Online Appendix for "Domestic Trade Frictions and Agriculture"
(Not for publication)

Sebastian Sotelo[1]

A Additional Robustness Checks

In this appendix I study how the main results change in the following extensions of the model: (i) shutting down domestic trade, (ii) allowing for costless domestic trade, (iii) introducing oligopolistic traders, and (iii) measurement errors in $\eta_{i,k}$.

A.1 No Domestic Trade

In this section, I explore the role of domestic trade in generating my results. A concern is that Peruvian agriculture is overwhelmingly oriented towards exports markets. In that case, the additional computational burden stemming from allowing domestic trade might be unjustified.

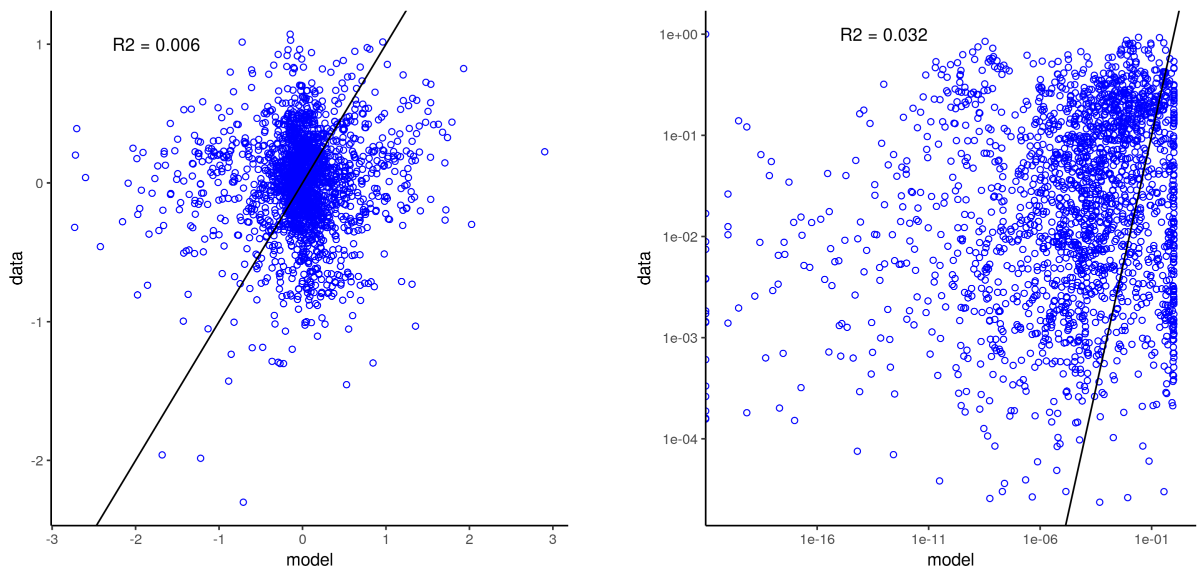
In this section, I show that not allowing for domestic trade substantially degrades the performance of the model. As an example of the consequences of taking this stance, I also show that the welfare impacts of the shock to foreign prices change substantially. Note that, for the preferences I use, the only way to induce zero consumption of a crop is to set its price to infinity. To avoid that problem, whenever region i cannot produce crop k in my baseline calibration (i.e, $A_{i,k} = 0$), I give that region a small endowment of that crop, setting $Q_{i,k} = \bar{Q}_{min}$ exogenously. In my calibrations, I set $\bar{Q}_{min} = 100$ tons, as a small value that allows for utility to be well-defined, but representing that, in fact, the crop is scarce in that region.

First, Figure A.1 displays the fit of prices and land shares in a calibration in which I shut down domestic trade, and allow only for international trade (this is the equivalent of Figure 4 in the main body of the paper.) The figure reveals that the model requires domestic trade to rationalize the data: in panel (a) the fit of within-crop variation in prices drops dramatically, while the model predicts substantially more specialization (both compared to the data and to the baseline calibration) as evidenced by the cluster of land shares simulated to be almost equal to one shown in panel (b).

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

Figure A.1: No Domestic Trade (Baseline Fit)

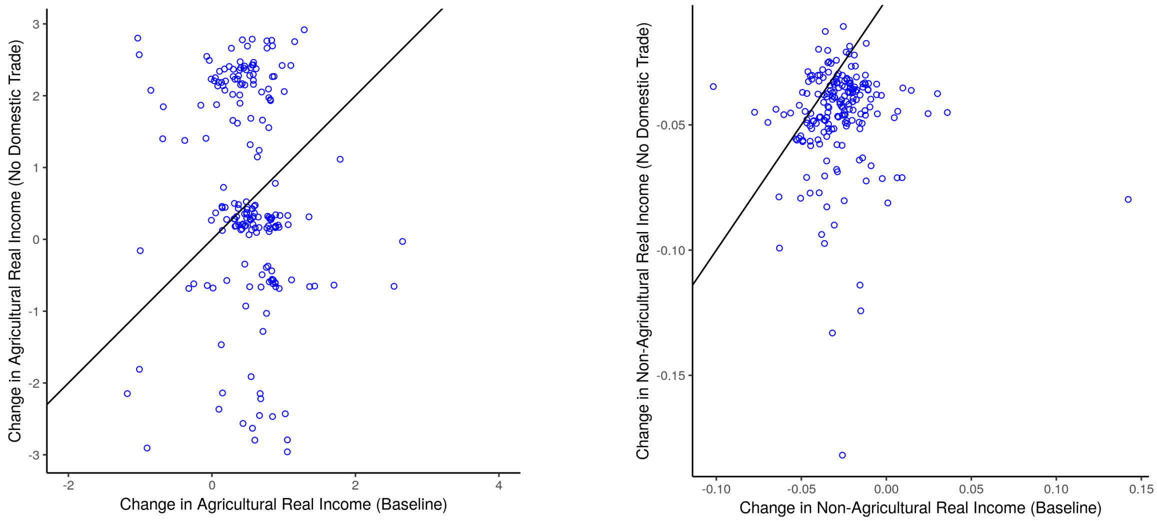
(a) Prices (b) Agricultural welfare



Next, Figure A.2 compares the changes in real farmer income and real income of non-agricultural workers induced by the international price shock in this calibration, relative to the baseline calibration. First, for each variable, there is a low correlation between the changes in the two calibrations. Second, on average, under the no-domestic-trade calibration the range in farmer income changes is wider than in the baseline. Third, non-farmer income in the no-domestic-trade calibration decreases more than under the baseline calibration.The reason is that if regions cannot trade with each other, the only relevant comparative advantage is relative to the rest of the world, and hence regions do trade with ROW. Once this happens, unlike in my baseline, most regions are directly affected by the foreign price changes in prices. Since the shock tends to increase crop prices, farmers benefit more, but the rise in the cost of living hurts non-agricultural workers. Moreover, since some farmer do specialize in this calibration in crops whose international price decreases, we observe more pronounced losses.

Figure A.2: Real Income Changes in the Baseline and No Domestic Trade Calibration

(a) Real Agricultural Income (b) Real Non-Agricultural Income



A.2 Costless domestic trade

This section investigates the role of domestic trade costs by studying a different extreme version of the model. In this calibration, I set domestic trade costs in agriculture to zero, $d_{ni,k} = 1$, $\forall n, i = 1, \ldots, I$, $\forall k = 1, \ldots, K$. By construction, this version of the model is unable to replicate the spatial dispersion of prices we observe in the data, since in the equilibrium arbitrage imposes a unique domestic price for each crop. Furthermore, the fit of land shares also worsens: as Figure A.3 shows, the data do not cluster around the 45-degree line, as they do in the baseline calibration (and a regression of data on simulation yields a slope of 0.21, with $R^2 = 0.25$, compared with a slope of 0.34 with an $R^2 = 0.28$ in the baseline). As I discussed in Section 7 land shares (and revenue shares) determine the first-order impact of any shock, and therefore condition the effects we infer from these shocks.

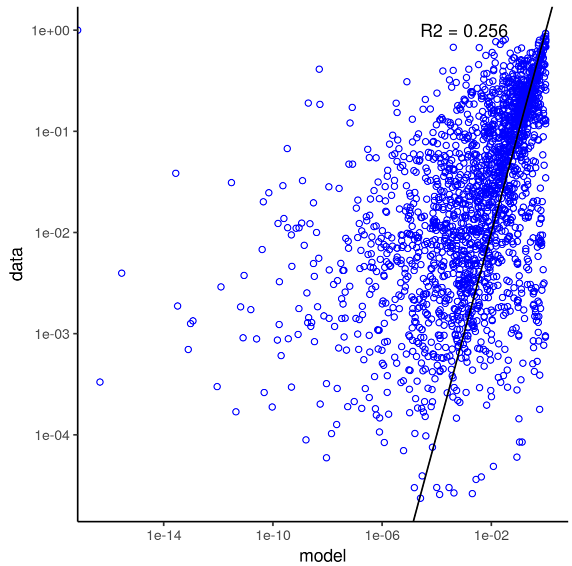
A.3 Oligopolistic traders

In this section, I show how to extend the framework to include oligopolistic traders (similar to Asturias, 2018). The results show how to suitably extend the no-arbitrage conditions that characterize the equilibrium (i.e., the corresponding complementarity slackness constraints) and points out how data on numbers of competitors can inform the model.

Suppose there is a set of potential firms, indexed by j, that can ship good k from i to n. Firms are heterogeneous in their transportation costs, given by $d^{(j)}_{ni,k}$, and compete a la Cournot. Firm j's profit-maximization problem is:

$$\max_{z^{(j)}_{ni}} p_{n,k}\left(C_{n,k}\right) z^{(j)}_{ni,k} - d^{(j)}_{ni,k} z^{(j)}_{ni,k} p_{i,k},$$

Figure A.3: Costless Domestic Trade (Fitting Land Shares)



where the price at the destination depends on total consumption of good k there, and where the demand of good k at n aggregates total shipments from all regions i:

$$C_{n,k} = a_k \left(\frac{p_{n,k}}{P_n}\right)^{1-\sigma} E \tag{1}$$
$$= \sum_{i=1}^{I} \sum_{j} z^{(j)}_{ni,k}.$$

The FOC of the firm is

$$\frac{\partial p_{n,k}}{\partial C_{n,k}} \frac{\partial C_{n,k}}{\partial z^{(j)}_{ni,k}} z^{(j)}_{ni,k} + p_{n,k} = d^{(j)}_{ni,k} p_{i,k}.$$

Using equation (1) and the fact that $C_{n,k} = \sum_{i,j} z_{ni,k}$, the FOC becomes:

$$\frac{p_{n,k} z^{(j)}_{ni,k}}{-\sigma C_{n,k}} + p_{n,k} = d^{(j)}_{ni,k} p_{i,k}.$$

Finally, defining $s^{(j)}_{ni,k} \equiv z^{(j)}_{ni,k}/C_{n,k}$ as the market share of firm j in market n, we obtain

$$p_{n,k}\left[1 - \frac{s^{(j)}_{ni,k}}{\sigma}\right] = d^{(j)}_{ni,k} p_{i,k}$$
$$\Rightarrow p_{n,k} = \frac{\sigma^{(j)}_{ni,k}}{\sigma^{(j)}_{ni,k} - 1} d^{(k)}_{ni,k} p_{i,k},$$

where

$$\sigma^{(j)}_{ni,k} \equiv \frac{\sigma}{s^{(j)}_{ni,k}}.$$

Firm j's elasticity of demand, $\sigma^{(j)}_{ni,k}$ is always larger than one, if $\sigma > 1$. The expression above shows that the wedge in prices between regions i and n will reflect not only the physical cost of shipping goods, but also the markup charged by the firms transporting the goods. The larger firm j's market share in the destination market and the lower σ, the lower firm j's effective elasticity (and the higher the markup). The total number of entrants in each origin-destination pair, $N_{ni,k}$, is determined endogenously by profitability.

To obtain the new no-arbitrage condition, suppose there are $N_{ni,k}$ entrants. Then, adding up the first-order conditions above across firms j, we obtain:

$$\begin{aligned}
\sum_{j=1}^{N_{ni,k}} p_{n,k}\left[\frac{\sigma - s^{(j)}_{ni,k}}{\sigma}\right] &= \sum_{j=1}^{N} d^{(j)}_{ni,k} p_{i,k} \\
p_{n,k}\sum_{j=1}^{N_{ni,k}}\left[1 - \frac{s^{(j)}_{ni,k}}{\sigma}\right] &= p_{i,k}\sum_{j=1}^{N} d^{(j)}_{ni,k} \\
p_{n,k}\left[1 - \frac{S_{ni,k}}{N_{ni,k}\sigma}\right] &= p_{i,k}\bar{d}_{ni,k},
\end{aligned}$$

which implies

$$p_{n,k} = \frac{\sigma N_{ni,k}}{\sigma N_{ni,k} - S_{ni,k}} \bar{d}_{ni,k} p_{i,k},$$

where $S_{ni,k} \equiv \sum_{j=1}^{N_{ni}} s^{(j)}_{ni,k}$ is the aggregate share of region i in n's consumption of k.

Therefore no-arbitrage conditions can be now re-written as:

$$p_{n,k} \leq \frac{\sigma N_{ni,k}}{\sigma N_{ni,k} - S_{ni,k}} \bar{d}_{ni,k} p_{i,k} \wedge z_{ni,k} = 0$$

at $N_{ni,k} = 1$, or

$$p_{n,k} = \frac{\sigma N_{ni,k}}{\sigma N_{ni,k} - S_{ni,k}} \bar{d}_{ni,k} p_{i,k} \wedge z_{ni,k} > 0$$

for an equilibrium number of entrants, $N_{ni,k}$, which make positive profits in equilibrium. If in the equilibrium $N_{ni} \to \infty$, this condition converges to the usual no-arbitrage condition.

A.4 Measurement error in $\eta_{i,k}$

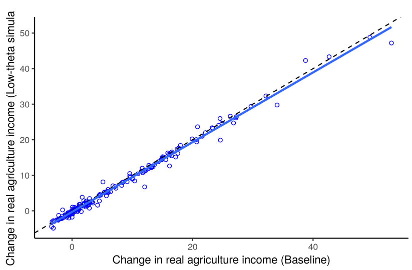
A potential problem with my strategy for estimating θ is that $\eta_{i,k}$, the regressor in equation (10), be measured with error. In the case of classical measurement error, this will bias the coefficient downward, which would bias θ upwards. As discussed in Section 7, θ controls the substitutability across crops in production, and therefore regulates the strength of price responses to shocks. If the estimate of θ is biased upwards, one would expect that regional crop supplies are too elastic, thus generating too small variation in prices (when these are determined by local equilibrium).

To study the robustness of my results to this threat, I provide an alternative estimate of θ and compare the results of using this estimates with those in my benchmark calibration. I start by re-estimating equation (10) with an IV strategy. I bring in an independent measure of land shares, $\eta_{i,k}^{census}$, coming from the Agricultural Census of 2012, and use it as an instrument for $\eta_{i,k}$. This strategy will yield a consistent estimate of θ, as long as the measurement errors are not correlated.[2] This strategy yields a lower estimate of $\theta = 1.22$, consistent with a degree of measurement error.

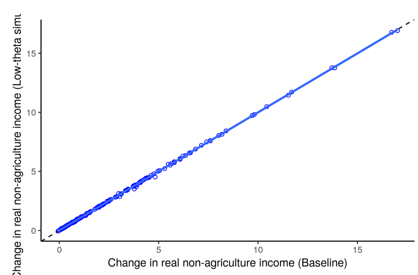
Next, I compare the results of using this value of θ to conduct my counterfactuals. Figure A.4, Panels (a) and (b) compare the baseline changes in agricultural real income to those in these alternative counterfactuals. The Figures show that the results are qualitatively unaffected and are quantitatively very similar to those of the baseline. I conclude that, given the values of other parameters and initial trade patterns, reducing the value of θ in the magnitude I propose has relatively small effects on my results.

Figure A.4: Real Income Changes using $\theta = 1.22$

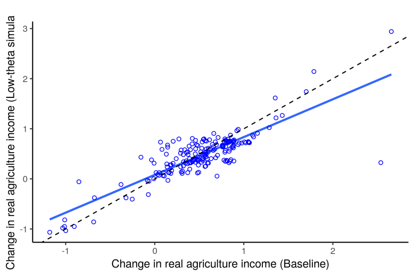
(a) Paving Roads: Changes in Agricultural Real Income



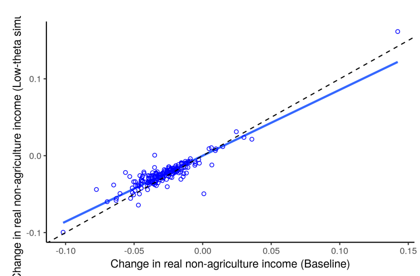
(b) Paving Roads: Changes in Non-agricultural Real Income



(a) Foreign Price Shock: Changes in Agricultural Real Income



(b) Foreign Price Shock: Changes in Agricultural Real Income



Notes: All panels display the 45-degree line (dashed) and a regression line (solid).

[2] See, for example, Cameron and Trivedi (2005) p. 908.

A.5 Farm size and revenue per unit of land

In this section I check whether non-market production, when interacted with distortions in land and labor markets, can explain the differences I observe in revenue per unit of land.

First, a canonical model of household production and consumption in development economics shows that, with complete markets, a household that maximizes its welfare will act as profit-maximizing on the production side. From that point of view, non-market production by itself does not pose a threat to inference, unless accompanied by market imperfections. Along those lines, in Online Appendix B.4 I show that in a version of the model where land owners make cropping decisions my analytic results remain unchanged.

A leading case where this separation breaks is that in which there are no land markets and households cannot supply labor freely outside of the farm (Bardhan and Udry, 1999, Ch 2). In that case, the basic model predicts that smaller farms will have higher labor to land ratios and, therefore, higher revenue per unit of land. As I show next, the evidence suggests that this mechanism cannot explain the variation in revenue per unit of land within regions that I observe in the data.

Using data from the agricultural census of 2012, in which I observe farms, I construct $R_i^f = \sum_k \eta_{i,k}^f \times (\text{rev per unit of land}_k)$, a measure of revenue per unit of land for each farm f in region i. R_i^f averages averages revenue per unit of land across crop k using farm-specific land shares as weights.[3] I then regress this measure, R_i^f, on farm size, h_i^f, controlling for region fixed effects (as in Fact 3). Table A.1 show that regressing R_i^f on farm sizes either does not statistically explain farm-specific revenue per unit of land (column 1, regressing on $\log h_i^f$) or quantitatively has no impact (column 2, regressing on h_i^f).[4]

Table A.1: Weighted Revenue per Unit of Land and Farm Size

	(1)	(2)
log farm size	-0.00124	
	(-0.59)	
farm size		0.0000191*
		(2.40)
N farms	1295026	1295026
R-sq	0.396	0.396

t statistics in parentheses

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

[3]The Census does not collect information on crop prices and therefore it is not possible to measure revenues using that data set. Therefore, for this calculation, I measure revenue per unit of land using the national statistics in agriculture described in Section 2.

[4]Given the coefficient in column 2, a one standard deviation increase in farm size h_i^f leads to a 0.007 standard deviation increase in revenue per unit of land.

B Proofs and Derivations

B.1 Optimal Farmer Behavior

In this section I characterize the optimal choices of farmers. I discuss the optimal choices of farmers for a plot ω, conditional on allocating a positive fraction of plot ω to the production of crop k, $(\phi_{i,k}(\omega) > 0)$, an event I denote $\omega \in \Omega_{i,k}$.

Then I move on to prove the three main Propositions in the paper. Using Assumptions #1 and #2, I first show how to calculate the fraction of land allocated to crop k. In equilibrium, $\omega \in \Omega_{i,k}$ implies $\phi_{i,k}(\omega) = 1$, so each plot ω is fully specialized in one crop k. Second, I derive the distributions of land quality, yields, revenue per unit of land, and rents, conditional on $\omega \in \Omega_{i,k}$. Third, I derive the optimal relation between land shares and revenue shares for a given region.

B.1.1 Basic Properties of Random Land Heterogeneity

The distribution of $\Lambda_{i,k}(\omega)$ is

$$F_{i,k}(\Lambda) = e^{-\tilde{\gamma}^{\theta} A_{i,k}^{\theta} \Lambda^{-\theta}},$$

which means that $\Lambda_{i,k}(\omega)$ is distributed like a Fréchet r.v with parameters $(\tilde{\gamma} A_{i,k}, \theta)$. Using the properties of the Fréchet distribution, the unconditional expectation of $\Lambda_{i,k}(\omega)$ is:

$$\begin{aligned} \mathbb{E}[\Lambda_{i,k}(\omega)] &= \tilde{\gamma} A_{i,k} \Gamma\left(1 - \frac{1}{\theta}\right) \\ &= A_{i,k} \end{aligned}$$

where in the last line we use the normalization $\tilde{\gamma}^{-1} = \Gamma\left(1 - \frac{1}{\theta}\right)$. We will exploit this normalization throughout the paper.

B.1.2 The Farmer's Cost Minimization Problem

The farmer seeks to maximize profits, expression (3) in the main body, by choosing $\{\phi_{i,k}(\omega), l_{i,k}(\omega), x_{i,k}(\omega), \omega \in \Omega_i, \text{ all } k\}$:

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

where total output of crop k is

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

An equivalent problem for the farmer is to maximize the profits obtained in each plot ω, and then add up the profits across those plots. The profits in plot ω are:

$$\sum_{k=1}^{K}\left[p_{i,k}\left(l_{i,k}\left(\omega\right)\right)^{\alpha}\left(x_{i,k}\left(\omega\right)\right)^{\beta}\left(\phi_{i,k}\left(\omega\right)\Lambda_{i,k}\left(\omega\right)\right)^{\gamma}-w_{i,A}l_{i,k}\left(\omega\right)-\rho_{i}x_{i,k}\left(\omega\right)-r_{i}\left(\omega\right)\phi_{i,k}\left(\omega\right)\right].$$

As is standard in trade theory, I will characterize the farmer's problem in terms of the unit cost function of producing each good, for a given plot ω. Suppose that $\omega\in\Omega_{i,k}$, so $\phi_{i,k}\left(\omega\right)>0$. The cost minimization problem is to choose $l_{i,k}\left(\omega\right)$, $x_{i,k}\left(\omega\right)$, $\phi_{i,k}\left(\omega\right)$ to minimize the total cost of production $\bar{q}_k$ units of crop k:

$$\min\left[w_{i,A}l_{i,k}\left(\omega\right)+\rho_{i}x_{i,k}\left(\omega\right)+r_{i}\left(\omega\right)\phi_{i,k}\left(\omega\right)\right]$$

$$s.t.$$

$$\left(l_{i,k}\left(\omega\right)\right)^{\alpha_k}\left(x_{i,k}\left(\omega\right)\right)^{\beta_k}\left(\phi_{i,k}\left(\omega\right)\Lambda_{i,k}\left(\omega\right)\right)^{\gamma_k}\geq\bar{q}_k$$

The solution to this problem is well-known, because the production function is Cobb-Douglas. This solution consists of a cost function and the conditional demands for inputs.[5]

We set up a Lagrangean

$$\mathcal{L}=w_{i,A}l_{i,k}\left(\omega\right)+\rho_{i}x_{i,k}\left(\omega\right)+r_{i}\left(\omega\right)\phi_{i,k}\left(\omega\right)-\mu\left[q_{i,k}\left(\omega\right)-\bar{q}\right]$$

and write down the first order conditions

$$\begin{aligned}
\left[l_{i,k}\left(\omega\right)\right]: \quad w_{i,A} &= \mu\alpha_k\left(l_{i,k}\left(\omega\right)\right)^{\alpha_k-1}\left(x_{i,k}\left(\omega\right)\right)^{\beta_k}\left(\phi_{i,k}\left(\omega\right)\Lambda_{i,k}\left(\omega\right)\right)^{\gamma_k}\\
\left[x_{i,k}\left(\omega\right)\right]: \quad \rho_i &= \mu\beta_k\left(l_{i,k}\left(\omega\right)\right)^{\alpha_k}\left(x_{i,k}\left(\omega\right)\right)^{\beta_k-1}\left(\phi_{i,k}\left(\omega\right)\Lambda_{i,k}\left(\omega\right)\right)^{\gamma_k}\\
\left[\phi_{i,k}\left(\omega\right)\right]: \quad r_i\left(\omega\right) &= \mu\gamma_k\left(l_{i,k}\left(\omega\right)\right)^{\alpha_k}\left(x_{i,k}\left(\omega\right)\right)^{\beta_k}\left(\phi_{i,k}\left(\omega\right)\right)^{\gamma_k-1}\left(\Lambda_{i,k}\left(\omega\right)\right)^{\gamma_k}
\end{aligned}$$

An easy way to solve this problem is to note that μ is the marginal cost and plug the first order conditions in the constraint –assuming it holds with equality–, after using them to solve for input demands.

$$\begin{aligned}
l_{i,k}\left(\omega\right) &= \frac{\mu\alpha_k q_{i,k}\left(\omega\right)}{w_{i,A}}\\
x_{i,k}\left(\omega\right) &= \frac{\mu\beta_k q_{i,k}\left(\omega\right)}{\rho_i}\\
\phi_{i,k}\left(\omega\right) &= \frac{\mu\gamma_k q_{i,k}\left(\omega\right)}{r_i\left(\omega\right)}
\end{aligned}$$

[5]Note that there is a continuum of plots of size 1 and for the constraint in the problem to be met, $\bar{q}_k$ must be small enough so the optimal value of $\phi_{i,k}\left(\omega\right)\leq 1$. Consistent with price taking behavior, the farmer does not take this constraint into account in his cost-minimization problem.

After substituting for input demands in the constraint, we solve for the marginal cost of producing crop k in plot ω, which equals average cost:

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

where $\bar{c}_k = \alpha_k^{-\alpha_k}\beta_k^{-\beta_k}\gamma_k^{-\gamma_k}$.

The demands for inputs per unit of land are independent of the total amount produced, and can be written as

$$\frac{l_{i,k}(\omega)}{\phi_{i,k}(\omega)} = \frac{\alpha_k}{\gamma_k}\frac{r_i(\omega)}{w_{i,A}} \tag{2}$$

for labor and

$$\frac{x_{i,k}(\omega)}{\phi_{i,k}(\omega)} = \frac{\beta_k}{\gamma_k}\frac{r_i(\omega)}{\rho_i} \tag{3}$$

for intermediate inputs. Note that these are also the quantities of labor and intermediates used when the plot is completely specialized, since the size of each plot is 1.

B.1.3 The Implications of Profit Maximization

Profit Maximization Having characterized the cost function, we can return to the profit maximization problem of the farmer. Let $q_{i,k}(\omega)$ be the quantity produced of crop k, in region i, using plot ω. Then the profit maximization problem in the plot is:

$$\max_{q_{i,k}(\omega)} p_{i,k}q_{i,k}(\omega) - c_{i,k}(\omega)\,q_{i,k}(\omega) \tag{4}$$

Since the technology exhibits constant returns to scale, we cannot determine the output by solving this problem. All we can say is that at an optimum, $p_{i,k} \le c_{i,k}(\omega)$, with equality of $q_{i,k}(\omega) > 0$. Furthermore, if $q_{i,k}(\omega) > 0$, we can derive a relationship between the equilibrium rental rate in plot $r_{i,}(\omega)$ and the price of crop k:

$$\begin{aligned} p_{i,k} &= \bar{c}_k\frac{w_{i,A}^{\alpha_k}\rho_i^{\beta_k}(r_i(\omega))^{\gamma_k}}{(\Lambda_{i,k}(\omega))^{\gamma_k}} \\ p_{i,k}(\Lambda_{i,k}(\omega))^{\gamma_k}\bar{c}_k^{-1} &= w_{i,A}^{\alpha_k}\rho_i^{\beta_k}(r_i(\omega))^{\gamma_k} \\ p_{i,k}(\Lambda_{i,k}(\omega))^{\gamma_k}\bar{c}_k^{-1}w_{i,A}^{-\alpha_k}\rho_i^{-\beta_k} &= (r_i(\omega))^{\gamma_k} \\ &\Rightarrow \\ r_{i,k}(\omega) &= \Lambda_{i,k}(\omega)\,p_{i,k}^{\frac{1}{\gamma_k}}\bar{c}_k^{-\frac{1}{\gamma_k}}w_{i,A}^{-\frac{\alpha_k}{\gamma_k}}\rho_i^{-\frac{\beta_k}{\gamma_k}} \\ &= \lambda_{i,k}\Lambda_{i,k}(\omega) \end{aligned}$$

where $\lambda_{i,k} = \bar{c}_k^{-\frac{1}{\gamma_k}} p_{i,k}^{\frac{1}{\gamma_k}} w_{i,A}^{-\frac{\alpha_k}{\gamma_k}} \rho_i^{-\frac{\beta_k}{\gamma_k}}$ is an index of the profitability of production of crop k in i, ignoring land quality.

Land Allocation The farmer behaves competitively, which ensures that all the difference between revenues and the labor and input costs is transferred to the landowners, such that the farmer earns zero profits. This is the same argument as one would make if there were a mass of competitive farmers bidding to rent land from the household. Therefore, a competitive farmer will choose crops such that the rental rate is the maximum that can be attained in that plot,

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

Note that, given our assumptions, this implies that there is complete specialization across all plots (those where specialization is incomplete are a measure zero.)

Let us denote the probability of that event happening by $\eta_{i,k}$,

$$\eta_{i,k} = \Pr\left[k = \arg\max_l \lambda_l \Lambda_{i,l}(\omega)\right].$$

Proposition 1. Profit maximization, together with Assumptions 1 and 2, implies that farmers in region i allocate a fraction $\eta_{i,k}$ of land to crop k.

Proof. We assumed that $\Lambda_{i,k}(\omega)$ is drawn from a Fréchet with parameters $(\tilde{\gamma} A_{i,k}, \theta)$. It follows that $\lambda_k \Lambda_{i,k}(\omega)$ is drawn from a Fréchet with parameters $(\tilde{\gamma} \lambda_k A_{i,k}, \theta)$. Using the properties of the Fréchet distribution,

$$\begin{aligned} \eta_{i,k} &= \frac{(\lambda_k A_{i,k})^\theta}{\sum_l (\lambda_l A_{i,l})^\theta} \\ &= \frac{(\lambda_k A_{i,k})^\theta}{\Phi_i^\theta} \end{aligned}$$

where we define $\Phi_i^\theta = \sum_l \left(\bar{c}_l^{-\frac{1}{\gamma_l}} p_{i,l}^{\frac{1}{\gamma_l}} w_{i,A}^{-\frac{\alpha_l}{\gamma_l}} \rho_i^{-\frac{\beta_l}{\gamma_l}} A_{i,l} \right)^\theta$.

Note that because we assume that there is a continuum of plots in each region i, $\eta_{i,k}$ also represents the aggregate fraction of land that is allocated to crop k, in region i. Note that $\eta_{i,k} > 0$ for all crops with $A_{i,k} > 0$. This outcome is due to the fact that the distribution of each $\Lambda_{i,k}(\omega)$ is unbounded from above, of which the Fréchet distribution is a special case.

Note also that we can henceforth assume that if $\omega \in \Omega_{i,k}$, then all land in ω is allocated to k, $\phi_{i,k}(\omega) = 1$. Assuming a continuous distribution for $\Lambda_{i,k}(\omega)$ guarantees that cases where the farmer is indifferent between two crops have zero probability. The interpretation is that each plot will become perfectly specialized in the production of a single crop k.

Finally, in equilibrium the fact that the farmer chooses k to maximize profits does not contradict the he earns zero profits on each plot in equilibrium. The land rent on that plot, $r_i(\omega)$, adjusts to ensure this is so. Because $r_i(\omega)$ is the maximum of a set of Fréchet r.v., we also obtain the distribution of rental rates. In fact, $r_i(\omega)$ is drawn from a Fréchet distribution with parameters (Φ_i, θ).

Optimal Demand for Labor and Intermediates Once we have the equilibrium value of $r_i(\omega)$ when $\omega \in \Omega_{i,k}$, equal to to $r_{i,k}(\omega)$, we can obtain the optimal labor-land and intermediates-land ratios, as a function of prices and land quality. Knowledge of optimal factor uses will allow us to calculate optimal yields and revenues per unit of land.

Using the first order conditions of the CMP, evaluated at $\phi_{i,k}(\omega) = 1$, to obtain the input demands per plot, if ω is specialized in k, and substituting $r_i(\omega)$ in these expressions:

$$\begin{aligned}
\frac{l_{i,k}(\omega)}{\phi_{i,k}(\omega)} &= \frac{\mu \alpha_k q_{i,k}(\omega)}{w_{i,A}} \frac{r_i(\omega)}{\mu \gamma_k q_{i,k}(\omega)} \\
l_{i,k}(\omega) &= \frac{\alpha_k}{\gamma_k w_{i,A}} r_i(\omega)
\end{aligned}$$

$$l_{i,k}(\omega) = \frac{\alpha_k}{\gamma_k w_{i,A}} \lambda_{i,k} \Lambda_{i,k}(\omega) \tag{5}$$

and, similarly,

$$x_{i,k}(\omega) = \frac{\beta_k}{\gamma_k \rho_i} \lambda_{i,k} \Lambda_{i,k}(\omega). \tag{6}$$

Revenue per Unit of Land Let $\psi_{i,k}(\omega)$ denote the optimal revenue per unit of land, conditional on $\omega \in \Omega_{i,k}$. Then substitute the optimal demands for labor (5) and intermediate inputs (6) in the production function (2), setting $\phi_{i,k}(\omega) = 1$ (which has the same interpretation as output per unit of land):

$$\begin{aligned}
\psi_{i,k}(\omega) &= p_{i,k} q_{i,k}(\omega) \\
&= p_{i,k} \left(\frac{\alpha_k}{\gamma_k w_{i,A}} \lambda_{i,k} \Lambda_{i,k}(\omega) \right)^{\alpha_k} \left(\frac{\beta_k}{\gamma_k \rho_i} \lambda_{i,k} \Lambda_{i,k}(\omega) \right)^{\beta_k} \left(\Lambda_{i,k}(\omega) \right)^{\gamma_k} \\
&= p_{i,k} \alpha_k^{\alpha_k} \beta_k^{\beta_k} \gamma_k^{\gamma_k - 1} w_{i,A}^{-\alpha_k} \rho_i^{-\beta_k} \lambda_{i,k}^{1-\gamma_k} \Lambda_{i,k}(\omega) \\
&= p_{i,k} \alpha_k^{\alpha_k} \beta_k^{\beta_k} \gamma_k^{\gamma_k - 1} w_{i,A}^{-\alpha_k} \rho_i^{-\beta_k} \left(\bar{c}_k^{-\frac{1}{\gamma_k}} p_{i,k}^{\frac{1}{\gamma_k}} w_{i,A}^{-\frac{\alpha_k}{\gamma_k}} \rho_i^{-\frac{\beta_k}{\gamma_k}} \right)^{-\gamma_k} \lambda_{i,k} \Lambda_{i,k}(\omega) \\
&= p_{i,k} \alpha_k^{\alpha_k} \beta_k^{\beta_k} \gamma_k^{\gamma_k - 1} w_{i,A}^{-\alpha_k} \rho_i^{-\beta_k} \left(\bar{c}_k p_{i,k}^{-1} w_{i,A}^{\alpha_k} \rho_i^{\beta_k} \right) \lambda_{i,k} \Lambda_{i,k}(\omega) \\
&= \frac{\lambda_{i,k} \Lambda_{i,k}(\omega)}{\gamma_k}.
\end{aligned}$$

Optimal Land Yield Let $y_{i,k}(\omega)$ denote the optimal land yield, still conditional on $\omega \in \Omega_{i,k}$. We obtain it as

$$\begin{aligned} y_{i,k}(\omega) &= \psi_{i,k}(\omega)/p_{i,k} \\ &= \frac{\lambda_{i,k}\Lambda_{i,k}(\omega)}{p_{i,k}\gamma_k} \end{aligned}$$

B.1.4 Expected yields, revenue and land rents

So far, we have characterized land rent, physical yields and revenues for a given plot, as a function of output and factor prices, and of that plot's random land quality $\Lambda_{i,k}(\omega)$ (under the assumption that $\omega \in \Omega_{i,k}$). To match the model to data, we must characterize the distributions of these variables conditional on $\omega \in \Omega_{i,k}$ actually happening. That is, we study how $r_{i,k}(\omega)$, $y_{i,k}(\omega)$ and $\psi_{i,k}(\omega)$ are distributed over the set of plots where it is optimal to grow crop k.

The Conditional Distribution of Land Quality $\Lambda_{i,k}(\omega)|\omega \in \Omega_{i,k}$ The main building block is the distribution of land quality, conditional on crop k being chosen, which I denote $G_{i,k}$

We need to calculate:

$$G_{i,k}(t) = \mathbb{P}\left[\Lambda_{i,k}(\omega) \le t | \lambda_{i,k}\Lambda_{i,k}(\omega) = \max_j \lambda_{i,j}\Lambda_{i,j}(\omega)\right].$$

After rearranging the conditioning event, we write this as

$$\begin{aligned} G_{i,k}(t) &= \frac{1}{\eta_{i,k}}\mathbb{P}\left[\frac{\lambda_{i,j}}{\lambda_{i,k}}\Lambda_{i,j}(\omega) \le \Lambda_{i,k}(\omega) \le t, \quad \forall j\right] \\ &= \frac{1}{\eta_{i,k}}\int_0^t \prod_{j\neq k}\mathbb{P}\left[\frac{\lambda_{i,j}}{\lambda_{i,k}}\Lambda_{i,j}(\omega) \le v\right] f_{i,k}(v)\,dv. \end{aligned}$$

Using the fact that $\Lambda_{i,j}(\omega)$, $\forall j$, is distributed like a Fréchet rv:

$$\begin{aligned} G_{i,k}(t) &= \frac{1}{\eta_{i,k}}\int_0^t \prod_{j\neq k}\exp\left(-\left(A_{i,j}^{-1}\frac{\lambda_{i,k}}{\lambda_{i,j}}v\right)^{-\theta}\right)\exp\left(-\left(A_{i,k}^{-1}v\right)^{-\theta}\right)\theta A_{i,k}^{\theta}v^{-\theta-1}dv \\ &= \frac{1}{\eta_{i,k}}\int_0^t \exp\left(-\lambda_{i,k}^{-\theta}v^{-\theta}\sum_j\left(A_{i,j}\lambda_{i,j}\right)^{\theta}\right)\theta A_{i,k}^{\theta}v^{-\theta-1}dv \\ &= \int_0^t \exp\left(-\lambda_{i,k}^{-\theta}v^{-\theta}\Phi_i^{\theta}\right)\theta\lambda_{i,k}^{-\theta}\Phi_i^{\theta}v^{-\theta-1}dv \\ &= \exp\left(-\lambda_{i,k}^{-\theta}\Phi_i^{\theta}t^{-\theta}\right) \\ &= \exp\left(-\left(\frac{\Phi_i}{\lambda_{i,k}}t\right)^{-\theta}\right), \end{aligned}$$

which proves that $\Lambda_{i,k}(\omega)|\omega \in \Omega_{i,k}$ is distributed like a Fréchet r.v. with parameters $(\tilde{\gamma}\Phi_i/\lambda_{i,k}, \theta)$.

The interpretation is as follows. If we went across all plots in the region for which $\omega \in \Omega_{i,k}$ and measured the land quality that is attained with an optimal allocation of land, that measure would be distributed according to $G_{i,k}(\omega)$.

Once we know the conditional distribution, it is also straightforward to show that

$$\mathbb{E}\left[\Lambda_{i,k}(\omega)\,|\omega \in \Omega_{i,k}\right] = \frac{\Phi_i}{\lambda_{i,k}}.$$

The Conditional Distribution of Land Rents Note that the land rent in plot ω, when $\omega \in \Omega_{i,k}$, is just $\Lambda_{i,k}(\omega)$ scaled by $\lambda_{i,k}$, as shown above:

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

Exploiting the properties of Fréchet r.v. once again, it is clear that the distribution of $r_i(\omega)\,|\omega \in \Omega_{i,k}$, inherits the properties of $G_{i,k}(t)$: $r_i(\omega)\,|\omega \in \Omega_{i,k}$ follows a Fréchet distribution, with parameters $(\tilde{\gamma}\Phi_i, \theta)$.

This result is important because it says that, no matter what crop k we are talking about, the distributions of $r_i(\omega)\,|\omega \in \Omega_{i,k}$ are identical. In particular, the average land rent in the plots of region i that grow crop k is

$$\begin{aligned}\mathbb{E}\left[r_i(\omega)\,|\omega \in \Omega_{i,k}\right] &= \tilde{\gamma}\lambda_{i,k}\mathbb{E}\left[\Lambda_{i,k}(\omega)\,|\omega \in \Omega_{i,k}\right] \\ &= \tilde{\gamma}\Phi_i. \end{aligned} \tag{7}$$

Note that the average return to land is equalized across crops within a region. This is a stronger result than we would obtain without the Fréchet assumption: optimal land allocation only guarantees that the return to land is equalized at the margin across crops, not on average. Also, note that both $r_i(\omega)$ and $\mathbb{E}\left[r_i(\omega)\,|\omega \in \Omega_{i,k}\right]$ are homogeneous of degree one in prices.

Similarly, the distributions of $y_{i,k}(\omega)\,|\omega \in \Omega_{i,k}$ and $\psi_{i,k}(\omega)\,|\omega \in \Omega_{i,k}$ can be obtained as an implication of the distribution of $\Lambda_{i,k}(\omega)$, too. In particular.

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

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

Proof. This follows from the optimal value of $y_{i,k}(\omega)$ and $\psi_{i,k}(\omega)$ and the previously derived conditional distribution of $\Lambda_{i,k}$.

Likewise,

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

and

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

Note that if $\alpha_k = \beta_k = 0$, then $\gamma_k = 1$, and we go back to the land-only world. Both in their model and in this one, the distribution of revenue per unit of land is identical across crops, within a region.

Case $\theta \to \infty$ In this case, heterogeneity vanishes and the logic of land allocation is the same as in a traditional Ricardian model: land is only allocated to the crops that maximize $\lambda_{i,k}A_{i,k}$, so, for example, if we observe in the data that $\eta_{i,k} > 0$, it means that $\lambda_{i,k}A_{i,k} = \max_l \left\{\lambda_{i,l}A_{i,l}\right\}$. This allows us to write:

$$\begin{aligned}
\lim_{\theta\to\infty}\mathbb{E}\left[y_{i,k}\left(\omega\right)|\omega\in\Omega_{i,k}\right] &= \lim_{\theta\to\infty}\frac{\Phi_i}{\gamma_k p_{i,k}} \\
&= \frac{1}{\gamma_k p_{i,k}}\lim_{\theta\to\infty}\Phi_i \\
&= \frac{\lambda_{i,k}A_{i,k}}{\gamma_k p_{i,k}}
\end{aligned}$$

and $y_{i,k}$ scales up with $A_{i,k}$, given prices.

Relation of Revenue Share and Land Share Let $\pi_{i,k}$ denote the revenue share of crop k in region i , that is

$$\pi_{i,k} = \frac{p_{i,k}q_{i,k}}{\sum_l p_{i,l}q_{i,l}}$$

Proposition 3. The revenue share is proportional to the land share, with the constant of proportionality reflecting the land intensity of the crop relative to the average

$$\pi_{i,k} = \frac{\gamma_k^{-1}\eta_k}{\sum_l \gamma_l^{-1}\eta_l}.$$

Proof

We can calculate total revenue from crop k as

$$\frac{1}{\gamma_k}\mathbb{E}\left[\lambda_k\Lambda_k\left(\omega\right)|\omega\in\Omega_k\right]\eta_k H_i.$$

Now, $\max_k \lambda_k\Lambda_k$ is drawn from a Fréchet with parameters $(\tilde{\gamma}\Phi_i, \theta)$, independent of which crop is the maximizer. Then

$$\begin{aligned}
\pi_{i,k} &= \frac{\gamma_k^{-1}\mathbb{E}\left[\lambda_k\Lambda_k\left(\omega\right)|\omega\in\Omega_{i,k}\right]\eta_k H_i}{\sum_l \gamma_l^{-1}\mathbb{E}\left[\lambda_l\Lambda_l\left(\omega\right)|\omega\in\Omega_{i,l}\right]\eta_l H_i} \\
&= \frac{\gamma_k^{-1}\Phi_i\eta_k}{\sum_l \gamma_l^{-1}\Phi_i\eta_l} \\
&= \frac{\gamma_k^{-1}\eta_k}{\sum_l \gamma_l^{-1}\eta_l}\square
\end{aligned}$$

B.1.5 Deriving cost-weighted cost shares

Proposition 3 implies that a cost-weighted average of land-shares in a given region is equal to the "harmonic" land-weighted average of those land-shares:

$$\begin{aligned}
\pi_{i,k}\gamma_k &= \frac{\eta_{i,k}}{\sum_l \gamma_l^{-1}\eta_{i,l}} \\
\sum_k \pi_{i,k}\gamma_k &= \sum_k \frac{\eta_{i,k}}{\sum_l \gamma_l^{-1}\eta_{i,l}} \\
&= \frac{1}{\sum_l \frac{\eta_{i,l}}{\gamma_l}} \\
&\equiv \bar{\gamma}_i
\end{aligned}$$

B.2 Aggregate Quantities

In this section I show how to aggregate the optimal behavior of the representative farmer across plots $\omega\in\Omega_i$. I derive the regional demand for labor and intermediates, as well as the total value of production of each crop and overall.

B.2.1 Regional Revenue and Output

First we calculate the total output and value coming from the production of crop k. Aggregate revenue is

$$\begin{aligned}
V_{i,k} &= \mathbb{E}\left[\psi_{i,k}\left(\omega\right)|\omega\in\Omega_{i,k}\right]\eta_{i,k}H_i \\
&= \gamma_k^{-1}\Phi_i\eta_{i,k}H_i \\
&= \gamma_k^{-1}\left(\lambda_{i,k}A_{i,k}\right)^{\theta}\Phi_i^{1-\theta}H_i
\end{aligned}$$

which allows us to calculate output in units of good k as

$$\begin{aligned}
q_{i,k} &= V_{i,k}/p_{i,k} \\
&= \gamma_k^{-1}\left(\lambda_{i,k}A_{i,k}\right)^{\theta}p_{i,k}^{-1}\Phi_i^{1-\theta}H_i.
\end{aligned}$$

We can further aggregate across goods to obtain the revenue function

$$\begin{aligned} V_i &= \sum_k V_{i,k} \\ &= \sum_k \gamma_k^{-1} \Phi_i \eta_{i,k} H_i \\ &= \Phi_i H_i \sum_k \gamma_k^{-1} \eta_{i,k} \\ &= \Phi_i H_i \sum_k \gamma_k^{-1} \eta_{i,k} \\ &= \frac{\Phi_i H_i}{\bar{\gamma}_i} \end{aligned}$$

which scales up with Φ_i.

B.2.2 Aggregate Demand for Labor and Intermediates

The plot-level demand for labor is, as shown before,

$$l_{i,k}(\omega) = \frac{\alpha_k}{\gamma_k w_{i,A}} \lambda_{i,k} \Lambda_{i,k}(\omega),$$

so the aggregate demand in region i, coming from crop k is

$$\begin{aligned} l_{i,k} &= \int_{\omega \in \Omega_{i,k}} l_{i,k}(\omega)\, d\omega \\ &= \mathbb{E}\left[l_{i,k}(\omega) | \omega \in \Omega_{i,k}\right] \eta_{i,k} H_i \\ &= \mathbb{E}\left[\frac{\alpha_k}{\gamma_k w_{i,A}} \lambda_{i,k} \Lambda_{i,k}(\omega) | \omega \in \Omega_{i,k}\right] \eta_{i,k} H_i \\ &= \frac{\alpha_k}{\gamma_k w_{i,A}} \Phi_i \eta_{i,k} H_i. \end{aligned}$$

Likewise, plot level demand for intermediates is

$$\begin{aligned} x_{i,k} &= \int_{\omega \in \Omega_{i,k}} x_{i,k}(\omega)\, d\omega \\ &= \mathbb{E}\left[\frac{\beta_k}{\gamma_k \rho_i} \lambda_{i,k} \Lambda_{i,k}(\omega) | \omega \in \Omega_{i,k}\right] \eta_{i,k} H_i \\ &= \frac{\beta_k}{\gamma_k \rho_i} \Phi_i \eta_{i,k} H_i. \end{aligned}$$

Summing across crops, within region i, we get

$$\begin{aligned} l_{i,A} &= \sum_k \frac{\alpha_k}{\gamma_k w_{i,A}} \Phi_i \eta_{i,k} H_i \\ &= \frac{\Phi_i H_i}{w_{i,A}} \sum_k \frac{\alpha_k}{\gamma_k} \eta_{i,k} \end{aligned}$$

and

$$\begin{aligned} x_i &= \sum_k \frac{\beta_k}{\gamma_k \rho_i} \Phi_i \eta_{i,k} H_i \\ &= \frac{\Phi_i H_i}{\rho_i} \sum_k \frac{\beta_k}{\gamma_k} \eta_{i,k} \end{aligned}$$

Something to note here is that these expressions, for example, the aggregate demand for labor depends on the wage directly through the $w_{i,A}$ term in the denominator, but also through the wage's indirect effect in the allocation of land across crops, which differ in the intensity with which they use labor. Thus, we cannot derive a closed form expression for the wage $w_{i,A}$ as a function of the labor demanded $l_{i,A}$. Defining $\bar{\alpha}_i$ as

$$\sum_k \alpha_k \pi_{i,k} = \bar{\alpha}_i,$$

we obtain

$$\begin{aligned} l_{i,A} &= \frac{\Phi_i H_i}{w_{i,A}} \sum_k \frac{\alpha_k}{\gamma_k} \eta_{i,k} \\ &= \frac{\Phi_i}{w_{i,A}} H_i \frac{\bar{\alpha}_i}{\bar{\gamma}_i} \\ &= \frac{\bar{\alpha}_i}{w_{i,A}} V_i \end{aligned}$$

since

$$\begin{aligned} \sum_k \frac{\eta_{i,k}}{\gamma_{i,k}} \alpha_k &= \sum_k \pi_k \left(\sum_l \eta_l / \gamma_l \right) \alpha_k \\ &= \frac{1}{\bar{\gamma}_i} \sum_k \pi_k \alpha_k. \end{aligned}$$

This expression for labor demand reflects the endogenous average labor and land shares (which depend, among other things, on input and output prices). For example, if the wage goes up, the amount of land allocated to relatively labor intensive crops will go down, and so will the demand for labor. That is on top of the decrease induced by an input mix that is less labor intensive, for all crops. Note that the last equation just delivers the usual Cobb-Douglas result for the share of labor in production. The reasoning to obtain

$$x_i = \frac{\bar{\beta}_i V_i}{\rho_i}$$

is analogous.

B.3 Competitive Equilibrium

Regions in Home take the prices in Foreign as given, and these prices remain unchanged regardless of how much is imported or exported. I provide next the definition of a competitive equilibrium which incorporates the implications of Assumptions 1 and 2, thus focusing on regional aggregates.

Definition 1. A competitive equilibrium consists of, for each region $i = 1, \ldots, I$:

(a) prices $p_{i,k}$ for all crops k ;

(b) wage rates $w_{i,M}$, $w_{i,A}$, and input prices ρ_i;

(c) final goods expenditure $E_{ii',TR}$, $i' \in \mathcal{W}$, 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 $k = 1, \ldots, K$, and the non-agricultural sector, $l_{i,NT}$, $l_{i,TR}$;

(e) trade flows: (e1) domestic $z_{ni,k}$, for all regions $n = 1, \ldots, I$ and crops $k = 1, \ldots, K$, (e.2) international $z_{Fi,k}$ and $z_{iF,k}$ for all crops $k = 1, \ldots, K$ (e.3) international $z_{iF,x}$ of the intermediate input X, such that,

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

$$\begin{aligned} C_{i,k} &= \left(\frac{p_{i,k}}{P_i}\right)^{-\sigma} \frac{b_i E_i}{P_i} \\ E_{ij,TR} &= \left(\frac{d_{ij,M} w_{j,M} / T_{j,TR}}{P_{i,TR}}\right)^{1-\varepsilon} b_{i,TR} E_i \\ E_{i,NT} &= b_{i,NT} E_i, \end{aligned}$$

where $P_{i,TR} = \left(\sum_{j \in \mathcal{W}} \left(d_{ij,TR} w_{ij,TR} / T_{j,TR}\right)^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}$;

(2) the inputs and outputs in (d) solve the agricultural producer'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 $n, i \in \mathcal{W}$, for 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 in (d) labor satisfies:

$$w_{i,M} \left(l_{i,NT} + l_{i,TR}\right) = \sum_{n=1}^{I} \left(\frac{d_{ni,TR} p_{ii,TR}}{P_{i,TR}}\right)^{1-\varepsilon} E_{n,TR} + E_{i,NT} + D_{ROW,i},$$

where $D_{ROW,i} = \left(w_{i,M} d_{Fi} / T_{i,TR}\right)^{1-\varepsilon} X_{ROW}$ and

$$w_{i,M} l_{i,NT} = E_{i,NT};$$

(5) in each region, local markets clear for agricultural labor, land, and crops:

$$
\begin{aligned}
L_{i,A} &= l_{i,A} \\
C_{i,k} &= q_{i,k} - \sum_{n\in\mathcal{W}} d_{ni,k} z_{ni,k} + \sum_{i'\in\mathcal{W}} z_{ii',k}, \quad \text{all } k = 1,\dots,K \\
z_{iF,x} &= x_{i,A} \\
L_{i,M} &= l_{i,NT} + l_{i,TR};
\end{aligned}
$$

(6) with the definition of E_i, this implies trade with Foreign is balanced: the value of exports is equal to the value of imports

$$\sum_{k=1}^{K} p_{F,k} \sum_{i=1}^{I} \frac{z_{Fi,k}}{d_{Fi,k}} + \sum_{i=1}^{I} D_{ROW,i} = \sum_{k=1}^{K} p_{F,k} \sum_{n=1}^{I} d_{nF,k} z_{nF,k} + \rho_F \sum_{n=1}^{I} d_{nF,x} x_n + \sum_{n=1}^{I} p_{nROW,M} C_{nROW,M}$$

To complete the description of the equilibrium, I drop the trade balance condition and choose a normalization, given by the price of foreign goods in the data.

B.4 Alternative Representations of the Farmer's Problem

The Farmer Seeks to Maximize Total Rents For simplicity consider only one input other than land. Also, let us drop the region index to simplify notation. We consider the case where: (i) the size of the plot is h, (ii) the farmer seeks to maximize the total revenue of the household, (iii) land quality is a Hicks-neutral productivity shock.

Suppose then that the technology for producing crop k in plot ω is

$$q_k(\omega) = \Lambda_k(\omega) L_{i,k}^{\alpha_k} (\phi_k(\omega))^{\gamma_k}$$

where $\alpha_k + \gamma_k = 1$. In equilibrium it will be that $\sum_{k=1}^{K} \phi_k(\omega) = h$. The solution to the CMP is the cost function

$$c_k(\omega) = \frac{\bar{c}_k w^{\alpha_k} r_k(\omega)^{\gamma_k}}{\Lambda_k(\omega)}$$

and the solution to the PMP is then

$$p_k \le c_k(\omega),$$

with equality if $\omega \in \Omega_k$. This yields a land rent for the plot

$$r_k(\omega) = \bar{c}_k^{-1/\gamma_k} w^{-\alpha_k/\gamma_k} p_k^{1/\gamma_k} \Lambda_k(\omega)^{1/\gamma_k}.$$

Now the farmer chooses land allocations and input use to maximize the land owner's total rent (which is what determines the land owner's income):

$$R = \sum_k r_k(\omega) \phi_k(\omega) h.$$

So the farmer chooses $\{\phi_k\}$ to maximize R. Note that, to be able to use the familiar results from the Fréchet, we need to assume that $\Lambda_k(\omega)$ is drawn from a Fréchet with parameters $(A_k, \theta/\gamma_k)$, such that $\Lambda_k(\omega)^{1/\gamma_k}$ is drawn from a Fréchet distribution with parameters $\left(A_k^{1/\gamma_k}, \theta\right)$. With those assumptions, $r_k(\omega)$, the rental rate of land if ω is used to grow k, is also drawn from a Fréchet distribution.

Now h is a constant, and so $r_k(\omega)h$ is also a Fréchet r.v. With probability 1 there is only one maximizer among the collection of random variables:

$$\{r_k(\omega)h\}_k$$

so, with probability one, the equilibrium is a corner solution with

$$\phi_k(\omega) = 1 \text{ iff } k = \arg\max_l \{r_l(\omega)h\}.$$

That being the case, the probability of the farmer choosing k is equal to η_k in the main text of the paper.

The Land Owner Chooses Inputs and Land What if the farmers act directly to maximize the total return to land in each plot? Then they want to maximize

$$\begin{aligned} R(\omega) &= \sum_k p_k q_k(\omega) - w_k l_k(\omega) \\ &= \sum_k \left(p_k \Lambda_k(\omega) l_k(\omega)^{\alpha_k} \phi_k(\omega)^{\gamma_k} - w l_k(\omega)\right) \end{aligned}$$

So choose $l_k(\omega)$ optimally conditional on $\phi_k > 0$

$$\begin{aligned} \alpha_k p_k \Lambda_k(\omega) l_k(\omega)^{\alpha_k - 1} \phi_k(\omega)^{\gamma_k} &= w \\ &\Rightarrow \\ l_k(\omega) &= \left(\frac{\alpha_k p_k \phi_k(\omega)^{\gamma_k} \Lambda_k(\omega)}{w}\right)^{\frac{1}{1-\alpha_k}} \end{aligned}$$

and plug it back in

$$R = \sum_k \left[p_k \Lambda_k(\omega) \left(\frac{\alpha_k p_k \phi_k(\omega)^{\gamma_k} \Lambda_k(\omega)}{w}\right)^{\frac{\alpha_k}{1-\alpha_k}} \phi_k(\omega)^{\gamma_k} - w l_k(\omega)\right]$$

Now, since $\alpha_k = 1 - \gamma_k$

$$R = \sum_k \left[\gamma_k \alpha^{\frac{\alpha_k}{\gamma_k}} p_k^{\frac{1}{\gamma_k}} w^{-\frac{\alpha_k}{\gamma_k}} \Lambda_k(\omega)^{1/\gamma_k} \phi_k(\omega)\right].$$

We can finally analyze the choice of crops. Note that, conditional on the realization of $\Lambda_k(\omega)^{1/\gamma_k}$, R is a sum of terms that are linear in ϕ_k, the land allocations to different uses. Therefore, the solution is found in a corner, setting $\phi_k = 1$ for the crop l such that

$$l = \arg\max_k \left(\gamma_k \alpha^{\frac{\alpha_k}{\gamma_k}} p_k^{\frac{1}{\gamma_k}} w^{-\frac{\alpha_k}{\gamma_k}} \Lambda_k^{\frac{1}{\gamma_k}} \right)$$

Since

$$\lambda_k = \alpha_k^{\frac{\alpha_k}{\gamma_k}} \gamma_k p_k^{\frac{1}{\gamma_k}} w^{-\frac{\alpha_k}{\gamma_k}}$$

This problem will yield the same solution in the main text.

C Homogeneous technology

The results drastically simplify when we assume $\gamma_k = \gamma$ for all k. In this case, our Propositions boil down to

1. Land Allocation:

$$\eta_{i,k} = \frac{\left(p_{i,k}^{1/\gamma} A_{i,k}\right)^{\theta}}{\tilde{\Phi}_i^{\theta}}$$

   with

$$\tilde{\Phi}_i = \left(\sum_k \left(p_{i,k}^{1/\gamma} A_{i,k}\right)^{\theta} \right)^{1/\theta}.$$

2. Revenue and yield per unit of land

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

$$\mathbb{E}\left[\psi_{i,k}(\omega) | \omega \in \Omega_{i,k}\right] = \kappa_y w_{i,A}^{-\alpha/\gamma} \rho_i^{-\beta/\gamma} \tilde{\Phi}_i$$

3. Land and revenue shares are equal

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

   for all i and k

Furthermore, we can show the revenue function for crop k simplifies to

$$V_{i,k} = \gamma^{-1} \bar{c}^{-\frac{1}{\gamma}} w_{i,A}^{-\frac{\alpha}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} \frac{p_{i,k}^{\frac{\theta}{\gamma}} A_{i,k}^{\theta}}{\tilde{\Phi}_i^{\theta}} \tilde{\Phi}_i H_i.$$

Total revenue is then

$$V_i = \gamma^{-1} \bar{c}^{-\frac{1}{\gamma}} w_{i,A}^{-\frac{\alpha}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} \tilde{\Phi}_i H_i,$$

and labor demand is

$$
\begin{aligned}
l_{i,A} &= \frac{\Phi_i H_i}{w_{i,A}} \sum_k \frac{\alpha}{\gamma} \eta_{i,k} \\
&= \frac{\Phi_i H_i}{w_{i,A}} \frac{\alpha}{\gamma}
\end{aligned}
$$

while total demand for intermediate inputs is

$$
\begin{aligned}
x_i &= \frac{\Phi_i H_i}{\rho_i} \sum_k \frac{\beta_k}{\gamma_k} \eta_{i,k} \\
&= \frac{\beta}{\gamma} \frac{\Phi_i H_i}{\rho_i}.
\end{aligned}
$$

This case allows for further simplification under the assumption that labor is immobile in agriculture. We can solve for agricultural wages from the labor demand, using market clearing

$$
\begin{aligned}
L_{i,A} &= \frac{\alpha}{\gamma} \bar{c}^{-\frac{1}{\gamma}} w_{i,A}^{-\frac{1-\beta}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} \tilde{\Phi}_i H_i \\
w_{i,A}^{\frac{1-\beta}{\gamma}} &= \frac{\alpha}{\gamma} \bar{c}^{-\frac{1}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} \tilde{\Phi}_i \frac{H_i}{L_{i,A}} \\
&= \bar{c}^{-\frac{1}{1-\beta}} \left(\frac{\alpha}{\gamma} \right)^{\frac{\gamma}{1-\beta}} \rho_i^{-\frac{\beta}{1-\beta}} \left(\tilde{\Phi}_i \frac{H_i}{L_{i,A}} \right)^{\frac{\gamma}{1-\beta}}.
\end{aligned}
$$

Substituting this expression for wages in the expression for V_i we get

$$
\begin{aligned}
V_i &= \gamma^{-1} \bar{c}^{-\frac{1}{\gamma}} \left(\bar{c}^{-\frac{1}{1-\beta}} \left(\frac{\alpha}{\gamma} \right)^{\frac{\gamma}{1-\beta}} \rho_i^{-\frac{\beta}{1-\beta}} \left(\tilde{\Phi}_i \frac{H_i}{L_{i,A}} \right)^{\frac{\gamma}{1-\beta}} \right)^{-\frac{\alpha}{\gamma}} \rho_i^{-\frac{\beta}{\gamma}} \tilde{\Phi}_i H_i \\
&= \gamma^{-1} \bar{c}^{-\frac{1}{\gamma} + \frac{\alpha}{\gamma(1-\beta)}} \left(\frac{\alpha}{\gamma} \right)^{-\frac{\alpha}{1-\beta}} \rho_i^{-\frac{\beta}{\gamma} + \frac{\alpha\beta}{\gamma(1-\beta)}} \left(\tilde{\Phi}_i H_i \right)^{1-\frac{\alpha}{1-\beta}} \\
&= \gamma^{-1} \left(\frac{\alpha}{\gamma} \right)^{-\frac{\alpha}{1-\beta}} \bar{c}^{-\frac{1}{1-\beta}} \rho_i^{-\frac{\beta}{1-\beta}} \left(\tilde{\Phi}_i H_i \right)^{\frac{\gamma}{1-\beta}} L_{i,A}^{\frac{\alpha}{1-\beta}} \\
&= \kappa_V \rho_i^{-\frac{\beta}{1-\beta}} \tilde{\Phi}_i^{\frac{\gamma}{1-\beta}} H_i^{\frac{\gamma}{1-\beta}} L_{i,A}^{\frac{\alpha}{1-\beta}}
\end{aligned}
$$

D Market Access and Productivity

To gain a clear understanding of how trade costs affect allocations and productivity, I discuss a stripped down version of the model. Suppose that land shares are the same, so $\gamma_k = \gamma$, for all crops. Then Propositions 1 trough 3 simplify quite a bit. The key distinction is that, since all crops have the same input shares, changes in the factor rewards do not affect the allocation of land across crops. Therefore, the land allocation is independent of factor rewards, and depends only on relative average land qualities and relative output prices.[6]

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

$$V_i \propto \rho_i^{-\frac{\beta}{1-\beta}} \tilde{\Phi}_i^{\frac{\gamma}{1-\beta}} H_i^{\frac{\gamma}{1-\beta}} L_{i,A}^{\frac{\alpha}{1-\beta}}. \tag{8}$$

where $\tilde{\Phi}_i^{\theta} = \sum_k \left(p_{i,k}^{1/\gamma} A_{i,k}\right)^{\theta}$. Equation (8) is the familiar revenue function. It relates the total revenue generated by region i to prices that are exogenous to the farmer and to the total stock of factors of production.

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

Equation (8) shows the sense in which the coefficient $\rho_i^{-\frac{\beta}{1-\beta}} \tilde{\Phi}_i^{\frac{\gamma}{1-\beta}}$ is a measure of productivity, or TFP. Keeping the coefficient constant, the total revenue of agricultural production has constant returns to scale in land and labor. Equation (8) also shows that in location i, agricultural productivity is higher because $\tilde{\Phi}_i$ is higher (capturing, in part, better land allocations) or because the relative price of intermediates, ρ_i, is lower.[7]

In the model, variation in ρ_i is entirely driven by transportation costs: imported intermediates will be more expensive in remote places. This is the first channel through which transport costs lower productivity. The elasticity of TFP with respect to the price of the intermediate input, keeping all other prices constant, is $-\beta/(1-\beta)$, which is higher the larger the cost share of intermediates. As shown before, however, input use depends on the price of the intermediate relative to the price of output. In the exchange between the farmer and the rest of the world, trade costs increase this relative price twice: once when the farmer ships his output to the closest port and once when he brings the intermediate input back to the farm.[8]

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

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

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

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

transport costs will induce a negative correlation between $p_{i,k}$ and $A_{i,k}$ across k, thus lowering $\tilde{\Phi}_i$.[9]

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

E First-Order Approximations to counterfactuals

E.1 Change in Value Added

We start by finding a first-order approximation to the change in value added (that is, the payments to labor and land) in response to a change in crop prices and intermediate price changes.

The problem of the farmers is equivalent to that of a planner that maximizes total revenue minus payments to intermediate inputs, or value added, subject to the total endowment of labor and land. That is, dropping region indexes for simplicity:

$$VA = \max_{\phi_k(\omega), l_k(\omega), x_k(\omega)} \sum_k \int_\Omega p_k \left[l_k(\omega)^{\alpha_k} x_k(\omega)^{\beta_k} \left(\phi_k(\omega) \Lambda_k(\omega)\right)^{\gamma_k} \right] - \rho x_k(\omega)$$

subject to

$$\begin{aligned} \sum_k \int_\Omega l_k(\omega) \, d\omega &= L_A \\ \sum_k \phi_k(\omega) &= 1, \quad \forall k. \end{aligned}$$

Using the envelope theorem, which allows us to ignore plots that switch uses, we calculate the total differential of value added when prices p_k and ρ change

$$dVA = \sum_k q_k dp_k - x_k d\rho$$

where $q_k = l_k(\omega)^{\alpha_k} x_k(\omega)^{\beta_k} \left(\phi_k(\omega) \Lambda_k(\omega)\right)^{\gamma_k}$ in equilibrium, and x_k and l_k is the equilibrium demand of intermediate inputs and labor, given in (5) and (6). Using the fact that payments to

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

factors obey the Cobb-Douglas form, as shown before

$$dVA = \sum_k \left(q_k dp_k - \beta_k p_k q_k \frac{d\rho}{\rho} \right).$$

And using $VA = \left(1 - \bar{\beta}\right) V$, where V is the total value of production, we write the change in value added in proportional terms:

$$\begin{aligned}
\frac{dVA}{VA} &= \sum_k \left(\frac{p_k q_k}{VA} \frac{dp_k}{p_k} - \beta_k \frac{p_k q_k}{VA} \frac{d\rho}{\rho} \right) \\
&= \sum_k \left(\frac{1}{1-\bar{\beta}} \pi_k \frac{dp_k}{p_k} - \beta_k \frac{1}{1-\bar{\beta}} \pi_k \frac{d\rho}{\rho} \right) \\
&= \frac{1}{1-\bar{\beta}} \sum_k \pi_k \left(\frac{dp_k}{p_k} - \beta_k \frac{d\rho}{\rho} \right) \\
&= \frac{1}{1-\bar{\beta}} \sum_k \pi_k \frac{dp_k}{p_k} - \frac{1}{1-\bar{\beta}} \frac{d\rho}{\rho} \sum_k \pi_k \beta_k \\
&= \frac{1}{1-\bar{\beta}} \sum_k \pi_k \frac{dp_k}{p_k} - \frac{\bar{\beta}}{1-\bar{\beta}} \frac{d\rho}{\rho}
\end{aligned}$$

E.2 Change in welfare

With this expression, it is easy to calculate the first-order approximation to the change in welfare of farmer households, which own land and labor. Welfare for them is:

$$W = \frac{VA}{P^b},$$

which implies

$$\begin{aligned}
\frac{dW}{W} &= \frac{dVA}{VA} - b \frac{dP}{P} \\
&= \frac{1}{1-\bar{\beta}} \sum_k \pi_k \frac{dp_k}{p_k} - \frac{\bar{\beta}}{1-\bar{\beta}} \frac{d\rho}{\rho} - b \sum_k s_k \frac{dp_k}{p_k} \\
&= \sum_k \left(\frac{1}{1-\bar{\beta}} \pi_k - b s_k \right) \frac{dp_k}{p_k} - \frac{\bar{\beta}}{1-\bar{\beta}} \frac{d\rho}{\rho}
\end{aligned}$$

F Data Appendix

In this section I provide more details on the data sets I use in the paper. For an overview of all data sources and samples, see Table H.1.

Regions. In matching the model to data, I use the administrative division of Peru to define the regions in the model. As of 2012, Peru is hierarchically divided into 24 departments, 194 provinces and 1838 districts. Each region i in Home in the model corresponds to one of the 194 provinces in Peru, although for some purposes, like the estimation of θ, I exploit the district level disaggregation. I use information on consecutive cross-sections, the exact years depending on the sample.

Crops. The National Statistics on Agriculture contain detailed information on over 180 crops grown in Peru. Besides being unmanageable from a computational perspective, many of these crops only account for tiny fractions of land and value of production. Therefore, I select the top 20 crops, according to their value of production in the years 2008-2011, and restrict the sample to those. The crops are listed in Table H.2. Section F.7 at the end of this Appendix explains how I match crops across different sources.

As shown in Table H.2 these crops do not account for all of production in the country. Further, the coverage varies across regions. Therefore, I re-scale the data on H_i as to keep the total value of production for each region unchanged. This procedure also leaves land shares $\eta_{i,k}$ and prices $p_{i,k}$ unchanged. The size H_i of each district is given by the total amount of land used for growing the crops in the sample. The 25th, 50th, and 75th percentiles of the resulting size distribution for provinces are $54.87km^2$, $134.87km^2$, and $252.17km^2$.[10]

F.1 National Statistics on Agriculture

These data are collected by the Peruvian Ministry of Agriculture (MINAG). Given the level of disaggregation, MINAG relies on local "qualified informers", who are often people trusted by the community, to gather the data.

The original data are disaggregated at the district level. For each district i, crop k, and year t, the data set contains information on farm-gate prices, $p_{i,k,t}$, physical yields, $y_{i,k,t}$, and land use, $\eta_{i,k,t}H_{i,k,t}$. I eliminate the time variation by averaging each variable at the district and crop level, and interpret these averages as the objects $p_{i,k}$, $y_{i,k}$ and $\eta_{i,k}H_i$ in the model. Using the data at the province level requires aggregation, which I do weighing appropriately by revenues or by land shares, depending on the variable.

I describe below two samples, which I use for different purposes depending on their relative strengths.

[10]The 25th, 50th, and 75th percentiles of the size distribution of districts are $1.88km^2$, $5.76km^2$ and $15.96km^2$. This distribution remains unchanged, since it is only used in the estimation of θ, which does not require the rescaling.

Wide sample

This is a cross-section of every district in Peru that produces agricultural goods. It is a balanced panel containing the years 2008-2011. Once I have an estimate of θ, I use this sample, aggregated at the province level, to estimate underlying land quality for each region and crop, $A_{i,k}$. Descriptive statistics for this sample are shown in Tables H.4, H.5 and H.6 in this Appendix.

Long sample

This is a sample of the districts contained in 4 out of 24 departments: Arequipa, Huánuco, La Libertad, Puno. They account for 22 percent of the total value of production in 2008. The advantage of this sample is that it includes the years 1997-2011, although the panel is unbalanced. I use it to get more precise estimates of the long-run equilibrium values of $\eta_{i,k}$ and $p_{i,k}$, which are necessary to improve the estimation of θ, the land heterogeneity parameter. Also, for the estimation of θ, which requires matching to the fine-grained GAEZ dataset, I use data at the district level of disaggregation.

F.2 Global Agro-Ecological Zones (GAEZ)

I describe the data set briefly –Costinot and Donaldson (2014) provide a more detailed discussion of it. The goal of the GAEZ project is to assess the agricultural potential for land cells in a fine grid of the World. FAO and IIASA have developed a methodology to estimate the potential land yield (see IIASA/FAO (2012)). That is, they estimate the land yield that would prevail if all land in a cell is entirely devoted to growing a crop. This method transforms information on land types, water resources and weather conditions into potential yields, through a model of agricultural production. Importantly, actual statistics on agricultural production are not inputs into the model. Hence, the database contains truly independent measures of potential agricultural productivity.

To access the data on potential land yields, the user must make a choice about management conditions: low, medium and high level of inputs. In estimation, I show the results of using low input levels, but nothing hinges on this: the results are virtually unchanged with medium input levels.

F.3 Freight rates, the transportation network and geography

The Transportation Network I use a digitized dataset for the universe of roads in Peru, provided by Ministerio de Transportes y Comunicaciones (MTC). The system is divided hierarchically in 3 parts: The 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. The Departmental roads (Red Vial Departamental), under the purview of each of the 24 departments of the administrative division, serve as an intermediate between National roads and the more local, Neighborhood roads (Red Vial Vecinal). The latter connect populated and production centers with departmental roads. See Ministerio de Transportes y Comunicaciones de Peru (2011)

Regardless of its place in the hierarchy, a road segment can be paved, graded or dirt. I use this characteristic later on to classify roads into high and low quality. The data set also contains a set of clearly identified roads that have not been built, but are being considered for being built (classified as "En Proyecto"). These are the roads I use in the Counterfactual contained in Section 7.2.

As documented by the Ministry of Transportations and Communications (Ministerio de Transportes y Comunicaciones de Peru, 2013), trains have historically, and until the present, been used almost exclusively for transportation of people or minerals, which is why I exclude them from the analysis of transportation of crops. Quoting and translating from the report: "During the year 2013, 38% of the public freight railroad service is conducted by [two firms]. The first transports mainly minerals, among them concentrated zinc, copper and zinc bars. The remaining 62% corresponds is a private service conduced by the firm Southern Peru Copper Corporation."[11]

The same report suggests that domestic freight transport by air is minimal as well. According to MTC, 11% of all freight air traffic (or 36, 190 tn) corresponds to domestic traffic (the rest is international traffic, a distinction that is irrelevant for my approach).

To the best of my knowledge, river trade is minimal as well, although I did not find in the same publication any direct measures. The only navigable rivers are in the jungle region (three out of 24 departments in my sample, which comprise 5.0% of the total workforce and 7.8% of total cultivated land). Still, available statistics about international fluvial trade suggest it is not particularly important: out of total exports through aquatic ports (12 262 metric tons), only 18 were carried through river ports.

Freight Rates I use a sample of freight rates between 46 pairs of districts, averaged over the years 2010-2013, where at least one of the districts in the pair belongs to the department of La Libertad. Most of the freight rates are expressed in terms of local currency per unit of weight. Others are measured in units (mostly animals) that need to be converted to weight units. The scope of the data is restricted this way because the source is the Direccion Regional de Agricultura de La Libertad (http://www.agrolalibertad.gob.pe/).

Altitude Altitude comes from the GTOPO30 project (https://lta.cr.usgs.gov/GTOPO30). It contains a raster of 30 arc seconds resolution.

F.4 National Household Surveys (ENAHO)

This survey is conducted by the Instituto Nacional de Estadistica e Informatica (INEI); it is Peru's main tool for learning about living standards. Every quarter, INEI samples households at random and applies a survey about income, expenditures, living conditions, etc. The household contains several modules; in this paper, I use the household expenditure module, which contains information

[11]Southern Peru is one of the main mining firms in Peru

on expenditures on disaggregated food items. INEI asks questions in this module to all households in the sample. Respondents give a detailed account of their expenditures on narrow food consumption categories, during the fifteen days previous to the day the survey was administered. The survey is conducted yearly, and to keep a consistent sample with the National Statistics, I use the years 2008 to 2011.

F.5 The International Price of Intermediate Inputs, ρ_F, and Crops, $p_{F,k}$

To calculate the international price of crops and intermediate inputs, I use data on agricultural trade between 2008 and 2011, coming from the Ministry of Agriculture, disaggregated at the HS10 level. Building on a classification employed by the Ministry of Agriculture[12], I match crops in the paper to HS10-level. This match is detailed in Table H.9

The International Price of Intermediates ρ_F To calculate the price of the good at the port, I construct a bundle of fertilizers and average their price. This is similar to assuming that, to obtain a unit of intermediate inputs, farmers combine all available fertilizers in fixed, equal proportions, as with a Leontieff production function. In Table H.7 I show the unit FOB price and the import quantity of each fertilizer I include in the input bundle. I take the simple average in the bottom row to be ρ_F.[13]

The International Price of Crops $p_{F,k}$ After matching the K crops in the paper to their corresponding HS10 codes, I construct unit values, corresponding to $p_{F,k}$, by dividing total flows by physical quantities. Table H.8 reports those values. The key observation is that, while it is true that a fraction of agricultural international trade comprises crops processed to different degrees, within each crop I identify prices gaps using only the crops that best reflect the unprocessed nature of crops at the farm gate and therefore set the right international prices in the calibration

Trade Flows by Crop and Customs Using the same data, I complement the match described previously with other HS10 codes for each crop, as to match the actual quantities of exports, but including more processed stages of output. I use these data to calibrate the parameters a_k, as described in the paper.

Using the same HS10 correspondence, I use data on exports by good and customs coming from the Association of Exporters (ADEX) to calculate the share of exports of each good going out through each port.

[12] This is the classification the Ministry uses in its Monthly Statistical Bulletin (Boletín Estadístico Mensual "EL AGRO EN CIFRAS")

[13] FAOSTAT data for the years 2008 and 2009 show: (i) the fertilizers included in the Table account for more than 95 percent of total imports of fertilizer, (ii) imported fertilizer is more than 99 percent of consumption, (iii) exports are about 3 percent of consumption (all measured by weight). Taken together, this evidence suggests that the assumption that intermediate inputs are imported from abroad is not too far from reality.

Non-Agricultural Trade Flows I obtain data on aggregate trade flows outside of agriculture from the 2008 Input-Output Tables for Peru.

F.6 Other Regional Data

Labor Force I use the National Population Census of 2007 to obtain the number of workers in agriculture and non-agriculture for each district, and aggregate them up to the province level. The distribution of workers across regions is given in Table H.3, panels (b) and (c).

Gross Domestic Product I obtain GDP data in Agriculture and Non Agriculture for each department in Peru from INEI.

F.7 A note on matching crops across data sets

Assembling the database requires matching crop definitions coming from three distinct data sets, each with its own nomenclature: (i) National statistics on agriculture (Ministry of Agriculture), (ii) Consumption module of the National Household Survey (ENAHO), and (iii) exports and imports of agricultural goods (published by the Ministry of Agriculture).

Source (i), the National statistics on agriculture, forms the basis of my list of crops, and it remains unchanged throughout. To match ENAHO to it, I use question "produc61" which, on the basis of the responses to questions p601A and p601X, match each item consumed to a list of finely disaggregated consumption goods. Then I match, one by one, the disaggregated categories to their closest counterparts in the National Statistics database. With this procedure, I was able to match the following goods: asparagus, avocado, banana, barley, cacao, cassava, coffee, dry bean, grape, maize (choclo), onion, orange, potato, rice, tangerine, wheat. The match is imperfect because some goods in the National Statistics are not frequently consumed by households in their unprocessed form.

The data I use on exports and imports (source (iii)), while published by the Ministry of Agriculture, are originally measured by the customs authority. Therefore, goods are classified at the HS10 level (the first 6 digits are in common with the rest of the world, while the latter 4 are specific to Peru.) The Ministry of Agriculture groups the relevant HS10 categories according to the crop they contain. Since this grouping sometimes combines raw crops with products with varying degrees of processing, I further disaggregate these groupings for the purposes of identifying prices, $p_{F,k}$, and export quantities, $z_{Fi,k}$ and $z_{iF,k}$.

To calculate prices, I only keep those HS10 categories that pertain to unprocessed crops. This selection excludes part of the total volumes of exports and imports (I retain approximately 60% of total export and import value of all agricultural exports and imports reported by the Ministry of Agriculture), but allows me to calculate an international price that is comparable to the farm-gate price in the national statistics. The categories I use for obtaining prices are shown in Table H.9.

To recover the correct quantities, I then add the rest of categories that the Ministry of Agriculture groups as belonging to a given crop. These additional categories are shown in Table H.10.

In both cases, the names are often close enough that the match is straightforward. Still, it is not a perfect procedure, and it requires a few judgment calls.

G Estimation Appendix

In this section, I describe the details of estimation that were left out of the paper for space considerations. I also derive results that I use in the paper to bring the model to the data.

G.1 Estimation of cost shares

Estimation of γ_k The basis for the estimation of γ_k is Proposition 3 in the main text. To estimate it, I regress

$$\log \pi_{i,k} = \log \eta_{i,k} + \iota_i + \iota_k + \epsilon_{i,k} \tag{9}$$

where i is a district, k is a crop and ι_i and ι_k are region and crop dummies (omitting a base crop l), and $\epsilon_{i,k}$ is motivated as measurement error on $\pi_{i,k}$. According to Proposition 3, the region fixed effect ι_i captures the normalization term $\log\left(\sum_l \gamma_l^{-1}\eta_{i,l}\right)$; while the crop fixed effect ι_k captures $\log(\gamma_k)$.

In regression (6), I include a full set of region fixed effects and I omit one crop fixed effect. The regression therefore identifies land intensities γ_k relative to a base category. To recover all levels, I normalize the revenue-weighted cost share of land for the country as a whole to 0.22, as in Dias Avila and Evenson (2010):

$$0.22 \quad = \quad \sum_k \underbrace{\left(\frac{\sum_i V_{i,k}}{\sum_{i'} V_{i'}}\right)}_{\text{k's total revenue share}} \quad \gamma_k$$

where $V_{i,k}$ is the total value of production in region i and crop k, and V_i region i's total value of production.

Note that in regression (9) we estimate a coefficient of 1.03 (std. err. 0.006) for $\eta_{i,k}$ – reassuringly, almost equal to 1.

Calibration of α_k and β_k To arrive at the normalization that the aggregate revenue-weighted cost share equals 0.22, I take the estimates for the input cost shares in Peru from Dias Avila and Evenson (2010), Table A.3a, for the period 1981-2001: cropland (22.17%), labor (56.23%), fertilizer and chemicals (7.02%), seeds (4.64%), mechanization (4.57%) and animal power (3.85%).

Interpreting the labor cost share in Dias Avila and Evenson (2010), Table A.3a, as the country-wide revenue-weighted cost share in the model suggests $\bar{\alpha} = 0.56$, and letting intermediate inputs include mechanization and animal power $\bar{\beta} = 0.22$. Thus, compared to the estimates in Hayami and Ruttan (1985), –later quoted in Restuccia, Yang, and Zhu (2008)–, who estimate that, for a sample of countries, the labor cost share is 0.42, the intermediate input share is 0.4 and the land share is 0.18, I use a higher labor cost share and a lower intermediate input share.

G.2 Obtaining the equations to estimate θ

Although potential land yield is not observed in equilibrium, the GAEZ data set is useful in the estimation of θ because it aids in handling the unobserved productivity terms, $A_{i,k}$. Suppose that, in equilibrium, we observe a region's prices $(p_{i,k})$, w_i, ρ_i. Then we can calculate the potential land yield $\mathbb{E}\left[y_{i,k}\left(\omega\right); w_i, \rho_i, p_{i,k}\right]$, that is, the yield that would prevail if all land were allocated to that crop, given the prices.

Lacking data on labor and intermediate input shares across crops, I impose that $\alpha_k/\beta_k = \nu$, where ν is crop independent, and since $\alpha_k + \beta_k = 1 - \gamma_k$, we have $\beta_k = \left(1-\gamma_k\right)/\left(\nu+1\right)$ and $\alpha_k = \nu\left(1-\gamma_k\right)/\left(\nu+1\right)$. Equipped with estimates of all land shares, $\hat{\gamma}_k$, now we can turn to optimal land allocation to estimate θ. The first step is to derive the unconditional land productivity. Start with the yield in crop k in plot ω is

$$y_{i,k}\left(\omega\right) = \frac{\lambda_{i,k}\Lambda_{i,k}\left(\omega\right)}{p_{i,k}\gamma_k}.$$

The unconditional expectation, $y_{i,k}\left(\omega\right)$, i.e., allocating all plots $\omega \in \Omega_i$ to crop k, is

$$\begin{aligned}\mathbb{E}\left[y_{i,k}\left(\omega\right)\right] &= \gamma_k^{-1}p_{i,k}^{-1}\lambda_{i,k}\mathbb{E}\left[\Lambda_{i,k}\left(\omega\right)\right]\\ &= \gamma_k^{-1}\bar{c}_k^{-\frac{1}{\gamma_k}}w_{i,A}^{-\frac{\alpha_k}{\gamma_k}}\rho_i^{-\frac{\beta_k}{\gamma_k}}p_{i,k}^{\frac{1-\gamma_k}{\gamma_k}}A_{i,k}\end{aligned}$$

Again, at those prices land would be allocated according to (4), so this is an object that, while never observed in equilibrium, corresponds to the statistics produced by the GAEZ project.

The GAEZ data. Next, I impose an additional restriction to make contact with the GAEZ data. I assume that the GAEZ *model* does not have spatial price variation in it (as explained below). I also assume there is a prediction error $u_{i,k}$, which allows us to make contact with the GAEZ data as

$$\tilde{y}_{i,k}^G = \gamma_k^{-1}\bar{c}_k^{-\frac{1}{\gamma_k}}\left(w_{i,A}^G\right)^{-\frac{\alpha_k}{\gamma_k}}\left(\rho_i^G\right)^{-\frac{\beta_k}{\gamma_k}}\left(p_{i,k}^G\right)^{\frac{1-\gamma_k}{\gamma_k}}A_{i,k}e^{u_{i,k}},$$

where $\tilde{y}_{i,k}^G$ is the GAEZ project's measure.

The estimating equation. To arrive at the estimation equation, I eliminate $A_{i,k}$ from the equation above. To do so, I use the land share equation (4):

$$\eta_{i,k} = \frac{\lambda_{i,k}^\theta A_{i,k}^\theta}{\Phi_i^\theta},$$

to solve for $A_{i,k}$ which yields:

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

Using this expression to substitute for the true $A_{i,k}$ in the equation that defines the GAEZ estimate, $y^G_{i,k}$, through the lens of the model, we obtain:

$$\tilde{y}^G_{i,k} = \gamma_k^{-1} \bar{c}_k^{-\frac{1}{\gamma_k}} \left(w^G_{i,A}\right)^{-\frac{\alpha_k}{\gamma_k}} \left(\rho^G_i\right)^{-\frac{\beta_k}{\gamma_k}} \left(p^G_{i,k}\right)^{\frac{1-\gamma_k}{\gamma_k}} \frac{\Phi_i}{\lambda_{i,k}} \eta_{i,k}^{\frac{1}{\theta}} e^{u_{i,k}},$$

so taking logs and rearranging, we obtain

$$\begin{aligned} \log \tilde{y}^G_{i,k} &= -\log \gamma_k - \frac{1}{\gamma_k} \log \bar{c}_k - \frac{\alpha_k}{\gamma_k} \log w^G_{i,A} - \frac{\beta_k}{\gamma_k} \log \rho^G_i + \frac{1-\gamma_k}{\gamma_k} \log p^G_{i,k} \\ &\quad + \log \Phi_i + \frac{1}{\theta} \log \eta_{i,k} + \frac{1}{\gamma_k} \log \bar{c}_k - \frac{1}{\gamma_k} \log p_{i,k} + \frac{\alpha_k}{\gamma_k} \log w_{i,A} + \frac{\beta_k}{\gamma_k} \log \rho_i \\ &\quad + u_{i,k}. \end{aligned}$$

GAEZ model has no spatial variation. As I anticipated, I further assume that there is no price variation in the GAEZ model. Doing so simplifies the equation to:

$$\begin{aligned} \log \tilde{y}^G_{i,k} &= \underbrace{-\log \gamma_k - \frac{(1-\gamma_k)\nu}{\gamma_k(1+\nu)} \log w^G_A - \frac{1-\gamma_k}{\gamma_k(1+\nu)} \log \rho^G + \frac{1-\gamma_k}{\gamma_k} \log p^G_k}_{=\iota_k} \\ &\quad + \underbrace{\log \Phi_i}_{\iota_i} + \frac{1}{\theta} \log \eta_{i,k} - \frac{1}{\gamma_k} \log p_{i,k} + \underbrace{\frac{(\nu \log w_{i,A} + \log \rho_i)}{1+\nu}}_{=\delta_i} \frac{(1-\gamma_k)}{\gamma_k} \\ &\quad + u_{i,k} \\ &\Rightarrow \\ \log\left(p_{i,k}^{\frac{1}{\gamma_k}} \tilde{y}^G_{i,k}\right) &= \frac{1}{\theta} \log \eta_{i,k} + \iota_k + \iota_i + \delta_i \frac{(1-\gamma_k)}{\gamma_k} + u_{i,k} \end{aligned}$$

where ι_i and ι_k are region and crop fixed effects, and δ_i is a region-specific coefficient for the regressor $(1-\gamma_k)/\gamma_k$ which is constructed from the first stage. In practice, this means that we construct the variable $(1-\gamma_k)/\gamma_k$ and multiply it by a set of regional dummies ι_i, so that we have, for each region i, the regressor:

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ (1-\gamma_1)/\gamma_1 \\ \vdots \\ (1-\gamma_K)/\gamma_K \\ \vdots \\ 0 \end{bmatrix}$$

and each one of these variables is associated with its own coefficient δ_i.

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

To match the GAEZ grid to the districts, I use the actual administrative division of Peru. The quartiles of the administrative district-size distribution are $93.8km^2$, $210.6km^2$ and $497.7km^2$. The quartiles for the distribution of harvested land are $1.88km^2$, $5.76km^2$, $15.96km^2$. Due to its projection, the cell size in the GAEZ grid is approximately $86.km^2$ at the Equator, but it grows larger at higher latitudes. The fact that the total agricultural land is usually much smaller than the total amount of land in a district justifies using the maximum GAEZ value for each district that contains more than one GAEZ cell. The crops included in the regression are those that are both observed in national statistics and in the GAEZ dataset: alfalfa, banana, barley, cacao, cassava, coffee, maize (yellow hard), onion, orange, potato, rice, wheat. Table H.12 shows the regression results.

G.3 Estimation of Iceberg Costs

The goal of this section is to produce an estimate of the iceberg trade costs between any two regions in Peru, and for each good in the data set.

The first step is to estimate a statistical model of transport costs: I project a sample of within-country price gaps on data about the quality and geography of the road that connects each origin-destination pair in the sample. Because data on geography and road quality are available for the whole country, I then use this estimated model to predict trade costs for all possible origin-destination pairs in Peru.

G.3.1 Preparing the road and geography data

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

Nodes. The following are nodes in the graph: (i) district capitals, (ii) locations where existing roads branch out or merge, (iii) locations where the quality of a road changes, (iv) if a road is the closest one to a district centroid, the point in the road that is closest to that centroid.

Arcs. Two nodes are connected by an edge if at least one of the following conditions is met: (i) there is a segment of road of any quality that connects them, (ii) the two nodes are district capitals at most 10 km. apart, (iii) one of the nodes is a district capital that would be disconnected from the rest of the graph without connection to another district capital. In case (ii) I use the straight-line distance and assign low quality to the connection. In case (iii) I use the straight-line distance to the closest district capital and assign low quality to the connection.

G.3.2 Estimation

I estimate a transport cost model, which will give estimates of the relative costs of traversing roads of different qualities and with different slopes. Let y_{ni} be the dependent variable (either price gaps or freight rates of shipping a kilogram of goods) from region i to region n. I estimate the following equation by NLS:

$$\mathbb{E}\left[\log y_{ni} | \text{geography, roads}\right] = \beta_0 + \beta_{distance} \log\left(\text{effective distance}_{ni}\left(\lambda\right)\right). \tag{10}$$

where $\beta_{distance}$ is the elasticity of the dependent variable to effective distance. For a given choice of the parameter vector λ, "effective distance$_{ni}$" is the lowest-cost path between regions n and i, calculated according to Dijkstra's algorithm, which minimizes the following weighted sum of road lengths:

$$\text{effective distance}_{ni}\left(\lambda\right) = \min_{R} \sum_{q \in Q} \sum_{\text{edge} \in E_q(R)} \left[h\left(\lambda_s s_{\text{edge}}\right) \cdot \left(\lambda_q \text{road length}_{\text{edge}}\right)\right]. \tag{11}$$

In equation (11), the effective distance between i and n is the weighted distance over route R on the network. For each road of quality $q \in Q$, route R contains edges $E_q\left(R\right)$. The cost of traversing a kilometer of road of quality q is λ_q and λ_s is the effect of traversing an edge with slope s_{edge}, captured through the function $h\left(\cdot\right)$. Without loss, we normalize the weight of high-quality distance, λ_{high}, to one.[14]

In practice, I set $Q = \{\text{high}, \text{low}\}$, where only paved roads have high quality, and I set $h\left(x\right) = \left(1 + x\right)$ to avoid non-linearities that depend on the way roads were segmented to make the graph. In all of these formulations I use the Matlab BGL Library (Gleich (2009)).

Letting $n = F$, and $y_{ni} = p_{F,coffee}/p_{i,coffee} - 1$ gives the results in the main body of the paper, contained in Table H.13(a).[15] Tables H.13(b) presents estimates of (10) in levels, and show that the results do not change much, especially for my main specification. Finally, the third column of both tables estimates the model allowing for the effect of slope of terrain, but the standard errors are too large to draw inference on them.[16]

Price gaps are observed at the district level, and that is also the level at which I construct

[14] In (11) the optimal road depends on the actual value of λ. The reason is that, given λ, Dijkstra's algorithm chooses among alternative ways of reaching n from i, over the graph, and these choices may change with the cost vector λ. In the extreme, if $h\left(\cdot\right) = 1$ and $\lambda_{\text{q}} = 1 \ \forall q \in Q$, the algorithm minimizes the simple road length between two points. As λ_q increases, for $q \neq \text{high}$, the algorithm gives priority to high-quality edges.

[15] The underlying model is that the iceberg trade cost from i to n follows the form: $d_{ni} = 1 + \exp\left(\beta_0 \text{distance}_{ni}^{\beta_{distance}}\right)$, which ensures that the iceberg cost is always weakly greater than one.

[16] Limao and Venables (1999) find empirical evidence for the role of infrastructure as a determinant of trade costs. Donaldson (2015) estimates that transporting goods on dirt roads increases transport costs by a factor of 7.9 relative to railroads. My estimates are larger, which possibly reflects that infrastructure plays a larger role in Peru.

the network described above. However, in the rest of the paper I identify a region with Peruvian provinces. This requires me to aggregate the predictions from equations (10) and (11).

G.4 International Barriers and Domestic Demand Parameters

In this section, I discuss how I calibrate the domestic demand parameters, a_k and international barriers τ_k. For any region i in Home, the cost of trading with Foreign has two components. The first is captured by the cost of trading with the closest international port. To find the closest international port, I select the three main sea ports by value traded, and find the closest port $o(i)$ to region i according to the predicted iceberg cost $\hat{d}_{o(i)} = \arg\min_j \hat{d}_{ji}$. The second component, denoted by τ_k below, is a barrier that prevents the traders from realizing the full price of goods at the port. Without it, the model is unable to reproduce the variation of crop prices within the country, as equilibrium prices would then deviate too little with respect to international prices.

I use information on net export quantities and international prices to calibrate a_k, $\forall k$. Specifically, I pick a_k to solve the following problem:

$$\min_{X_{ag}, a_k} \left(\frac{\mathrm{NX}_k^d \tilde{d}_{F,k} \tau_k - q_k^d + c_k^d \left(a_k, X_{ag}\right)}{q_k^d} \right)^2 \mathbb{I}\left[\mathrm{NX}_k^d > 0\right] \quad +$$

$$\left(\frac{\mathrm{NX}_k^d / \left(\tilde{d}_{F,k}\tau_k\right) - q_k^d + c_k^d \left(a_k, X_{ag}\right)}{c_k^d \left(a_k, X_{ag}\right)} \right)^2 \mathbb{I}\left[\mathrm{NX}_k^d < 0\right] \quad +$$

$$\left(\frac{q_k^d - c_k^d \left(a_k, X_{ag}\right)}{q_k^d} \right)^2 \mathbb{I}\left[\mathrm{NX}_k^d = 0\right]$$

subject to $\sum_k a_k = 1$. In this problem, $c\left(a_k, X_{ag}; p_{F,k}^d\right)$ is consumption of k according to the model, and is calculated making use of information on average farm-gate prices $p_{F,k}^d$, preference parameters a_k and X_{ag}, a measure of real consumption in agricultural goods; NX_k^d are observations on aggregate net exports of crop k, q_k^d are data on aggregate output of crop k, p_k^d are median farmgate prices for crop k, $\tilde{d}_k$ is the minimal cost of trading internationally among regions that produce k. The intuition is to choose values of X_{ag}, a_k that allow a simple version of the model – where there are no costs to trade domestically but there are uniform costs to trade abroad – to get as close as possible to matching data on net exports. Note that I weigh each term in this objective function by the size of the industry. I choose the international wedges τ_k, such that $\tau_k = \max\left\{1, p_{F,k}/\left(p_k^d \tilde{d}_{F,k}\right)\right\}$ if $NX_k > 0$ and $\tau_k = \max\left\{1, p_{F,k}^d/\left(p_k^d/\tilde{d}_{F,k}\right)\right\}$ if $NX_k < 0$.

Note that there is an asymmetry in the treatment of the international wedges in the calibration: they enter in the difference between domestic and international prices, but not in the difference between domestic flows and international flows. The reason is that, for a few crops, even with small trade costs $\tilde{d}_{F,k}$, $\mathrm{NX}_k^d \tilde{d}_{F,k} > q_k^d$, which implicitly states that consumption of that crop is negative. However, there are large differences between international, $p_{F,k}^d$, and domestic average prices, p_k^d, which explains the need for the τ_k wedges. This calibration is similar to having a tax on exports and imports that doesn't get rebated to consumers.

Figure H.4(a) compares the full simulation results of net exports (normalized by gross output in agriculture) with those numbers in the data. This correlation is not explicitly targeted in the in the calibration. The Figure shows that the fit is good, with the exception of avocado. The reason is that there is large within country heterogeneity in the price avocado, which the model is unable to generate; coupled with a large elasticity of supply for that crop, this generates excessive specialization. Figure H.4(b) shows that the model is quite close to fitting net exports as a fraction of each crop's output, however. The results of this procedure are reported in Tables H.17 and H.18. With these values of a_k and τ_k, I compute for each good, the following measure of international trade costs

$$\hat{d}_{iF,k} = \tau_k \hat{d}_{o(i)}.$$

G.5 Land Quality Parameters, $A_{i,k}$

The estimation proceeds in two steps. First, we obtain the value of Φ_i in equilibrium. Second, using that value, we use the land equation to back out the value of $A_{i,k}$ that rationalizes land allocations given prices.

Total Value of Production and Baseline Value of Φ_i. First, recall that the equilibrium value of agricultural production in region i is

$$V_i = \tfrac{H_i}{\bar{\gamma}_i} \Phi_i \quad .$$

Because V_i, $\pi_{i,k}$ are observable, and we have estimates of γ_k, we can back out Φ_i from the data as:

$$\Phi_i = \tfrac{\bar{\gamma}_i V_i}{H_i} \quad . \tag{12}$$

Equation (12) uses the aggregation properties of the model to infer region i's aggregate land productivity from data on its land share of income, total value of production, and land endowment. The variable Φ_i is informative about the aggregate level of productivity, as shown in the distribution of yields and revenues in Proposition 2. A higher value of $\bar{\gamma}_i V_i$ relative to H_i and will lead us to infer a higher land quality for all crops in i, because it means that region i produces more value for a given factor use.

Estimation of $A_{i,k}$. To estimate the $A_{i,k}$ parameters I rely on the model structure. As I have discussed in detail in Section 4, Assumption #2 imposes strong restrictions on what data are informative about land quality. The only way to learn about the relative values of the parameters $A_{i,k}$ is by comparing land allocations across crops, within a region. In contrast, data on revenue per unit of land and physical yields are informative about the common component of all $A_{i,k}$ within a region. Recall that relative values of $A_{i,k}$ are not directly observable, except for the limiting case where $\theta \to \infty$, in which the observed land yield of crop k fully reveals $A_{i,k}$.

My approach, which extracts the model parameters using data on the endogenous variables, is an alternative to the use of external measures of productivity. Costinot, Donaldson, and Smith (2016) and Costinot and Donaldson (2014), for example, use directly the potential quality measures produced by the GAEZ project. Their method has the benefit that the productivity measures are independent of the model, insofar as the researcher only needs to choose how to interpret the productivity data. Its main shortcoming is that, although constructed with extreme care, the GAEZ measures are an imperfect measure of actual land quality. For my application, there is an additional complication: GAEZ does not estimate potential productivity data for some goods that are important in my database.

Just like aggregate output and endowments are informative of a common component of land quality for all crops in region i, data on prices and land allocations are informative about the relative land qualities within that region. Using equation (4) to solve for $A_{i,k}$ we obtain:

$$A_{i,k} = \eta_{i,k}^{\frac{1}{\theta}} \frac{\Phi_i}{\lambda_{i,k}}. \tag{13}$$

To take this expression to data, we use

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

together with

$$w_{i,A} = \frac{\bar{\alpha}_i V_i}{L_{i,A}}$$

to construct the baseline value of

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

We can take these expressions to data because $p_{i,k}$, $\eta_{i,k}$, V_i and $L_{i,A}$ are measured directly, and equation (12) tells us how to measure Φ_i with the regional aggregates. It is clear that these estimates, $\hat{A}_{i,k}$, are independent of the numeraire in the data, since equation (13) is homogeneous of degree zero in prices.

Let us take a moment to interpret this equation. As already said, the statistic Φ_i shifts all estimates of $A_{i,k}$ proportionally, based on how much output is produced in i, compared to its endowments. A large value of $\eta_{i,k}$ requires a higher land quality for crop k, relative to the other crops, to rationalize it. But we must also net out the effect of the profitability of growing that crop in i, $\lambda_{i,k}$, which also tends to generate a large land allocation to crop k.[17]

There is an alternative interpretation of equation (13) that will help understand the results of the simulations in Sections 6 and 7. The estimate of $A_{i,k}$ combines information on prices and land allocations. In this estimation, variation in land allocations is more important, relative to price

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

variation, the larger $1/\theta$. Given γ, a lower value of θ (high heterogeneity) gives less importance to land allocations in the estimation of $\hat{A}_{i,k}$.

Finally, while in the theory section I have specified labor as inelastically supplied to each sector, the calibration does not require taking a stance on labor mobility since, using the model, one is able to back out implied wages in agriculture and non-agriculture, $w_{i,M}$ and $w_{i,A}$.

Sample and Results My primary goal in obtaining $A_{i,k}$ is coverage, so I sacrifice precision in the estimation to be able to obtain an estimate for every region in the country. Hence, I use the wide sample of national statistics, which contains repeated cross-sections from 2008-2011 and covers the whole country. Data on $p_{i,k}$, and $\eta_{i,k}$ are averages across time. I use the corresponding data on land yields, $y_{i,k}$, to construct the total value of production in region i, $V_i = \sum_{k \in K} p_{i,k} y_{i,k} \eta_{i,k} H_i$.

Table H.15 shows, for each crop k, the summary statistics of the estimates of $A_{i,k}$. The estimates vary substantially between crops, reflecting the fact that the price of a ton of output also varies much between crops.

G.6 Elasticity of substitution σ

ENAHO surveys each year a random sample of households for a sample of regions i. I treat each household as randomly sampled from the model, and match its consumption to the goods k used in the simulation. The observations that I can match between ENAHO and the National Statistics on agriculture account for 8.3% of total household expenditure in the survey. To minimize the role of extreme observations, for each crop I trim the top and bottom 1% observations of the unit values.

As discussed in the main body of the paper, my main IV strategy relies on the GAEZ estimates of potential productivity, $\tilde{y}^G_{i,k}$. Column 1 of Table H.16 shows OLS results of regressing expenditure shares on log unit values, for comparison purposes. Column 2 shows the the first stage coefficient and F-statistic, which suggests GAEZ productivities are relevant instruments. Column 3 shows the second stage, which is the number reported in the paper. Columns 6 and 7 repeat the estimation using the logarithm of GAEZ productivities as instruments. These results confirm the relevance of the instrument, and point to a higher elasticity of substitution $\sigma = -2.804$. While the log form for the instrument is appealing, all zero estimates in GAEZ are dropped in this regression, which decreases the number of observations.

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

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

Intuitively, if $\eta_{i,k} = 0$, the region cannot produce crop k so, unless its importing costs are too high, the supply of the crop in question will be affected by the price of delivering the crop from abroad, approximated by, $p_{F,k} + f_{i,k}$. On the other hand, if $\eta_{i,k} > 0$, region i produces some amount of crop k, and, provided trade costs are not too high, will export it, so the price will be close to $p_{F,k} - f_{i,k}$. Note, according to the model, whether $\eta_{i,k}$ is positive is entirely exogenous, and controlled by whether $A_{i,k}$ is positive. Combining the information on trade costs and international prices is crucial to generate enough variation in the data such that household and crop fixed effects can be included in the estimation. Columns (4) and (5) of Table H.16 show the results of this strategy. The results point to a $\sigma = -2.164$, slightly less than my main estimates, but confirm the direction of the OLS bias.

G.7 Manufacturing Productivity $T_{i,TR}$ and $T_{i,NT}$

To estimate the manufacturing productivity parameters $T_{i,M}$, I use data on value added per worker outside of agriculture. Under the assumption of free mobility within the non-agricultural sector M, the wage $w_{i,M}$ implied by the data is given by

$$w_{i,M} = \frac{VA_{i,TR} + VA_{i,NT}}{L_{i,TR} + L_{i,NT}},$$

where $VA_{i,j}$ is value added in region i, sector j. Unfortunately, sectoral value added data are only available at the department level, which is a higher level of aggregation than provinces, my unit of observation. To make use of the available data, I assume that all regions i within a department have the same equilibrium wage $w_{i,M}$ in the data.

I assign $T_{i,TR}$ such that the, given all data on production in the agricultural sector, and data on factor endowments, the model is able to replicate the equilibrium wages $w_{i,M}$ I observe in the data. Finally, $T_{i,TR}$ is chosen as to replicate the allocation of labor within the non-agricultural sector.

H Appendix Tables and Figures

H.1 Summary Statistics of Agriculture Data Set

Table H.1: Description of Data Sets

Name	Source	Unit of Observation	Contents	Time	Coverage
National Statistics on Agriculture	Ministry of Agriculture of Peru (MINAG)	(Crop, District) pairs	-Land use -Physical land yields -Price	-Wide Sample: 2008-2011 -Long Sample: 1999-2011	-Wide Sample: 24 departments which produce agricultural goods -Long Sample: 4 selected departments.
Global Agro-Ecological Zones (GAEZ)	FAO-IIASA	5 arc-minute x 5 arc-minute grid of the country, by crop	Potential land yield, if all the land in the pixel is used in that crop	Average for 1961- 1990	All of Peru
National Household Survey (ENAHO)	Instituto Nacional de Estadistica e Informatica (INEI)	Household	Food Consumption Module	2008-2011	Nationally Representative Samples
Transport Network	Ministry of Transport and Communications of Peru (MTC)	Each road in the national system of roads (National, Departmental, Rural)	Exact location, shape, length and road quality of each road	As of 2011	All of Peru
Altitude	Shuttle Radar Topography Mission	3 arc-second by 3 arc-second grid of the country	Altitude of each cell	-	All of Peru
Freight Rates	Direccion Regional de Agricultura de La Libertad	Sample of 45 pairs of districts where at least one district belongs to La Libertad	Cost in LCU of shipping different goods, per corresponding unit of measure	Years 2011-2013	At least one of the trading districts belongs to the department of La Libertad
Employment	National Population and Housing Census	1838 Districts in Peru	Number of people working in agriculture as main or secondary activity	2007	All of Peru
Manufacturing GDP	INEI	Department	Total GDP in LCU	Years 2008-2011	All of Peru
Imports and Exports	MINAG	Aggregate	HS10 export and import quantities and prices	Years 2008-2011	All of Peru

Table H.2: Crops Included

Crop	Rank	Cumulative Share of Production Value
rice	1	0.14
potato	2	0.26
coffee	3	0.37
maize (yellow hard)	4	0.42
alfalfa	5	0.46
asparagus	6	0.50
banana	7	0.54
cassava	8	0.57
maize (amilaceo)	9	0.59
grape	10	0.61
cotton branch	11	0.63
onion	12	0.65
maize (choclo)	13	0.66
dry bean	14	0.68
avocado	15	0.69
wheat	16	0.70
cacao	17	0.71
orange	18	0.72
barley grain	19	0.73
tangerine	20	0.74

For a crop ranked k, "Cumulative Share of Production Value" refers to the first k crops' share of production, out of the total value produced between 2008 and 2011, evaluated at 2011 prices.

Table H.3: Summary Statistics of Land Shares, $\eta_{i,k}$

(a) Arable Land (Hectares)

	count	mean	std	min	25%	50%	75%	max
Land	194.00	18257.49	17632.09	0.00	5487.25	13487.29	25217.16	117681.50

(b) Agricultural Labor

	count	mean	std	min	25%	50%	75%	max
Agricultural Labor	194.00	1221.43	972.59	29.10	483.72	977.40	1714.38	5071.50

(c) Non-Agricultural Labor

	count	mean	std	min	25%	50%	75%	max
Non-Agricultural Labor	194.00	3881.36	23608.11	67.50	372.95	848.10	1892.73	325405.40

Table H.4: Summary Statistics of Land Shares, $\eta_{i,k}$ (fraction of total land)

	count	mean	std	10%	25%	50%	75%	90%
alfalfa	194.00	0.11	0.22	0.00	0.00	0.01	0.08	0.40
asparagus	194.00	0.01	0.07	0.00	0.00	0.00	0.00	0.00
avocado	194.00	0.01	0.02	0.00	0.00	0.00	0.00	0.02
banana	194.00	0.04	0.09	0.00	0.00	0.00	0.03	0.17
barley grain	194.00	0.09	0.11	0.00	0.00	0.03	0.18	0.25
cacao	194.00	0.01	0.05	0.00	0.00	0.00	0.00	0.02
cassava	194.00	0.03	0.07	0.00	0.00	0.00	0.02	0.07
coffee	194.00	0.05	0.13	0.00	0.00	0.00	0.00	0.17
cotton branch	194.00	0.01	0.06	0.00	0.00	0.00	0.00	0.02
dry bean	194.00	0.04	0.05	0.00	0.00	0.01	0.05	0.10
grape	194.00	0.01	0.04	0.00	0.00	0.00	0.00	0.01
maize (amilaceo)	194.00	0.11	0.13	0.00	0.00	0.05	0.19	0.29
maize (choclo)	194.00	0.02	0.04	0.00	0.00	0.00	0.02	0.08
maize (yellow hard)	194.00	0.09	0.13	0.00	0.00	0.02	0.14	0.27
onion	194.00	0.01	0.03	0.00	0.00	0.00	0.01	0.02
orange	194.00	0.01	0.02	0.00	0.00	0.00	0.00	0.02
potato	194.00	0.17	0.20	0.00	0.00	0.12	0.23	0.41
rice	194.00	0.09	0.18	0.00	0.00	0.00	0.07	0.32
tangerine	194.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
wheat	194.00	0.09	0.12	0.00	0.00	0.02	0.14	0.26

Table H.5: Summary Statistics of Log Prices, $\log p_{i,k}$ (local currency)

	count	mean	std	10%	25%	50%	75%	90%
alfalfa	140.00	-1.60	0.52	-2.19	-1.81	-1.41	-1.29	-1.16
asparagus	22.00	0.88	0.21	0.70	0.79	0.85	0.94	1.18
avocado	127.00	0.13	0.59	-0.86	-0.24	0.19	0.62	0.82
banana	112.00	-0.74	0.41	-1.27	-1.03	-0.72	-0.47	-0.27
barley grain	127.00	-0.03	0.19	-0.29	-0.18	0.01	0.10	0.18
cacao	56.00	1.51	0.38	0.87	1.24	1.69	1.77	1.86
cassava	115.00	-0.50	0.46	-1.22	-0.70	-0.42	-0.17	-0.03
coffee	67.00	1.57	0.38	0.88	1.33	1.62	1.90	1.97
cotton branch	44.00	0.78	0.47	0.08	0.39	0.98	1.10	1.24
dry bean	153.00	0.87	0.31	0.41	0.73	0.92	1.07	1.20
grape	47.00	0.46	0.35	0.05	0.23	0.41	0.69	0.91
maize (amilaceo)	139.00	0.48	0.29	0.17	0.28	0.50	0.65	0.74
maize (choclo)	115.00	-0.28	0.33	-0.66	-0.48	-0.24	-0.05	0.07
maize (yellow hard)	139.00	-0.21	0.21	-0.51	-0.32	-0.20	-0.10	0.01
onion	99.00	-0.33	0.28	-0.65	-0.54	-0.30	-0.19	0.05
orange	121.00	-0.50	0.41	-1.06	-0.80	-0.51	-0.20	-0.00
potato	152.00	-0.43	0.22	-0.67	-0.59	-0.43	-0.27	-0.19
rice	79.00	-0.04	0.33	-0.53	-0.21	-0.04	0.13	0.33
tangerine	45.00	-0.51	0.48	-1.14	-0.82	-0.59	-0.07	0.06
wheat	138.00	0.21	0.19	-0.02	0.06	0.18	0.36	0.49

Table H.6: Summary Statistics of Log Physical Yields, $\log y_{i,k}$ (kg/ha)

	count	mean	std	10%	25%	50%	75%	90%
alfalfa	140.00	-3.45	0.46	-4.08	-3.79	-3.40	-3.20	-2.92
asparagus	22.00	-4.84	0.36	-5.29	-5.12	-4.85	-4.53	-4.41
avocado	127.00	-4.73	0.41	-5.19	-5.00	-4.68	-4.46	-4.20
banana	112.00	-4.58	0.53	-5.21	-4.83	-4.55	-4.37	-3.95
barley grain	127.00	-6.72	0.33	-7.07	-6.93	-6.78	-6.57	-6.25
cacao	56.00	-7.25	0.31	-7.66	-7.49	-7.15	-7.06	-6.90
cassava	115.00	-4.43	0.38	-4.80	-4.65	-4.44	-4.21	-4.05
coffee	67.00	-7.23	0.30	-7.60	-7.40	-7.25	-7.00	-6.94
cotton branch	44.00	-6.26	0.55	-6.96	-6.72	-6.16	-5.89	-5.63
dry bean	153.00	-6.71	0.33	-7.08	-6.91	-6.75	-6.50	-6.30
grape	47.00	-4.51	0.60	-5.21	-4.96	-4.49	-3.98	-3.76
maize (amilaceo)	139.00	-6.59	0.45	-7.07	-6.92	-6.69	-6.41	-5.96
maize (choclo)	115.00	-4.82	0.49	-5.43	-5.19	-4.81	-4.51	-4.13
maize (yellow hard)	139.00	-5.77	0.57	-6.43	-6.22	-5.91	-5.36	-4.84
onion	99.00	-4.10	0.62	-5.08	-4.55	-4.01	-3.59	-3.36
orange	121.00	-4.77	0.54	-5.40	-5.13	-4.77	-4.42	-4.22
potato	152.00	-4.41	0.42	-4.82	-4.64	-4.53	-4.20	-3.70
rice	79.00	-5.31	0.62	-6.23	-5.88	-5.05	-4.86	-4.66
tangerine	45.00	-4.62	0.58	-5.35	-4.94	-4.70	-4.49	-3.72
wheat	138.00	-6.62	0.44	-7.01	-6.92	-6.74	-6.48	-5.98

Table H.7: The Intermediate Input Bundle

Input Name	Unit Price (LCU / Kg)	Import Quantity (Tons)
Urea	1.140	1412085.28
Diammonium phosphate (DAP)	1.561	509983.82
Ammonium sulfate	0.657	465571.80
Potassium chloride	1.453	241862.40
Ammonium nitrate	1.236	204021.46
Potassium sulphate	2.201	117853.96
Magnesium and potassium sulphate	0.999	87836.48
Superphosphate	1.215	9609.98
Average	1.205	

Table H.8: Summary Statistics of International Prices, $p_{F,k}$ (local currency)

	Price
alfalfa	0.00
asparagus	6.38
avocado	4.54
banana	1.80
barley grain	0.93
cacao	8.76
cassava	0.00
coffee	11.11
cotton branch	1.26
dry bean	2.89
grape	6.53
maize (amilaceo)	0.00
maize (choclo)	0.00
maize (yellow hard)	0.75
onion	0.74
orange	2.00
potato	0.00
rice	0.37
tangerine	2.73
wheat	0.90

A value of 0.00 indicates the price for a crop that Peru does not trade internationally in raw form.

Table H.9: HS-10 correspondence to Crops (Prices)

Crop	HS10 Code	Description
alfalfa	1214100000	Alfalfa pellets
asparagus	0709200000	Asparagus, fresh or refrigerated
avocado	0804400000	Avocado, fresh or dried
banana	0803001100	Banana for cooking, fresh
banana	0803001200	Banana fresh
banana	0803001300	Musa Acuminata
banana	0803001900	Other banana, fresh or dried
banana	0803002000	Banana dried or fresh
banana	0803901100	Banana fresh
barley grain	1003001000	Barley for seed
barley grain	1003009000	Barley not for seed
cacao	1801001100	Cacao for seed
cacao	1801001900	Other cacao grain, whole or broken, raw
cacao	1801002000	Cacao grain, whole or broken, toasted
cassava	0714100000	Cassava roots fresh, refrigerated, frozen or dried
coffee	0901119000	Coffee not toasted, not decaffeinated, others
cotton branch	5201001000	Cotton - fiber length larger than 3492 mm
cotton branch	5201002000	Cotton - fiber length larger than 2857 mm
cotton branch	5201003000	Cotton - fiber length larger than 2222 mm
cotton branch	5201009000	Cotton - fiber length larger than 28.57 mm
dry bean	0713311000	Beans sp Vigna Mungo (L) Hepper or Vigna Radiata (L) Wilckez, for seed
dry bean	0713319000	Beans sp Vigna Mungo (L) Hepper or Vigna Radiata (L) Wilckez, not for seed
dry bean	0713321000	Beans sp Phaseoulus or Vigna Angularis, for seed
dry bean	0713329000	Beans sp Phaseoulus or Vigna Angularis, not for seed
dry bean	0713331100	Common beans (Phaseoulus Vulgaris), for seed
dry bean	0713331900	Other common beans for seed, except black
dry bean	0713339100	Black bean, not for seed
dry bean	0713339200	Canario bean, not for seed
dry bean	0713339900	Other common beans, not for seed
dry bean	0713359000	Other wild beans or caupi
dry bean	0713391000	Other beans for seed
dry bean	0713399100	Pallares (Phaseoulus Lunatus)
dry bean	0713399200	Castilla beans (Vigna Unguiculata)
dry bean	0713399900	Other beans
grape	0806100000	Fresh grapes
maize (yellow hard)	1005901100	Maize yellow hard
onion	0703100000	Onions fresh refrigerated
orange	0805100000	Oranges, fresh or dried
orange	0805202000	Tangelo (Citrus reticulata x citrus paradisis)
potato	0701100000	Potato for seed, fresh or refrigerated
potato	0701900000	Other potato, fresh or refrigerated
rice	1006300000	Rice, semi-white or wite, including polished or glazed
tangerine	0805201000	Tangerines, fresh or dried
wheat	1001101000	Hard wheat, for seed
wheat	1001109000	Hard wheat, not for seed
wheat	1001190000	Other hard wheat, not for seed
wheat	1001901000	Other wheat, for seed
wheat	1001902000	Other wheat
wheat	1001903000	Morcajo (wheat plus rye)

Table H.10: HS-10 correspondence to Crops (Quantities)

Crop	HS10 Code	Description
asparagus	0710801000	Asparagus frozen
asparagus	2005600000	Asparagus prepared or in preserve, not frozen
cacao	1802000000	Peels and other residuals of cacao
cacao	1803100000	Cacao butter, grease not removed
cacao	1803200000	Cacao butter, grease totally or partially removed
cacao	1804001100	Cacao butter
cacao	1804001200	Cacao butter
cacao	1804001300	Cacao butter
cacao	1804002000	Cacao fat and oil
cacao	1805000000	Powder cacao without additives
cacao	1806100000	Powder cacao with additives
cacao	1806201000	Other cacao preparations, without additives
cacao	1806209000	Other cacao preparations, in bars or blocks
cacao	1806310000	Chocolates and preparations, in blocks, tablets or bars, no filling
cacao	1806311000	Preparations in blocks, filled, without additives
cacao	1806319000	Other chocolates and preparations in blocks, tablets or bars, filled
cacao	1806320000	Other chocolates and preparations in blocks, tablets or bars, not filled
cacao	1806900000	Other Chocolates and food preprations that contain cacao
coffee	0901211000	Toasted coffee, not decaffeinated, grain
coffee	0901212000	Toasted coffee, not decaffeinated, ground
coffee	0901220000	Toasted coffee, decaffeinated
dry bean	2005510000	Beans shelled, prepared or in conserve, not frozen
grape	0806200000	Dry grapes, including raisins

H.2 Additional Estimation and Simulation Tables

Table H.11: More evidence on spatial differences in prices (local currency)

Crop	Region dummies				Urban dummy			
	Excluding income		Including income		Excluding income		Including income	
	F statistic	p value	F statistic	p value	Coefficient	s.e	Coefficient	s.e
alfalfa								
asparagus	49.160	0.000	9261.149	0.000	0.336	0.078	0.279	0.082
avocado	367.949	0.000	302.087	0.000	0.363	0.008	0.289	0.008
banana	225.344	0.000	208.876	0.000	0.123	0.003	0.091	0.003
barley grain	31709.330	0.000	1410.410	0.000	0.402	0.013	0.285	0.015
cacao	18.362	0.000	4182578.000	0.000	0.302	0.092	0.301	0.086
cassava	590.747	0.000	2347.150	0.000	0.246	0.004	0.189	0.005
coffee	102.337	0.000	522.045	0.000	0.194	0.022	0.184	0.022
cotton branch								
dry bean	151.542	0.000	125.030	0.000	0.190	0.005	0.153	0.005
grape	113848.500	0.000	307.192	0.000	-0.016	0.006	-0.034	0.006
maize (amilaceo)								
maize (choclo)	200.788	0.000	262.682	0.000	0.335	0.007	0.246	0.008
maize (yellow hard)								
onion	98.850	0.000	99.475	0.000	-0.016	0.003	-0.037	0.003
orange	364.937	0.000	329.860	0.000	0.128	0.004	0.083	0.005
potato	259.760	0.000	175.609	0.000	0.242	0.003	0.161	0.003
rice	168.931	0.000	166.006	0.000	0.054	0.002	0.014	0.002
tangerine	155.203	0.000	147.595	0.000	0.156	0.005	0.105	0.006
wheat	231.343	0.000	76.201	0.000	0.604	0.007	0.454	0.008

Unit values are not available for crops that are not consumed directly by households. In that case, the corresponding row is empty.

Table H.12: Estimation of Inverse Heterogeneity θ

	(1)
	$\log \tilde{y}_{ik} p_{ik}^{1/\gamma_k}$
$\log \eta_{ik}$	0.603***
	(0.0690)
Crop FE	Yes
Region FE	Yes
Region x $\frac{1-\gamma_k}{\gamma_k}$	Yes
Observations	453
Adjusted R^2	0.918

Standard errors in parentheses

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

Table H.13: Estimation of Transportation Models (Based on coffee price gaps)

(a) Specification in logs

	Constrained Model	Road Quality Model	Road Quality and Slope Model
log effective distance β_{dist}	0.348	0.473	0.473
	(0.115)	(0.069)	(0.153)
high quality λ_{hi}	1.000	1.000	1.000
	—	—	—
low quality λ_{lo}	1.000	11.548	11.548
	—	(5.330)	(7.957)
slope λ_s	0.000	0.000	0.000
	—	—	(34.020)
Intercept β_0	0.210	-0.104	-0.104
	(0.049)	(0.061)	(0.123)
N	332	332	332
R-squared	0.192	0.351	0.351

Bootstrapped standard errors in parentheses

(b) Specification in levels

	Constrained Model	Road Quality Model	Road Quality and Slope Model
log effective distance β_{dist}	0.528	0.633	0.633
	(0.186)	(0.155)	(0.272)
high quality λ_{hi}	1.000	1.000	1.000
	—	—	—
low quality λ_{lo}	1.000	11.459	11.459
	—	(10.466)	(14.471)
slope λ_s	0.000	0.000	0.000
	—	—	(14.318)
Intercept β_0	0.431	-0.028	-0.028
	(0.046)	(0.140)	(0.107)
N	332	332	332
R-squared	0.231	0.404	0.404

Bootstrapped standard errors in parentheses

Table H.14: Estimation of Transportation Models (Based on freight rates)

(a) Specification in logs

	Constrained Model	Road Quality Model	Road Quality and Slope Model
log effective distance β_{dist}	0.358	0.399	0.399
	(0.104)	(0.078)	(0.149)
high quality λ_{hi}	1.000	1.000	1.000
	—	—	—
low quality λ_{lo}	1.000	44.740	44.740
	—	(34.382)	(48.325)
slope λ_s	0.000	0.000	0.000
	—	—	(2.852)
Intercept β_0	-1.469	-2.010	-2.010
	(0.230)	(0.218)	(0.370)
N	46	46	46
R-squared	0.404	0.755	0.755

Bootstrapped standard errors in parentheses

(b) Specification in levels

	Constrained Model	Road Quality Model	Road Quality and Slope Model
log effective distance β_{dist}	0.308	0.919	0.919
	(0.094)	(0.172)	(0.182)
high quality λ_{hi}	1.000	1.000	1.000
	—	—	—
low quality λ_{lo}	1.000	36.544	36.544
	—	(14.512)	(7.803)
slope λ_s	0.000	0.000	0.000
	—	—	(26.084)
Intercept β_0	-1.253	-2.080	-2.080
	(0.239)	(0.399)	(0.585)
N	46	46	46
R-squared	0.301	0.876	0.876

Bootstrapped standard errors in parentheses

Table H.15: Summary Statistics of the Estimates of Land Quality, $\hat{A}_{i,k}$

	count	mean	std	10%	25%	50%	75%	90%
alfalfa	140	3.630388e+01	3.486289e+01	1.524197e+00	6.913883e+00	3.911770e+01	4.679384e+01	7.262951e+01
asparagus	22	1.170378e-08	5.079878e-24	1.170378e-08	1.170378e-08	1.170378e-08	1.170378e-08	1.170378e-08
avocado	127	3.482371e-06	4.251938e-21	3.482371e-06	3.482371e-06	3.482371e-06	3.482371e-06	3.482371e-06
banana	112	7.303294e-03	7.097508e-03	3.171708e-03	3.171708e-03	3.171708e-03	1.098977e-02	1.985968e-02
barley grain	127	4.645417e-04	5.046777e-04	4.290997e-05	1.128846e-04	3.453349e-04	5.582779e-04	1.371461e-03
cacao	56	1.055720e-07	7.813299e-08	3.013011e-08	3.013011e-08	9.015227e-08	1.778844e-07	2.051020e-07
cassava	115	7.418659e-03	6.773331e-03	1.396418e-03	3.495364e-03	3.495364e-03	1.245552e-02	1.852217e-02
coffee	67	8.631909e-08	7.419904e-08	7.199417e-09	2.608333e-08	5.192090e-08	1.488786e-07	1.988152e-07
cotton branch	44	9.533441e-06	6.854604e-21	9.533441e-06	9.533441e-06	9.533441e-06	9.533441e-06	9.533441e-06
dry bean	153	7.201240e-06	2.549444e-21	7.201240e-06	7.201240e-06	7.201240e-06	7.201240e-06	7.201240e-06
grape	47	6.186096e-08	0.000000e+00	6.186096e-08	6.186096e-08	6.186096e-08	6.186096e-08	6.186096e-08
maize (amilaceo)	139	1.017276e-04	2.584293e-19	1.017276e-04	1.017276e-04	1.017276e-04	1.017276e-04	1.017276e-04
maize (choclo)	115	1.407872e-04	3.266841e-19	1.407872e-04	1.407872e-04	1.407872e-04	1.407872e-04	1.407872e-04
maize (yellow hard)	139	8.816300e-04	8.774615e-04	3.613700e-05	1.632306e-04	5.756270e-04	1.525179e-03	2.196089e-03
onion	99	1.214701e-02	1.184831e-02	1.132381e-03	5.217986e-03	1.290757e-02	1.290757e-02	1.980601e-02
orange	121	8.307131e-04	6.878109e-04	6.570713e-05	4.450889e-04	4.450889e-04	1.257689e-03	1.828284e-03
potato	151	1.018773e-03	1.098937e-03	3.604862e-05	2.515808e-04	7.805389e-04	1.142663e-03	2.888733e-03
rice	79	6.841441e-04	6.383558e-04	2.098341e-04	2.353166e-04	2.353166e-04	1.252442e-03	1.700897e-03
tangerine	45	1.167522e-04	5.482267e-20	1.167522e-04	1.167522e-04	1.167522e-04	1.167522e-04	1.167522e-04
wheat	138	2.802876e-04	3.088465e-04	2.367452e-05	6.535354e-05	2.121918e-04	3.179492e-04	8.057018e-04

Table H.16: Instrumental Variable Estimation of Elasticity of Substitution, σ

	(1) OLS	(2) First stage	(3) GAEZ level	(4) First stage	(5) Freight instrument	(6) First stage	(7) GAEZ log
log unit value	0.144***		-1.385***		-1.616***		-1.804***
	(0.00417)		(0.0416)		(0.0413)		(0.0875)
GAEZ productivity		-0.0376***					
		(0.000498)					
logp_intl_inst				0.214***			
				(0.00269)			
log GAEZ productivity						-0.0261***	
						(0.000625)	
R-sq	0.415	0.520	0.339	0.660	0.207	0.517	0.169
First st. F stat	.		5705.1		6301.2		1742.0

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

Table H.17: Estimates of crop-specific preferences, a_k

crop	a_k
rice	18.543
wheat	12.780
maize (yellow hard)	11.202
potato	11.168
cotton branch	9.892
cacao	6.967
maize (amilaceo)	6.503
dry bean	6.408
cassava	3.242
onion	2.320
barley grain	2.151
banana	2.115
alfalfa	1.893
grape	1.807
maize (choclo)	1.756
orange	0.598
avocado	0.249
tangerine	0.207
coffee	0.100
asparagus	0.100

Table H.18: Estimates of international barriers, τ_k

crop	τ_k
alfalfa	
asparagus	2.395
avocado	3.298
banana	3.153
barley grain	1.000
cacao	1.007
cassava	
coffee	1.000
cotton branch	1.801
dry bean	1.016
grape	3.809
maize (amilaceo)	
maize (choclo)	
maize (yellow hard)	1.000
onion	1.000
orange	2.928
potato	
rice	2.286
tangerine	4.325
wheat	1.087

Blank spaces indicate crops that are not traded internationally.

Table H.19: Baseline consumption and revenue shares

(a) Consumption, $s_{i,k} \times 100$

Crop	Average	C. Variation.
potato	20.62	0.36
alfalfa	17.70	0.44
rice	14.73	0.46
maize (yellow hard)	11.30	0.39
wheat	6.63	0.40
cassava	5.89	0.48
banana	5.02	0.46
onion	3.36	0.44
maize (amilaceo)	3.29	0.36
maize (choclo)	2.34	0.41
barley grain	2.25	0.45
dry bean	1.99	0.35
cotton branch	1.89	0.42
orange	1.22	0.34
grape	0.76	0.36
cacao	0.49	0.51
tangerine	0.27	0.44
avocado	0.24	0.42
asparagus	0.02	0.41
coffee	0.01	0.50

(b) Revenue, $\pi_{i,k} \times 100$

Crop	Average	C. Variation.
potato	22.14	1.08
alfalfa	15.14	1.75
coffee	8.17	2.20
maize (amilaceo)	6.35	1.68
rice	6.17	2.27
maize (yellow hard)	5.62	1.29
wheat	5.51	1.36
cassava	4.81	2.83
onion	4.44	3.08
banana	3.97	2.62
avocado	3.28	3.93
maize (choclo)	3.16	2.94
barley grain	2.74	1.41
dry bean	1.81	1.91
grape	1.46	5.66
asparagus	1.19	7.23
cotton branch	1.00	4.50
cacao	0.95	3.31
orange	0.85	3.31
tangerine	0.20	5.18

Table H.20: Paving Roads: Average Change in Land Yields Across Crops

Crop	Average Yield Change
alfalfa	6.07
rice	5.03
orange	4.79
maize (yellow hard)	4.41
tangerine	4.32
dry bean	4.28
banana	3.99
maize (choclo)	3.97
onion	3.87
cassava	3.77
maize (amilaceo)	2.91
potato	2.84
avocado	2.60
cacao	2.41
wheat	2.34
cotton branch	2.16
barley grain	1.47
grape	1.07
coffee	0.89
asparagus	−0.42

Table H.21: Production patterns of top 25% regions with largest trade cost reductions (Road Building Policy)

Crop	Avg. $\pi_{i,k}$
potato	20.37
alfalfa	12.88
cassava	11.08
coffee	9.50
rice	7.10

Table H.22: Paving Roads: Contributions to Changes in Real Income of Top Winners and Losers

(a) Top 10% regions with largest real income gains

Crop	Avg. $\pi_{i,k}$	Avg. $\Delta p_{i,k}$	Avg. $\frac{1}{1-\beta_i}\pi_{i,k}\Delta p_{i,k}$	Avg. $b_i s_{i,k}\Delta p_{i,k}$
wheat	4.60	4.05	0.92	0.04
coffee	12.73	2.80	0.75	0.00
potato	14.60	0.88	0.73	0.03
cassava	9.14	1.39	0.61	0.01
alfalfa	13.73	1.68	0.60	0.03

(b) Top 10% regions with largest real income losses

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	38.97	−0.67	−0.38	−0.01
maize (yellow hard)	7.45	−1.00	−0.15	−0.01
banana	6.12	−0.90	−0.12	0.00
cassava	2.15	−0.72	−0.09	0.00
wheat	4.43	−0.59	−0.09	0.00

H.3 Additional Estimation and Simulation Figures

Figure H.1: Distribution of Agriculture Expenditure Share b_i across Regions

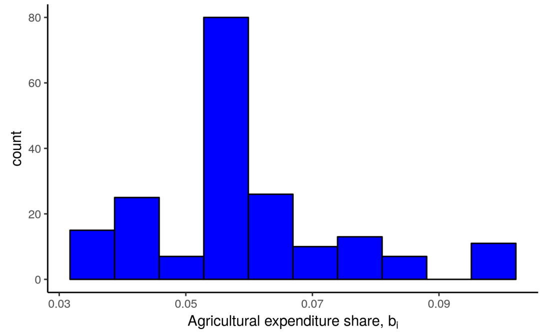


Figure H.4: Net exports as fraction of crop output (model and data)

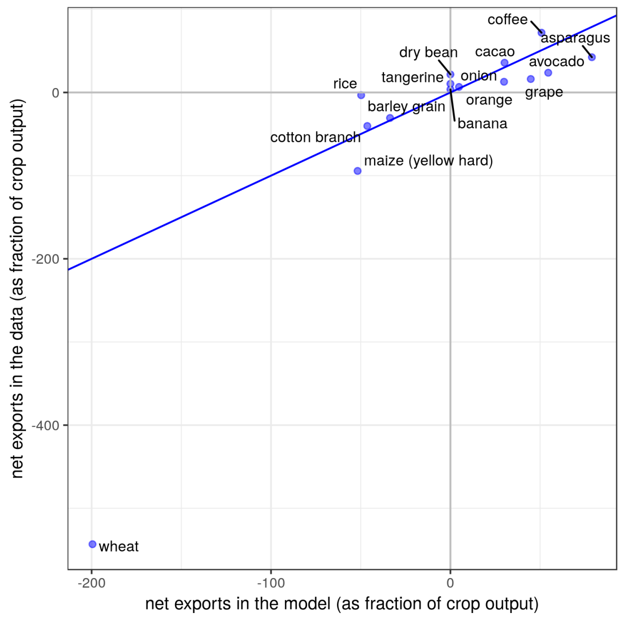
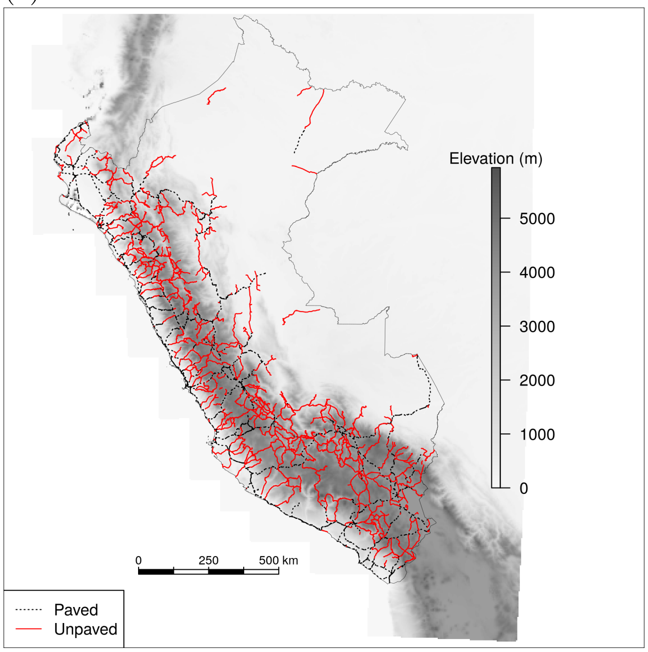
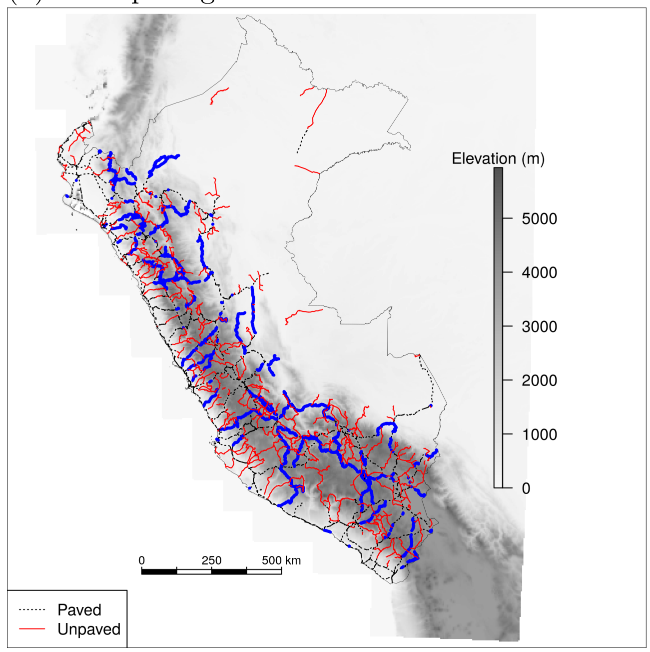


Figure H.2: Maps of the Infrastructure Policies

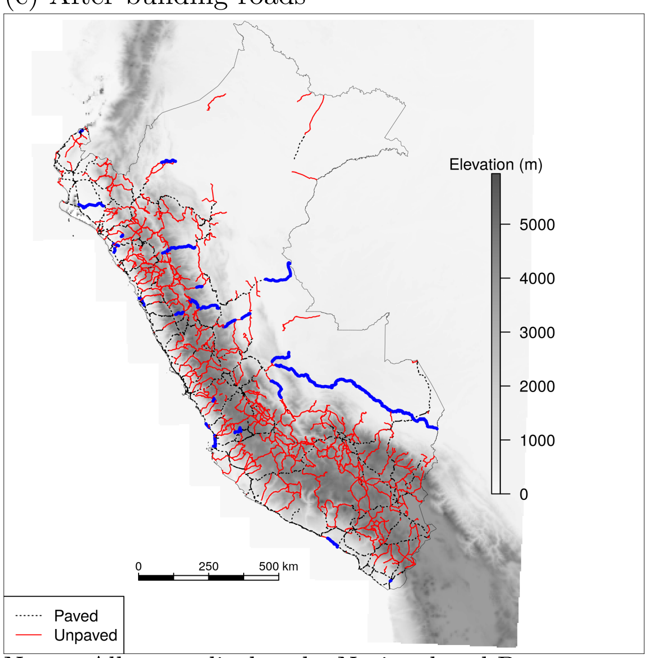
(a) Before



(b) After paving roads



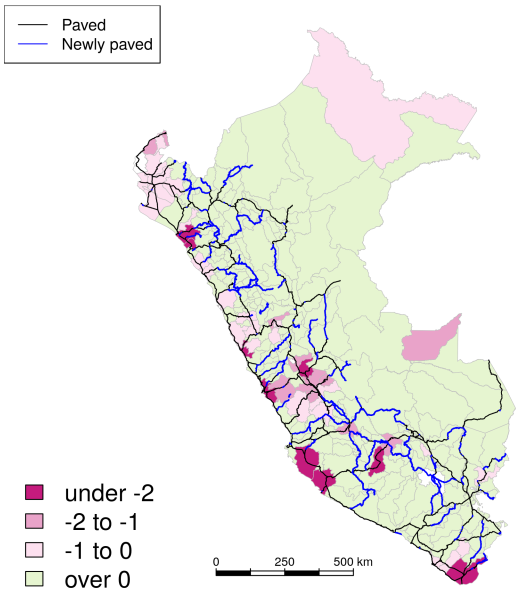
(c) After building roads



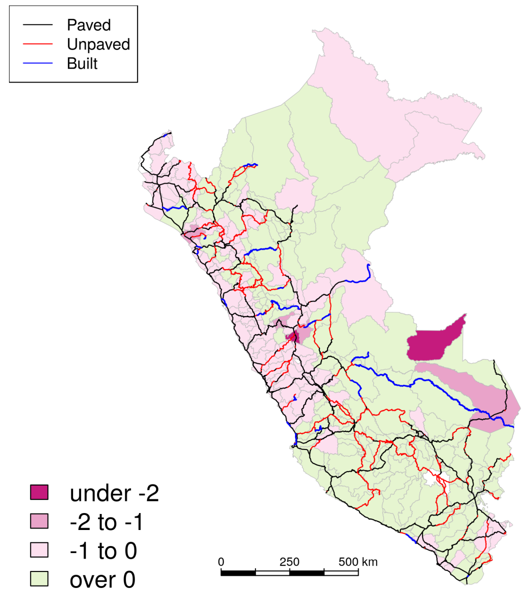
Notes: All maps display the National and Departmental road systems. They do not include the Rural road system.

Figure H.3: Spatial Distribution of Welfare Policy Responses

(a) Paving Highways in the National System



(b) Building New Roads



Notes: All maps display only the National road systems.

Figure H.6: Paving Roads: Counterfactual Change in Prices and Land Shares

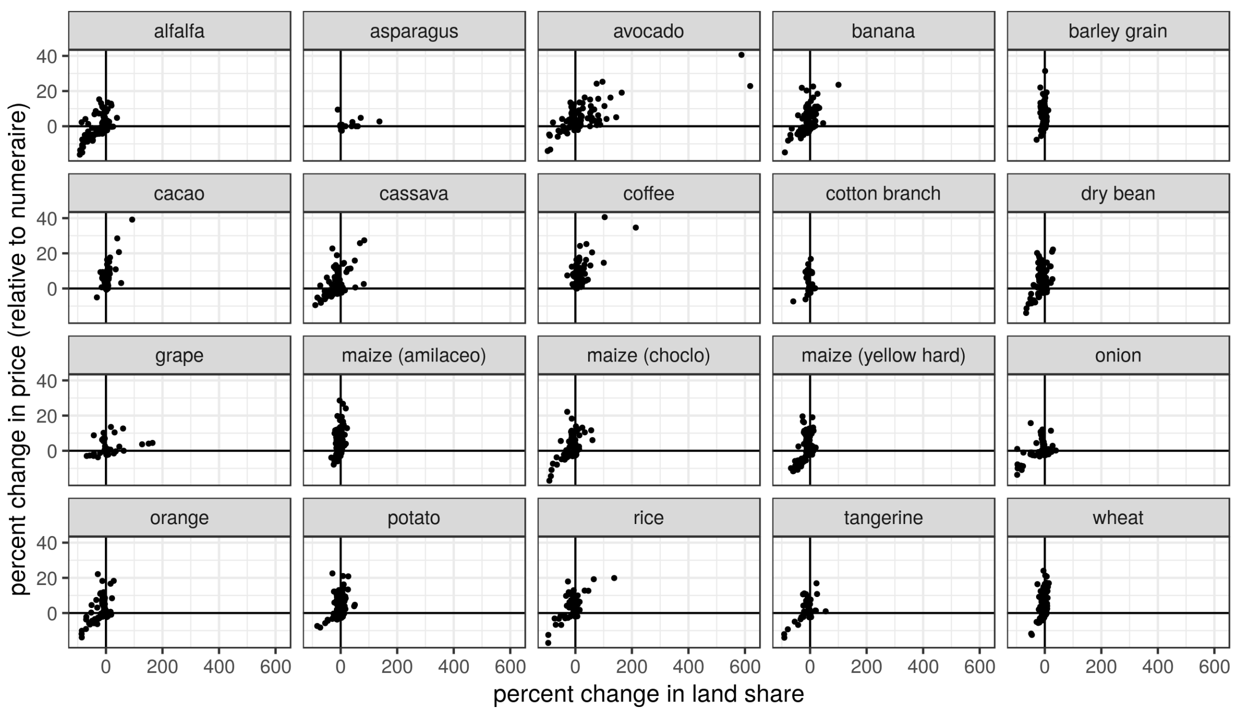
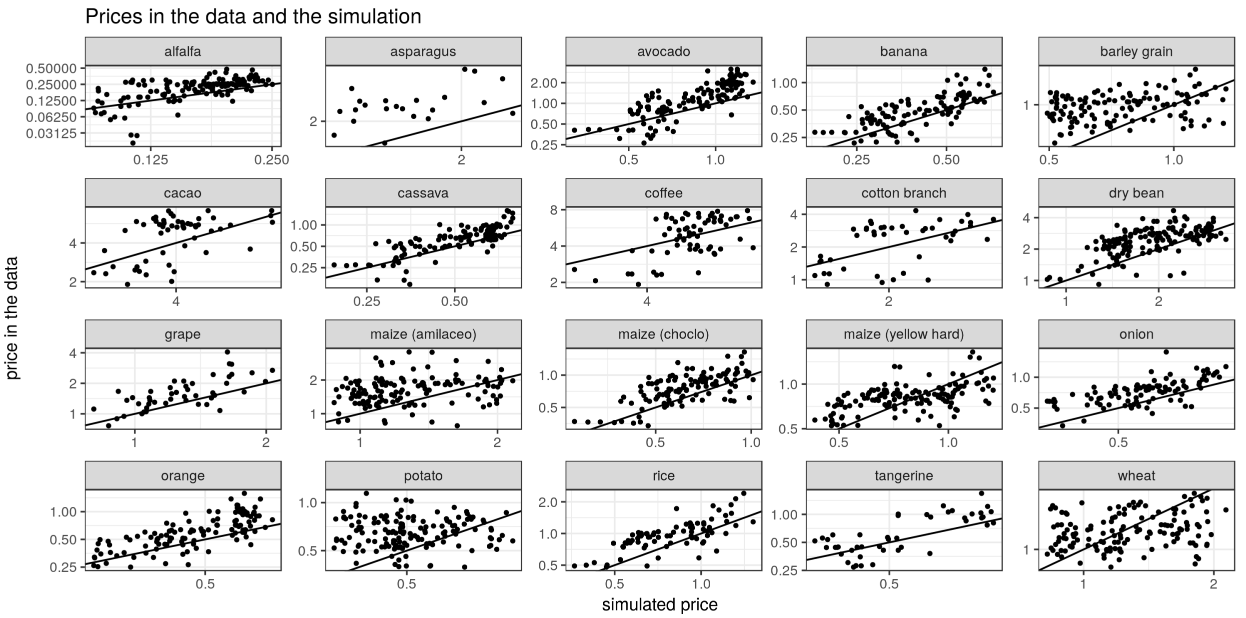


Figure H.5: Fitting Prices and Land Shares

(a) Prices in the data and the simulation



(b) Land shares in the data and the simulation

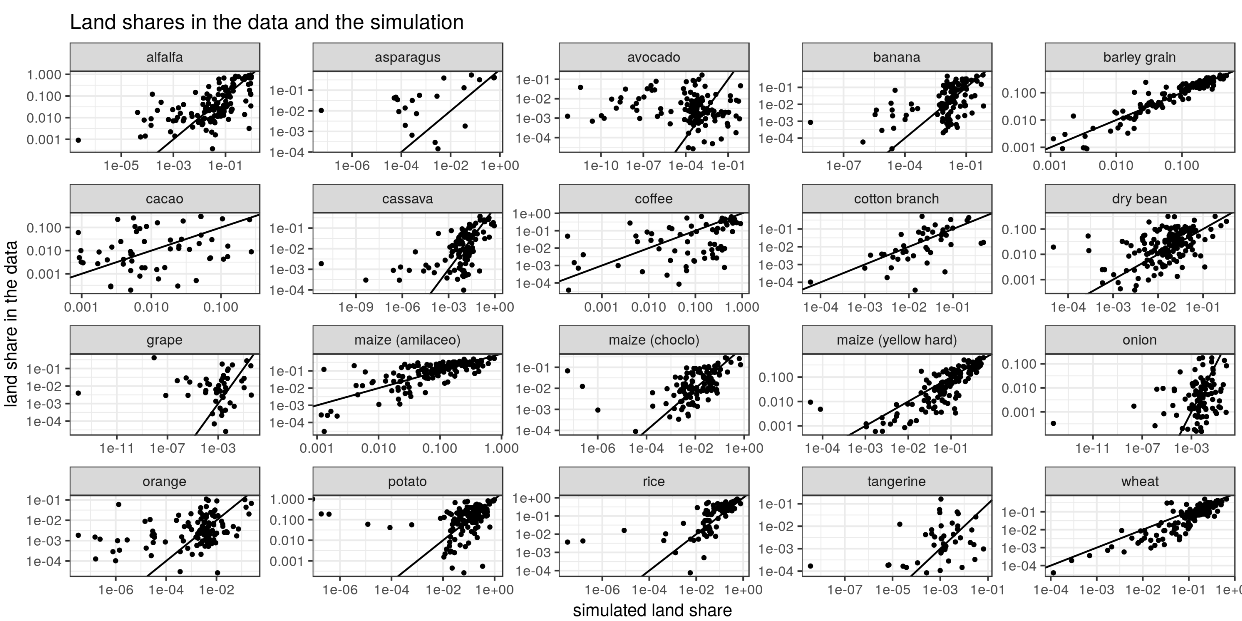


Figure H.7: Paving Roads: Distribution of Changes in Prices and Land Shares

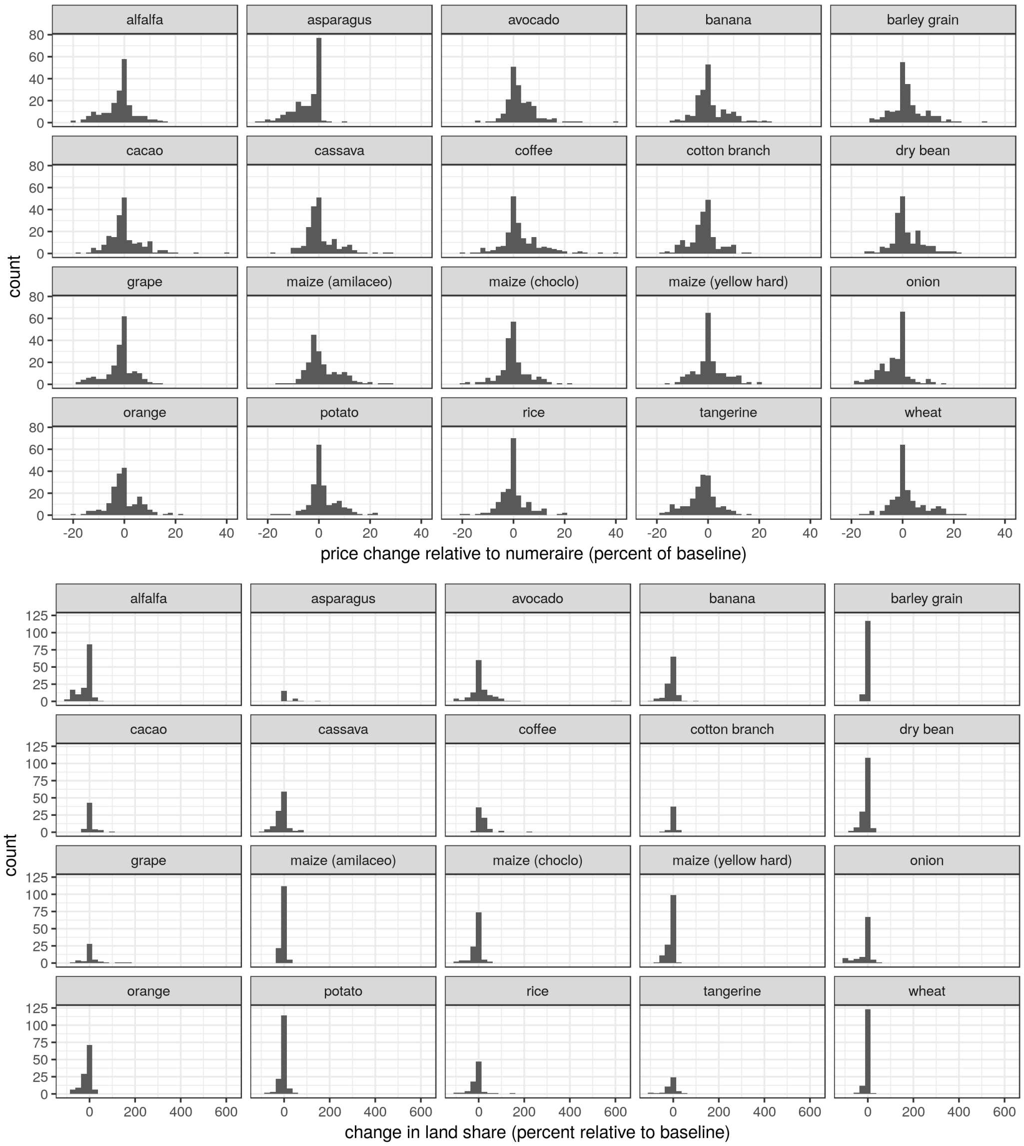


Figure H.8: Paving Roads: Net Export Growth

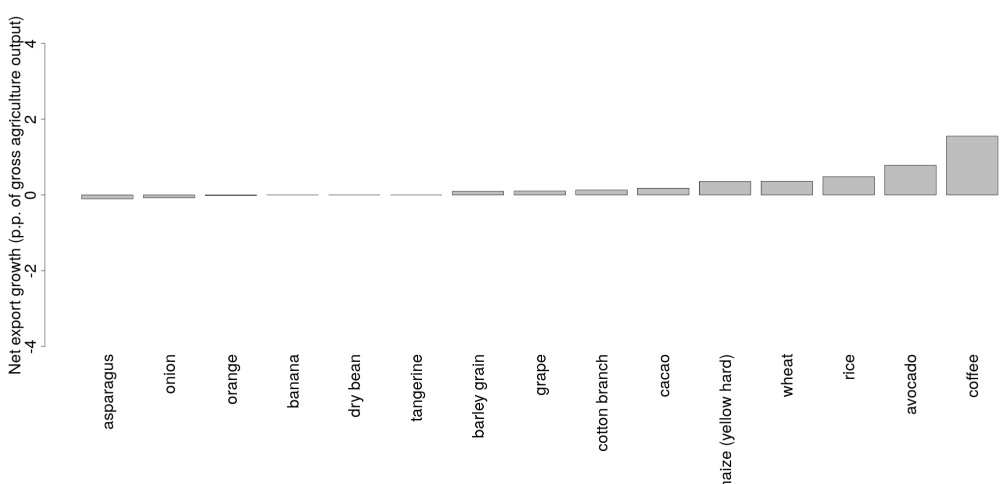
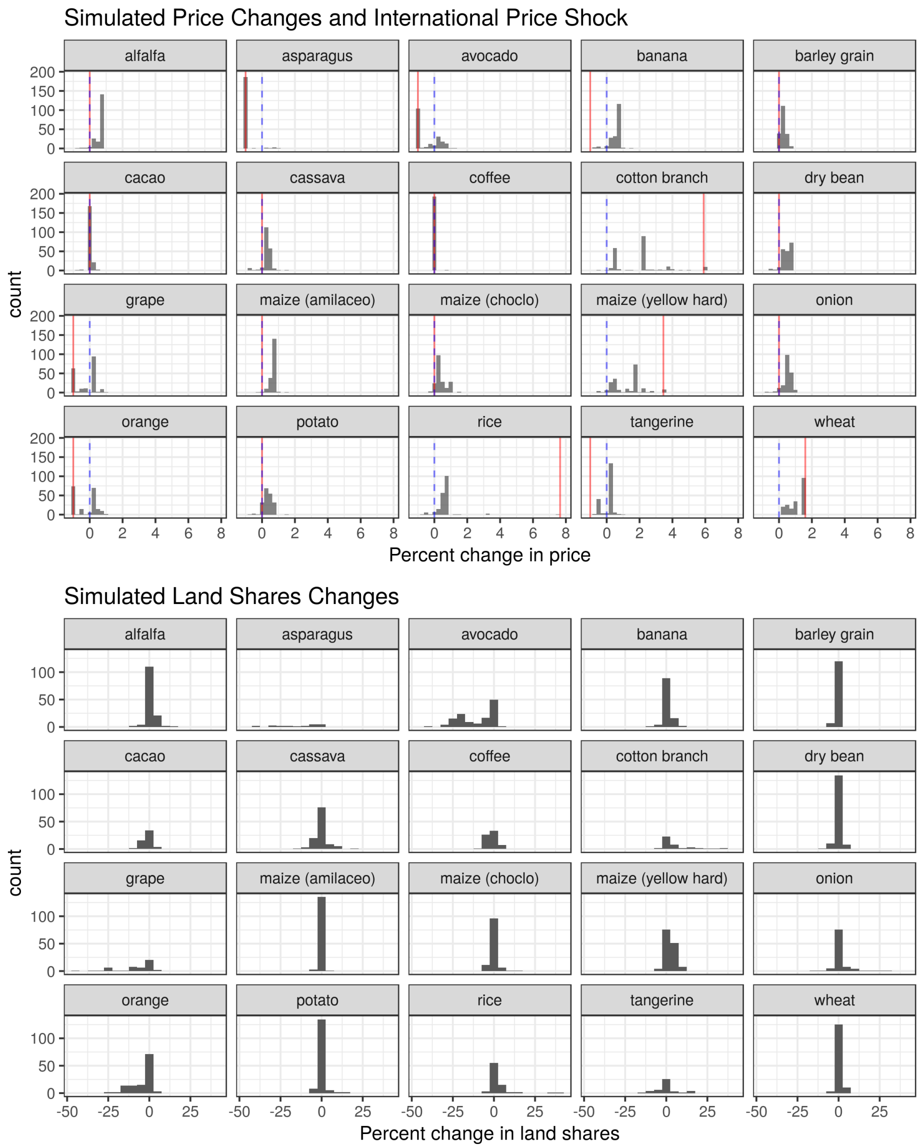


Figure H.9: International Price Shock: Distribution of Changes in Prices and Land Shares

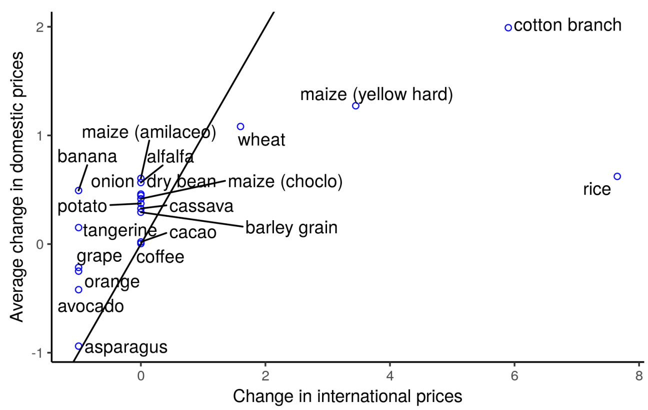


Notes. Red lines indicate the magnitude of the international shock.

Figure H.10: International Price Shock: Price Response and Cost of Trading to Ports

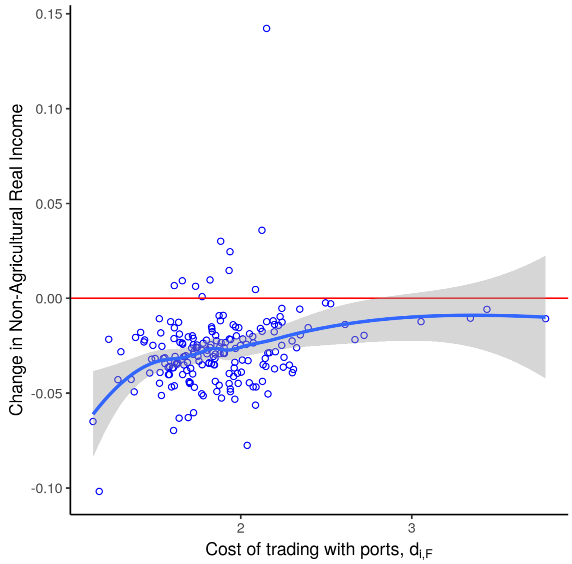
Notes. Red lines indicate the magnitude of the international shock.

Figure H.11: International Price Shock: Imperfect Average Pass-Through



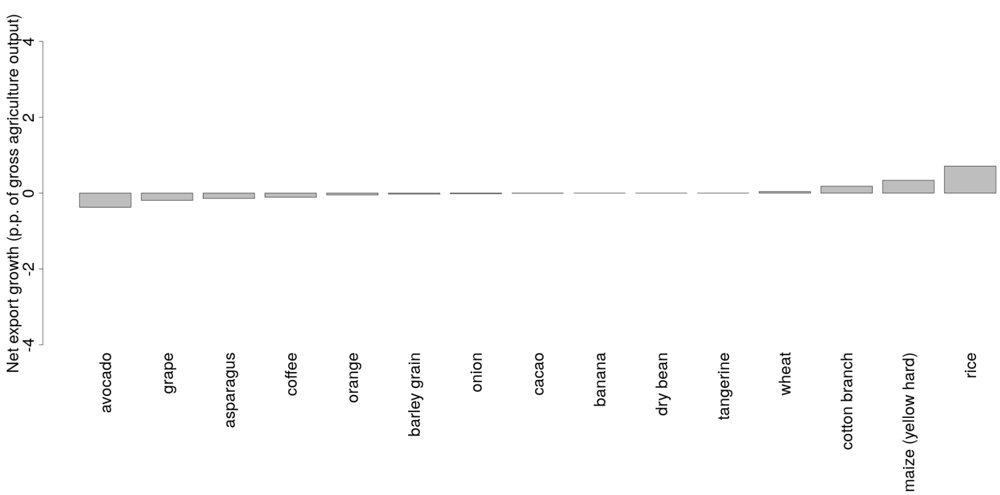
Notes: The figure displays the 45-degree line

Figure H.12: Non-Agricultural Real Income Change following an International Price Shock and Cost of trading with ROW



Notes: The Figure shows a locally weighted regression line of changes in non-agricultural real income on cost on trading with ports.

Figure H.13: International Price Shock: Net Export Growth



References

Asturias, J. (2018): "Endogenous Transportation Costs," .

Bardhan, P., and C. Udry (1999): *Development Microeconomics*. Oxford University Press.

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

Costinot, A., and D. Donaldson (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.

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.

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

Gleich, D. F. (2009): "Models and Algorithms for PageRank Sensitivity," Ph.D. thesis, Stanford University, Chapter 7 on MatlabBGL.

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

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

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

Ministerio de Agricultura y Riego del Peru (2008-2011): "Boletin Estadistico Mensual. EL AGRO EN CIFRAS," .

Ministerio de Transportes y Comunicaciones de Peru (2011): "D.S. N 012-2013-MTC," .

——— (2013): *Anuario Estadistico 2013*.

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.