y Comunicaciones de Peru, 2011a), which creates heterogeneous barriers to the transportation of agricultural goods.

I will link this variation in the status of highways and roads, together with distance over the road network, to trade costs in the model. The characteristics of the road system will thus contribute to spatial price dispersion in equilibrium.<sup>11</sup>

## 4 Specialization, Input Use, and Trade

Guided by these facts, I study the link between trade frictions, agricultural productivity, and welfare in a model of factor allocation and trade based on comparative advantage. In the model, the Home country consists of many regions that differ in terms of their population and land endowment. Within a region, the quality of land to grow different crops varies across plots. In equilibrium, plots are allocated according to comparative advantage. On average, some regions are relatively better suited than others for growing particular crops. Moreover, crops use land with different intensities. These sources of comparative advantage produce trade between regions and with the rest of the world, and drive patterns of specialization across regions.

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

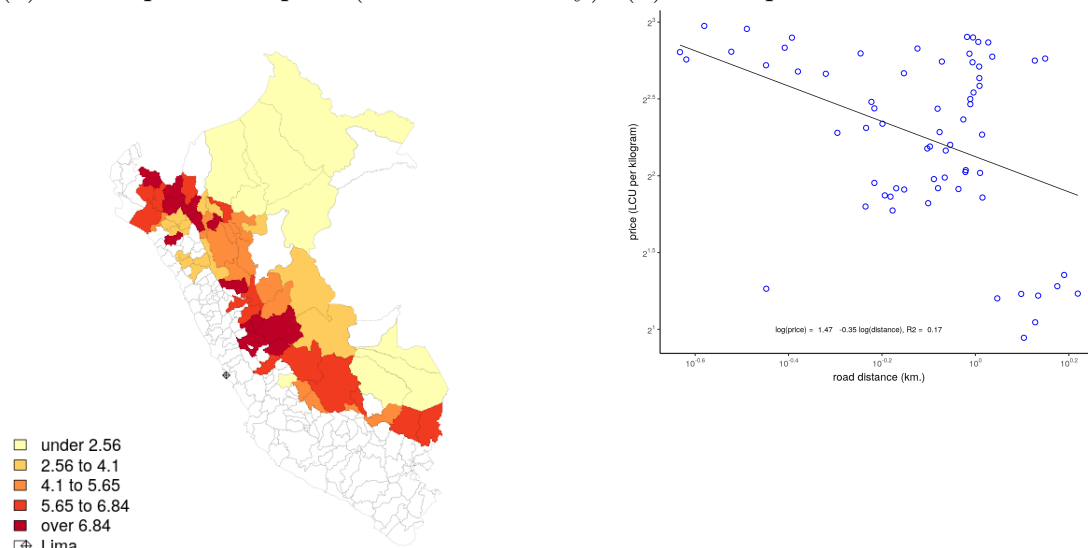
The assumption that land is heterogeneous reflects that in reality, the suitability of a location to grow a crop depends on the quality of the soil, altitude, weather, etc. To make contact with observed land allocations, I introduce assumptions on technology and the distribution of crop-specific land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

---

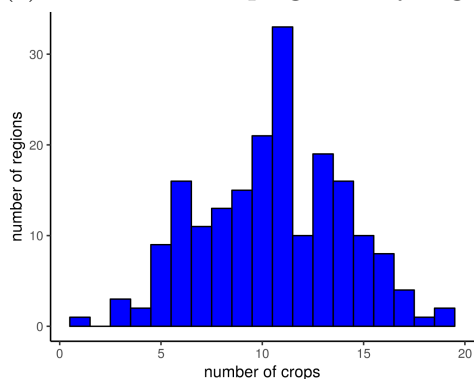
<sup>11</sup>As documented in the Annual Statistics of the Ministry of Transportation (2013), the majority of domestic trade in agriculture is carried out by roads, rather than by trains, boats or airplanes. Therefore, I limit my analysis to this part of the transportation infrastructure. See the Online Appendix for a detailed discussion.

Figure 1: Four Motivating Facts

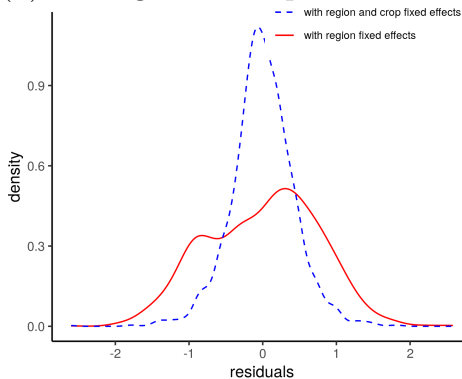
(a) Coffee prices in space (in local currency) (b) Coffee prices decline with distance to Lima



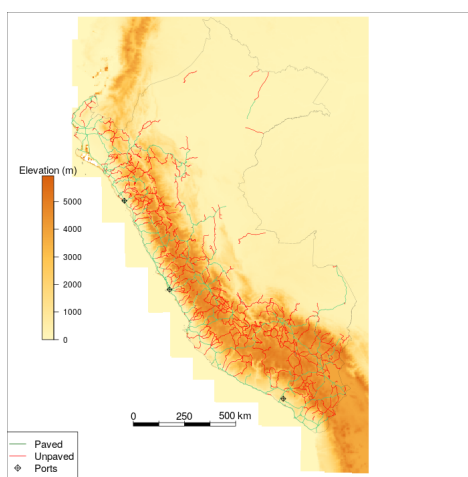
(c) Number of crops grown by region



(d) Average revenue per unit of land



(e) Roads and geography in Peru



Notes. Panels (a) and (b) use average prices over 2008-2011. Blank regions indicate that coffee is not grown there. “Road distance” is calculated using the road network. In Panel (e) Peru’s road system is divided into three levels: National, Departmental and Rural roads (only the first two are plotted to avoid clutter.)

In what follows, I model trade costs using the iceberg formulation, which avoids specifying the details of the transportation sector. The computational tractability of the model, however, would remain with an explicit transportation sector and additive trade costs.<sup>12</sup>

## 4.1 Environment

**Geography and Commodities.** I divide the world into Home—the focus of attention—and Foreign. Home consists of regions indexed  $n, i = 1, \dots, I$ . I denote Foreign by  $i = F$ , and the set of all regions by  $\mathcal{W} = \{1, \dots, I, F\}$ . There are  $k = 1, \dots, K$  homogeneous agricultural goods (crops, for short). There is also an intermediate input  $x$  used in agricultural production, which is imported from Foreign. The rest of the economy is aggregated in a “non-agricultural” sector, denoted by  $M$ , which includes a traded and a non-traded component.

**Agents.** In each region  $i$ , there are three types of agents. First, the representative consumer owns land and supplies labor; he rents out his factor inputs and purchases consumption goods in local markets. The second and third agents are the representative farmer, who hires factors in local markets, and the non-agricultural firms, which also hire labor in local markets. Both farmers and firms can sell their output in all markets.

**Endowments.** The representative consumer supplies inelastically his endowment of heterogeneous land, which consists of a continuum of plots. I denote the set of plots by  $\Omega_i$ , and all plots, indexed by  $\omega$ , have size one. The total amount of land in the region is  $H_i = \int_{\Omega_i} d\omega$ . The household in region  $i$  also supplies labor inelastically to agriculture,  $L_{i,A}$ , and to the non-agricultural sector,  $L_{i,M}$ .<sup>13</sup>

**Trade costs.** Domestic and international trade are subject to 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  $d_{nn,k} = 1$ ,

---

<sup>12</sup>Since I focus on how to solve the model numerically, as I show in the Online Appendix, one can extend this approach to imperfect competition in the transportation sector while retaining many of the properties of the equilibrium. Allen and Atkin (2016) propose a related approach to homogeneous commodity trade and transportation, which yields an analytical characterization of the equilibrium given their assumptions on the productivity of traders and on the trade cost formulation.

<sup>13</sup>In line with recent research (e.g., Restuccia, Yang, and Zhu, 2008; Tombe, 2015; Gollin, Lagakos, and Waugh, 2013; and Swiecki, 2014), in my data there are large differences in value added per worker across sectors, which put the assumption of labor immobility much closer to the data (see also Dix-Carneiro, 2014). In the data, value added per worker is 90 percent, 70 percent, and 50 percent lower in agriculture relative to non-agriculture in the 10th, 50th, and 90th percentiles across departments in the country. This suggests that differences in worker productivity are not driven by spatial variation, since they persist even within regions. In Section 7, I explore the robustness of my results to specifying a labor supply based on worker selection, akin to Lagakos and Waugh (2013).

all  $n, k$ . I also assume that costs are symmetric, so  $d_{ni,k} = d_{in,k}$ , and I impose the triangle inequality, i.e.,  $d_{ni,k} \leq d_{nj,k} \times d_{ji,k}$ . Non-agricultural goods are subject to costs  $d_{ni,M}$ .

**Preferences.** The consumer in region  $n$  spends a constant fraction  $b_n$  of income on a constant-elasticity aggregate of crops:

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

where  $\sigma > 0$  is the elasticity of substitution across crops and  $\sum_{k=1}^K a_k = 1$ , with  $a_k > 0$ . A fraction  $b_{i,TR}$  of expenditure goes to traded non-agricultural goods from each region  $i \in \mathcal{W}$ , which are differentiated according to their origin and combined in a constant-elasticity bundle:

$$C_{n,M} = \left( \sum_{i \in \mathcal{W}} C_{ni,M}^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}},$$

where  $\varepsilon$  is Armington the elasticity of substitution. The consumer spends the rest his of income on non-traded, non-agricultural goods.<sup>14</sup>

**Prices in Domestic and International Markets.** Each region in 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 non-agricultural labor,  $\rho_i$  the price of the intermediate input,  $p_{i,k}$  the price of crop  $k$ , and  $r_i(\omega)$  the rental rate of plot  $\omega$ . The Foreign price of crop  $k$  is  $p_{F,k}$ , and Foreign is also the only producer of the intermediate input, which costs  $\rho_F$  there. Agents take international prices as given.<sup>15</sup>

<sup>14</sup>The homotheticity assumption allows me to attribute all income to a single representative consumer, but the model will miss expenditure changes induced by large changes in income. A recent literature explains how non-homotheticity helps reconcile trade models with observations on international trade. See Fieler (2011), Markusen (2013), and Fajgelbaum and Khandelwal (2016). Atkin (2013) shows how local abundance shapes preferences and the benefits of trade.

<sup>15</sup>The assumption that Foreign is the only producer of intermediate inputs is a good representation of reality in the case of Peru. There, between 2008 and 2011, nearly 100 percent of the fertilizer used in production was imported, according to FAOSTAT. Appendix D.2 shows that intermediate input use declines with distance from the closest port. The model rationalizes this finding through higher prices, due to trade costs. I assume foreign prices are constant, because for the crops included in my sample, Peru is a relatively small player in world markets: Using FAOSTAT data for 2008, the crop for which Peru's share of world production is largest is asparagus, with a 4.6 percent share, followed by avocado and coffee (with 4.0 and 3.2 percent).

## 4.2 Production

**Farmers.** The following two assumptions about the production function and the distribution of land quality allow me to take the model to the data.

**Assumption 1.** *The technology to grow crop  $k$  exhibits constant returns to scale. It combines labor, the intermediate input, and land. The suitability of plot  $\omega$  in region  $i$  for producing crop  $k$  is captured by an efficiency shifter  $\Lambda_{i,k}(\omega) \geq 0$ ,*

$$q_{i,k}(\omega) = (l_{i,k}(\omega))^{\alpha_k} (x_{i,k}(\omega))^{\beta_k} (\phi_{i,k}(\omega) \Lambda_{i,k}(\omega))^{\gamma_k} \quad (2)$$

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

Fixing the cost shares at the plot level aids in obtaining tractable expressions for regional supply. It also implies that yield—the average product of land—is a choice that responds to the prices of inputs, a fact I that guides the estimation of the model in Section 5.<sup>16</sup>

The next assumption ensures that we obtain a simple structural equation for the allocation of land and revenues across crops.

**Assumption 2.** *The vector of land qualities for producing different crops  $k$  in region  $i$ , plot  $\omega$ ,  $\Lambda_{i,k}(\omega)$ , is i.i.d Fréchet with parameters  $(\tilde{\gamma}A_{i,k}, \theta)$ , i.e.,  $G_{i,k}(\Lambda) = e^{-\tilde{\gamma}^\theta A_{i,k}^\theta \Lambda^{-\theta}}$ .<sup>17</sup> In a region in which growing crop  $k$  is impossible, set  $A_{i,k} = 0$ .*

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

The representative farmer in region  $i$  rents land, hires labor, and buys the imported intermediate input. He decides how to allocate plots of land across crops and how much labor and intermediate inputs to use in each plot. Formally, the producer's problem is to

<sup>16</sup>The empirical cross-country literature on agriculture usually allows for a degree of substitutability across inputs. See, for example, Hayami and Ruttan (1985) and, more recently, Restuccia, Yang, and Zhu (2008).

<sup>17</sup>The normalization,  $\tilde{\gamma} \equiv [\Gamma(1 - \frac{1}{\theta})]^{-1}$ , where  $\Gamma(\cdot)$  is the Gamma function, simplifies the algebra later on. Allowing for certain types of correlation across plots within a region does not change the results, and only requires redefining some variables. See, for example, Eaton and Kortum (2002), footnote 14. See also Ramondo and Rodriguez-Clare (2013) for a multivariate extension of the Fréchet distribution.

choose  $\{\phi_{i,k}(\omega), l_{i,k}(\omega), x_{i,k}(\omega), \omega \in \Omega_i, \text{ all } k\}$ , to maximize profits,

$$\max \left\{ \sum_{k=1}^K p_{i,k} q_{i,k} - \int_{\Omega_i} \sum_{k=1}^K [w_{i,A} l_{i,k}(\omega) + \rho_i x_{i,k}(\omega) + r_i(\omega) \phi_{i,k}(\omega)] d\omega \right\}, \quad (3)$$

where total output of crop  $k$  is

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

for all  $k$ .

To connect the model to data on land shares, revenue shares, and yields across crops, I exploit Assumptions 1 and 2, which yield simple results for farmer behavior, which we turn to next. The three propositions that follow condense the model's empirical predictions, taking as given the equilibrium prices and factor rewards.<sup>18</sup>

As in standard trade theory, it is quite useful to work with unit cost functions to describe the farmer's choices. In doing so, we treat each plot as a separate factor, since the rental rate  $r_i(\omega)$  is plot specific. Given the technology in (2), the unit cost function, which measures the cost of producing a unit of crop  $k$  in plot  $\omega$ , is:

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

Let  $\omega \in \Omega_{i,k}$  denote that  $\omega$  is used to grow  $k$ .<sup>19</sup> Then profit maximization pins down a relation between rental rates, crop prices, and factor rewards, conditional on  $\omega \in \Omega_{i,k}$ ,

$$r_{i,k}(\omega) = \left( \bar{c}_k^{-1} p_{i,k}^1 w_{i,A}^{-\alpha_k} \rho_i^{-\beta_k} \right)^{1/\gamma_k} \Lambda_{i,k}(\omega).$$

Defining  $\lambda_{i,k} \equiv \bar{c}_k^{-1/\gamma_k} p_{i,k}^{1/\gamma_k} w_{i,A}^{-\alpha_k/\gamma_k} \rho_i^{-\beta_k/\gamma_k}$ , which we can interpret as the rental rate per efficiency unit of land associated with growing  $k$  in  $\omega$ , we obtain:

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

A competitive farmer will choose crops such that the rental rate is the maximum that

---

<sup>18</sup>Within-region plots are not directly observed in my data set, and I therefore only use the model's predictions for regional aggregates. Note also that farm sizes are undetermined in the model, and that all allocations in the model are efficient given the transportation technology. The model therefore is silent about the relation between farm size, labor productivity, and misallocation (e.g., Adamopoulos and Restuccia, 2014).

<sup>19</sup>We define  $\bar{c}_k \equiv \alpha_k^{-\alpha_k} \beta_k^{-\beta_k} \gamma_k^{-\gamma_k}$ .



can be attained in that plot; that is,

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

Because of Assumptions 1 and 2, only one crop maximizes rents for plot  $\omega$ . Those plots in which specialization is incomplete have measure zero.<sup>20</sup> Proposition 1 below derives the resulting land shares for region  $i$ .

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

$$\eta_{i,k} = \frac{(\lambda_{i,k} A_{i,k})^\theta}{\Phi_i^\theta}, \quad (4)$$

where

$$\Phi_i = \left( \sum_{l=1}^K (\lambda_{i,l} A_{i,l})^\theta \right)^{\frac{1}{\theta}}. \quad (5)$$

Equation (4) states that the relative use of land in crop  $k$  increases with the rental rate per efficiency unit of land,  $\lambda_{i,k}$ , and with average land quality  $A_{i,k}$ . The effect of all other crop prices and productivities are captured in the equilibrium statistic  $\Phi_i$ , as formalized in Proposition 2 below. The partial elasticity of land use to price is  $\theta/\gamma_k$ . When  $\theta$  is large, land is more homogeneous, and a given increase in the crop's price produces a larger shift in its land use. With a smaller land cost share  $\gamma_k$ , moreover, land heterogeneity is less of a limiting factor in expanding a crop's land use. Equation (4) is key in the identification of the parameter  $\theta$ —which governs equilibrium price adjustments in response to shocks—when combined with data on potential yields from the GAEZ project. It also aids in the determination of average land quality, conditional on data on land shares and prices.

While I do not observe rental rates directly in the data, I do observe the land yield and revenue per unit of land across crops in all regions. Proposition 2 below shows that data on physical yields or revenue per unit of land can only inform us about aggregate land quality in a region, because the average quality of the land supplied to a crop is inversely related to the amount of land supplied to that crop.

---

<sup>20</sup>All agricultural technologies have constant returns to scale at the plot level, and all factors are paid their marginal products, so all farmers 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. Also, as more land is allocated to a crop, the average quality of land used in that crop decreases. Hence, at the regional level, an increase in the amount of labor, intermediate inputs, and land allocated to the production of a crop does not increase its output in the same proportion. The Online Appendix shows that a model in which the landowner makes cropping decisions yields the same aggregate behavior as this formulation.

**Proposition 2.** *Let  $y_{i,k}(\omega)$  denote yield and  $\psi_{i,k}(\omega)$  denote revenue per unit of land for plot  $\omega$ . In equilibrium, the average yield of crop  $k$  is  $\mathbb{E}[y_{i,k}(\omega) | \omega \in \Omega_{i,k}] = \Phi_i / (\gamma_k p_{i,k})$ , while the average revenue per unit of land is  $\mathbb{E}[\psi_{i,k}(\omega) | \omega \in \Omega_{i,k}] = \Phi_i / \gamma_k$ .*

Propositions 1 and 2 summarize how each region will adjust to differences in relative prices and relative land qualities. To illustrate, consider an exogenous increase in the relative price of some crop  $k$ ,  $p_{i,k}$ , keeping all other prices constant. As its land share increases (Proposition 1), the average quality of land used in crop  $k$  must decrease. But the increase in this crop's price—though potentially countered by the decrease in its physical yield—must ultimately increase its revenue per unit of land, since the statistic  $\Phi_i$  increases. Conversely, for the rest of the crops, land use declines, which leads to an increase in average land quality, revenues and yields (This is seen directly from the increase in the statistic  $\Phi_i$ .) Assumptions 1 and 2 ensure that the proportional increase in revenues for crop  $k$  is identical to that of the rest of the crops. The result that increasing the amount of land in a given use decreases its productivity is not specific to this model—but assumptions 1 and 2 put constraints on the exact magnitudes of these changes.

Proposition 2 shows that average land quality,  $A_{i,k}$ , does not translate directly into observed differences in land yields or revenues. To see this, note that in the expressions for the expected yields and revenues—the theoretical counterparts of data on yields and revenues—the values of  $A_{i,k}$  are buried in the  $\Phi_i$  statistic. Thus, the model imposes the strong restriction that observed land yields and revenues per unit of land are not informative about unobserved land quality.<sup>21</sup>

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

$$\pi_{i,k} = \frac{p_{i,k} q_{i,k}}{\sum_{k'=1}^K p_{i,k'} q_{i,k'}}.$$

**Proposition 3.** *Within a region, the land share and the revenue share that crop  $k$  commands are equalized, up to a crop-specific constant*

$$\pi_{i,k} = \frac{\gamma_k^{-1} \eta_{i,k}}{\sum_{l=1}^K \gamma_l^{-1} \eta_{i,l}}. \quad (6)$$

---

<sup>21</sup>As shown in the Online Appendix, the model also implies that the reward to land is equalized on average across crops, and not only at the margin, since  $\mathbb{E}[r_i(\omega) | \omega \in \Omega_{i,k}] \propto \Phi_i$ . This is a strong restriction, which is unlikely to hold exactly in the data, although I do not have data on returns to land across crops to contrast it to.

A crop's revenue share,  $\pi_{i,k}$ , will be large relative to its land share,  $\eta_{i,k}$ , when its land cost share,  $\gamma_k$ , is low since land will optimally be combined with more of the other inputs.<sup>22</sup>

**Non-agricultural sector** Non-agricultural production is linear in labor. Region  $i$ 's productivity in the regionally differentiated good is  $T_{i,TR}$ , while labor productivity in the non-traded good is  $T_{i,NT}$ .

### 4.3 Regional Supply and Input Demands

To close the model in general equilibrium, we first characterize aggregate supply in each region. Aggregating across all plots used to grow crop  $k$  in region  $i$ , we obtain the regional supply:

$$q_{i,k} = \gamma_k^{-1} p_{i,k}^{-1} (\lambda_{i,k} A_{i,k})^\theta \Phi_i^{1-\theta} H_i \quad (7)$$

where, recall,  $\lambda_{i,k} = \bar{c}_k^{-1/\gamma_k} p_{i,k}^{1/\gamma_k} w_{i,A}^{-\alpha_k/\gamma_k} \rho_i^{-\beta_k/\gamma_k}$  measures rental rate per efficiency unit of land. The output of crop  $k$  is increasing in its price, with a constant partial elasticity of  $\theta/\gamma_k - 1$ , and decreasing in the price of labor and intermediates, with constant elasticities that also depend on the land intensity of the crop. Therefore,  $\theta$  and  $\gamma_k$  regulate the output response to price changes and, as I will discuss later, control the transmission of price shocks together with the elasticity of substitution in consumption  $\sigma$

Aggregating revenue across crops for this region, we obtain the total value of production in agriculture:

$$V_i = \frac{\Phi_i H_i}{\bar{\gamma}_i}, \quad (8)$$

where  $\bar{\gamma}_i = \sum_k \gamma_k \pi_{i,k}$  is the average cost share of land in region  $i$ . This cost share is endogenous, since it depends on the revenue shares of the crops grown in region  $i$ .

Turning to input demands, by defining appropriate aggregate cost shares we can write labor and intermediate input demand in a familiar way. Letting  $\bar{\alpha}_i$  be the average cost share of labor, we obtain regional labor demand,  $l_{i,A} = \bar{\alpha}_i V_i / w_{i,A}$ .<sup>23</sup> A wage increase reduces the amount of land allocated to relatively labor-intensive crops, but it also induces a less labor-intensive input mix for every crop. Similarly, letting the average intermediate input cost share be  $\bar{\beta}_i$ , we can express the aggregate input demand as  $x_{i,A} = \bar{\beta}_i V_i / \rho_i$ .

<sup>22</sup>Propositions 1 through 3 provide extensions of the results in Costinot, Donaldson, and Smith (2016) and Fajgelbaum and Redding (2014) to the case of heterogeneous factor intensities across crops. General versions of these results have been derived by Costinot and Vogel (2015). I highlight here the implications of the model that are most relevant for quantification later.

<sup>23</sup>The average cost shares of labor and the intermediate input are given by  $\bar{\alpha}_i = \sum_k \alpha_k \pi_{i,k}$  and  $\bar{\beta}_i = \sum_k \beta_k \pi_{i,k}$ .

## 4.4 Demand

The representative consumer's expenditure,  $E_n$ , equals the household's income from all sources,  $E_n = w_{n,A}L_{n,A} + w_{n,M}L_{n,M} + \int_{\Omega_n} r_n(\omega) d\omega$ .

Given income and prices, the solution to the representative consumer's problem is standard. The consumer in region  $n$  spends a share  $s_{n,k} \equiv b_n a_k (p_{n,k}/P_n)^{1-\sigma}$  of income on crop  $k$ , where  $P_n = \left(\sum_{k=1}^K a_k p_{n,k}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$  is the price of the agricultural bundle. As I show in Section 6,  $\sigma$  governs the sensitivity of the intensive margin of trade to trade costs, and therefore plays a key role in transmitting price shocks across regions.

Region  $n$ 's share of total expenditure on traded non-agricultural goods produced in  $i$  is  $b_{n,TR} \times (d_{ni,M} w_{i,M} / T_{i,TR})^{1-\varepsilon} / P_{n,TR}^{1-\varepsilon}$ , where  $P_{n,TR}$  denotes the price index in that sector, given by  $P_{n,TR} = \sum_{i \in \mathcal{W}} (d_{ni,M} w_{i,M} / T_{i,TR})^{1-\varepsilon}$ .

## 4.5 Equilibrium

I provide here a summary definition of a competitive equilibrium (see the Online Appendix for all details).

**Definition.** A competitive equilibrium consists of, for each region  $i = 1, \dots, I$ : (a) prices  $p_{i,k}$  for all crops  $k$ ; (b) wage rates  $w_{i,M}$ ,  $w_{i,A}$  and input prices  $\rho_i$ ; (c) final goods expenditure  $E_{i',TR}$ ,  $i' \in \mathcal{W}$ , and  $E_{i,NT}$  and consumption  $C_{i,k}$  for all crops  $k$ ; (d) input demands  $l_{i,A}$ ,  $x_{i,A}$  and outputs  $q_{i,k}$  for all crops  $k = 1, \dots, K$  and the non-agricultural sector,  $l_{i,NT}$ ,  $l_{i,TR}$ ; (e) trade flows: (e1) domestic  $z_{ni,k}$  for all regions  $n = 1, \dots, I$  and crops  $k = 1, \dots, K$ , (e2) international  $z_{Fi,k}$  and  $z_{iF,k}$  for all crops  $k = 1, \dots, K$  (e3) international  $z_{iF,x}$  of the intermediate input  $x$ ; such that (1) the quantities in (c) solve the consumer's problem, given income and prices; (2) the inputs and outputs in (d) solve the farmer's problem, given prices; (3) the agricultural goods prices in (a) come from the cheapest supplier

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

with equality if  $z_{ni,k} > 0$ , for all regions  $n, i \in \mathcal{W}$  and all crops  $k$ ; the intermediate input prices are

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

for all regions  $i$  in Home; (4) the labor demand for non-agricultural goods in (d) solve the non-agricultural firms' problem; (5) in each region, local markets clear for agricultural labor, land, and crops; and (6) with the definition of  $E_i$ , trade with Foreign is balanced. Domestic prices follow the normalization given by international prices,  $p_{F,k}$ .

**Equilibrium trade and zero trade flows.** In this model, regions trade for two reasons: land productivity differences and relative factor abundance. On the one hand, as in a Ricardian model, if region  $i$  is relatively better at growing crop  $k$ , as captured by a relatively higher average land quality  $A_{i,k}$ , it will tend to produce and export that crop. This is clearest in the limiting case in which plots are homogeneous and land cost shares are the same across crops ( $\theta \rightarrow \infty$  and  $\gamma_k = \gamma, \forall k$ ), which brings us to a Ricardian world with many goods and three factors. On the other hand, if region  $i$  is relatively abundant in land, it will tend to specialize in goods that use land intensively (i.e., goods with a high land cost share  $\gamma_k$ ). In fact, the limiting case of homogeneous plots and no differences in average land quality ( $\theta \rightarrow \infty$  and  $A_{i,k} = A_i, \forall i, k$ ) is similar to a Heckscher-Ohlin model with many goods and three factors. On top of these forces that produce regional trade, within-region land heterogeneity adds curvature to the production possibility frontier of each region, controlling how land allocations change with changes in relative prices.

Finally, note that zero trade flows, for any crop, between any pair of regions, are a possibility in this framework (and are common in equilibrium, as I show once I quantify the model). This possibility arises from treating crops as a finite set of homogeneous goods so that, for each crop and destination, only a small number of suppliers, possibly zero, may attain the lowest cost.

## 5 Connecting the Model and Agricultural Data

In this section, I estimate the model’s key parameters. First, I estimate the cost shares of land  $\gamma_k$ , exploiting the equilibrium relation between land and revenue shares (6). Second, I estimate the heterogeneity parameter  $\theta$  by interpreting national statistics on land allocation and exogenous yield estimates from the GAEZ through the lens of the model. Third, I estimate a simple model of transportation costs, following the approach in Donaldson (2015), and use it to produce measures of transport costs for all origin-destination pairs. Next, I use household expenditure data to estimate the elasticity of substitution between crops in demand. Finally, I briefly explain how I calibrate other parameters. All additional details are included in the Online Appendix.

Estimation and simulation of the model requires taking a stance on how to aggregate the data sets described in Section 2. Unless otherwise noted, a region  $i$  in the model corresponds to one of 194 provinces according to Peru’s 2012 administrative division, and data are regional averages for the years 2008-2011. A crop  $k$  is one of the top 20 crops by value of production.<sup>24</sup>

---

<sup>24</sup>Peru is divided into 24 departments. These departments are further divided into 194 provinces and more than 1,800 districts. The identification of crops  $k$  across data sets is explained in the Online Appendix,

**Factor Cost Shares:  $\gamma_k$ ,  $\alpha_k$ , and  $\beta_k$ .** To estimate the cost shares of land,  $\gamma_k$ , I construct land and revenue shares,  $\eta_{i,k}$  and  $\pi_{i,k}$ , using national statistics. I regress  $\log \pi_{i,k}$  on  $\log \eta_{i,k}$  and a full set of crop and region fixed effects (omitting one crop), since Proposition 3 shows that crops whose revenue shares are systematically higher than their land shares also have lower land cost shares,  $\gamma_k$ . This regression identifies land intensities relative to a base crop, so I set the aggregate revenue-weighted cost share of land to 0.22 to pin down the levels of all  $\gamma_k$ . Lacking information on labor and intermediate use across crops, I constrain the ratio of labor and intermediate input shares to be constant across crops and equal to 2.5 (see Dias Avila and Evenson, 2010, Table A.3a). Together with the estimates of  $\gamma_k$ , this assumption pins down  $\alpha_k$  and  $\beta_k$ . Based on these shares, I calibrate  $\bar{\alpha} = 0.55$ ,  $\bar{\gamma} = 0.22$  and  $\bar{\beta} = 1 - \bar{\gamma} - \bar{\alpha} = 0.23$ .

Figure 2(a) shows sizable differences in the estimated coefficients  $\gamma_k$ , as one would expect if some crops are more land intensive than others. Fruit crops, for example, appear to have lower land intensities than grains. Note, however, that contrary to what Proposition 3 states, crop fixed effects do not fully account for the differences between  $\pi_{i,k}$  and  $\eta_{i,k}$  within a region, which means that there is variation in revenue shares that the model will not be able to capture. Nevertheless, the residual standard deviation of a regression of  $\log \pi_{i,k}$  on  $\log \eta_{i,k}$  shrinks substantially when crop fixed effects are included (from 0.80 to 0.54), suggesting that extending the model to include heterogeneous technologies helps account for a large fraction of the variation in the data.<sup>25</sup>

**Heterogeneity,  $\theta$ .** Two equations stemming from the model form the backbone of the estimation. First, I use the model to interpret the GAEZ potential yield measures, which combine two types of information: purely exogenous climatic conditions and management techniques. Since these measures correspond to the yield that would be attained if a grid cell were fully devoted to growing a particular crop, I start by calculating the yield,  $\tilde{y}_{i,k}$ , that would be obtained using labor and inputs optimally (given factor prices), but allocating all land in region  $i$  to the production of crop  $k$ :

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

---

together with other details excluded from the main body of the paper. The crops in the sample account for 74 percent of total value and 67 percent of the total area under cultivation. Throughout the analysis, internationally traded crops are those for which the absolute value of net exports is at least 1 percent of domestic production.

<sup>25</sup>The F-statistic associated with the set of crop fixed effects in the regression that identifies  $\gamma_k$  is 202.76, which rejects the null that  $\gamma_k = \gamma$  for all  $k$  at standard levels of significance. Previous work has found that crops differ in their labor intensities. For example Vollrath (2011) finds, as I do, that wheat and other dry cereals are less labor intensive than rice (p. 345).

To connect this object to the data, I further assume that there exist prices  $p_k^G$ ,  $w_A^G$ ,  $\rho^G$  that rationalize the management technique assumptions used by IIASA and FAO to construct the GAEZ data set and that an error,  $u_{i,k}$ , captures the possibility that the link with the GAEZ data,  $\tilde{y}_{i,k}^G$ , is imperfect:

$$\tilde{y}_{i,k}^G = \zeta_k A_{i,k} e^{u_{i,k}}. \quad (9)$$

Doing so connects the GAEZ data to land quality  $A_{i,k}$ , which I do not observe directly.<sup>26</sup>

The second key equation is the derived demand for land (4), which also relates land shares to land quality,  $A_{i,k}$ . Using (4) to substitute the unobserved  $A_{i,k}$  in (9), I relate GAEZ yields to observed land allocations and prices:

$$\tilde{y}_{i,k}^G = \zeta_k \frac{\Phi_i}{\lambda_{i,k}} \eta_{i,k}^{\frac{1}{\theta}} e^{u_{i,k}}.$$

Using the assumption that  $\alpha_k/\beta_k$  is a constant  $\nu$  for all crops, and taking logs, I obtain the following intermediate expression:

$$\log \tilde{y}_{i,k}^G = \log \zeta_k + \log \Phi_i + \frac{1}{\theta} \log \eta_{i,k} - \frac{1}{\gamma_k} \log \bar{c}_k - \frac{1}{\gamma_k} \log p_{i,k} - \frac{\alpha_k + \beta_k}{\gamma_k} \frac{1}{1 + \nu} (\nu \log w_{i,A} + \log \rho_i) + u_{i,k}.$$

Finally, noting that  $\gamma_k = 1 - (\alpha_k + \beta_k)$ , and collecting terms, we obtain the estimating equation:

$$\log \left( \frac{1}{p_{i,k}^{\frac{1}{\gamma_k}} \tilde{y}_{i,k}^G} \right) = \frac{1}{\theta} \log \eta_{i,k} + \iota_k + \iota_i + \delta_i \frac{1 - \gamma_k}{\gamma_k} + u_{i,k}. \quad (10)$$

where  $\iota_k$  and  $\iota_i$  are dummies that absorb components that do not vary simultaneously at the region-crop level. To estimate this regression, I use my estimated  $\gamma_k$  to construct the left-hand side of the model. I also use these estimates to construct the regressor  $(1 - \gamma_k)/\gamma_k$ , which I multiply by a region-specific coefficient  $\delta_i$ .

An interpretation of Equation (10) is that we use the model to compute a prediction of  $\tilde{y}_{i,k}^G$ , which is a noisy measure of potential yields. When,  $\eta_{i,k}$ , the land share of region  $i$  devoted to a crop  $k$ , is high, it is either because the region is productive in that crop or because the price of that crop is high. Given output prices, using Equation (9) we would then

---

<sup>26</sup>Here,  $\zeta_k \equiv \gamma_k^{-1} \bar{c}_k^{-\frac{1}{\gamma_k}} (w_A^G)^{-\frac{\alpha_k}{\gamma_k}} (\rho^G)^{-\frac{\beta_k}{\gamma_k}} (p_k^G)^{\frac{1-\gamma_k}{\gamma_k}}$ . My interpretation of the GAEZ data is contingent on the model of production. For a different interpretation, see Costinot, Donaldson, and Smith (2016), who assume that the plot-level elasticity of substitution between labor and land is zero. 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, it does not take into account the spatial variation of those relative prices.

predict a large GAEZ estimate of potential productivity, which we treat as a noisy measure of  $A_{i,k}$ . In the model, a high elasticity  $\theta$  makes land shares very responsive to price and productivity, and hence a change in land share corresponds to a smaller difference in price or inferred yield (see Equation 4) Note that we estimate  $\theta$  without specifying preferences, which makes this approach compatible with many specifications of demand, and not just CES.

Data on  $p_{i,k}$  and  $\eta_{i,k}$  come from the Peruvian Ministry of Agriculture. I estimate (10) on a long sample of national statistics that averages more than 10 years of data and contains information for four departments at the district level (which ensures a better spatial match with GAEZ). I estimate a coefficient on land allocation of 0.603 (std. err. 0.069), which implies an estimate  $\hat{\theta} = 1.658$ . Figure 2(b) relates  $\log\left(p_{i,k}^{\frac{1}{\gamma_k}} \tilde{y}_{i,k}^G\right)$  to  $\log \eta_{i,k}$  after removing the other regressors in Equation (10), thus showing the variation that identifies  $\theta$  ( $R^2 = 0.3$ ). This value implies a large partial elasticity of land allocation with respect to price for the average crop,  $\theta/\bar{\gamma} \approx \frac{1.658}{0.22} = 7.54$ , as well as a large output response to prices. A high value of  $\theta$  reflects that heterogeneity is limited within small regions. It also reflects the fact that by using variation across regions, I am identifying a long-run elasticity.

There are two threats to identification, both of which bias the estimator of  $\theta$  upward. The first threat—which stems from thinking of (10) as an estimable land demand equation—is that through local market equilibrium, unobserved jumps in anticipated yields specific to a crop and location would simultaneously reduce the price and increase land shares. Such shocks would bias the estimated coefficient toward zero, hence leading to an overestimate of  $\theta$ . These unobserved shocks could come, for example, from short-run shocks induced by weather. Averaging the data for each crop  $k$  and region  $i$ , I mitigate the effect of such transitory shocks. Furthermore, I estimate (10) using the sample of crops that are traded internationally to mitigate potential endogeneity induced by local market equilibrium.<sup>27</sup> The second threat is measurement error in  $\eta_{i,k}$ . The Online Appendix handles this threat by bringing in additional measurements of  $\eta_{i,k}$  (from the 2012 Agricultural Census) and shows that my main results go through qualitatively and quantitatively, albeit with a lower value of  $\theta$ , which both threats would suggest is correct.

**Transport Costs.** Since I do not observe a full set of within-country trade flows, I cannot back out the levels of trade costs that rationalize observed regional trade. Instead I construct a network using geo-coded information on the complete set of roads and altitudes in Peru, and

---

<sup>27</sup>While using time-series variation is plausible, it requires additional assumptions on how the  $A_{i,k}$  parameters change over time. It also requires assumptions on the timing of farming decisions relative to the time prices are realized.



use it to predict trade costs for each origin-destination pair in the country (see Donaldson, 2015). Using coffee, a crop that is almost completely exported, I estimate the relative costs of traversing roads of different quality by fitting the following model:

$$\mathbb{E} [\log (p_{n,coffee}/p_{i,coffee} - 1) | \text{geography, roads}] = \beta_0 + \beta_{distance} \log [\text{effective distance}_{ni}(\lambda)], \quad (11)$$

letting  $p_{n,coffee}$  be the price at the port and  $n = \text{Lima}$ , which contains the capital and the main port.<sup>28</sup> The assumption here is that all producing regions  $i$  export some coffee, which ends up in either Lima or Foreign. In Equation (11),  $\beta_{distance}$  and  $\beta_0$  translate effective distance into an iceberg cost, approximated by the price gap. For a given choice of the transport cost parameter  $\lambda$ , “effective distance $_{ni}$ ” is the lowest-cost path between regions  $n$  and  $i$ , calculated according to Dijkstra’s algorithm. The algorithm chooses the path  $R$  that minimizes

$$\text{effective distance}_{ni}(\lambda) = \min_R \sum_{q \in Q} \sum_{\text{edge} \in E_q(R)} [\lambda_q \text{distance}_{\text{edge},q}]. \quad (12)$$

Equation (12) weighs all edges of road quality  $q$  used in route  $R$ ,  $\text{edge} \in E_q(R)$ , by the cost of traversing them,  $\lambda_q$ . In practice, I set  $Q = \{\text{high, low}\}$  and let paved roads be “high” quality. Table 3 compares two versions of the model: (i) a model that constrains  $\lambda_q = 1, \forall q$ , and (ii) a model that constrains only  $\lambda_{high} = 1$ . Taking into account the quality of the road improves the estimation, increases the correlation of data and predictions, and increases the effect of distance from 0.35 to 0.47. I also estimate that in terms of effective distance, unpaved roads are 11.5 times more expensive to transit than paved ones.<sup>29</sup>

To obtain iceberg costs for every pair of regions, I perform the following additional steps: (i) use Equations (11) and (12) to predict  $d_{ni,k}$  for all pairs of districts (constant across  $k$ ), (ii) aggregate trade costs at the province level, and (ii) add an international wedge,  $\tau_k$ , to calculate the costs of trading abroad. Figure 3(a) shows the resulting trade costs with Lima, while Table 4 presents summary statistics. Trade costs range from small in coastal regions

---

<sup>28</sup>Price gaps measure trade costs if region  $i$  actually exports to region  $n$ . According to FAOSTAT, 86 percent of coffee output is exported, which suggests that it is reasonable to assume that the main port is a destination for coffee production.

<sup>29</sup>Equations (11) and (12) are estimated on district data to reflect the precision of the road network data. The Online Appendix considers alternative formulations. First, it shows that the estimates are similar—although the effect of distance is more pronounced—when formulating Equation (11) in levels. Second, it shows that the relative costs  $\lambda_q$  are ordered in the same way, but with larger magnitudes, when using freight rates for a sample of 46 origin-destination pairs as an alternative, but incomplete, measure of trade costs. Finally, it shows how to incorporate altitude in Equation (12), but since it is estimated quite imprecisely, I ignore its role in what follows. I use the estimates discussed above, since they provide conservative but plausible trade costs.

to close to prohibitive for eastern provinces.

Table 3: Estimates of the Transportation Model

	Constrained Model	Road Quality Model
log effective distance $\beta_{dist}$	0.348 ( 0.115)	0.473 ( 0.069)
high quality $\lambda_{hi}$	1.000 —	1.000 —
low quality $\lambda_{lo}$	1.000 —	11.548 ( 5.330)
Intercept $\beta_0$	0.210 ( 0.049)	-0.104 ( 0.061)
N	332	332
R-squared	0.439	0.592

Bootstrapped standard errors in parentheses

Notes: Bootstrapped standard errors in parentheses.

**Demand Elasticity,  $\sigma$ .** I estimate the demand equation implied by CES preferences over crops,

$$\log(s_{i,k,t,h}^{ENAH O}) = \iota_k + \iota_h + \iota_t + (1 - \sigma) \log v_{i,k,t,h}^{ENAH O} + \epsilon_{i,k,t,h}^{ENAH O}, \quad (13)$$

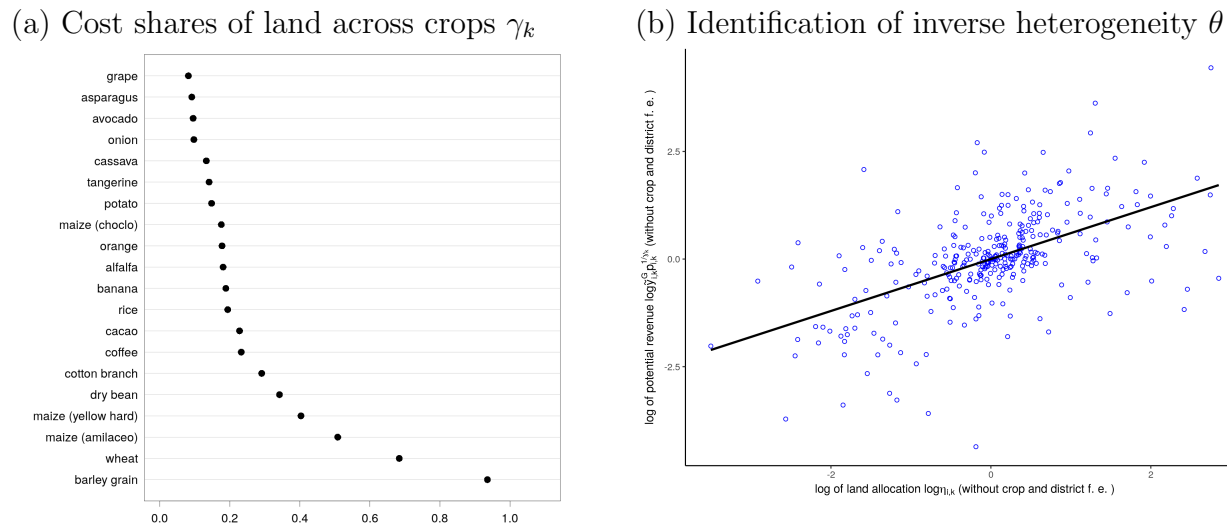
using detailed information from ENAHO on household expenditures and unit values. In (13),  $s_{i,k,t,h}^{ENAH O}$  is the expenditure share and  $v_{i,k,t,h}$  is the unit value (expenditure divided by quantity) of crop  $k$  for household  $h$  at time  $t$  in region  $i$ . The error  $\epsilon_{i,k,t,h}^{ENAH O}$  reflects that expenditure in each category and physical quantities are measured with error. Since unit values are measured as the ratio of expenditure and quantity, these measurement errors create a bias concern (see Deaton, 1997).

I therefore instrument unit values using the GAEZ potential productivity estimates,  $\tilde{y}_{i,k}^G$ . The instrument captures the idea that places that are relatively bad for growing a crop will tend to have higher prices. The estimate in the second stage is  $\hat{\sigma} = 2.385$ , while OLS yields  $\hat{\sigma}^{OLS} = 0.14$ . These results are consistent with measurement error that induces a positive correlation between expenditure shares and unit values, and with the instruments being orthogonal to those errors.<sup>30</sup> As I discuss in detail later, because it controls the sensitivity of demand to crop prices,  $\sigma$  governs the elasticity of trade flows with respect to trade barriers

<sup>30</sup>Results are similar when using region, instead of household, fixed effects. A value of  $\sigma = 2.386$  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. An alternative is to instrument prices using information on freight rates and international prices. Doing so yields a similar estimate of  $\sigma = 2.804$ . The Online Appendix shows these alternative results.

and the transmission of prices across markets.

Figure 2: Estimation of technology parameters



**Other key parameters.** I provide here a summary of how I obtain the remaining parameters, and relegate a more detailed explanation to the Online Appendix. First, to obtain the  $A_{i,k}$  I rely on price and land allocation data from national statistics, combined with my estimates of the cost share parameters. Using Equation (4), I back out the  $A_{i,k}$  parameters that rationalize land allocations given crop and factor price data. In the equilibrium, these  $A_{i,k}$  parameters determine a region’s comparative advantage and therefore are key in the determination of land and revenue shares, which control the first-order impact on welfare and productivity of changes in prices.

Next, I jointly calibrate international barriers,  $\tau_k$ , and domestic demand parameters,  $a_k$ . I choose these parameters to rationalize the difference between domestic and international prices,  $p_{F,k}$ , as well as aggregate apparent consumption of each crop (given production and net export data). I calibrate  $b_i$  to match the aggregate value of production in agriculture, while also matching the relative sizes of agricultural consumption across regions, which come from household consumption surveys. Relative expenditure shares within the non-agricultural sector,  $b_{i,TR}$ , ensure that the model replicates the size of the non-traded sector in each region. In equilibrium, the  $a_k$  parameters determine consumption expenditure shares in each crop and, together with the crop expenditure shares  $b_i$ , control the impact of shocks to prices on the cost of living.

Labor endowments are observed in the 2007 population census, while land endowments,  $H_i$ , are total harvested area from the national statistics on agriculture.<sup>31</sup> I choose regional

<sup>31</sup>I treat  $H_i$  as an exogenous endowment, shutting down the link between access to markets and the margin

non-agricultural productivities,  $T_{i,TR}$  and  $T_{i,NT}$ , to match regional value added per worker in that sector. Following Allen and Arkolakis (2014), I adopt an Armington elasticity of  $\varepsilon = 9$  for the non-agricultural sector.<sup>32</sup> Finally, I consider the three main ports in the country as the ones through which international trade is conducted.

## 6 Baseline Simulation

In this section I discuss key features of the simulated model and how they determine the effects of the trade shocks that I will study in the final sections of the paper.<sup>33</sup> I start by showing that the model is able to replicate the facts that motivate my approach (Section 3). I then show that the model produces a sparse pattern of domestic trade. Appendix D shows that the model accounts for untargeted data, including a sample of domestic trade flows, patterns of intermediate input use, and shares of net exports through different ports.

Intuitively, the main reason the model does not perfectly fit the agricultural data is that I have used variation in prices and land shares to choose only the  $A_{i,k}$  parameters, which means that I only choose half as many parameters as observations. Other reasons include that values of  $d_{ni,k}$  are constant across crops and subject to sampling variation, and that preferences are simple and independent of income.

---

of total cultivated land. The link between deforestation and roads has been studied, for example, by Chomitz and Gray (1996). While this effect is probably important from a societal point of view, I ignore it and assume that the low quality of rain forest soil for agriculture will have a small impact on the productivity or welfare of farmers.

<sup>32</sup>This value implies a trade elasticity larger than the one estimated, for example, by Simonovska and Waugh (2014) using variation across countries (approximately 4). I choose a larger elasticity to reflect that the goods produced by small regions within a country are probably less differentiated than across countries.

<sup>33</sup>To compute the equilibrium described in Section 4, I use Knitro 10.2, a numerical solver that handles complementarity constraints. Since the solver only produces approximate equilibria, I classify as zeros those flows for which  $z_{ni,k}$  is less than the convergence criterion parameter I provide the solver,  $10^{-6}$ .

Figure 3: Trade Costs in Space

(a) Baseline iceberg trade cost to Lima

(b) Average reduction in iceberg trade costs (%)

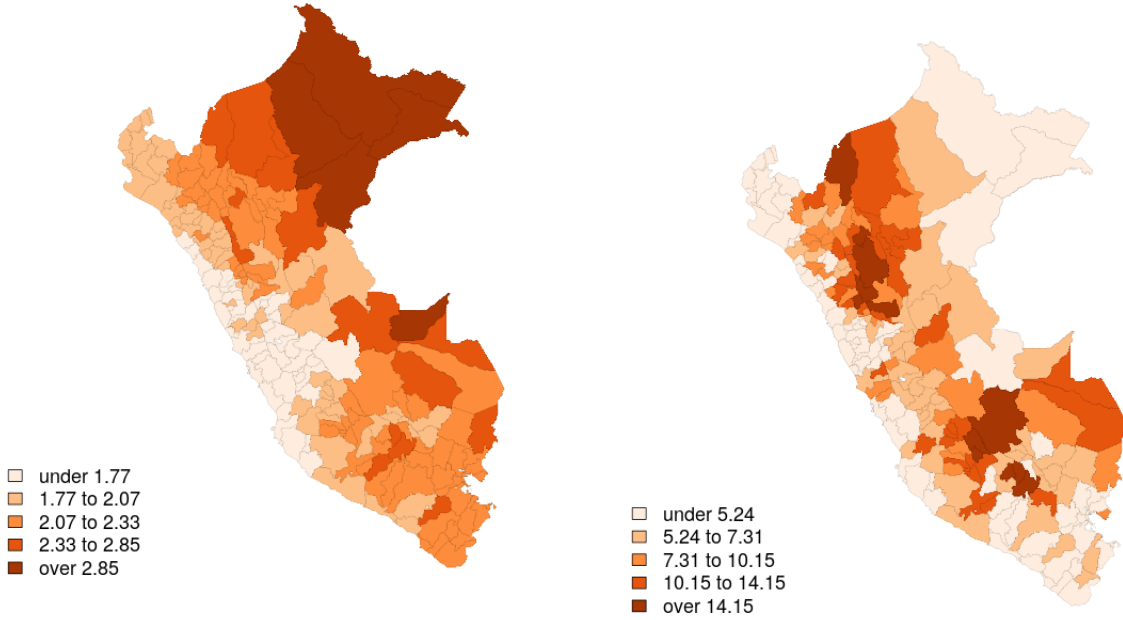


Table 4: Summary Statistics of the Estimates of Iceberg Trade Costs,  $\hat{d}_{ni}$

	mean	std	10%	25%	50%	75%	90%	max
iceberg	2.34	0.41	1.87	2.09	2.32	2.54	2.76	4.35

Notes: The table presents the distribution, across regions, of trade costs  $\hat{d}_{ni}$  averaged across trading partners.

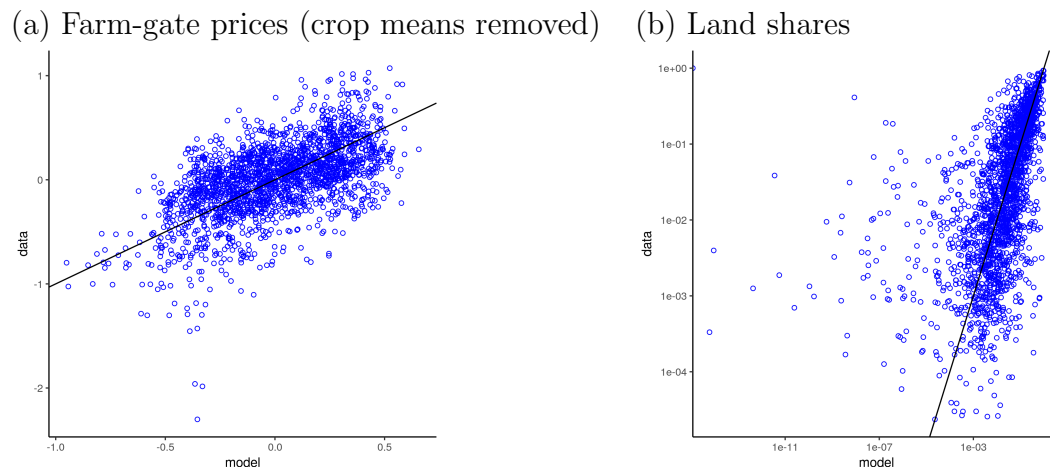
**Farm-Gate Prices (Facts 1 and 4).** Figure 4(a) shows that the simulations capture the variation in log farm-gate prices after removing crop fixed effects. Doing so focuses only on the spatial dispersion of prices—which I am trying to explain—and correctly ignores the fact that on average, some crops are more expensive than others. Note that because I am using national statistics on production, I only include farm-gate prices in this comparison. This is a tougher test of the model, since it ignores the fact that a crop will be pricier in regions that do not produce it; this is a price difference the model also generates. Finally, in the specific case of coffee, a regression of prices on effective distance recovers quantitatively similar coefficients to those shown as motivation in Section, although with a somewhat smaller slope 3 (see Figure A.9).<sup>34</sup> Note that absent trade costs, there would be no spatial variation in

<sup>34</sup>Replicating Table 1 with simulated data is not possible, since the model does not generate within-region price dispersion. Appendix Figure A.9, however, shows that the model captures well the mean difference of farm-gate prices across crops relative to Lima (which contains the largest urban center) that I observe in the data.

prices.

**Land Allocation (Fact 2).** The model captures key aspects of the distribution of  $\eta_{i,k}$ . Appendix Table A.2 shows that the degree to which specialization is incomplete in the model is similar to the data (e.g., the distribution of median shares across regions is close to that in Table 2). Furthermore, the model also captures the pattern of specialization *across* crops and regions: Figure 4(b) compares land shares  $\eta_{i,k}$  in the data and model. The predicted land shares clearly cluster around the 45-degree line, especially for larger land shares. The model predicts somewhat high specialization relative to the data, which is important since, as I argue in Section (7), the allocation of land is one of the main determinants of the welfare and productivity effects of shocks. Recall that these moments are controlled by  $\theta$  and  $\gamma_k$ , which govern the sensitivity of land allocation to prices, as well as by  $A_{i,k}$  (see Equation (4)). For example, if  $\theta \rightarrow \infty$ , each region would specialize in only one crop.

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



Notes: Each panel shows the 45-degree line. In Panel (a), a regression yields a slope of 0.77, with an  $R^2$  of 0.34. In Panel (b), a regression yields a slope of 0.34, with an  $R^2$  of 0.28.

**Revenue per unit of land (Fact 3).** As Proposition 2 shows, the model rationalizes within-region differences in revenue per unit of land via differences in land intensity,  $\gamma_k$ . In the model, these parameters also control revenue shares,  $\pi_{i,k}$ . Sections 3 and 5 showed that variation in both of these variables is substantially accounted for by allowing crops to use land with different intensities. Capturing variation in revenue shares is important, since they govern the importance of shocks to each region. And since land intensities  $\gamma_k$  govern the elasticity of supply of each crop (see Equation 7), they partially control the magnitudes of

changes in prices.<sup>35</sup>

**Trade flows, sparsity, and non-analytic gravity.** Finally, I turn to the quantitative results on domestic trade flows. I discuss two distinctive trade patterns in my baseline quantification: (i) a sparse trade matrix and (ii) a gravity-like relation for the intensive margin of trade.<sup>36</sup> Using a sample of domestic trade flows collected by the Ministry of Agriculture, Appendix D shows that predictions for the intensive and extensive margins find support in the data, even though these features were not targeted directly in the estimation.<sup>37</sup>

First, the simulated bilateral matrix is quite sparse: Of all possible domestic trade links, there are positive trade flows in only 0.4 percent. The average region has 15 positive trade links (either as an importer or exporter). A few large regions, however, import from many regions simultaneously (such as Lima, with 5 percent non-zero flows).

Second, for positive trade flows, the model generates a numerical relation between trade flows and trade barriers, reminiscent of the gravity equation. Using simulated data, I regress log trade flows on log trade barriers (controlling for origin-crop and destination fixed effects) to recover an elasticity of  $-0.73$ . Behind this elasticity is the fact that in neoclassical models the intensive margin of trade may satisfy gravity, with an elasticity equal to  $1 - \sigma$  (Deardorff, 1998). The intensive margin is also present in this model, although the coefficient I recover is lower than  $1 - \sigma = -1.386$ .<sup>38</sup> In different contexts, previous research on agricultural trade has estimated somewhat higher values for these trade elasticities. By focusing on substitution among crops in consumption, the model ignores additional margins of adjustment to price changes, which would yield a higher elasticity.<sup>39</sup>

---

<sup>35</sup>Furthermore, land intensities  $\gamma_k$  are informative of the distributional effect of shocks within the agricultural sector (see Section 9 and Appendix C.2).

<sup>36</sup>Current trade workhorses belonging to the class studied by Arkolakis, Costinot, and Rodriguez-Clare (2012) do produce analytic gravity equations. Models in this class include Eaton and Kortum (2002), the quantitative version of Melitz (2003) developed by Chaney (2008), and the Armington version of the model in this paper, due to Costinot, Donaldson, and Smith (2016). The model I present does not belong in that class. For a recent discussion of theory and empirical work on gravity equations, see Head and Mayer (2014).

<sup>37</sup>Appendix D also discusses the model’s performance for two other untargeted moments: regional use of intermediate inputs and the export share of each port.

<sup>38</sup>Deardorff studies equilibria in which each destination  $n$  purchases crop  $k$  from a single producer  $i$ ; in my simulations, large urban regions do not satisfy this assumption. Moreover, he studies equilibrium in which each region specializes in the production and export of different goods, while in my setup there are more regions than goods.

<sup>39</sup>For example, Donaldson (2015) finds an average elasticity with respect to trade costs of 3.8 across 17 crops in colonial India. Tombe (2015) estimates it to be 4.1 for agriculture as a whole, using international data. Further, the elasticity of simulated trade flows to effective distance is  $-.13$ . This value captures the negative effect of distance on trade, although it is substantially lower than  $-1$ . This low value is not surprising, given Deardorff’s result. For example, at the median trade cost of 2.32, the elasticity of trade costs with respect to distance is  $.273$  (Equation 11), which would imply a coefficient of  $(1 - \sigma) \times .273 \approx -.378$ —substantially below one. Note that the implied elasticity of  $.273$  is comparable to previous findings. For colonial India, Donaldson (2015) finds an elasticity of trade costs with respect to effective distance of  $.25$ . Hummels (2001)

## 7 The Effect of Improving Market Access

To understand the effects of market-access policies on measured productivity and welfare, I consider two counterfactual scenarios that are grounded in policy considerations in Peru. In the first, I simulate the paving of all previously unpaved highways in the National Highway System (Red Vial Nacional), which constitute the main axes in Peru’s road system. In the second counterfactual scenario, I study the effect of building roads that do not exist but are part of the Ministry of Transportation’s expansion plans. Online Appendix Figure H.2 presents maps corresponding to both policies.<sup>40</sup>

### 7.1 Paving Highways in the National System

As stated by the Ministry of Transportation and Communications (2017), a goal of the government of Peru is to pave all roads in the National Highway System—i.e., the main roads in the system—by 2021. In my data set, such a policy requires paving approximately 9,700 km. of roads (or 10 percent of the total network), and implies a median reduction in trade costs of 6.3 percent (3.6 percent standard deviation). The effect is asymmetric, and trade costs that were initially low are essentially unaffected because they were generated by traversing high quality highways. Figure 3(b) shows that the regions in which trade costs change the most tend to be removed from the coast and ports, although not all interior regions benefit equally.

In what follows, I examine the effects of this policy on measured productivity and welfare, focusing on its distributional effects within and across sectors. As I explain below, I use measures of value added to aggregate productivity changes across crops. In turn, to quantify the welfare effects of the policy, I study changes in the real income that accrues to factors working in agriculture separately from the real income in the rest of the economy. In both cases, I measure real income relative to the cost of living in each region, given by  $P_i^{b_i} P_{i,M}^{1-b_i}$ .

**Effects of the policy on measured agricultural productivity and real income.** To discuss productivity, note that since each region can grow many crops, we need a multi-crop index of productivity. One such index is regional value added in wheat equivalents, which measures the productivity of labor and land in constant units. This simple measure is analytically tractable (which I exploit below), and has previously been used in agricultural

---

finds an elasticity of ad valorem freight rates to distance of .27 using international trade data.

<sup>40</sup>In both cases, I calculate trade cost reductions according to the estimates of the transport cost model in Section 5 and simulate the model again, keeping all other parameters constant. The exact hat algebra strategy in Dekle, Eaton, and Kortum (2008)—used, for example in Caliendo and Parro (2015), Parro (2013), and Ossa (2014)—is not applicable here due to the lack of analytic characterization of trade flows.



productivity measurements.<sup>41</sup> Figure 5(a) relates, for each region, the change in agricultural productivity and the average reduction in trade costs, both of which are induced by the policy. While regions that experience large reductions in trade costs also experience large productivity increases, the impact is limited for regions not directly affected by the policy. Measured productivity in the average region increases by 3.8 percent and by 4.9 percent in the aggregate.<sup>42</sup>

To understand what drives this result, consider the following first-order approximation to the change in measured productivity following the policy:

$$\Delta \text{Productivity}_i = \frac{1}{1 - \bar{\beta}_i} \sum_k \pi_{i,k} \Delta p_{i,k} - \frac{\bar{\beta}_i}{1 - \bar{\beta}_i} \Delta \rho_i - \Delta p_{i,\bar{k}}, \quad (14)$$

where  $\Delta$  denotes log changes and  $\bar{k}$  denotes the unit of measurement of productivity. Recall also that in region  $i$ ,  $p_{i,k}$  and  $\pi_{i,k}$  are the price and the revenue share of crop  $k$ , and  $\rho_i$  is the price of the intermediate input.<sup>43</sup> Equation (14) shows that this productivity measure tends to increase when crops with large revenue shares become more valuable and when inputs become cheaper, while changes in revenue shares do not have a first-order effect on productivity (see Costinot and Vogel, 2015). However, for some regions trade cost reductions are large enough that this first-order approximation misses substantial gains due to factor reallocation across crops (see Appendix Figure A.10).

Note that inference on productivity will depend on the unit of measure, due to changes in relative prices. Since here changes in farm-gate prices reflect output that is not lost in transit due to better transportation, it is not clear that one should remove these price changes when measuring productivity improvements. Still, an alternative is to value output in units whose price does not vary across equilibria, which I do using international prices,  $p_{F,k}$ . According to this metric, aggregate productivity grows by approximately 3.8 percent.<sup>44</sup>

Now consider changes in real income earned by labor and land employed in agriculture. Figure 5(b) shows that the policy generates winners and losers, and farmers in those regions

---

<sup>41</sup>See, for example, Hayami and Ruttan (1985). In the case of homogeneous technologies across crops,  $\gamma_k = \gamma$ , this index coincides with  $\Phi_i$ . Further discussion of the properties of this index is provided in the Online Appendix.

<sup>42</sup>According to this measure, productivity declines in some regions in which the price of wheat increases in response to the reform. Productivity in the median region grows by 1.1 percent.

<sup>43</sup>See the Online Appendix for proofs of this and the following first-order approximations. Online Appendix H presents additional statistics for the simulations in this section and the following.

<sup>44</sup>For domestically traded crops, I use average prices at the baseline. This type of measurement is closer to recent fixed-price productivity measures, e.g., in Restuccia, Yang, and Zhu (2008). A direct measure of productivity that does not rely on aggregation is physical yields for each crop. Table H.20 in Online Appendix H shows that following this reform, yields increase for almost every crop. This yield increase reflects access to cheaper intermediate inputs, but also the offsetting effect of increases in land allocation toward crops whose relative price increases.

that face small trade cost reductions lose the most. The median farmer gains 2.7 percent of real income (despite experiencing an approximately 6 percent trade cost reduction). A farmer at the 25th percentile of the welfare change distribution gains only 0.1 percent, while a farmer at the 75th percentile gain approximately 13.2 percent. Although the policy reduces domestic trade costs for all regions, more than 20 percent of farmers lose, and sizable losses are not uncommon: The farmer in the 5th percentile of the welfare gains distribution loses 2.4 percent.<sup>45</sup>

The following first-order approximation to changes in farmers' real income,  $W_i$ , clarifies the sources of this result:

$$\Delta W_i = \sum_k \left( \frac{1}{1 - \bar{\beta}_i} \pi_{i,k} - b_i s_{i,k} \right) \Delta p_{i,k} - (1 - b_i) \Delta P_{i,M} - \frac{\bar{\beta}_i}{1 - \bar{\beta}_i} \Delta \rho_i, \quad (15)$$

For region  $i$ , an increase in the price of crop  $k$  will tend to improve welfare when the crop's revenue share is large and the consumption share is small, as measured by the difference  $(1 - \bar{\beta}_i)^{-1} \pi_{i,k} - b_i s_{i,k}$ .

In specific cases, one can show that to a first-order approximation, changes in productivity and welfare are proportional to the reductions in trade costs.<sup>46</sup> More generally, however, Equations (14) and (15) show that the key to understanding measured productivity and welfare effects are (i) the baseline patterns of specialization and consumption of each region and (ii) the magnitude of changes in crop and input prices, both of which are driven by key assumptions and parameters in the model, as I explain next.

First, baseline patterns of specialization (i.e., revenue and consumption shares,  $\pi_{i,k}$  and  $s_{i,k}$ ) are shaped by comparative advantage, in the form of Ricardian differences in land productivity and of relative factor endowments. Within-region heterogeneity in land quality, inversely related to  $\theta$ , limits the degree of specialization. This is shown clearly by Proposition 3: Specialization captured by land shares,  $\eta_{i,k}$ , together with differences in land intensity,  $\gamma_k$ , determine equilibrium revenue shares. Finally, trade costs (domestic and international) are important to generate the right pattern of incomplete specialization, because they lower

---

<sup>45</sup>For the country as a whole, the compensating variation associated with this policy is 88.5 thousand USD per kilometer paved. This figure is roughly double the one associated with the policy of building roads, detailed in Appendix B. One reason is that National Highways tend to serve as the main axes of the road system, and therefore improvements to it have a larger aggregate impact; another is that the road building policy tends to target remote areas that trade less and are less populated, so the first-order impact of the policy is limited.

<sup>46</sup>For example, suppose a region fully specializes in one crop and imports everything else. Using Equation (14),  $\bar{\beta}_i = 0.22$ , the median trade cost reduction of 6.3 percent, and the median change in the price of wheat equal to 0.12 percent, one predicts a change in measured productivity of  $\Delta d (1 + \bar{\beta}_i) / (1 - \bar{\beta}_i) \times 100 - 0.12 = 9.7$  percent. Of this change, approximately 2 percentage points (or one fifth of the total change) come from intermediate input use.

the prices of export crops in regions far from ports, which decreases the share of exportable crops in revenues.

Second, to understand the price changes induced by the policy, start by noting that a decrease in trade costs entails an increase in the efficiency of Home’s exchange with the rest of the world—as well as an increase in the domestic supply of crops—because less output is lost in transit. Table 5 shows which crops command the largest revenue shares in the regions in which trade costs drop the most. As shown in the table, those regions tend to concentrate their production in internationally traded crops (such as coffee and wheat) and on some domestically-traded, high-consumption goods (such as potato and amilaceo).<sup>47</sup> This is key to what follows, because these are the crops whose supply will increase in response to the policy. More specifically, we expect that increased exports to the rest of the world are accompanied by increased prices of export crops in these regions, and we also expect that the increased supply of domestically traded crops will drive their prices down everywhere else. Finally, increased Foreign imports of certain crops further contribute to domestic price declines, as I explain below.<sup>48</sup>

Table 5: Paving Roads: Production Patterns of Regions the with Largest Trade Cost Reductions

Crop	Avg. $\pi_{i,k}$
potato	17.06
coffee	15.54
maize (amilaceo)	10.05
alfalfa	9.59
wheat	7.01

Notes: The table shows average revenue shares across regions in which the average reduction in trade costs is in the top quartile of the distribution.

To demonstrate how these improved trade opportunities and competition effects work quantitatively, Table 6 presents the crops that the make largest individual contributions to real income, as suggested by Equation (15). Panel (a) focuses on regions in the top decile of the distribution of real income gains, which are precisely the regions that experience large trade cost reductions. The panel shows, for each crop, the average across regions of (i) changes in prices, (ii) contributions to changes in income, and (iii) contributions to changes in consumption prices. Column 1 shows that these regions specialize in crops whose supplies

<sup>47</sup>Online Appendix Table H.19 shows that consumption is highly concentrated in potato, rice, and yellow maize.

<sup>48</sup>On aggregate, agricultural exports grow by 15 percent, while agricultural imports shrink by 3 percent. The slight decrease in agricultural imports reflects an increased domestic supply of import-competing crops (such as wheat, which accounts for 26 percent of baseline imports) and the substitution away for other high-consumption crops whose price declines

we expected to increase, given the production patterns presented in Table 5. Moreover, the reduction in trade costs experienced by these regions translates, as the second column shows, into increases in these crops' prices. The price increases, together with each region's patterns of specialization, lead to large contributions to real income increases.<sup>49</sup>

Panel (b) focuses instead in regions at the bottom decile of the distribution of real income changes (i.e., losers). An important fact emerges here: These regions tend to specialize in the same crops as the winners but, as we anticipated, prices instead decrease due to increased competition. Moreover, note that if region  $i$  loses, it is both because of tougher competition in  $i$ 's own market and also in other markets to which  $i$  used to export: Among the losers, exports to other regions shrink up to 23 percent, depending on the crop (not shown in the table). The key takeaway is that internal geography and comparative advantage play an essential role in the determination of prices. For crops that contribute to the largest losses, prices are determined by regional, instead of international, equilibria. Three of these crops—potatoes, amilaceo, and alfalfa—are not traded internationally; for the others, although potentially tradable, there is in fact little trade in them at the baseline. Note, moreover, that these regions also have large revenue shares in goods whose imports from Foreign increased (such as bananas and onions), which puts additional downward pressure on prices.<sup>50</sup>

Note that for both winning and losing regions the expenditure share in agriculture,  $b_i$  is small, so income effects dominate the effect of prices on consumption. Finally, since trade costs drop in the winning regions, the accompanying reduction in  $P_{i,M}$ , the non-agricultural price index, reinforces the positive income shock due to increased productivity, which amplifies the real income gains. But because trade cost reductions in losing regions are small, the effect of the policy on  $P_{i,M}$  is also more muted. In fact, for winners, the non-agricultural price index drops by an average of 3.3 percent, while for losers it barely moves.

---

<sup>49</sup>Winners of this policy tend to be households in the middle of the real income distribution. The reason is that, empirically, National highways tend to be close to richer areas, while the poorest areas tend to be served by Departmental and Rural roads. Therefore, the policy benefits relatively richer regions among those that were served by low-quality roads.

<sup>50</sup>Farmers in coastal regions are not the only losers. Since the policy reduces trade costs relatively more in certain regions, several interior regions lose as well. As remote regions get connected, previous producers of *domestic* crops face tougher competition. Contrast this result with Cosar and Fajgelbaum (2016), where fixed factors in coastal locations lose from improving market access in remote locations. In their paper, however, it is worker migration into remote locations that produces the result. Figures H.3(a) and H.3(b) in the Online Appendix document the spatial distribution of welfare changes.

Table 6: Paving Roads: Contributions to Changes in Real Income for Winners and Losers

(a) Regions in top decile of real income change distribution (Winners)				
Crop	Avg. $\pi_{i,k}$	Avg. $\Delta p_{i,k}$	Avg. $\frac{1}{1-\beta_i} \pi_{i,k} \Delta p_{i,k}$	Avg. $b_i s_{i,k} \Delta p_{i,k}$
coffee	19.72	12.02	4.45	0.00
potato	15.85	7.49	2.83	0.12
banana	9.46	9.38	2.07	0.06
maize (amilaceo)	6.56	9.96	1.46	0.03
wheat	6.96	5.64	1.37	0.05

(b) Regions in bottom decile of real income change distribution (Losers)				
Crop	Avg. $\pi_{i,k}$	Avg. $\Delta p_{i,k}$	Avg. $\frac{1}{1-\beta_i} \pi_{i,k} \Delta p_{i,k}$	Avg. $b_i s_{i,k} \Delta p_{i,k}$
potato	39.56	-2.13	-1.10	-0.02
banana	9.82	-1.76	-0.35	-0.01
onion	6.51	-2.18	-0.18	0.00
alfalfa	7.63	-1.81	-0.17	-0.02
maize (amilaceo)	7.90	-2.35	-0.16	0.00

Notes: In Panel (a), each column shows the simple average across regions in the top decile of the distribution of welfare changes in response to the policy. In Panel (b), each column shows the simple average across regions in the bottom decile of the distribution of welfare changes in response to the policy.

In general equilibrium, shocks to regional export and import prices spread through expenditure switching and land switching. Consider, for example, the transmission of import prices through expenditure switching. Appendix A shows analytically that since crops are substitutes in consumption ( $\sigma > 1$ ), an exogenous increase in the price of crop  $j$  imported in region  $i$  directs spending away from crop  $j$ , thereby shifting local demand for other crops. The magnitude of this shift increases with the expenditure shares of crop  $j$  in consumption and with the elasticity of substitution in consumption  $\sigma$ . The Appendix also shows that a higher value of the elasticity  $\theta/\gamma_k$  translates more of a given demand shift into an increase in the local price of crop  $k$  rather than an increase in its local production. An analogous reasoning holds for land reallocation. An exogenous increase in the price of a crop exported by region  $i$  diverts land away from other crops; with a large elasticity  $\sigma$ , this translates into smaller price effects and larger quantity effects to other crops.

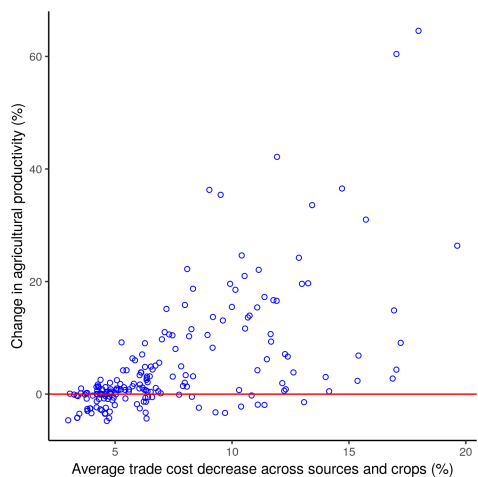
**Effects of the policy on non-agricultural real income.** Finally, Figure 5(c) shows that where farmers win, non-agricultural real income increases substantially less than agricultural income (in proportional terms), even though non-agricultural firms experience the same improvement in their export costs as farmers do. Since both farmers and non-agricultural producers face the same consumption prices, the reason for this asymmetric outcome lies in income changes.

The expansion capacity of non-agricultural firms is limited. While farmers are ultimately able to sell crops to Foreign at a fixed price, non-agricultural producers face a downward-

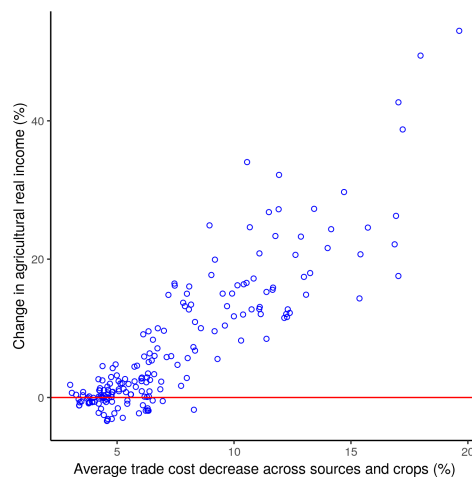
sloping demand there; they also face increased competition from remote regions in domestic markets, including their own. In both cases, demand elasticity is determined by regional differentiation and quantitatively governed by  $\epsilon$ .<sup>51</sup>

Figure 5: Paving Roads: Counterfactual Changes in Measured Productivity and Real Income

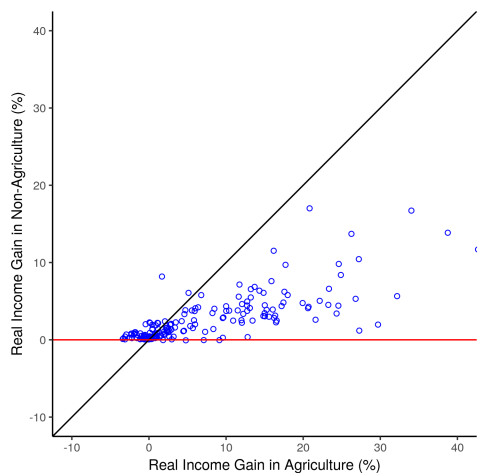
(a) Measured agricultural productivity



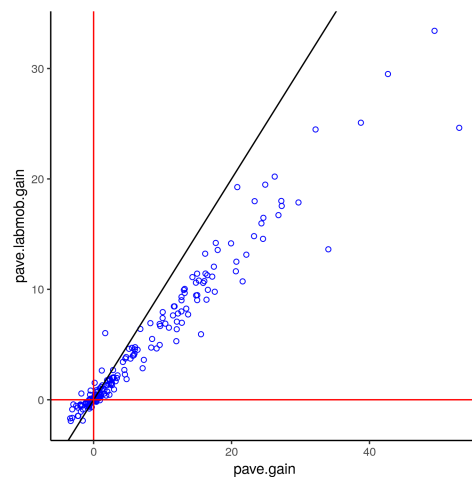
(b) Agricultural real income



(c) Real income: Non-agriculture rel. to agriculture



(d) Agricultural real income (mobile labor)



Notes: Panels (c) and (d) display the 45-degree line.

<sup>51</sup>Note that if all of the non-agricultural sector were non-traded,  $b_{i,TR} = 0$ , the change in non-agricultural wages,  $\Delta w_{i,M}$ , would be proportional to changes in farmers' income, so there would be no distributional effect. Note, as well, that if the only source of demand for non-agricultural output is Foreign (i.e.,  $b_i = 1$ ), then the change in non-agricultural wages would be a factor  $(\epsilon - 1)/\epsilon < 1$  of the reduction in the cost of trading with Foreign,  $\Delta d_{i,F}$ .

## 7.2 Building New Roads

I consider now the effect of building new roads according to the Ministry’s expansion plans. I focus here on the effect of the policy on the real income of farmers and discuss the policy in more detail in Appendix B. This policy requires constructing approximately 2,300 km of new roads in a spatially targeted fashion, totaling a 2.5 percent expansion of the road system. The reduction in trade costs is focused and small: It implies a median reduction in trade costs of 0.5 percent. In this case, real agricultural income for the median farmer increases by 0.02 percent, while farmers at the 25th percentile of the welfare gains distribution lose 0.1 percent and those at the 75th percentile experience 0.25 percent gains.<sup>52</sup> More than a quarter of farmers, however, lose from the policy. As before, competition effects are key to understanding these results. Regions whose access to markets improves as a result of the policy produce large amounts of goods not traded internationally (such as potato). As these regions’ supply to the rest of the country increases, agricultural prices drop everywhere, affecting mainly other regions that specialize in those non-traded crops.

## 8 The Effect of a Shock to International Prices

Finally, I study a shock to international trading opportunities by examining how a change to international crop prices spreads throughout the domestic economy. This exercise also shows that the degree of domestic trade integration is critical for understanding the effects of such a shock. Specifically, I study an international price shock in line with a set of World Bank’s simulations of potential effects of the Doha trade talks. The shock entails a mild increase in the price of cereals and cotton, as well a small decrease in the price of fruit crops.<sup>53</sup>

The median farmer benefits (0.5 percent real income gain) and the median non-agricultural worker loses (0.05 percent real income loss), but these measures mask substantial heterogeneity in both sectors: Farmer gains range from -1.2 to 2.7 percent of real income and non-agricultural gains range from -0.1 to 0.1 percent.

To understand this heterogeneity, consider again the first-order approximation to changes in welfare (Equation (15)). As before, baseline specialization in consumption and production—which reflect comparative advantage and how it is limited by land heterogeneity within

---

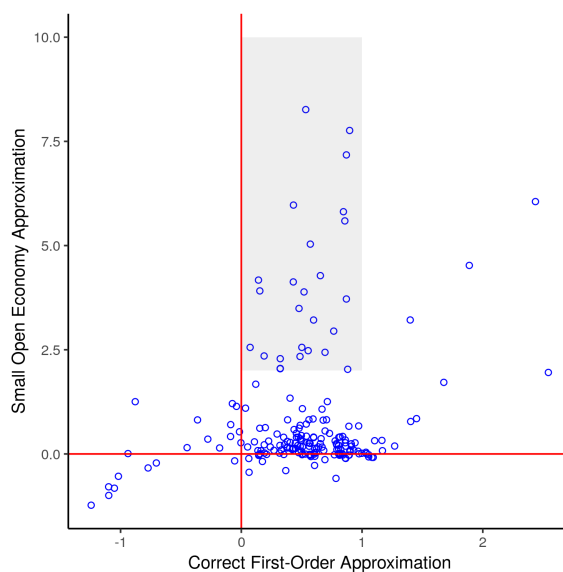
<sup>52</sup>Because this is a focused policy there are a few large winners and losers. At the 90th percent of the distribution, farmers win close to 2 percent, while at the 10th they lose 0.5 percent.

<sup>53</sup>I feed in the simulated changes in world prices,  $\Delta p_{F,k}$ , reported by Hertel and Ivanic (2006), keeping all other parameters constant. For each crop, I average the import and export price changes for the world, relative to a bundle of rich country manufacturing exports (Tables 3.8 and 3.9 in Hertel and Ivanic, 2006). Then I classify the crops in my data according to their categories. The largest price changes are for rice (7.7 percent), wheat (1.6 percent), other cereals (3.45 percent), and plant fibers (5.9 percent). I assume these price changes apply relative to the numeraire.

regions—are two of the main determinants of the first-order impact of price changes. But it is also important to realize that foreign shocks do not transmit directly to domestic prices, because many regions do not directly trade these crops with Foreign.

The role of price changes can be summarized as follows: Evaluate (15) invoking the small open economy assumption that international prices apply directly to all regions at Home (i.e., at  $\Delta p_{i,k} = \Delta p_{F,k}$ ), and compare the result to what we obtain by evaluating it instead at the simulated equilibrium price changes. Because the shocks to foreign prices are small, the prediction involving equilibrium prices is quite accurate (see Appendix Figure A.10(b)). However, Figure 6 shows that although both predictions are positively correlated, the small open economy prediction is often quantitatively off. Moreover, it gets the sign of welfare changes wrong in about 23 percent of the regions.

Figure 6: International Price Shock: Agricultural Real Income Change (Model and Small Open Economy Approximation)



The small open economy assumption does not yield good predictions because domestic trade costs tend to insulate remote regions from shocks to international prices. This is clearly seen by focusing on regions in the shaded area of Figure 6, in which the small open economy assumption is quantitatively off. For those regions, Table 7(a) lists the top 5 crops according to their average *expected* contribution to income, based on the small open economy assumption. As the table shows, these regions tend to be specialized in the production of crops whose foreign price increases (that is, they have high revenue shares  $\pi_{i,k}$  in crops for which  $\Delta p_{F,k}$  is large), so one would expect these crops to make large contributions to income,  $1/(1 - \bar{\beta}_i) \pi_{i,k} \Delta p_{F,k}$ . The last two columns of the table show that, instead, the



actual contribution of these crops is much smaller than expected, because the foreign price shock is transmitted only imperfectly to regional prices.<sup>54</sup>

Abstracting from general-equilibrium adjustments through regional factor and goods markets, a shock to the foreign price of crop  $k$  will only have a direct effect on  $p_{i,k}$  if region  $i$  directly imports or exports that crop from Foreign. Thus, for crops whose foreign price increases in this counterfactual (such as rice, cotton, and yellow maize), domestic prices only capture the full increase close to ports; elsewhere the price increase is smaller. In fact, despite the iceberg trade cost formulation, because not every region exports or imports each crop from Foreign, the pass-through in the aggregate is imperfect.<sup>55</sup>

Table 7: International Price Shock: Contributions to Changes in Real Income for Selected Regions

(a) Regions in which the SOE assumption fails quantitatively					
Crop	Avg. $\pi_{i,k}$	$\Delta p_{F,k}$	Avg. $\frac{1}{1-\beta_i}\pi_{i,k}\Delta p_{F,k}$	Avg. $\Delta p_{i,k}$	Avg. $\frac{1}{1-\beta_i}\pi_{i,k}\Delta p_{i,k}$
rice	34.77	7.65	3.41	0.57	0.25
maize (yellow hard)	14.49	3.45	0.64	0.76	0.12
cotton branch	2.93	5.90	0.22	0.64	0.02
wheat	2.33	1.60	0.05	1.38	0.04
alfalfa	5.64	0.00	0.00	0.58	0.04

(b) Regions in top decile of real income change distribution (Winners)					
Crop	Avg. $\pi_{i,k}$	$\Delta p_{F,k}$	Avg. $\frac{1}{1-\beta_i}\pi_{i,k}\Delta p_{F,k}$	Avg. $\Delta p_{i,k}$	Avg. $\frac{1}{1-\beta_i}\pi_{i,k}\Delta p_{i,k}$
potato	55.92	0.00	0.00	0.90	0.64
rice	18.09	7.65	1.78	1.37	0.61
cotton branch	12.11	5.90	0.91	3.24	0.48
maize (yellow hard)	9.51	3.45	0.42	1.87	0.21
maize (choclo)	10.30	0.00	0.00	0.95	0.13

Notes: In Panel (a), each column shows the simple average across the regions in the area highlighted in Figure 6. In Panel (b), each column shows the simple average across regions in the top decile of the distribution of welfare changes in response to the international price shock.

To understand how the rest of the agricultural sector reacts to the shock, focus on the top 10 percent of regions in the distribution of real income changes. Table 7(b) shows the top five crops according to their simulated contributions to changes in income. The foreign price of three of these crops—cotton, yellow maize, and rice—increases substantially, while these winning regions tend to be specialized in those crops (Columns 1 and 2). The table also shows that these regions capture a larger fraction of the foreign price increase because

<sup>54</sup>For these regions, average trade costs with Foreign,  $d_{iF,k}$ , are larger than the rest of the country—close to the 75th percentile of the distribution of the costs of trading with Foreign.

<sup>55</sup>Online Appendix Figure H.9 displays the distribution of price changes across regions for each crop and shows that only a few regions experience changes in  $p_{i,k}$  equal to the exogenous shocks to  $p_{F,k}$ . Online Appendix Figure H.10 shows that regions in which  $\Delta p_{i,k} \approx \Delta p_{F,k}$  tend to be close to ports. Finally, Online Appendix Figure H.11 shows that aggregate pass-through is imperfect.

they are better connected to ports.<sup>56</sup> Importantly, some of the winners were specialized in the production of non-traded crops (such as potato), for which the price increase is entirely indirect, and about half as large as the increase, for example, in the regional price of yellow maize (and about a quarter of the foreign price shock). The reason, as explained above, is the substitutability of crops in consumption and production. Empirically, the foreign shock price is largest for high-consumption cereals, which then increases the price of other non-traded crops through expenditure switching. Moreover, the increase in foreign prices diverts land away from these crops, which also increases their price.

Finally, as mentioned previously, non-agricultural workers experience real income losses on average, since the shock tends to increase the price of the crops they consume. The effect is larger again for workers close to ports. On average, a 1 percent reduction in baseline trade costs with Foreign leads to a 0.05 percentage point increase in these workers' real income losses (see Online Appendix Figure H.12).

## 9 Robustness

This section considers how results from Sections 7 and 8 change in three alternative calibrations. First, I show that allowing for labor mobility across sectors leads to an expansion of agricultural employment in regions whose market access improves. Second, I show that, within the agricultural sector, labor tends to benefit more than land in a calibration in which I separate ownership of both. Finally, I explain how my quantitative results change in a calibration in which land shares are homogeneous.

**Paving roads when labor is mobile within regions.** Over long periods of time, the assumption of fixed labor supplies becomes less plausible. To understand the implications of allowing for labor mobility across sectors, I study an environment in which worker heterogeneity induces a positively sloping sectoral labor supply. Appendix C.1 explains how I incorporate worker heterogeneity in the style of Lagakos and Waugh (2013) into the model, and how I calibrate preferences and productivities such that initial labor allocations match those in the baseline calibration (and the data).

Figure 5(d) compares changes in agricultural real income in the baseline specification to those obtained when allowing workers to sort across sectors. Two main findings emerge. First, there is a strong positive association between changes in real agricultural income in both scenarios. Second, allowing for worker mobility across sectors weakens the distributional

---

<sup>56</sup>For these regions, average trade costs with Foreign,  $d_{iF,k}$ , are comparable to the 30th percentile of the overall distribution of trade costs with Foreign.

effects of the policy: In the agricultural sector, losers tend to lose less, while winners also tend to win less. Part of the reason is that, as argued above, the agricultural sector is able to take advantage of trade shocks more easily than the rest of the economy, which creates an incentive for labor to move to agriculture. In addition, export crops tend to be more labor intensive, which means that agricultural labor benefits more than land in regions in which trade costs decrease, which strengthens the incentive for labor to move into agriculture (recall Figure 2(a)). Thus, as shown in Appendix Figure A.2, the policy induces further specialization in agriculture in regions in which trade costs dropped the most.

**Distributional effects for land and labor.** Appendix C.2 presents a calibration of the model in which landownership is distinct from ownership of labor. It shows that among winners, this policy benefits labor more than land because export crops tend to be labor intensive. Thus, within the agricultural sector, landowners gain proportionally less than labor (see Figure A.3). The Appendix also shows that these distributional effects occur only when cost shares are heterogeneous across crops, because otherwise both land and labor earn a fixed share of total agricultural income.

**The quantitative role of heterogeneous land intensity.** Appendix C.3 studies agricultural income changes induced by the international price shock in Section 8, under an alternative calibration in which all crops are produced using the same technologies,  $\gamma_k = \gamma$ . Figure A.4 shows that compared to the baseline calibration, winners win more and losers lose less. The main reason, as the figure shows, is that this alternative calibration eliminates heterogeneity in crop supply elasticities—which, in turn, govern the equilibrium adjustment of prices to external shocks.

## 10 Conclusion

The main message of this paper is that trade costs within countries matter. Together with comparative advantage, trade costs determine regional agricultural specialization and ultimately govern the local impact of policies such as building roads or world trade liberalization, which aim at improving the lives of farmers in developing countries

Infrastructure policy—which reduces trade costs—can improve welfare and productivity for some farmers, but can also hurt others in equilibrium. Specifically, in the context of Peru, a simulated road paving policy leads to a 2.7 percent increase in the median farmer’s welfare. These gains, however, are concentrated in regions in which market access improves the most: More than 20 percent of farmers lose due to increased competition from remote suppliers,

even though the policy reduces trade costs for all regions. The losers are concentrated in coastal and inland regions that specialize in the production of crops less traded internationally (such as potato or certain types of maize) and in regions in which market access improved only modestly.

Foreign price shocks that would arise from a decline of world tariffs have limited pass-through to local prices within Peru. In response to an approximately 5 percent increase in the relative price of cereals and fibers, winners tend to be in regions that specialize in those crops and that are also close to ports. But farmers in other well-connected regions benefit as well, even if they specialize in crops that are traded only domestically. Through equilibrium expenditure and land reallocation, the prices of these crops increase by about one quarter of the Foreign price increase in cereals and fibers.

## References

- ADAMOPOULOS, T. (2011): “Transportation Costs, Agricultural Productivity, And Cross-Country Income Differences,” *International Economic Review*, 52(2).
- ADAMOPOULOS, T., AND D. RESTUCCIA (2014): “The Size Distribution of Farms and International Productivity Differences,” *American Economic Review*, 104(6), 1667–97.
- ALLEN, T., AND C. ARKOLAKIS (2014): “Trade and the Topography of the Spatial Economy,” *The Quarterly Journal of Economics*, 129(3), 1085–1140.
- ALLEN, T., AND D. ATKIN (2016): “Volatility and the Gains from Trade,” Working Paper 22276, National Bureau of Economic Research.
- ARKOLAKIS, C., A. COSTINOT, AND A. RODRIGUEZ-CLARE (2012): “New Trade Models, Same Old Gains?,” *American Economic Review*, 102(1), 94–130.
- ATKIN, D. (2013): “Trade, Tastes, and Nutrition in India,” *American Economic Review*, 103(5), 1629–63.
- BEHRMAN, J. R., AND A. B. DEOLALIKAR (1989): “Is Variety the Spice of Life? Implications for Calorie Intake,” *The Review of Economics and Statistics*, 71(4), 666–72.
- CALIENDO, L., AND F. PARRO (2015): “Estimates of the Trade and Welfare Effects of NAFTA,” *Review of Economic Studies*, 82(1), 1–44.
- CHANEY, T. (2008): “Distorted Gravity: The Intensive and Extensive Margins of International Trade,” *American Economic Review*, 98(4), 1707–21.

- CHOMITZ, K. M., AND D. A. GRAY (1996): “Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize,” *The World Bank Economic Review*, 10(3), 487–512.
- COSAR, A. K., AND P. D. FAJGELBAUM (2016): “Internal Geography, International Trade, and Regional Specialization,” *American Economic Journal: Microeconomics*, 8(1), 24–56.
- COSTINOT, A., AND D. DONALDSON (2012): “Ricardo’s Theory of Comparative Advantage: Old Idea, New Evidence,” *American Economic Review, Papers and Proceedings*, 102(3), 453–58.
- (2014): “How Large are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-1997,” .
- COSTINOT, A., D. DONALDSON, AND C. SMITH (2016): “Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World,” *Journal of Political Economy*, 124, 205–248.
- COSTINOT, A., AND J. VOGEL (2015): “Beyond Ricardo: Assignment Models in International Trade,” *Annual Review of Economics*, 7(20585), 31–62.
- COURANT, P. N., AND A. V. DEARDORFF (1992): “International Trade with Lumpy Countries,” *Journal of Political Economy*, 100(1), 198–210.
- DEARDORFF, A. (1998): “Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?,” in *The Regionalization of the World Economy*, NBER Chapters, pp. 7–32. National Bureau of Economic Research, Inc.
- DEATON, A. (1997): *The analysis of household surveys: a microeconometric approach to development policy*. The World Bank.
- DEKLE, R., J. EATON, AND S. KORTUM (2008): “Global Rebalancing with Gravity: Measuring the Burden of Adjustment,” *IMF Staff Papers*, 55(3), 511–540.
- DIAS AVILA, A. F., AND R. E. EVENSON (2010): *Total Factor Productivity Growth in Agriculture: The Role of Technological Capital* vol. 4 of *Handbook of Agricultural Economics*, chap. 72, pp. 3769–3822. Elsevier.
- DISDIER, A.-C., AND K. HEAD (2008): “The Puzzling Persistence of the Distance Effect on Bilateral Trade,” *The Review of Economics and Statistics*, 90(1), 37–48.
- DIX-CARNEIRO, R. (2014): “Trade Liberalization and Labor Market Dynamics,” *Econometrica*, 82(3), 825–885.

- DONALDSON, D. (2015): “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” *American Economic Review*, Forthcoming.
- DONOVAN, K. (2018): “Agricultural Risk, Intermediate Inputs, and Cross-Country Productivity Differences,” .
- 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.
- FABER, B. (2014): “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System,” *The Review of Economic Studies*, 81(3), 1046–1070.
- FAJGELBAUM, P., AND S. REDDING (2014): “External Integration and Internal Development: Evidence from Argentina, 1870-1914,” NBER Working Papers 20217, National Bureau of Economic Research, Inc.
- FAJGELBAUM, P. D., AND A. K. KHANDELWAL (2016): “Measuring the Unequal Gains from Trade,” *The Quarterly Journal of Economics*, p. Forthcoming.
- 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 (2013): “The Agricultural Productivity Gap,” *The Quarterly Journal of Economics*.
- GOLLIN, D., AND R. ROGERSON (2014): “Productivity, transport costs and subsistence agriculture,” *Journal of Development Economics*, 107(C), 38–48.
- 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 (2014): *Gravity Equations: Workhorse, Toolkit, and Cookbook* vol. 4 of *Handbook of International Economics*, chap. 3, pp. 131–195. Elsevier.
- HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): “Estimating Trade Flows: Trading Partners and Trading Volumes,” *The Quarterly Journal of Economics*, 123(2), 441–487.

- HERTEL, T. W., AND M. IVANIC (2006): “Assessing the World Market Impact of Multilateral Trade Reforms,” in *Poverty & the WTO. Impacts of the Doha Development Agenda*. Palgrave Macmilland and the World Bank.
- HUMMELS, D. (2001): “Toward a Geography of Trade Costs,” .
- 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.
- MARKUSEN, J. R. (2013): “Putting per-capita income back into trade theory,” *Journal of International Economics*, 90(2), 255–265.
- MELITZ, M. J. (2003): “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71(6), 1695–1725.
- MESQUITA MOREIRA, M., J. BLYDE, C. VOLPE, AND D. MOLINA (eds.) (2013): *Too far to export: domestic transport cost and regional export disparities in Latin America and the Caribbean*. Inter-American Development Bank.
- MINISTERIO DE TRANSPORTES Y COMUNICACIONES DE PERU (2011a): *Anuario Estadístico 2011*.
- (2011b): “D.S. N 012-2013-MTC,” .
- (2013): *Anuario Estadístico 2013*.
- (2017): “Plan Operativo Institucional,” Discussion paper.
- OSSA, R. (2014): “Trade Wars and Trade Talks with Data,” *American Economic Review*, 104(12), 4104–46.
- 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.
- RAUCH, J. E. (1991): “Comparative Advantage, Geographic Advantage and the Volume of Trade,” *Economic Journal*, 101(408), 1230–1244.

- 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.
- SIMONOVSKA, I., AND M. E. WAUGH (2014): “The elasticity of trade: Estimates and evidence,” *Journal of International Economics*, 92(1), 34–50.
- STOREYGARD, A. (2016): “Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa,” *The Review of Economic Studies*, 83(3), 1263–1295.
- SWIECKI, T. (2014): “Intersectoral Distortions and the Welfare Gains from Trade,” Discussion paper.
- TOMBE, T. (2015): “The Missing Food Problem,” *American Economic Journal: Macroeconomics*, 7(3), 1–33.
- VOLLRATH, D. (2011): “The agricultural basis of comparative development,” *Journal of Economic Growth*, 16, 343–370.
- WORLD BANK (2007): *World Development Report 2008 : Agriculture for Development*. The World Bank.



# Appendix

## A Price transmission across crops

In this Section I study how regional market prices respond to an exogenous shock to the price of one crop. I first study changes in revenues and expenditures, then move on to exploit regional equilibria to understand the changes in prices. Finally, I study how the adjustment depends on key elasticities  $\theta$  and  $\sigma$ .

### A.1 Change in the equilibrium price of crop $k$

**Change in revenue and expenditure in  $k$ .** To simplify the calculations, I study the following special case: (i) intermediate inputs and land are the only factors of production and (ii) land intensities are homogeneous across crops  $\gamma_k = \gamma$ . For clarity, drop the region index (all analysis applies to a particular region  $i$ ).

In response to a shock in prices,  $\Delta p_k$  and  $\Delta \rho$ , the log-change in total revenue coming from crop  $k$ , is:

$$\Delta V_k = \frac{\theta}{\gamma} \Delta p_k + \left( \frac{1}{1-\beta} - \frac{\theta}{\gamma} \right) \sum_l \eta_l \Delta p_l - \frac{\beta}{1-\beta} \Delta \rho$$

and after grouping the terms:

$$\Delta V_k = \left( (1 - \eta_k) \frac{\theta}{\gamma} + \frac{\eta_k}{1-\beta} \right) \Delta p_k - \left( \frac{\theta}{\gamma} - \frac{1}{1-\beta} \right) \sum_{l \neq k} \eta_l \Delta p_l - \frac{\beta}{1-\beta} \Delta \rho.$$

Note that  $(1 - \eta_k) \frac{\theta}{\gamma} + \frac{\eta_k}{1-\beta} > 0$  and  $\frac{1}{1-\beta} - \frac{\theta}{\gamma} < 0$ , since  $1 - \beta = \alpha + \gamma$  and  $\theta \geq 1$ .

Next, the change in expenditure in  $k$  is

$$\Delta (p_k C_k) = (1 - \sigma) (1 - s_k) \Delta p_k - (1 - \sigma) \sum_{l \neq k} s_l \Delta p_l + \chi \frac{1}{1-\beta} \sum_l \eta_l \Delta p_l$$

where  $\chi$  is the fraction of total spending in the region that comes from agriculture value added.

**Equilibrium in crop  $k$ .** For equilibrium in local market of crop  $k$  to hold,  $\Delta p_k C_k = \Delta V_k$ . Assuming  $\chi$  is small enough to ignore:

$$\begin{aligned} \left( (1 - \eta_k) \frac{\theta}{\gamma} + \frac{\eta_k}{1 - \beta} \right) \Delta p_k + \left( \frac{1}{1 - \beta} - \frac{\theta}{\gamma} \right) \sum_{l \neq k} \eta_l \Delta p_l - \frac{\beta}{1 - \beta} \Delta \rho = \\ (1 - \sigma) (1 - s_k) \Delta p_k - (1 - \sigma) \sum_{l \neq k} s_l \Delta p_l \end{aligned}$$

Now, suppose  $p_j$  changes exogenously. We will focus on the direct effect of this change on crop  $k$ 's price (i.e., we ignore all other effects on other crops  $l \neq j, k$ .) We also set  $\Delta \rho = 0$ , although it is clear that it will tend to increase the price of  $p_k$  (because  $\Delta \rho > 0$  shifts back the supply). Collecting terms:

$$\left[ (1 - \eta_k) \frac{\theta}{\gamma} + \frac{\eta_k}{1 - \beta} + (\sigma - 1) (1 - s_k) \right] \Delta p_k = \left[ \left( \frac{\theta}{\gamma} - \frac{1}{1 - \beta} \right) \eta_j + (\sigma - 1) s_j \right] \Delta p_j + \frac{\beta}{1 - \beta} \Delta \rho,$$

solving for  $\Delta p_k$ :

$$\Delta p_k = \left[ (1 - \eta_k) \frac{\theta}{\gamma} + \frac{\eta_k}{1 - \beta} + (\sigma - 1) (1 - s_k) \right]^{-1} \left\{ \left[ \left( \frac{\theta}{\gamma} - \frac{1}{1 - \beta} \right) \eta_j + (\sigma - 1) s_j \right] \Delta p_j + \frac{\beta}{(1 - \beta)} \Delta \rho \right\}$$

and setting  $\Delta \rho = 0$ :

$$\Delta p_k = \left[ \frac{1}{1 - \beta} + (1 - \eta_k) \left( \frac{\theta}{\gamma} - \frac{1}{1 - \beta} \right) + (\sigma - 1) (1 - s_k) \right]^{-1} \left[ \left( \frac{\theta}{\gamma} - \frac{1}{1 - \beta} \right) \eta_j + (\sigma - 1) s_j \right] \Delta p_j$$

which shows that the coefficient of  $\Delta p_j$  on  $\Delta p_k$  is always positive. The magnitude of the effect is increasing in  $\eta_k$ ,  $\eta_j$ ,  $s_k$ , and  $s_j$ , since they determine larger shifts along the supply and demand of  $j$  and larger shifts in the supply and demand of  $k$ .

## A.2 Dependence of $\Delta p_k / \Delta p_j$ on $\theta$ and $\sigma$

First we ask whether  $\partial(\Delta p_k / \Delta p_j) / \partial \theta < 0$ . The sign of the derivative is equal to

$$\begin{aligned} \text{sign} \left\{ \frac{\eta_j}{\gamma} \left[ (1 - \eta_k) \frac{\theta}{\gamma} + \frac{\eta_k}{1 - \beta} + (\sigma - 1) (1 - s_k) \right] - \frac{(1 - \eta_k)}{\gamma} \left[ \left( \frac{\theta}{\gamma} - \frac{1}{1 - \beta} \right) \eta_j + (\sigma - 1) s_j \right] \right\} = \\ \text{sign} \left\{ \eta_j \left[ \frac{1}{1 - \beta} + (\sigma - 1) (1 - s_k) \right] - (\sigma - 1) (1 - \eta_k) s_j \right\} = \end{aligned}$$

so it will be negative if

$$\eta_j \left[ \frac{1}{(1-\beta)(\sigma-1)} + (1-s_k) \right] < (1-\eta_k) s_j.$$

That is, the transmission depends negatively on  $\theta$  when the demand shock in  $k$  dominates the supply shock in  $k$  (both induced by market  $j$ ).

Second, we ask whether  $\partial(\Delta p_k/\Delta p_j)/\partial\sigma < 0$ . The sign of the derivative is equal to

$$\begin{aligned} \text{sgn} \left\{ s_j \left[ (1-\eta_k) \frac{\theta}{\gamma} + \frac{\eta_k}{1-\beta} + (\sigma-1)(1-s_k) \right] - (1-s_k) \left[ \left( \frac{\theta}{\gamma} - \frac{1}{1-\beta} \right) \eta_j + (\sigma-1) s_j \right] \right\} = \\ \text{sgn} \left\{ s_j (1-\eta_k) \frac{\theta}{\gamma} + s_j \frac{\eta_k}{1-\beta} - \left( \frac{\theta}{\gamma} - \frac{1}{1-\beta} \right) (1-s_k) \eta_j \right\} \end{aligned}$$

and, conversely, the transmission decreases in  $\sigma$  if supply shocks dominate.

**Special case:**  $K = 2$ . Noting that if  $j$  and  $k$  are the only two crops in production and consumption,  $\eta_k = 1 - \eta_j$  and  $s_k = 1 - s_j$ , and we can further specialize to

$$\frac{\Delta p_k}{\Delta p_j} = \frac{\left[ \left( \frac{\theta}{\gamma} - \frac{1}{1-\beta} \right) \eta_j + (\sigma-1) s_j \right]}{\left[ \frac{1}{1-\beta} + \left( \frac{\theta}{\gamma} - \frac{1}{1-\beta} \right) \eta_j + (\sigma-1) s_j \right]}$$

and

$$\text{sgn} \{ \partial(\Delta p_k/\Delta p_j)/\partial\theta \} = \text{sgn} \left\{ \eta_j \left( \frac{1}{1-\beta} \right) \right\}$$

which is always positive if  $\eta_j \neq 0$ , while

$$\text{sgn} \{ \partial(\Delta p_k/\Delta p_j)/\partial\sigma \} = \text{sgn} \left\{ s_j \left( \frac{1}{1-\beta} \right) \right\}$$

which is always positive if  $s_j \neq 0$ .

## B Building new roads

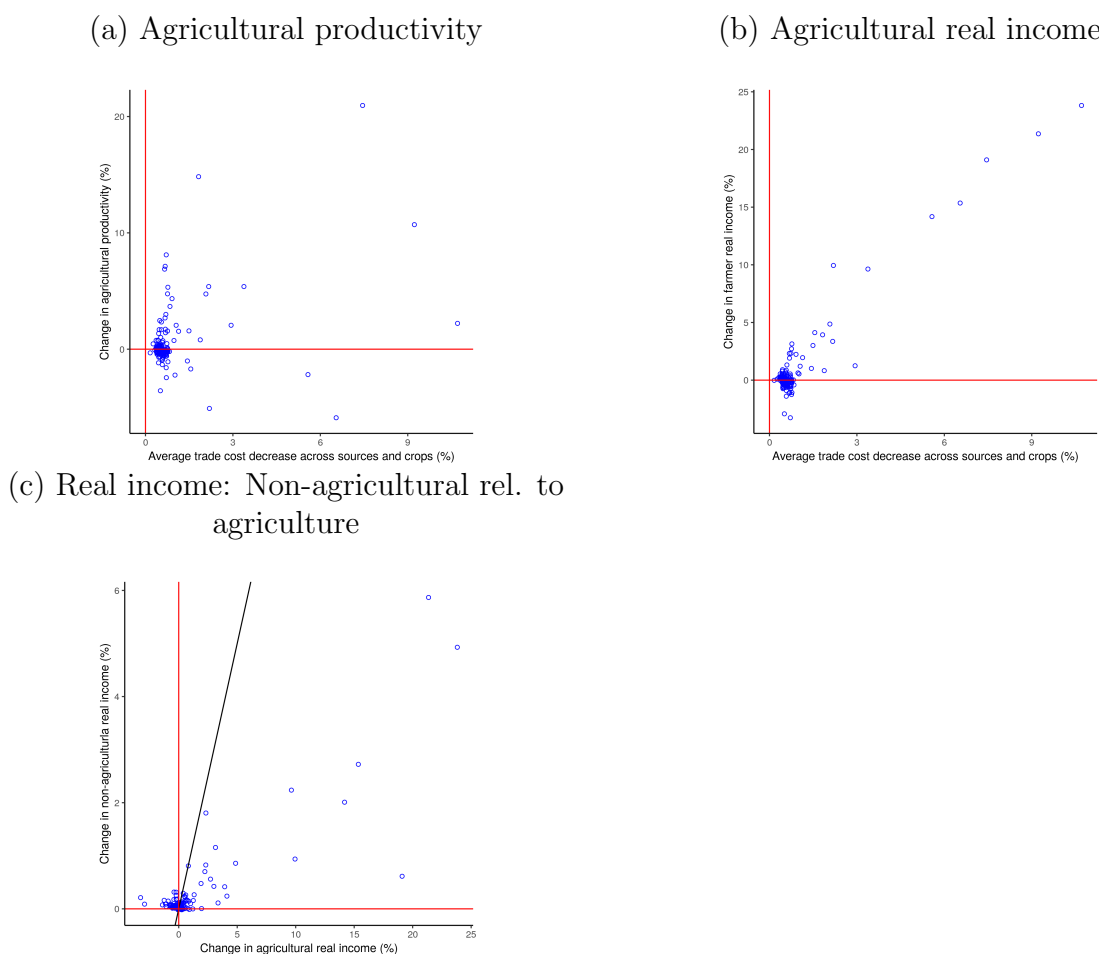
In this section, I provide details on the counterfactual summarized in Section 7.2. As noted in the main text, this is a more targeted policy that nevertheless has substantial equilibrium effects.<sup>57</sup> In aggregate, there is a 0.5 percent increase in productivity, measured in wheat equivalents. The gains, however, are concentrated in the few regions close to the improved

<sup>57</sup>For the country as a whole, the compensating variation associated with this policy is 42.8 thousand USD per kilometer of new paved roads.

roads: Farmers at the 75th percentile have a 0.22 percent productivity increase, while those at the median experience a 0.1 percent decrease (Figure A.1, Panel (a)). Online Appendix Tables H.21 and H.22 contain the breakdown of real income gains across crops for top winners and losers. The tables highlight the role of non-traded crops, such as potato, whose price decreases and hurts regions that specialize in that crop and are not close to new roads.

Finally, note that non-agricultural workers tend to gain everywhere – but gain less than farmers where trade costs drop the most, because non-agricultural producers cannot expand as easily.

Figure A.1: Building New Roads: Counterfactual Change in Productivity and Real Income



## C Robustness

In this section I discuss the robustness of the main results to two alternative model specifications. First, I discuss the quantitative importance of allowing for labor mobility across

sectors. Second, I distinguish landowners from non-landowners and study the distributional effects of shocks within agriculture.

## C.1 Labor mobility with heterogeneous workers

Assume that each worker draws a preference shock for working in the agricultural and the non-agricultural sectors, distributed according to:

$$\begin{aligned}\Pr[Z_A(i) \leq z] &= \exp(-B_A z^{-\omega}) \\ \Pr[Z_M(i) \leq z] &= \exp(-B_M z^{-\omega})\end{aligned}$$

E.g., a worker with taste shock  $Z_A(i)$  obtains utility  $w_{i,A}/P_i Z_A(i)$  from working in the agricultural sector in region  $i$ . Workers obtain the same level of utility from working on any crop within the agricultural sector, so they are assigned to the production of each crop at random. Observing market wages,  $w_{i,M}$  and  $w_{i,A}$ , workers sort optimally across the two sectors.

This setup is akin to that of Lagakos and Waugh (2013), in that unobserved heterogeneity explains equilibrium sorting with differential wages across sectors. It differs in that the heterogeneity is in preferences, not productivity. The reason I depart from their setup is that assuming unobserved preference shocks provides a straightforward match to the data on wages and labor allocations. In addition, the focus here is to understand labor mobility across sectors, not the effect of selection on aggregate productivity.

Given wages  $w_{i,M}$  and  $w_{i,A}$ , the regional labor supplies of workers in each sector for region  $i$  are:

$$L_{i,A} = \frac{B_{i,A} w_{i,A}^\omega}{\Xi_i^\omega} L_i$$

and

$$L_{i,M} = \frac{B_{i,M} w_{i,M}^\omega}{\Xi_i^\omega} L_i$$

where  $\Xi_i \equiv (B_{i,A} w_{i,A}^\omega + B_{i,M} w_{i,M}^\omega)^{1/\omega}$ , and  $L_i$  is the total number of workers in the region. The average utility attained by workers in  $i$  is  $\Xi_i$ , equalized for agriculture and non-agriculture.

With this formulation, much of the structure in the main specification carries over. In particular, the total value added in non-agriculture in equilibrium is

$$V_{i,M} = w_{i,M} l_{i,M},$$

while equilibrium payments to workers in agriculture are still  $l_{i,A} w_{i,A} = \bar{\alpha}_i V_i$ .

To calibrate this version of the model, we need to know  $B_{i,A}$ ,  $B_{i,M}$ , and  $\omega$ . I make the assumption that  $\omega = 4$ , which is in the range of values of the parameters for the distribution of ability in agriculture and non-agriculture from Lagakos and Waugh (2013), ignoring the correlation structure of the shocks. Then, note that  $B_{i,A}$ ,  $B_{i,M}$ , the two new labor allocations,  $L_{i,M}^{het}$ ,  $L_{i,A}^{het}$ , and  $\Xi_i$  satisfy the following equations:

$$\frac{L_{i,M}^{het}}{L_{i,A}^{het}} = \frac{B_{i,M}}{B_{i,A}} \left( \frac{w_{i,M}^{base}}{w_{i,A}^{base}} \right)^\omega \quad (16)$$

$$\Xi_i = \left( B_A (w_{i,A}^{base})^\omega + B_M (w_{i,M}^{base})^\omega \right)^{1/\omega} . \quad (17)$$

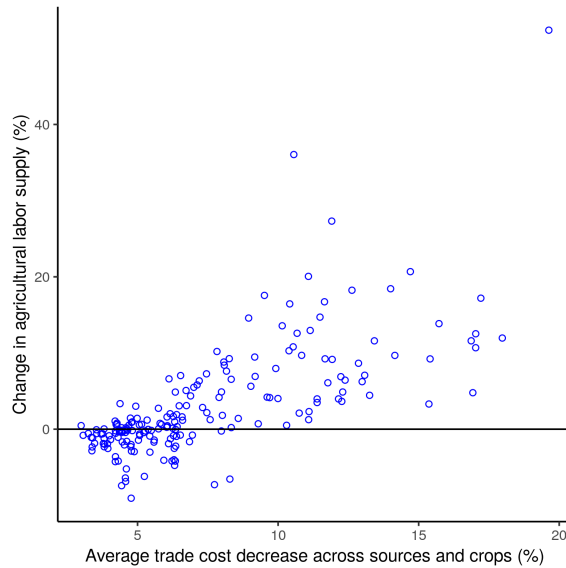
To proceed with the calibration, I normalize  $B_{i,A} = 1$ , which implies choosing units for utility, given the heterogeneity in preferences, and I back out

$$B_{i,M} = \frac{L_{i,M}}{L_{i,A}} \left( \frac{w_{i,A}^{base}}{w_{i,M}^{base}} \right)^\omega ,$$

where  $L_{i,M}$  and  $L_{i,A}$  are the labor supplies in the data, and  $w_{i,A}^{base}$  and  $w_{i,M}^{base}$  are wages in the baseline calibration. The utility  $\Xi_i$  is determined in equilibrium.

Figure A.2 below shows that the share of labor in agriculture grows in regions in which the infrastructure shock from Section 7.1 has the largest effect.

Figure A.2: Paving Roads: Labor Reallocation across Sectors



## C.2 Landownership

In this extension of the model I consider how landownership shapes the distributional effects of policies. I separate households into three types and bring in data on landownership from the Peruvian Agricultural Census of 2012 to inform this specification. I show that because export crops are labor intensive, labor tends to gain more than land. Thus, where agriculture gains more, landowner households gain proportionally less than non-landowner households.

The setup is identical to the main specification in the paper, with the exceptions I list below. First, suppose in each region  $i$  there are: (i) households that own non-agricultural labor, endowed with  $L_{i,M}$ ; (ii) households that own agricultural inputs (land and labor), endowed with  $L_{i,A}^{owner}$  and  $H_i$ ; and (iii) households that only own agricultural labor, endowed with  $L_{i,A}^{non-owner}$ . Furthermore, denote the fraction of agricultural workers who own land as  $\mu_i \equiv L_{i,A}^{owner} / (L_{i,A}^{owner} + L_{i,A}^{non-owner})$ .

Then, in equilibrium, total income in landowner households is

$$\begin{aligned} E_i^{owner} &= \bar{r}_i H_i + w_{i,A} L_{i,A}^{owner}, \\ &= \bar{r}_i H_i + \mu_i w_{i,A} L_{i,A} \\ &= (\bar{\gamma}_i + \mu_i \bar{\alpha}_i) V_i \end{aligned}$$

while total income for non-landowner households is

$$\begin{aligned} E_i^{non-owner} &= w_{i,A} L_{i,A}^{non-owner} \\ &= (1 - \mu_i) \bar{\alpha}_i V_i. \end{aligned}$$

Following my main specification, I consider homothetic preferences for consumption. Because preferences are homothetic, breaking down ownership in this way does not alter the equilibrium, since regional demands depend only on total expenditure, which still adds up to  $E_i = \bar{r}_i H_i + w_{i,A} L_{i,A} + w_{i,M} L_{i,M}$ . Therefore, both baseline and counterfactual prices and quantities remain unchanged relative to my main specification, which facilitates comparison.

Start with a version of the model with homogeneous land intensities across crops. In such a world, both wages and land income are proportional to the total value of production:

$$\begin{aligned} w_{i,A} L_{i,A} &= \alpha V_i \\ \bar{r}_i H_i &= \gamma V_i. \end{aligned}$$

Using the new ownership structure, it follows that:

$$E_i^{owner} = (\gamma + \mu_i \alpha) V_i$$

$$E_i^{non-owner} = (1 - \mu_i) \alpha V_i.$$

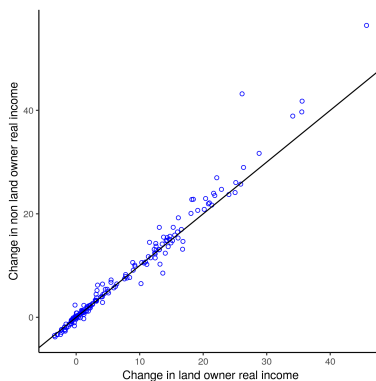
These expressions highlight that in proportional terms, all agents working in farming gain or lose the same after a shock. Therefore, the channel of ownership plays no role in the distributional effect of policies or shocks within a region in a world with homogeneous technologies.

Instead, in a world with different land intensities in production, differences in gains between landowners and non-owners will arise from the correlation between price changes and the intensity of land in production. That is, if the price of a land-intensive crops (i.e., high  $\gamma_k$ ) increases more than that of other crops, landowners will tend to win more than non-owners from that change.

To quantify this channel, I use data from the 2012 Peruvian Agricultural Census to measure the fraction of farmers who operate on their own land,  $\mu$ . Since the model does not offer a theory of farms, mapping to the data is not straightforward. I therefore assume that  $L_{i,A}$  represents the universe of interviewees in region  $i$ , and the fraction that reports working on their own land will give me a quantification of  $\mu_i$ .

Figure A.3(a) compares the change in real income per worker in both types of agricultural households. The figure shows that as real income changes get large, non-landowners gain more in proportional terms. Empirically, this finding is driven by the fact that, where the shock implies a large reduction in trade costs, comparative-advantage crops tend to be labor intensive (see Table H.19 and Figure 2(a)).

Figure A.3: Paving roads: Real Income Changes and Land Ownership





### C.3 The Quantitative Role of Heterogeneous Land Intensity

In this Section, I examine the quantitative implications of allowing for different land cost shares,  $\gamma_k$ , in the production of crops. In particular, I study how counterfactual changes in agricultural income depend on the calibration of these parameters, in the counterfactual in which international prices change. I take this approach for three reasons: (i) it provides a summary measure of how calibrations shape the agricultural sector's outcomes; (ii) as shown in Section 7, the consumption effects of changes in agricultural prices are modest; and (iii) for each region, a shock to international prices entails movements along the PPF, whose effects depend on the crop supply elasticities, which are the objects we want to study.

In what follows, I decompose the total difference in agricultural income gains across calibrations into differences in baseline revenue shares and differences in counterfactual price changes. The starting point of this decomposition is a first-order approximation to changes in regional agricultural income in a given calibration  $j$ :

$$\Delta \text{Income}_i^j = \sum_k \frac{1}{1 - \bar{\beta}_i^j} \pi_{i,k}^j \Delta p_{i,k}^j - \frac{\bar{\beta}_i^j}{1 - \bar{\beta}_i^j} \Delta \rho_i^j,$$

for  $j = B$  (baseline calibration) or  $A$  (for the alternative, homogeneous technologies calibration). Figure A.4 shows that there is a positive correlation between the gains in both counterfactuals. The figure also shows that average income gains are larger in the alternative calibration. This is especially true of losing regions, which lose less in the alternative calibration.

To understand these differences, I study  $\Xi_i \equiv \Delta \text{Income}_i^A - \Delta \text{Income}_i^B$ , which I rewrite as:

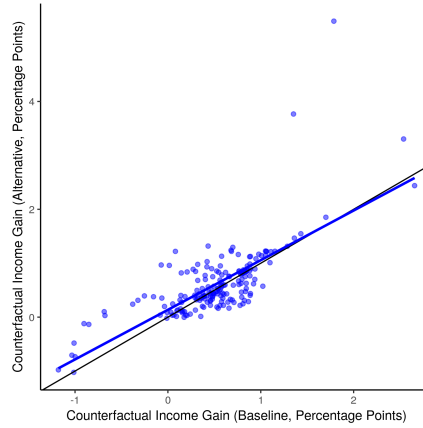
$$\begin{aligned} \Xi_i &= \underbrace{\sum_k \left( \frac{1}{1 - \bar{\beta}_i^A} \pi_{i,k}^A - \frac{1}{1 - \bar{\beta}_i^B} \pi_{i,k}^B \right) \Delta p_{i,k}^A - \left( \frac{\bar{\beta}_i^A}{1 - \bar{\beta}_i^A} - \frac{\bar{\beta}_i^B}{1 - \bar{\beta}_i^B} \right) \Delta \rho_i^A}_{\equiv \text{Contribution of Initial Shares}} \\ &+ \underbrace{\sum_k \frac{1}{1 - \bar{\beta}_i^B} \pi_{i,k}^B (\Delta p_{i,k}^A - \Delta p_{i,k}^B)}_{\equiv \text{Contribution of Price Changes}} \end{aligned}$$

As the equation above shows, the total difference across calibrations,  $\Xi_i$ , stems from two sources.<sup>58</sup> First, since the model does not fit the data perfectly, both calibrations feature different baseline revenue shares,  $\pi_{i,k}$ . Second, in response to the same exogenous shock,

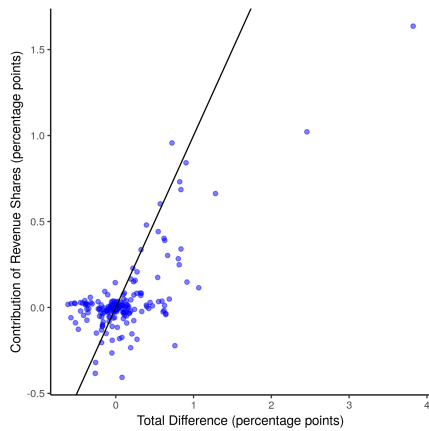
<sup>58</sup>In both calibrations the counterfactual change in the price of intermediate inputs is the same:  $\Delta \rho_i^A = \Delta \rho_i^B$ . Note also that  $\Delta \rho_i = 0$  for the international price shock counterfactual and that in the alternative calibration the share of intermediate inputs,  $\bar{\beta}_i = \beta$ , is the same across regions.

Figure A.4: Decomposition of Counterfactual Gains from International Price Shock Counterfactual (Baseline and Alternative Calibrations)

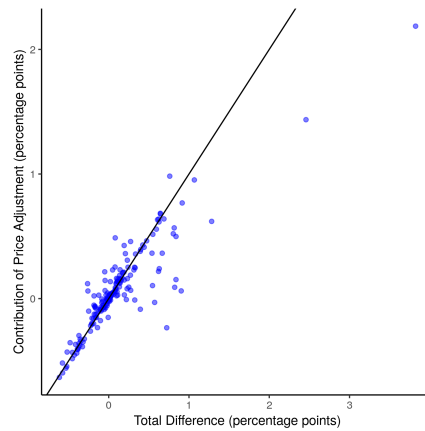
(a) Total Change in Agricultural Income



(b) Contribution of Baseline Shares



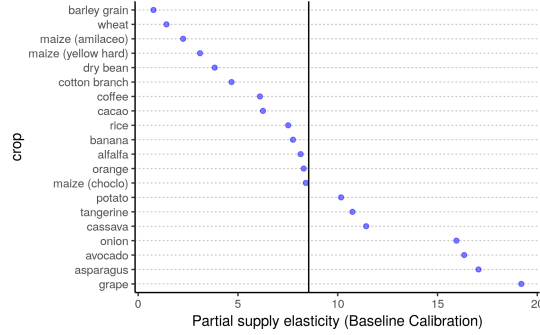
(c) Contribution of Price Changes



Notes: All panels include the 45-degree line

counterfactual price changes vary across calibrations, because the partial elasticity of crop supply is given by  $\theta/\gamma_k - 1$ . Panels (b) and (c) of Figure A.4 plot each of these two components against the total difference across calibrations,  $\Xi_i$ . The panels show that for the large majority of regions, the different results across calibrations,  $\Xi_i$ , come from the equilibrium adjustment of prices. This is intuitive: With a high elasticity of supply, expenditure changes induced by international price shocks translate to smaller price adjustments (or larger movements along the PPF). In going from the baseline to the alternative calibration, the partial elasticities of supply increase for some crops, while they decrease for others (see Figure A.5).

Figure A.5: Crop Supply Elasticities (Baseline and Alternative Calibrations)



Notes: The figure shows, for each crop, the partial elasticity of crop supply  $\theta/\gamma_k - 1$  in the baseline calibration as points. The vertical line displays the homogeneous supply elasticity for all crops in the alternative calibration.

## D Untargeted moments

In this Section, I examine in the data three implications of the theory that I did not target directly in quantifying it: (i) the extensive and intensive margins of domestic trade, (ii) the use of intermediate inputs, and (iii) the relative importance of each port in exporting crops.

### D.1 Domestic trade flows

In this section, I examine the model’s quantitative implications for domestic trade flows. To do so, I collect a new sample of domestic trade flow data and assess the model against it. I start by describing the trade data set and verifying its quality by comparing its properties to those established in the literature. Then I confront the model’s key predictions for the extensive and intensive margin of trade with the data. I conclude that the model’s predictions are largely consistent with the data, although these data were not targeted at all in the calibration.

Trade data are collected by the Ministry of Agriculture under its “System of Prices and Provisions” (SISAP). This measures the volumes and origins of all cargo entering wholesale markets in Lima – the main and most populous city in the country with close to a third of the country’s population. The fact that this is the most complete domestic trade data for Peru (to the best of my knowledge) is in line with one of the motivations of the paper: The scant availability of domestic trade flows in many developing countries.

The data set records the volume (in tons) of a crop  $k$  flowing to Lima from province  $i$  in year  $t$ . Note that province is the same level of aggregation I use in taking the model to the data, which facilitates the comparisons in this section. Furthermore, the data collection manual reveals that data collectors are instructed to specifically ask what is the real origin

of the shipment, not the intermediate stops.<sup>59</sup> Throughout, I focus on average trade flows by crop and origin, during the period 2008-2011, which coincides with the sample I use in all of my calculations.

As a check on data quality, I provide some statistics about this new data set that confirm that it behaves similar to other trade data sets. A key object the trade literature has studied is the elasticity of trade flows with respect to distance. Therefore, I regress log trade flows on log effective distance, and obtain an elasticity of -0.86.<sup>60</sup> Note that this elasticity is close to what other papers report. For example, Disdier and Head (2008) estimate an average of -0.89 across 103 papers, spanning decades, that measure this elasticity using international trade data. In a context more similar to mine, Donaldson (2015) estimates a larger value of -1.6. Moreover, to verify that distance also affects the extensive margin of trade, I estimate a linear probability model for observing positive domestic trade flows. The coefficient on distance is -0.065 and significant at standard levels. The model does not provide analytic guidance as to what this coefficient should be, but strongly suggests that it should be negative (as would other theories that explicitly generate zeros, such as Helpman, Melitz, and Rubinstein, 2008, and Eaton, Kortum, and Sotelo, 2012).

Having checked the quality of the data, I compare it with two key simulated trade patterns. First, I verify that the model captures the frequency of zeros and its relation to geography, as observed in the data. Second, I verify that conditional on being positive, trade flows are positively correlated in the model and the data. These comparisons show that even for a relatively small sample of trade flows and even without targeting them explicitly, the model is able to capture key aspects of the data.

First, recall that one implication of the model is that zero trade flows could arise in equilibrium. I show that in fact, zeros are prevalent in the data, and in a magnitude similar to what the model predicts. Because I include 20 crops and 194 regions in my sample, there are potentially  $3,880 = 194 \times 20$  potential trade links. Of these, the data record only 280 positive trade flows (or 7 percent of all possible links). In my baseline quantification the model predicts 5 percent non-zero trade flows to Lima (188 links).

Second, I show that the model's predictions for this extensive margin line up with the data. As shown in Table A.1, the model correctly predicts 96 percent of the zero trade flow observations. The model does less well when predicting positive trade flows, but still correctly predicts 19 percent of the positive trade flows in the data.

---

<sup>59</sup>When the transporter does not know the origin, it is classified as "unknown." I exclude these observations from the sample.

<sup>60</sup>Since there is a single destination, I control for crop output at the origin instead of the origin fixed effects that are usual in gravity specifications. The results are essentially unaltered without averaging the data, but clustering errors at the origin level.

Table A.1: Zero and Positive Trade Flows to Lima (Data and Model)

Positive trade flow (data)	Positive trade flow (simulation)		
	0 No.	1 No.	Total No.
0	3,457	132	3,589
1	235	56	291
Total	3,692	188	3,880

Notes: The table compares zero and positive trade flows in a sample of trade flows to Lima and in the baseline simulation. Rows indicate counts of positive trade flows in the data and columns indicate counts of positive trade flows in the simulation.

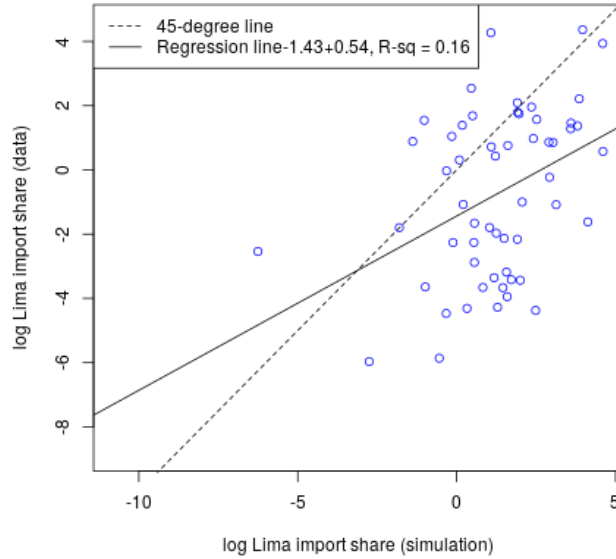
Moreover, a linear probability regression for positive trade flows on effective distance yields a coefficient of -0.030 (0.004); recall that the same coefficient in the data is -0.065. Thus, the likelihood of two regions trading decreases with their distance (and with the same order of magnitude in the data and simulation).

Third, moving on to the intensive margin, the model also does a good job of predicting the trade flows, conditional on their being positive (see Figure A.6). A regression of observed log domestic trade shares on log simulated trade shares yields a coefficient of 0.54 and an  $R^2$  of 0.16.<sup>61</sup>

---

<sup>61</sup>The data do not contain prices for the trade flows at the destination. To avoid units of measurement of different crops from driving this relation, I express shipments as shares of Lima imports (see, e.g., Eaton, Kortum, and Sotelo, 2012). The association is even larger when using trade flows instead (Figure A.11). Table A.3 shows that the prediction is also positively associated with the data if one specifies this regression in levels (OLS or Poisson).

Figure A.6: Positive Trade Flows to Lima (Data and Model)



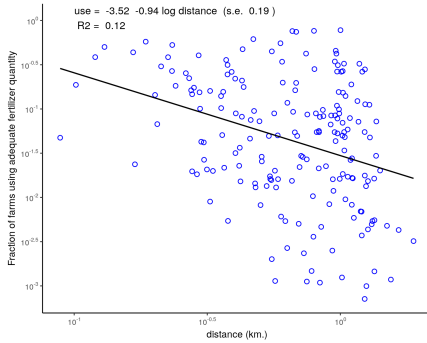
In sum, the model captures well the domestic trade patterns observed in the data, at both the extensive and intensive margins. This is especially important if one considers that the model is quite parsimonious in its specification of demand and trade costs.

## D.2 Intermediate Input Use

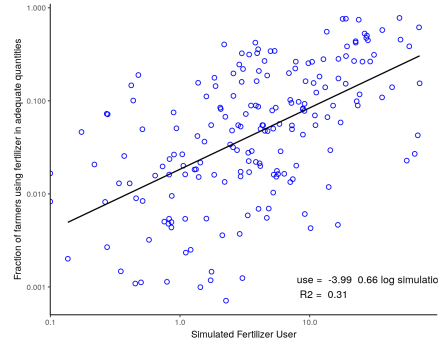
In this section I document, using data from the 2012 Peruvian Agricultural Census, that intermediate input use decreases noticeably with distance from Lima, the main port. In Figure A.7 (a), I relate the fraction of farmers who report “using adequate amounts of fertilizer” to the distance from Lima. The model rationalizes this relation by, first, assuming that most modern inputs are imported into Peru (see footnote 15), and second, inducing their price to increase as they are shipped from the port to the location where they are to be used. Panel (b) shows a clear positive correlation between this fraction of farmers who report adequate use of intermediate inputs in the data and the model-generated input use, showing that the mechanism in the model has the potential to capture actual variation in intermediate input use. In interpreting this result, however, one should exercise caution, since the quantity in the model does not correspond exactly to the measurement in the Census.

Figure A.7: Intermediate Input Use (Data and Model)

(a) Fertilizer use and distance



(b) Intermediate input use: Data and Model

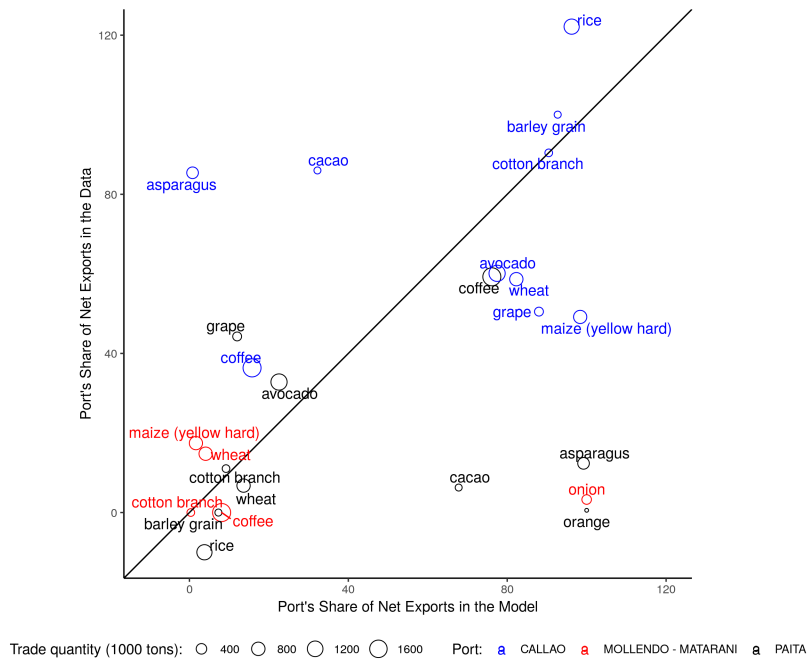


### D.3 Ports' Share of Net Exports

In this section I verify that the model is able to capture the patterns of net exports, by crops and ports, that we observe in the data. To do so, I bring in new customs data on exports by crop for the years 2008 to 2011 and compute, for each crop, the fraction of net exports accounted for by each of the ports in the model (out of the ports in the country). Then I compare these shares to those implied in my baseline simulation. Figure A.8 shows that the model shares are broadly in line with what we observe in the data. A regression of data shares on observed share (using simulated trade flows as weights) yields an  $R^2$  of 0.39 (with a slope of 0.56), meaning that only on the basis of geography, roads, and comparative advantage, the model is able to capture in large part the patterns of export routes.<sup>62</sup>

<sup>62</sup>The unweighted regression yields a slope of .35, with  $R^2 = 0.15$ .

Figure A.8: Ports' Share of Exports by Crop



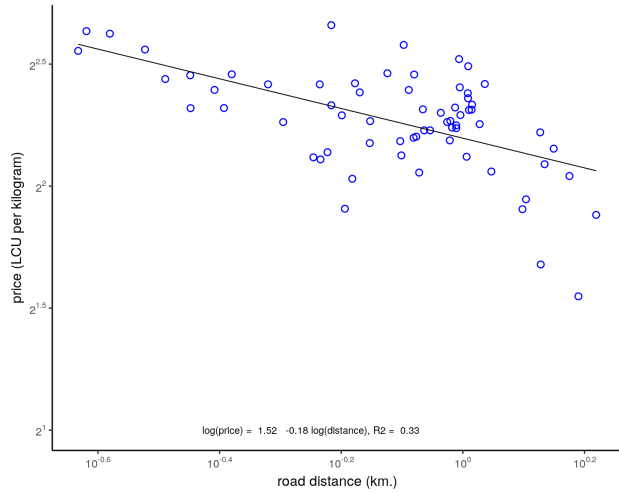
Notes: The figure displays the 45-degree line.



# E Additional Tables and Figures

Figure A.9: Additional Evidence on Fitting Prices at the Baseline

(a) Coffee prices decline with distance to Lima (simulation)



(b) Farm-gate price differences relative to Lima (data and simulation)

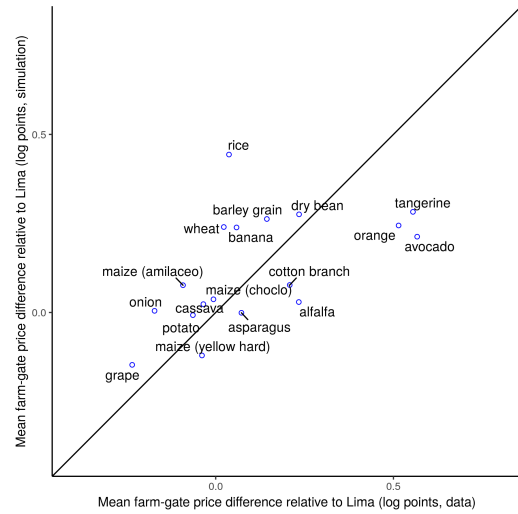
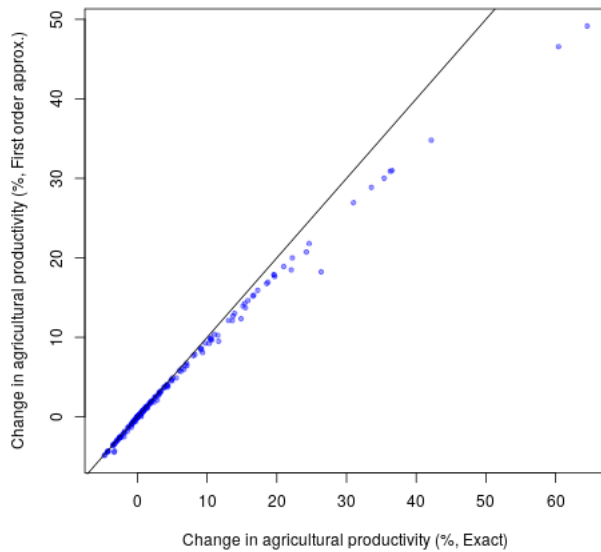


Figure A.10: Comparison of Exact Changes and Predictions Based on First-order Approximations

(a) Productivity (Paving Roads)



(b) Real Income (International Price Shock)

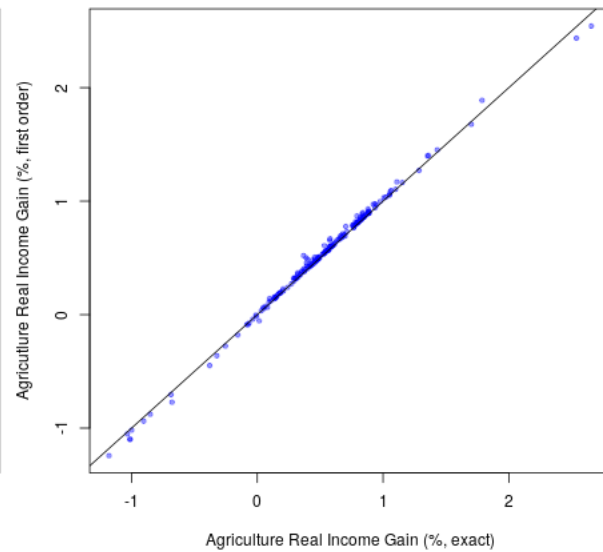


Figure A.11: Positive Trade Flows to Lima (Levels, Data and Model)

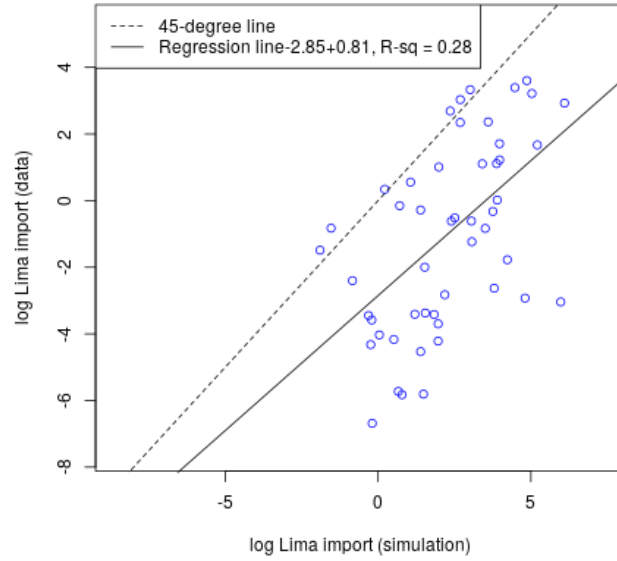


Table A.2: Distribution of Largest, Median, and Smallest Land Shares across Regions (Simulation)

	min	1st quartile	median	3rd quartile	max
largest share	0.204	0.339	0.426	0.586	0.977
median share	0.000	0.016	0.030	0.063	0.288
smallest share	0.000	0.000	0.001	0.003	0.158

Notes: The table shows, across columns, the simulated distribution across regions of the largest land shares, the median land shares, and the smallest land shares.

Table A.3: Intensive Margin: Domestic Trade Model Predictions

	(1)	(2)	(3)
log simulated Lima import shr	0.543** (0.168)		0.679*** (0.0387)
simulated Lima import shr		0.209*** (0.0210)	
R-sq	0.161	0.0409	
Pseudo-R-sq			0.226
Observations	56	2328	142

Standard errors in parentheses

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

Notes: The table shows the results of regressing positive trade flows in the data on their simulated counterparts. Each observation is a crop-origin pair. Columns (1) and (2) show the results of OLS regressions. Column (3) shows the results of a Poisson PML regression.