Borradores de ECONOMÍA

Market access, agricultural productivity and selection into trade: evidence from Colombia

By: Margarita Gafaro
Heitor S. Pellegrina

No. 1050
2018

BANCO DE LA REPÚBLICA COLOMBIA

- Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá -

MARKET ACCESS, AGRICULTURAL PRODUCTIVITY AND SELECTION INTO TRADE: EVIDENCE FROM COLOMBIA*

Margarita Gafaro
Banco de la Republica de Colombia

Heitor S. Pellegrina
NYU Abu Dhabi

Abstract

We study the impact of the selection of farmers into trade on agricultural productivity using new data on the universe of farms in Colombia. To guide our analysis, we formulate a spatial economy model where better market access induces high skill farmers to switch from subsistence to cash crops that are traded in urban centers. We estimate reduced form effects of market access using distance to historical settlements as an instrument and calibrate our model according to these effects. Structural estimates indicate that the selection of farmers into trade have a large effect on agricultural productivity.

Keywords: Agricultural Trade, Spatial Economics, Selection into trade

JEL classification: F14, J43, N56, O13, Q12, Q17, R14

*We thank Andres Felipe Fajardo Ramirez for outstanding research assistance. We would like to thank participants at the NYUAD Economic Development Seminar, Banco de la Republica Seminar and TIGN-LACEA. We are also grateful to Camila Casas and Marcela Eslava for their comments. Email: mgafargo@banrep.gov.co, heitor.pellegrina@nyu.edu.

Acceso a mercados, productividad agrícola y selección: Evidencia para Colombia*

Margarita Gáfaro — *Banco de la República de Colombia*

Heitor S. Pellegrina — *NYU Abu Dhabi*

Resumen

En este documento estudiamos la elección que hacen los agricultores entre la producción de cultivos comerciales con altos costos fijos y la producción de cultivos de subsistencia con menores costos de producción. Formulamos un modelo de economía espacial en el que un mejor acceso al mercado induce a agricultores de mayor habilidad a pasar de la producción de cultivos de subsistencia a la producción de cultivos que se comercializan en los centros urbanos. Probamos las predicciones del modelo utilizando datos para Colombia que provienen del Tercer Censo Nacional Agropecuario. Explotamos la variación exógena en la ubicación de los asentamientos indígenas en el siglo XVI para estimar efectos de forma reducida del acceso al mercado sobre la decisión de producir cultivos comerciales. Utilizamos los resultados de estas estimaciones para calcular los efectos sobre la productividad agrícola de la existencia de altos costos fijos en la producción de cultivos comerciales. Los restados sugieren que las barreras a la entrada que imponen estos costos fijos en la producción de cultivos comerciales tienen un efecto importante sobre la productividad del sector agrícola en Colombia.

Palabras clave: Comercio en agricultura, Economí espacial, selección en comercio

ClasificaciónJEL: F14, J43, N56, O13, Q12, Q17, R14

* Agradecemos a Andrés Felipe Fajardo Ramírez por su excelente asistencia de investigación. También agradecemos a Camila Casas, Marcela Eslava y a los asistentes al seminario en desarrollo económico del Departamento de Economía de NYUAD, al seminario interno del Banco de la República y al seminario TIGN-LACEA por sus comentarios. Email: mgafargo@banrep.gov.co, heitor.pellegrina@nyu.edu.

1 Introduction

In developing countries, a large share of the population is employed in low productivity farms producing subsistence crops, and this accounts for most of the aggregate differences in labor productivity between countries (Caselli, 2004; Restuccia et al., 2008). Still, in some regions of developing countries, farmers trade their crops in regional and international markets and are highly productive. In this paper, we study the impact of the selection of farmers into trade on agricultural productivity, a mechanism that has been shown to be key in the manufacturing sector (Eaton et al., 2011; Pavcnik, 2002), but that has not been investigated in the agricultural context.[1]

To do so, we bring new data on the universe of farms in Colombia and, to guide our empirical analysis, develop a spatial economy model with selection of farmers into trade. In our model, as in Melitz (2003), an improvement in market access induces farmers with higher skills to switch from subsistence to cash crops that can be traded in urban centers, while simultaneously forcing farmers with moderate skills to leave subsistence and become rural workers. Our model serves two empirical purposes. First, we use it to inform our inspection of the data and the evaluation of the reduced form impact of market access on the selection of farmers into trade. Second, based on our reduced form estimates, we quantify the structural parameters of the model to measure the impact of the selection of farmers into trade on agricultural productivity.

The set up of our model consists of a country with rural localities that vary in terms of their proximity to urban centers where farmers can sell their crops. There is a rural population that chooses in which rural locality they live and whether to become farmers or rural workers based on heterogeneous farming skills as in Lucas Jr (1978). If they become farmers, they can either produce food with low fixed costs, or pay additional fixed costs to produce cash crops that can be sold in urban centers. The country has a comparative advantage in the production of the cash crop and the relative price of the cash crop is higher closer to urban centers. As such, the population has incentives to produce close to urban centers, but congestion forces due to the immobility of land leads to a spatial dispersion of production.

The spatial patterns of selection into trade that emerge from the model are intuitive. First, in localities on the outskirts of urban centers, all farmers produce the cash crop because the relative price of the cash crop is sufficiently above its fixed costs. Second, in localities farther from urban centers, high skill farmers producing the cash crop coexist with moderately skill farmers producing food for subsistence. In these localities, the relative price

[1] See Bernard et al. (2007) for a literature review on firm heterogeneity in international trade.

of the cash crop is not sufficient to compensate the additional fixed costs of the cash crop operation for all farmers. Third, in remote regions, the relative price of the cash crop is sufficiently low and all farmers have incentives to produce food.

The solution to the spatial equilibrium of the model delivers a tractable expression for the gains from trade. In particular, we break down the gains from moving from an economy where all rural localities are in autarky to a trading economy into three terms: (1) one that that captures the full gains from removing both the geographic trade costs and the extra fixed costs of producing the cash crop, (2) a loss function that captures only the losses from geographic trade costs, but not the extra fixed costs of producing the cash crop, and (3) an additional loss function that captures the extra costs of producing the cash crop and its interaction with the geographic trade costs. When we remove these additional fixed costs in the production of the cash crop, our expression collapses to a pure comparative advantage model like the one formulated in Fajgelbaum and Redding (2014) where geographic trade costs are the only source of friction in the economy.

With the theory in mind, we turn to the analysis of the data. First, we construct our measure of selection into trade using farm-level data on the market destination of the agricultural output.[2] Our baseline results use a broader definition of trading that considers both regional and international markets, but we also experiment with a series of alternative definitions based on narrower definitions of trading and based on the nature of crops being produced.[3] Second, we construct a measure of market access based on the travel distance of every rural locality to the nearest State capital, which is where the largest regional urban center is typically located. We also test alternative measures such as travel time and distance to ports. Importantly, Colombia is a large country with several medium-sized urban centers and sizable domestic trade costs due to its irregular geography, which provides us with large spatial variation in market access.

We start by showing that, in contrast with the strong positive correlation between firm size in terms of labor employment and selection into exporting found in the manufacturing sector (Arkolakis, 2010; Eaton et al., 2011), the unconditional relationship between farm size in terms of land use and selection into trade is weakly negative. We show, however, that when we look within rural localities where farmers face similar land rents, the relationship between selection into trade and farm size is strongly positive. We take this contrast between the unconditional correlation and the within municipality one as evidence that the spatial dimension of the economy is key for the selection into trade in the context of agriculture,

[2]For example, farmers respond whether they are selling each of their products to cooperatives, international markets, wholesalers, etc.

[3]For example, we choose from the list of crops those that are typically associated with trade such as coffee and sugarcane and those that can be used for consumption such as yam and corn.

since land is immobile and a key factor of production. To complement this analysis, we show similar patterns when we use investments in soil improvements as a *proxy* for the fixed costs of production.

After our initial inspection of the data, we evaluate the reduced form impact of market access on selection into trade and investments in soil improvements. To deal with reverse causality and the endogenous formation of urban centers, we propose an instrumental variable that uses the distance of modern urban centers to pre-Columbian settlements. Our identification is based on the argument that, despite dramatic changes in the Colombian agriculture due to the introduction of several new crops, the location of settlements before these transformations is still related with the location of modern urban centers, but it is unrelated to the economic activity generated by these new crops. To address the possibility of violation of the exclusion restriction, we introduce a rich set of controls for geography and institutional characteristics of municipalities. Our instrumental variable estimates show that increasing the distance to urban centers by 100 *km* decreases the share of farmers that select into trade and the share of farmers investing in soil improvements both by 7 percentage points. In addition, in the spirit of the papers evaluating the sources of comparative advantages with a reduced form approach (Levchenko, 2007; Nunn, 2007; Rajan and Zingales, 1998), we show a positive impact of market access on the specialization of a municipality in crops that are more likely to be traded.

To close our analysis, we combine moments from our stylized facts with our reduced form estimates to estimate a minimal version of our model that can provide estimates of the impact of selection into trade on agricultural productivity. We assume that farming skill distribution is log-normal, and use moments from the farm size distribution and the share of farmers selecting into trade to estimate the parameters associated with the costs and benefits of producing the cash crop. We then calibrate the decay in the cash crop premium to match our reduced form estimates of the impact of market access on selection into trade. Our structural estimates indicate that removing the additional fixed costs of producing the cash crop would increase the aggregate agricultural productivity up to 10%.

Related Literature. We contribute to growing research on agricultural trade and economic development (Costinot and Donaldson, 2014; Donaldson, 2018; Donaldson and Hornbeck, 2016; Fajgelbaum and Redding, 2014; Pellegrina, 2017; Sotelo, 2016; Tombe, 2015). Relative to this literature, we are the first to show the empirical relationships between farm size and selection into trade, as well as to introduce heterogeneity in farming skills into a spatial economy model. In doing so, our paper relates to a small set of papers formulating trade models with occupational choices (Antràs et al., 2006; Dhingra and Tenreyro, 2017; Kehoe et al., 2016). Closest to our analysis, Dhingra and Tenreyro (2017) study how the monopolistic power of traders affect the transmission of international price shocks to farmers. Unique to our model is the existence of an explicit spatial dimension in the economy, which allows us to rationalize the relationship between farm size and the selection into trade in the data.

Our paper also relates to vast literature on the determinants of agricultural productivity (Adamopoulos and Restuccia, 2014; Donovan, 2017,1; Foster and Rosenzweig, 1995, 2011; Gollin et al., 2013,1; Karlan et al., 2014; Krishna and Sheveleva, 2017; Lagakos and Waugh, 2013; Restuccia et al., 2008). Within this literature, our paper speaks directly to papers evaluating the relationship between farm size distribution and agricultural productivity (Adamopoulos and Restuccia, 2014; Adamopoulos et al., 2014; Foster and Rosenzweig, 2011). We contribute to these papers by bringing trade and showing how the spatial dimension influences the farm size distribution. On a similar line, Krishna and Sheveleva (2017) formulate a theoretical model to study how search costs and the lack of contract enforcement keeps farmers from producing cash crops, but their model also abstracts from the spatial dimension of the economy.

By proposing an instrumental variable for market access based on the proximity of modern urban centers to pre-Colombian settlements, we draw from a large literature that explores historical regional characteristics to evaluate the effects of transportation infrastructure (Duranton, 2015; Duranton et al., 2014; Duranton and Turner, 2012; Faber, 2014; Michaels, 2008; Morten and Oliveira, 2016). Similar to our instrumental variable approach, Duranton (2015) explores historical roads in Colombia to study the impact of current roads on trade. Our empirical analysis complements these papers by highlighting a specific mechanism through which market access affects agricultural productivity.

2 Model

2.1 Setup

We consider a small open economy that consists of rural localities indexed by i and urban centers arbitrarily distributed on a map. Rural localities produce agricultural goods that can be traded with urban centers. Urban centers host trading firms that sell agricultural goods to consumers across urban centers and the rest of the world. We assume that the price of agricultural goods are the same across urban centers and that farmers in rural localities always trade their goods with the nearest urban center. Let $i = 0$ be the rural locality on the outskirts of an urban center and i^F be the locality at the agricultural frontier. Each rural locality is endowed with an homogeneous mass of land L_i. We order rural localities in a continuum $i \in [0, i^I]$.

There are two agricultural goods indexed by k: a cash crop C and a food crop F. We consider the cash crop to be goods such as coffee and sugarcane, which are typically used as inputs in the food industry, or goods such as fresh flowers, which are mainly produced for international markets, and the food crop to be goods such as potatoes, yam, and cassava, which can be used for subsistence consumption. The price of these crops at urban centers is exogenous and given by $\{p_k^*\}_{k=\{C,F\}}$. There is an iceberg trade cost that increases with an exponential decay δ, so that for each unit of k that rural locality i sells to an urban center, only $exp(-\delta d_{0i})$ arrives, where d_{0i} is the distance from i to the nearest urban center.

There is a rural population of mass N that can only work in agriculture and each rural locality has a representative landowner. Individuals in this population choose in which locality i they live and whether to become a worker or a farmer based on their farming skill φ, which is drawn from a distribution $G(\varphi)$ over R^+. Individuals make these choices in two stages. In the first stage, they choose where to live based on an expected real income u_i before observing their farming skill φ. In the second stage, they learn about their farming skill φ and choose between becoming a rural worker or to set up a farm.

Next, we present the preferences and the production functions to solve for the equilibrium. We first solve the model for a local equilibrium in each rural locality i given exogenous prices $\{p_{ki}\}_{k=\{C,F\}}$ and a population mass N_i. We then solve for the spatial equilibrium where N_i is endogenous.

Preferences. Individuals and landlords spend their income locally, consume only the food crop and are risk-neutral. Landowners earn land rents r_i. An individual with a farming skill φ can earn wage w_i as a worker or profit $\pi_{ki}^*(\varphi)$ as a farmer. Let $m_i(\varphi)$ be the nominal income of individual φ, then his utility $u_i(\varphi)$ is

$$u_i(\varphi) = \frac{m_i(\varphi)}{p_{Fi}}, \tag{1}$$

and the utility of landowners is given by $u_i = r_i/p_{Fi}$.

Technology and occupational choices. An invidual φ in locality i can hire f_k workers to set up a farm and produce output $q_{ki}(\varphi)$ with

$$q_{ki}(\varphi) = a_{ki}\varphi^{1-\gamma}l_{ki}(\varphi)^{\gamma},$$

where $l_{ki}(\varphi)$ is the use of land, a_{ki} is an efficiency term, and γ is the span-of-control term.[4] The cash-crop C has a larger fixed cost ($f_C > f_F$). We think of these fixed costs as the

[4] We assume the land-intensity to be the same for both crops to abstract from Heckshler-Ohlin forces of specialization.

minimum land preparation, marketing or administrative costs required to establish a farm. Given this production function, a farmer chooses $l_{ki}(\varphi)$ to maximize

$$\pi^*_{ki}(\varphi) \equiv \max_{l_{ki}(\varphi)} \{p_{ki} a_{ki} \varphi^{1-\gamma} l_{ki}(\varphi)^{\gamma} - r_i l_{ki}(\varphi) - w_i f_k\},$$

which yields the demand for land

$$l^*_{ki}(\varphi) = \varphi \left(\frac{p_{ki} a_{ki} \gamma}{r_i}\right)^{\frac{1}{1-\gamma}}, \tag{2}$$

and the net profit

$$\pi^*_{ki}(\varphi) = \varphi \Gamma \left(\frac{p_{ki} a_{ki}}{r_i^{\gamma}}\right)^{\frac{1}{1-\gamma}} - w_i f_k, \tag{3}$$

where $\Gamma \equiv \gamma^{\frac{\gamma}{1-\gamma}} - \gamma^{\frac{1}{1-\gamma}}$. Based on their net profit $\pi_{ki}(\varphi)^*$ and opportunity cost w_i, individuals choose the occupation that maximizes their nominal income $m_i(\varphi)$. Since profits $(\pi^*_{ki}(\varphi))$ increase linearly with farming skills, in equilibrium there will be a cutoff skill above which individuals become farmers. In what follows, we refer to Φ_{ki} as the set of individuals that choose to become a farmer and produce crop k.

Comparative advantages. We allow the crop-specific efficiencies $\{a_{Ci}, a_{Fi}\}$ to vary across rural localities, but subject to the same relative efficiency $a = a_{Ci}/a_{Fi}$ as in Fajgelbaum et al. (2016). Rural localities have a comparative advantage in the production of the cash crop C, so that $a_{Ci} > a_{Fi}$.

2.2 Local equilibrium

Definition 1. (Local Equilibrium) Given $\{N_i,\ \{p_{ki}, a_{ki}, f_k\}_{k=\{C,F\}},\ \gamma\}$, a local equilibrium in rural locality i consists of labor and land demand $\{f_k, l^*_{ki}(\varphi)\}_{k=\{F,C\}}$, occupational choices $\{\Phi_{ki}\}_{k=\{F,C\}}$, and factor prices $\{w_i, r_i\}$ such that:

1. for any $k \in \{C, F\}$, the set of farmers Φ_{ki} satisfies

$$\Phi_{ki} = \{\varphi \in R^+ | \pi^*_{ki}(\varphi) \geq \max\{w_i, \pi^*_{Ci}(\varphi), \pi^*_{Fi}(\varphi)\}\}; \tag{4}$$

2. rural population maximizes utility

$$E(u_i(\varphi)) \leq u^*, \text{ if } N_i > 0; \tag{5}$$

3. land market clears

$$\sum_{k=\{F,C\}} N_{ki} \int_{\varphi \in \Phi_{ki}} l^*_{ki}(\varphi) g(\varphi | \varphi \in \Phi_{ki}) d\varphi = L_i; \tag{6}$$

4. population is fully employed

$$\sum_{k=\{F,C\}} N_{ki}(1+f_k) = N_i; \tag{7}$$

5. trade is balanced.

Next, we characterize three types of local equilibria. First, a local equilibrium where all farmers produce the cash crop C, which we call the integrated economy. Second, a local equilibrium where farms producing the food crop F and the cash crop C coexist in a rural locality, which we define as the partially integrated economy. Third, we characterize a local equilibrium where all farmers produce food and are indifferent between trading and living in subsistence, which we call the autarky economy.

2.2.1 Local equilibrium 1: Integrated economy (I)

In this local equilibrium, the rural locality is fully integrated with urban centers and all farmers produce the cash crop C. We present the main equations characterizing the equilibrium and provide a full derivation of the model in the appendix. We first solve for wages and occupational choices assuming the existence of this local equilibrium. We then present a proposition that defines when this local equilibrium holds.

We start by calculating the cutoff above which individuals choose to become farmers $\bar{\varphi}_i^I$. This cutoff is given by $\pi^*_{Ci}(\bar{\varphi}_i^I) = w_i$, the skill of the individual who is indifferent between being a farmer or a worker. Isolating $\bar{\varphi}_i^I$ in this equation gives

$$\bar{\varphi}_i^I = w_i(1+f_C) \Big/ \Gamma \rho_i \left(\frac{p_{Fi} a_{Fi}}{r_i^\gamma} \right)^{\frac{1}{1-\gamma}}, \tag{8}$$

where we defined a cash crop premium

$$\rho_i \equiv \left(\frac{p_{Ci} a_{Ci}}{p_{Fi} a_{Fi}} \right)^{\frac{1}{1-\gamma}}. \tag{9}$$

The term ρ_i will be central in our analysis. It captures the increase in gross profits that a farmer obtains by producing the cash crop instead of producing food.

Using the full employment condition in (7) and the fact that $N_{Ci} = N_i(1 - G(\bar{\varphi}_i^I))$, we obtain

$$(1 - G(\bar{\varphi}_i^I))(1 + f_C) = 1. \tag{10}$$

This equation provides the cutoff $(\bar{\varphi}_i^I)$ as an implicit function of f_C. From the land market clearing condition (6), we get

$$N_i \rho_i \left(\frac{p_{Fi} a_{Fi} \gamma}{r_i}\right)^{\frac{1}{1-\gamma}} (1 - G(\bar{\varphi}_i^I)) E(\varphi | \varphi > \varphi_i^I) = L_i, \tag{11}$$

Inserting the expression (8) and (10) into (11), we get

$$\frac{w_i N_i}{\alpha_i^I (1-\gamma)} = \frac{r_i L_i}{\gamma},$$

where

$$\alpha_i^I \equiv \frac{\bar{\varphi}_i^I}{E(\varphi | \varphi > \bar{\varphi}_i^I)}$$

is the share of payments to labor relative to profits. The term φ_i^I in the nominator is associated with the compensation of the farmer who is indifferent between becoming a worker, and the term in the denominator is associated with the average compensation of managers. Using the equation above we can write the agricultural GDP in i as $Y_i = w_i N_i / \alpha_i^I (1 - \gamma) = r_i N_i / \gamma$. Therefore, the local economy behaves as if it had a Cobb-Douglas production function with an endogenous factor share of labor given by α_i^I.

Substituting the land rent r_i in the equation above back into the cutoff in (8) gives

$$w_i = \alpha_i^I (1 - \gamma) p_{Fi} a_{Fi} (h_i^I)^{1-\gamma} \left(\frac{L_i}{N_i}\right)^{\gamma},$$

where we defined

$$h_i^I \equiv \frac{\rho_i E(\varphi | \varphi > \bar{\varphi}_i^I)}{1 + f_C}. \tag{12}$$

The term h_i^I is central in our analysis. The nominator gives the average productivity of farms adjusted by the cash crop premium. The denominator gives the average employment of labor (workers and managers) per farm. Therefore, this term captures the average productivity of labor in locality i.

Using the expression for wages, we can obtain the equation for expected real income

$$E(u_i^I) = (1 - \gamma) a_{Fi} \left(h_i^I\right)^{1-\gamma} \left(\frac{L_i}{N_i}\right)^{\gamma}. \tag{13}$$

The equation above has several intuitive properties. It decreases with the ratio of labor to land (N_i/L_i), with γ controlling the magnitude of this congestion effect. It increases with the average productivity of farmers (h_i^I) and with the productivity of land (a_{Fi}). With this equation, we conclude the characterization of the local equilibrium where the economy is fully integrated with urban centers. Below, we present a proposition that states when this equilibrium holds. The proof for all the propositions are in the appendix.

Proposition 1. *A rural locality i is an integrated economy if and only if the cash crop premium is larger than the ratio of the fixed costs plus the opportunity cost of farmers. That is, when $\rho_i \geq \Delta$, where $\Delta \equiv \left(\frac{1+f_C}{1+f_F}\right)$.*

The intuition for proposition 1 is simple: when the cash crop premium is large enough so that it covers the additional cost of setting up a cash crop farm given by Δ, which includes the fixed costs as well as its opportunity cost, then no farmer chooses to produce food. Next, we present the partially integrated economy, which will be the local equilibrium where the decision to become a cash crop producer depends on the farming skill φ and the market conditions in i.

2.2.2 Local equilibrium 2: Partially integrated economy (P)

In a partially integrated economy, farms producing the cash crop C coexist with farmers producing food F. The proposition below defines when this local equilibrium holds.

Proposition 2. *A rural locality i is a partially integrated economy if and only if $\Delta \geq \rho_i \geq 1$. If a rural locality i is a partially integrated equilibrium, then there is a partition of φ such that individuals with skills above $\tilde{\varphi}_i^P$ produce the cash crop, those with skills between $\tilde{\varphi}_i^P$ and $\bar{\varphi}_i^P$ produce food, and those with skills below $\bar{\varphi}_i^P$ become rural workers.*

In this local equilibrium, even though $\rho_i \geq 1$, only farmers with higher skills obtain enough gross profits to compensate the additional fixed costs of setting up a cash crop operation. Farmers with moderate skills do not produce the cash crop, but they still obtain enough gross profits to set up a farm that produces food.

The cutoff to become a farmer $\bar{\varphi}_i^P$ is obtained by equalizing $\pi_{Fi}^*(\bar{\varphi}_i^P) = w_i$, the skill that makes the farmer producing food indifferent to becoming a worker,

$$\bar{\varphi}_i^P = w_i(1+f_F) \Big/ \Gamma \left(\frac{p_{Fi}a_{Fi}}{r^\gamma}\right)^{\frac{1}{1-\gamma}}. \tag{14}$$

The cutoff that makes the farmer producing the cash crop C indifferent to producing food

(F) is obtained by equalizing $\pi^*_{Ci}(\tilde{\varphi}^P_i) = \pi^*_{Fi}(\tilde{\varphi}^P_i)$. This gives the following expression for $\bar{\varphi}^P_i$

$$\tilde{\varphi}^P_i = \bar{\varphi}^P_i \left(\frac{\Delta - 1}{\rho_i - 1}\right). \tag{15}$$

The full employment condition gives

$$(1 - G(\tilde{\varphi}^P_i))(1 + f_C) + (G(\tilde{\varphi}^P_i) - G(\bar{\varphi}^P_i))(1 + f_F) = 1. \tag{16}$$

Equations (14), (15) and (16) provide implicitly the solution for $\bar{\varphi}^P_i$ and $\tilde{\varphi}^P_i$ as a function of f_C, f_F and ρ_i.

The land market clearing gives

$$N_i \left(\frac{p_{Fi} a_{Fi} \gamma}{r_i}\right)^{\frac{1}{1-\gamma}} \Big((\rho_i - 1)(1 - G(\tilde{\varphi}^P_i))E(\varphi|\varphi > \tilde{\varphi}^P_i) + (1 - G(\bar{\varphi}^P_i))E(\varphi|\varphi > \bar{\varphi}^P_i)\Big) = L_i.$$

As in the integrated economy, we can manipulate the land market clearing to get

$$\frac{w_i N_i}{\alpha^P_i (1 - \gamma)} = \frac{r_i L_i}{\gamma},$$

where

$$\alpha^P_i \equiv \frac{\bar{\varphi}^P_i((1 - G(\tilde{\varphi}^P_i)(\Delta - 1) + (1 - G(\bar{\varphi}^P_i)))}{\Big((1 - G(\tilde{\varphi}^P_i))(\rho_i - 1)E(\varphi|\varphi > \tilde{\varphi}^P_i) + (1 - G(\bar{\varphi}^P_i))E(\varphi|\varphi > \bar{\varphi}^P_i)\Big)}.$$

Again, following the steps used for in the integrated economy, the expected real income can be expressed by

$$E(u^P_i) = (1 - \gamma) a_{Fi} \left(h^I_i\right)^{1-\gamma} \left(\frac{L_i}{N_i}\right)^{\gamma}. \tag{17}$$

where

$$h^P_i \equiv \frac{\Big((1 - G(\tilde{\varphi}^P_i))(\rho_i - 1)E(\varphi|\varphi > \tilde{\varphi}^P_i) + (1 - G(\bar{\varphi}^P_i))E(\varphi|\varphi > \bar{\varphi}^P_i)\Big)}{(1 + f_F)((1 - G(\tilde{\varphi}^P_i)(\Delta - 1) + (1 - G(\bar{\varphi}^P_i))}.$$

The nominator in h^P_i captures the total mass of skills employed by farmers unadjusted by the cash crop premium $((1 - G(\bar{\varphi}^P_i))E(\varphi|\varphi > \bar{\varphi}^P_i))$ and adds to this mass the additional efficiency coming from farmers who select into trade $((1 - G(\tilde{\varphi}^P_i))(\rho_i - 1)E(\varphi|\varphi > \tilde{\varphi}^P_i)$. Similarly, the denominator captures the employment of workers in all farms $(1 - G(\bar{\varphi}^P_i)(1 + f_F)$ and adds to this the additional employment of workers in cash crop farms $(1 + f_F)(1 - G(\tilde{\varphi}^P_i)(\Delta - 1)$. Therefore, as in h^I_i, the term h^P_i captures the average productivity of labor (workers and managers) in locality i.

In this equilibrium, when the cash crop premium ρ_i increases, the ratio of the cutoffs $\tilde{\varphi}^P_i / \bar{\varphi}^P_i$ shrinks (see equation 15). Therefore, the share of farmers producing food decreases relative to producers of the cash crop. In addition, we show in the appendix that there is a reduction in the overall share of individuals becoming farmers, with the surviving farms having a higher average productivity. This effect is analogous to the impact of selection into exporting on aggregate productivity highlighted by Melitz (2003). We summarize this result with the following proposition.

Proposition 3. *If a rural locality i is a partially integrated economy and N_i is exogenous, an increase in the cash crop premium ρ_i increases the share of farmers producing the cash crop C (decreases $\tilde{\varphi}_i^P$) and decreases the share of individuals who become farmers (increases $\bar{\varphi}_i^P$). As a consequence, the aggregate productivity of locality i increases due to a reallocation of factors of production towards the production of the cash crop.*

2.2.3 Local equilibrium 3: Autarky economy (A)

If the cash crop premium is lower than 1, then all farmers have incentives to produce food and stay in autarky. Following the same procedure that we used previously, the expected real income is given by

$$E(u_i^A) = (1-\gamma)a_{Fi}\left(h_i^A\right)^{1-\gamma}\left(\frac{L_i}{N_i}\right)^{\gamma}, \tag{18}$$

where

$$h_i^A \equiv \frac{(1-G(\bar{\varphi}_i^A))E(\varphi|\varphi > \bar{\varphi}_i^A)}{1+f_F}. \tag{19}$$

We close the analysis of the local equilibria with Figure 1, which shows the region of parameters associated with each local equilibrium. When the cash crop premium is sufficiently low, all farmers produce food and economies are in autarky. When the cash crop premium increases, we enter in the region of partially integrated economies. The share of farmers producing food shrinks until we reach the region where the conditions for a full integration with urban centers are met. Next, we turn to the analysis of the spatial equilibrium where N_i is endogenous.

2.3 Spatial equilibrium

Definition 2. (Spatial Equilibrium): Given $\{\{p_k^*\}_{k=\{C,F\}},\ N,\ L_i\}$, a spatial equilibrium consists of a labor and land demand $\{f_k, l_{ki}^*(\varphi)\}_{k=\{F,C\}}$, occupational choices $\{\Phi_{ki}\}_{k=\{F,C\}}$, and factor prices $\{w_i, r_i\}$ such that (i) a local equilibrium holds for all $i \in [0, i^F]$, (ii) prices $\{p_{ki}\}_{k=\{C,F\}}$ satisfies a non-arbitrage condition where $p_{ki}/p_{ki'} \leq exp(\delta d_{ii'})$ for $k \in \{C,F\}$ for all $i \in [0, i^F]$ and (iii) the expected real income u^* adjusts so that we have national full employment $\int_0^{i^F} N_i d_i = N$.

With this definition of spatial equilibrium, we can prove the existence of a partition of localities that separate the country in different regions according to the type of local equilibrium. Figure 2 shows the partition of regions that emerges in spatial equilibrium, which corresponds to the regions of skills in Figure 1.[5]

Proposition 4. *There is a unique spatial equilibrium and, if $i=0$ is an integrated economy and i^F is in autarky, there exists a partition of i such that all municipalities at a distance smaller than i^{IP} are integrated economies, a i^{PA} such that all municipalities in between i^{IP} and i^{PA} are partially integrated economies, and all municipalities in between i^{PA} and i^F are in autarky.*

Next, we characterize the aggregate productivity of the economy given in spatial equilibrium. To do so, we explore the fact that, since this is a perfectly competitive economy with no market imperfections, the first welfare theorem guarantees that the local equilibrium in each locality i maximizes the local GDP. This allow us to define the following expression for agricultural TFP

$$h_i \equiv (1-\gamma)a_{Fi}(\max\{h_i^I, h_i^P, h_i^A\})^{1-\gamma}. \tag{20}$$

This equation shows that the agricultural TFP in rural locality i is a combination of an exogenous overall quality of land a_{Fi} and an endogenous term related to the composition of farming skills.[6]

Using the equation 20 and the welfare equalization $E(u_i)=u^*$, isolate the total labor N_i in the expression for expected real income and substitute it back into the national full employment condition to obtain

$$N = \left(\int_0^{i^F} L_i \left(\frac{h_i}{u^*} \right)^{\frac{1}{\gamma}} di \right).$$

The equation above also gives

$$u^* = N^{-\gamma} \left(\int_0^{i^F} (h_i)^{\frac{1}{\gamma}} di \right)^{\gamma}. \tag{21}$$

[5] If a rural locality i is in autarky, any positive price p_{Fi}^A and p_{Ci}^A that satisfies $p_{ki}^A \le p_{ki'} exp(\delta d_{ii'})$ for any $i' \neq i$ is an equilibrium. For our figures, we just assume that the economies in autarky have the relative prices of i^{PA}, which would satisfy the non-arbitrage condition.

[6] Note that, with the definition for h_i, we can express the GDP in rural locality i as

$$Y_i = h_i N_i^{1-\gamma} L_i^{\gamma}$$

.

This expression shows that the expected real income of the economy is a geometric average of TFPs across rural localities.

2.4 Gains from trade

With the full characterization of the spatial equilibrium, we can now study the gains from trade in the model. We evaluate the gains from moving from an autarky economy to an economy where rural localities trade with urban centers. We focus on the effects of trade on expected real income u^*. Since the GDP of an economy is proportional to the expected real income, the impact of trade on the expected real income is proportional to the impact on the overall economy.

Following Fajgelbaum et al. (2016), we derive the expression for the gains from trade by considering two extremes. First, an autarky economy where geographic trade costs $\delta \to \infty$ are too large and no locality trades with urban centers. Second, an economy where there are no fixed costs of exporting ($f_C \to f_F$) and no geographic trade costs ($\delta = 0$). Let u^A be the expected real income in the economies in autarky and $\bar{u}$ the one for the economy with no barriers to trade. Then we have

$$u^A = h^A \left(\frac{A}{N}\right)^{\gamma},$$

where $h^A = h_i^A$ and $A \equiv \int_0^{i^I} (a_{Fi})^{\frac{1}{\gamma}} L_i d_i$ represents the total efficiency of the land. Since $f_F < f_C$, h^A captures the largest average employment of skill per farm in the absence of the cash crop premium. For $\bar{u}$, we have

$$\bar{u} = \rho_0 h^A \left(\frac{A}{N}\right)^{\gamma},$$

where ρ_0 is the highest cash crop premium. Note that in the economy with no barriers to trade, we have the largest cash crop premium ρ_0 as well as the largest proportion of individuals becoming farmers, which is given by the cutoff skill from the autarky economy.

Assuming that $L_i = L$ to simplify our expressions and using the expected real income u^* in (21), the gains from moving from autarky to trade can be written as

$$\frac{u^*}{u^A} = L(\delta, \Delta)\Omega(\delta)\frac{\bar{u}}{u^A}, \tag{22}$$

where

$$L(\delta,\Delta) \equiv \left(\frac{\int_0^{i^F} \left(h_i\right)^{\frac{1}{\gamma}} d_i}{\int_0^{i^F} \left(\rho_i h^A\right)^{\frac{1}{\gamma}} d_i} \right)^{\gamma},$$

and

$$\Omega(\delta) \equiv \left(\int_0^{i^F} \left(\frac{\rho_i}{\rho_0}\right)^{\frac{1}{\gamma}} d_i \right)^{\gamma}.$$

The function $L(\delta,\Delta)$ captures the efficiency loss related to the selection of farmers into trading, and Ω is the efficiency loss generated uniquely by the geographic trade cost. Both functions are between 0 and 1 and discount the gains from removing all the barriers to trade ($\bar{u}/u^A$). Note that, when $\Delta \rightarrow 1$, then $L(\delta,\Delta) \rightarrow 1$ and the gains from trade collapses to a formula where the only source of friction comes from trade costs as the one presented in Fajgelbaum et al. (2016). Therefore, the pure comparative advantage model is a special case of our framework.

3 Data and Agriculture in Colombia

With the theory in mind, we move to our empirical analysis. In this section, we briefly describe the agricultural sector in Colombia, present our data and provide an initial inspection of the empirical relationships between selection into trade, crop choice and production costs.

3.1 Agriculture and market access in Colombia

Agriculture is a key sector of the Colombian economy: it accounts for roughly 6% of its GDP and 14% of its exports. The country has 1.2 million km^2, being twice as large as France. Its tropical geographic location, along with a variety of altitudes, gives Colombia a wide variety of temperatures and soil conditions suitable for the production of many crops. Farmers produce both cash crops that are purchased by the local industry and international companies, as well food crops that are mostly consumed locally. International trade plays a key role in Colombian agriculture. The country is among the world's top exporters of coffee, African palm oil and bananas.

Transporting goods from rural areas to urban centers in Colombia is difficult and expensive. The civil conflict that took place in the country between the 1960s and 2016 hindered the development of rural roads for decades. In addition, as a result of the hilly topology, connectivity is difficult even in areas with access to roads. Figure 3 presents the road network in Colombia, as well as the driving time to the closest *departamento* capital via this network. Figure A1 in appendix E presents the driving distance from each municipality to Bogota, the capital of the country. It shows that it takes about 17 hours to drive from Bogota to Cartagena in the northern coast. The driving distance between these two cities is 1,042 kilometers (647.5 miles), which is equivalent to the driving distance between Atlanta and Miami, but in the U.S. it takes half the time to complete this route.

Perhaps as a consequence of this uneven geography, the urban system is relatively well spread across regions when compared to other countries (Duranton, 2016). It has several influential medium to large cities, but, within regions, the population is still highly concentrated in urban centers. For example, about 45% of the Colombian population lives in the *departamento* capitals. Besides hosting a large part of the consumers of agricultural goods, most of the cooperatives and agricultural trading firms with international connections are located in these capitals.

3.2 The data and the elements of the model

We organized the Colombian agricultural census of 2014 for our empirical analysis. We consider a rural locality in our model as the municipalities in the data. Therefore, we aggregate the data to the municipality level for part of our exercises. After cleaning, our final dataset contains about 860.000 farms in 1200 municipalities distributed across 33 *departamentos*, which is the broader political division of Colombia.

Using information on the destination of the agricultural production of each farmer, we define trading farmers as those who sell their products to cooperatives, intermediaries in international markets, wholesalers, food retailers, and the industry.[7] This measure takes into account both international and intraregional trade. We experiment with several alternative definitions in robustness checks. As a *proxy* for the fixed costs of production, we use information on whether farmers invest in soil improvements.[8] In table A1 of appendix E, we show that the differences across crops in the share of farms with machinery and in the share of managers with a high school degree are small, which led us to focus on the investments in soil improvements.[9]

For simplicity, our model considers two types of crops. In our empirical analysis, we

[7]This measure of trading has a correlation of -0.24 with the use of the output for farmers' own consumption.

[8]We consider as soil improvement the use of fertilizers and corrections for soil acidity versus other practices such as burning the land, religious practices or no application.

[9]The geography of Colombia limits the production of highly mechanized crops as observed in most regions of Brazil, the US or Canada. This likely explains why we find limited variation across farmers in the use of machinery.

look at the 15 most important crops in Colombia according to planted area.[10] Table 1 provides summary statistics and shows that there are large differences across crops in their commercial-orientation. Table 1 also shows that there is large variation across crops in the investments in soil improvements. Consistent with our model, crops with a larger proportion of the land with soil improvements are also the crops with a higher share of farmers trading internationally or intra-regionally.

Our measure of market access is the travel distance from each municipality to the nearest *departamento* capital, which corresponds to the distance to urban centers d_{0i} in our model. We calculate this distance using Google Maps. For some municipalities that are not connected to the road network, we calculate the Euclidian distance to the closest municipality connected to the network within a 50 kilometer radius, and add this euclidean distance to the travel distance of the connected municipality and the nearest state capital.[11] Figure 3 presents the road network in Colombia and our measure of distance to urban centers. Colder colors in the map represent the municipalities that are closer to urban centers.

In some of our estimates, we include control variables that characterize the institutional context of each municipality, these variables are provided by the CEDE-Data Center of Universidad de los Andes. We also include in our analysis information on the elevation, the ruggedness of the terrain, the availability of rivers, and the climatological conditions in each municipality. We calculated these variables combining municipality shape files provided by the Colombian statistical unit (DANE) with satellite imagery from ASTER GDEM and climate zoning maps provided by the National Institute of Hydrology and Meteorology (IDEAM). Appendix A.1 describes further details about these datasets and the calculations of the geographical variables.

3.3 Initial inspection of the data

Selection into trade and crop choice. Figure 5 shows that there are large differences across Colombian municipalities in the share of land with soil improvements and the share of farmers selecting into trading. We have 11% of municipalities where all farmers trade with urban markets, 15% of municipalities in autarky and in the vast majority of municipalities we have both farms selecting into trading and producing food, which corresponds to the partially integrated economies in our model.

[10] These crops account for 82% of the total planted area. In the appendix we present some results with the whole sample of crops and show that the patterns we find in the data do not depend on the choice of crops that we include in the sample.

[11] This is the case for 66 municipalities out of 1120.

The distributions presented in figure 5 come from variations in the proportion of crops being produced in each municipality, and also from variations within crops. We use the following equation to decompose these sources of variation

$$\frac{N_i^X}{N_i} = \sum_k \left(\lambda_{ik} - \bar{\lambda}_k\right) \frac{N_{ik}^X}{N_{ik}} + \sum_k \bar{\lambda}_k \frac{N_{ik}^X}{N_{ik}}, \tag{23}$$

where N_i^X is the total number of farmers trading in county i, N_i is the total number of farmers in i, N_{ik}^X is the total number of farmers selling to urban centers in county i and crop k, λ_{ik} is the share of farmers in county i producing crop k and $\bar{\lambda}_k$ is the average share of farmers producing crop k across municipalities.[12] The hollow bars in figure 5 represent the distribution of the share of farmers selecting into trading when we assume that there is no variation in the share of farmers producing different crops ($\lambda_{ik} = \bar{\lambda}_k$). In this case, the variance in the share of farmers trading reduces by one tenth. We obtain similar results when we look at the share of land with soil improvements. These patterns show that crop selection is closely associated with soil improvements and selection into trade.[13]

Selection into trade and farm size. Figure 6 shows a weak negative relationship between the share of farmers selecting into trade and farm size in terms of land use, and a strong negative relationship between soil investments and farm size.[14] This result contrasts with the typical positive correlation between firm size and selection into exporting in the manufacturing sector typically found by the trade literature. However, Figure 7 shows that, when we take into account the fact that the price of land is different across municipalities by dividing farm size and selection into trade according to the municipality average, the relationship between these variables become strongly positive. This suggests that the spatial dimension of the economy is key for analyzing the relationship between farm size and selection into trade.

[12]One potential concern is that, if the municipality has no production of a certain crop, we do not observe its share of farmers exporting. In unreported figures, we collapse the data into three groups according to the export-orientation and select only the municipalities in which we observe information on the share of farmers exporting. Conclusions are largely unaffected when we use this alternative procedure.

[13]It is important to highlight that there are also differences within crops. For example, there are different ways of producing vegetables. However, even within crops, we typically observe farmers producing higher quality types that are trade-oriented closer to cities. While our framework focuses on the broader differences that exist between crops. It also applies to the smaller differences that may exist within crops.

[14]In the appendix, we also show that the share of farmers selecting into trading is negatively associated with the share of farmers consuming part of their production.

4 Market Access and Selection into Trade

This section evaluates the reduced form effects of market access suggested by the model. First, we evaluate the direct impact of market access on the selection of farmers into trading. Second, we study the effect of market access on the patterns of crop specialization. Third, we discuss the potential alternative mechanisms explaining our results.

4.1 The impact of market access on the selection of farmers into trade

To begin our analysis, panel A in figure 8 shows the relationship between the share of farmers trading and the share of farmers investing in soil improvements. The share of farmers trading is around 70% in municipalities up to 200 km away from urban centers, but it falls to almost 30% for municipalities that are more than 400 km away. The figure also shows that this relationship is substantially weaker when we remove the variation in the share of farmers producing different crops using equation (23), which indicates that variation in the types of crops is an important driver of these relationships. Panel B shows the same qualitative patterns for the investments in soil improvements.

We estimate the effect of distance to urban centers with the following equation

$$y_i = \theta d_{0i} + \boldsymbol{x}_i' \boldsymbol{\lambda} + \eta_d + \epsilon_i \tag{24}$$

where y_i is the share of land with soil improvements or the share of farmers selling to urban centers and international markets in municipality i, d_{0i} is the travel distance from municipality i to the nearest urban center, $\boldsymbol{x}_i$ is a vector of covariates, η_d are *departamento* fixed effects, and ϵ_i is the unobserved component.[15] We are interested in the parameter θ, which captures the effect of travel distance on the patterns of production in i.

According to our theory, θ should be negative. Indeed, table 2 presents OLS estimates of equation (24) and shows a large negative effect: increasing the distance to urban centers by 100 km decreases by 4 percentage points the share of farmers trading. The coefficient increases with the inclusion of geography controls. When we look at the share of land with soil improvements, results are similar: the share of land with soil improvements decreases by 8.5 percentage points when there is an increase in distance of 100 km. Using the decomposition presented in equation (23), column 7 of table 2 shows the effect when we assume that all

[15] We use travel distance to the urban center instead of log of distance since we obtain better fit of the model in most specifications. In the appendix, we show that our results are robust to alternative specifications of distance.

municipalities have the same share of crops. The magnitude of the link between market access and the dependent variable decreases substantially in this case.

Note that differences in the overall productivity of a municipality can be included in ϵ_i as long as it affects equally the productivity of cash and food crop producers. According to our model, this would lead to a higher density in the municipality, but not in a larger share of farmers trading or investing in soil improvements. Given this remark, we address two potential sources of bias that could generate a negative θ for reasons that are unrelated to market access. First, it is possible that proximity to urban centers is correlated with agricultural productivity shocks that are specific to the cash crop. In this case, the effect of market access would be explained by higher productivity in cash crops rather than the access to the demand from urban centers. To minimize this source of bias, we include several geographic characteristics of the municipalities and fixed effects for nearest urban centers in the regression. Column 2 in table 2 shows that results are robust to the inclusion of these controls.

Second, there is a possibility of reverse causality where agricultural productivity in cash crops may affect market access. If municipalities with larger production of the cash crops have more resources to invest in transportation, we would observe better market access in these municipalities. This would lead us to overstate the impact of market access. However, if road investments are targeted to regions with more dynamic non-agricultural sectors, such as manufacturing or mining, or to regions with higher rates of poverty and lower agricultural productivity, we would understate its impact. In addition, it is possible that the location of urban centers themselves emerged as a result of the dynamism in the production of cash crops. In Colombia, 18 out of 33 state capitals were legally constituted between 1850 and 1950, which is also the period of the onset of agricultural exports in the country. Some of these cities emerged as a result of the expansion of the agricultural frontier during this period. This is the case of some cities in the current coffee growing region, which resulted from the adoption and expansion of coffee production in the western states in the late 1890s (Parson, 1961), and of the cities in the eastern plains, which emergence was triggered by the surplus from the agricultural activities developed in the late 1900s. To tackle these sources of reverse causality, we proceed with an instrumental variable approach.

Instrumental variable estimates. To deal with the possibility of a reverse in the estimation of equation 24, we explore variations in market access given by the location of pre-hispanic settlements in Colombia. In particular, we use the euclidean distance from each municipality to the closest indigenous settlement between 1510 and 1561 as an instrument for current distance to urban centers.[16]

Many of the *departamento* capitals that exist today in Colombia correspond to pre-hispanic settlement. Before the arrival of the Spaniard colonizers, Colombia was populated by a diversity of indigenous communities. Some of them were nomadic communities, like the ones that occupied the inter-Andean valleys and the Amazon region. Others established villages and developed more complex social and cultural structures. This is the case of the communities in the Atlantic coast and in the Andean region (Melo, 1995). As part of their strategy to maintain social and economic control of the colonies, Spaniards took over these indigenous settlements and established, in the same locations, the headquarters of the colonial government. After the independence most of these population nuclei experienced an accelerated process of growth. Still today, most of these cities host the national and local governments, and gather large shares of the country's population.

Our identification assumption is that the factors that determined the location of indigenous settlements are uncorrelated with unobserved productivity shocks that are specific to cash crops. This assumption is likely to be satisfied to the extent that pre-colonial agriculture focused on the production of native staple crops, and the cash crops that are important nowadays were introduced many decades (or centuries depending on the crop) after the establishment of these indigenous settlements. Upon the arrival of the colonizers, Colombian agriculture changed dramatically. New crops and new techniques of production were introduced by the Spaniards, during colonial times, and by local and international entrepreneurs, after the independence. For example, sugarcane was introduced from the Caribbean Islands in the late 1500s, and the Spaniards established large plantations in the southwestern states. Coffee arrived in the 1700s but only started being produced commercially in the late 1800, after the independence. Organized production of bananas started in the late nineteenth century triggered by the participation of international companies, and African palm oil arrived at the country in the twentieth century. Despite this revolution in agriculture, due to the path dependence, the current system of cities still presents clear traits of the earlier system of cities existent before the arrival of the Spaniards.

Figure 4 presents the location of indigenous settlements between 1510 and 1561 and the location the current state capital. Lighter colors represent an earlier year of establishment of the city. The map shows that there is a strong correlation between the location of indigenous settlements and the current location of *departamento* capitals.[17] We specify the following

[16] This information comes from historical records complied by Melo (1995) and available in the municipality data from CEDE.

[17] Figure A2 in appendix E provides further evidence of the strong positive relation between our instrument and our measure of market access.

first stage equation

$$d_{0i} = \theta^{FS} d_i^{IV} + \boldsymbol{x}_i' \boldsymbol{\lambda}^{FS} + \eta_d^{FS} + \epsilon_i^{FS} \tag{25}$$

where d_i^{IV} is distance to pre-Columbian settlements. Our identification assumption is that $E(\epsilon_i^{FS}\epsilon_i|\boldsymbol{x}_i', \eta_d) = 0$.

Columns 4 and 5 of table 2 show the IV estimates of equation (24) using the first stage above. Panel C shows that distance to indigenous settlements is a strong predictor of current market access and it passes standard weak instruments test. The IV estimates are larger than the OLS estimates, which could be a result of an attenuation bias due to classic measurement error of d_i or an indication of road infrastructure being targeted to regions that do not produce cash crops.

One concern with our instrument is that it may not satisfy the exclusion restriction if the distance to pre-Columbian cities affect the share of farmers exporting through channels that are unrelated to market access. We address this issue by including a rich set of controls for alternative causal channels. First, note that we already control for several geographic characteristics that could be correlated with distance to pre-Columbian cities in column 2 and 5 of table 2. Second, to account for the possibility that distance to pre-Columbian settlements is associated with the formation of institutions that are more conducive to the production of cash crops, we brought several variables related to the current institutional characteristics of rural localities: the intensity of the armed conflict in the municipality, the area planted with coca crops, the year of foundation of the municipality, which captures the fact that earlier municipalities might have more stable local institutions, the public investments in sewage and piped water, and the presence of the state agricultural bank[18], which captures the funding for agricultural activities. Our results are robust to the inclusion of these controls.

Robustness checks. In appendix E we present a number of robustness checks related to the choice of variables and crops in our main empirical analysis. Table A2 presents the estimation results when we do not restrict the sample to the 15 most important crops. As shown, the patterns presented in this section hold for the whole sample of crops.

We also evaluate if our results are sensitive to changes in the definition of trading and changes in the definition of market access. First, we use a narrow definition of selection into trade that considers only the products sold to international markets and cooperatives (see panel A of table A3). Second, we use a broader definition of exporting, which includes not only selling to cooperatives, international markets, wholesales and food retailers but also farm gate sales and sales in the municipality market (see panel C of table A3). In both

[18]Banco Agrario provides 90% of the credit to small farmers in Colombia.

cases, the negative relationship between distance to urban centers and selection into trade is robust to specification of variable defining trading.

Alternatively, we classify products between cash and food crops depending on crop-specific characteristics, rather than the behavior of the farmers producing each crop. In particular, we define as cash crops those that are typically demanded by the industry or by international markets. These include: rice, exportable banana[19], cocoa, coffee, sugarcane, yellow corn, African palm oil, pineapple, roses and carnations.[20] We classify the remaining crops (potato, yam, yucca, white corn, local banana, kidney beans, and plantain) as food crops. Results are presented in table A4 in appendix E. Again, distance to markets reduces the selection into the production of cash crops.

We also check if our results are robust to changes in the definition of market access. We calculate the travel distance to main ports in Colombia or any city that is identified as a custom city exporting goods. This provides a narrower definition of markets access based on the idea of integration with international markets. Results for our main specification using the broad and narrow definitions of market access, as well as our measure of soil improvements, are presented panels B, D, F of table A3 in appendix appendix E. The negative relationship between market access and selection into trading remains statistically significant.

4.2 The impact of market access on the patterns of crop specialization

An additional implication of the model is that distance to urban centers should affect the patterns of crop specialization. To test this prediction, we follow the approach adopted in empirical trade papers (Levchenko, 2007; Nunn, 2007; Rajan and Zingales, 1998) and estimate the following equation

$$\log(y_{ic}) = \theta^S s_c d_{0i} + \boldsymbol{f_c'} \boldsymbol{x}_i' \boldsymbol{\lambda} + s_c \boldsymbol{z}_i' \boldsymbol{\gamma} + \boldsymbol{x}_{ic}' \boldsymbol{\beta} + \eta_i + \eta_c + \epsilon_{ic}, \tag{26}$$

where y_{ic} represents total sales, use of land or number of farmers in municipality i and crop c, and s_c is the share of farmers trading crop c or the share of farmers with soil investments also in crop c. We are interested in the parameter θ^S. This parameter indicates whether municipalities that are closer to markets specialize in the production of crops with higher s_c.

[19]the data allows as to distinguish between the variety of bananas that are sold in international markets and the varieties that are consumed locally

[20]Our analysis with the 15 most important crops does not include fresh-cut flowers because the area planted with these crops is small relative to the area in other crops, however, given their importance in primary sector exports in Colombia, we include these crops here.

The second term, $\boldsymbol{f_c'}\boldsymbol{x_i'}$ interacts a vector of characteristics from the municipality with characteristics of crop c that are not captured by the model such as human capital and machinery. The third term interacts s_c with a vector of observable characteristics of the municipality. This term captures the correlation of market access with other observed characteristics that benefit the production of crop c. The fourth term $\boldsymbol{x'_{ic}}$ includes characteristics that are crop and municipality specific. In addition, by incorporating municipality and crop fixed effects, η_i and η_c, our specification controls for unobserved shocks that shift the dependent variable in any given crop across municipality, and in any given crop over municipalities. ϵ_{ic} is the unobserved component.

Table 3 shows OLS estimates of equation (26) for three different dependent variables: the number of farms, total sales and total area in each crop. Odd numbered columns show the baseline results without controls. The coefficients are negative and statistically significant, indicating that municipalities farther from urban centers are less specialized in trading crops. Even numbered columns control for all the controls specified above. Following the literature, we present the explanatory variables in standardized coefficients. An increase of one standard deviation in the distance decreases by 0.2 to 0.3 percent the specialization in trading crops. These estimates rise with the instrumental variable estimates based on the interaction $z_c d_{0i}^{IV}$. The first stage estimates of this interaction are presented in appendix E, table A5. In table A6 of the appendix we show that our results are sensitive to alternative definitions of trading behavior. Additionally, the appendix presents the effect of the interaction on crop specialization in the extensive margin.

4.3 Can alternative models explain the data?

We close this section by discussing how alternative models would rationalize the following two facts: (1) between municipalities, smaller farms close to urban centers are more likely to select into trade and specialize in the production of cash crops, (2) within municipalities, smaller farms produce food and are less likely to select into trade.

First, consider an agricultural trade model without heterogeneity in farming skills where the food crop is land-intensive relative to the cash crop.[21] In this model, municipalities farther from urban centers would have lower land rents and, therefore, a competitive advantage in the production of food. This would rationalize fact (1). However, within municipalities the price of factors are the same, and this model would generate a positive correlation between

[21] A model like Sotelo (2016) and Donaldson (2018) could be extended to provide predictions about farm size as shown in Pellegrina (2017). In this case, activities that are land-intensive (high γ) tend to have large farms if the price of factors are the same across crops and the factor share of intermediate inputs is not substantially different across crops.

farm size and the production of food, which goes agains fact (2).

Second, consider now a model where the cash crop is land intensive relative to food, which is consistent with measures of land intensity in the data.[22] In this case, municipalities close to urban centers would produce the cash crop only if trade costs were sufficiently high to compensate the disadvantage of the higher price of land. Under this condition, this model would explain fact (1) and (2), but it would have to bring additional assumptions to rationalize the relationship between farm size and investments in soil improvements.

Our model can rationalize both facts (1) and (2) without relying on any assumption on land intensity or the magnitude of trade costs, and also delivers additional predictions about costs of production that we find to be consistent with land investments in the data. We take this combination of facts as evidence of the existence of the mechanism incorporated into our model.

5 Impacts on Agricultural Productivity

This section explores the facts and reduced form estimates presented in the previous section to calibrate the structural parameters of our model. The goal is to recover the minimal set of structural parameters and moments that allow us to estimate the impact of selection into trade on agricultural productivity. We focus on the facts about the selection into trade, the heterogeneity in farming skills and the influence of the spatial dimension on these choices. We estimate $L(\delta, \Delta)$ and $\Omega(\delta)$ presented in the theoretical section, which are related to the impact of selection and geographic trade costs on agricultural productivity, and the impact of changing Δ and δ on agricultural productivity.

5.1 Recovering the structural parameters

We assume that the distribution $G(\varphi)$ is log-normal, which is consistent with the data on farm size distribution. We normalize the overall productivity of the economy to 1 by assuming that $\mu = 1$, $a_{Fi} = 1$ and $f_F = 1$, and set $\gamma = 0.3$, which is in the range of values obtained in the literature.[23] This leaves us with 4 parameters: the geographic trade cost (δ), the

[22]In the appendix we look at data on revenues per hectare in Colombia, which is a typical measure of land intensity. For example, Sotelo (2016) and Pellegrina (2017) use revenues per hectare as a moment condition for the estimation of the cost share of land in the production function. This measure for Colombia indicates that the production of cash crops is more intensive in land than the production of food

[23]We experiment with $\gamma = 0.2$ and $\gamma = 0.4$ and found very similar results. We also experiment with alternative values of f_F, which interacts with the shape of the distribution of skills, and also found very similar results.

maximum cash-crop premium (ρ_0), the dispersion of skills (σ) and the additional fixed costs of cash crops (Δ).

We use three moments from the data to calibrate Δ, ρ_0 and σ. We construct these moments using data from the counties that are below 25 km from any urban center and assume that they face ρ_0.[24] We use the share of farmers producing cash crop relative to total farmers

$$m_1(\Delta,\delta,\rho_0) = \left(m_0^1 - \frac{1-G(\tilde{\varphi}_0^P)}{1-G(\bar{\varphi}_0^P)} \right),$$

where m_0^1 is the median share of farmers selecting into trade. Second, we use

$$m_2(\Delta,\delta,\rho_0) = \left(m_0^2 - \rho_0 \frac{E(\tilde{\varphi}_0^P|\varphi > \tilde{\varphi}_0^P)}{E(\tilde{\varphi}_0^P|\tilde{\varphi}_0^P > \varphi > \bar{\varphi}_0^P)} \right),$$

where m_0^2 is the median ratio of average farm size among cash crop producers relative to food producers. We derive this moment by manipulating the expression for land use (2). Third, we use

$$m_3(\Delta,\delta,\rho_0) = \left(m_0^3 - \rho_0^2 \frac{V(\tilde{\varphi}_0^P|\varphi > \tilde{\varphi}_0^P)}{V(\tilde{\varphi}_0^P|\tilde{\varphi}_0^P > \varphi > \bar{\varphi}_0^P)} \right),$$

where m_0^2 is the median ratio of the variance of farm size among cash crop producers relative to food producers.[25]

We therefore have 3 moments and 3 parameters. Due to the non-linearity of these moments, we were not able to provide a formal proof of the identification, but the calibration is stable across different initial guesses. We estimate the parameters according to three definitions of selection into trade: our baseline measure, and the broader and narrow measures described in the robustness check section. For the baseline measure, we find Δ to be 1.18, which indicates that the fixed costs of producing the cash crop is 18% higher than the costs of producing food. For ρ_0, we find a cash crop premium 11%. For the dispersion in farming skills σ, we find a value of 0.93. By bootstrapping, we find small standard errors for Δ, $\rho 0$ and σ.

We then calibrate δ in the model to match the reduced form impact of market access on the share of farmers producing cash crop. We match this moment according to column (4) in Table (2). We find a δ of 0.08 for every 1000 km. With this parameter, a locality that is 200 km from an urban center has a cash crop premium of 1.09, and in a locality that is 400 km from an urban center the cash crop premium of 1.08.

[24]In addition, we drop municipalities with less than 50 farmers to avoid outliers.

[25]We use the medians because outliers in these ratios had a large effect on the ratios.

5.2 Results

Using the parameters estimated above, we simulate an economy using the actual distances and share of land use in each municipality as observed in the data. We compute the real income in three scenarios: (1) no fixed costs for the production of the cash crop ($\Delta = 1$), (2) no trade costs or fixed costs for exporting ($\Delta = 1$ and $\delta = 0$) and the baseline economy. Using these estimates, we find that $L(\Delta, \delta) = 0.95$ and $\Omega(\delta) = 0.99$ (see table 5) using the baseline measure. For the broader measure, the selection mechanism goes to 0.91, indicating the selection into trade with regional markets is more relevant than the barriers to international markets. In general, the isolated effect of the decay, which is captured by $\Omega(\delta)$ is small. Importantly, we note that the moment that we use to estimate δ gives us the decay in the cash crop premium, but it does not represent the overall decrease in geographic trade costs. In addition, table 6 shows the productivity gains from reducing the barriers to trade to zero. It suggests that the scope for gains from trade are substantially larger with reductions in fixed costs of producing the cash crop relative to geographic trade costs.

6 Conclusion

Recently, there has been growing research on agricultural trade and economic development (Donaldson, 2018; Donaldson and Hornbeck, 2016; Fajgelbaum and Redding, 2014; Pellegrina, 2017; Sotelo, 2016; Tombe, 2015). Here, we studied the impact of the selection of farmers into trade on agricultural productivity, a mechanism that has been shown to be key for productivity in the manufacturing sector, but that has not been explored in the agricultural context. We presented new facts about farm size distribution and selection into trade using a rich dataset containing the universe of farms in Colombia, and showed that the spatial dimension of the economy is key for our understanding of the relationship between farm size and agricultural trade. We developed a spatial economy model that accounts for the facts that we presented. We calibrated our model according to the reduced form impact of market access on farmers' selection into trade and find substantial productivity gains from removing the costs of selection into trade with regional markets. Future work may extend our model to study specific factors generating the costs of selection into trade, for example, the costs of search associated with finding buyers or the imperfections in credit markets.

References

Adamopoulos, T. and Restuccia, D. (2014). The size distribution of farms and international productivity differences. *The American Economic Review*, 104(6):pp. 1667–1697.

Adamopoulos, T., Restuccia, D., et al. (2014). Land reform and productivity: A quantitative analysis with micro data. *Unpublished manuscript, University of Toronto, December*, pages 1667–97.

Antràs, P., Garicano, L., and Rossi-Hansberg, E. (2006). Offshoring in a knowledge economy. *The Quarterly Journal of Economics*, 121(1):31–77.

Arkolakis, C. (2010). Market penetration costs and the new consumers margin in international trade. *Journal of political economy*, 118(6):1151–1199.

Bernard, A. B., Jensen, J. B., Redding, S. J., and Schott, P. K. (2007). Firms in international trade. *Journal of Economic perspectives*, 21(3):105–130.

Caselli, F. (2004). Accounting for cross-country income differences. Working Paper 10828, National Bureau of Economic Research.

Costinot, A. and Donaldson, D. (2014). How large are the gains from economic integration: Theory and evidence from u.s. agriculture, 1880-1997. *Working paper*.

Dhingra, S. and Tenreyro, S. (2017). Piggy-back exporting, intermediation, and the distributional gains from trade in agricultural markets.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A market access approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Donovan, K. (2017). Agricultural risk, intermediate inputs, and cross-country productivity differences. *Working Paper*.

Duranton, G. (2015). Roads and trade in colombia. *Economics of Transportation*, 4(1-2):16–36.

Duranton, G. (2016). Determinants of city growth in colombia. *Papers in Regional Science*, 95(1):101–131. PIRS-FA-2015-1463.R1.

Duranton, G., Morrow, P. M., and Turner, M. A. (2014). Roads and trade: Evidence from the us. *Review of Economic Studies*, 81(2):681–724.

Duranton, G. and Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, 79(4):1407–1440.

Eaton, J., Kortum, S., and Kramarz, F. (2011). An anatomy of international trade: Evidence from french firms. *Econometrica*, 79(5):1453–1498.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3):1046–1070.

Fajgelbaum, P. and Redding, S. J. (2014). External integration, structural transformation and economic development: Evidence from argentina 1870-1914. Technical report, National Bureau of Economic Research.

Fajgelbaum, P. D. et al. (2016). Internal geography, international trade, and regional specialization. *American Economic Journal: Microeconomics*, 8(1):24–56.

Foster, A. D. and Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6):1176–1209.

Foster, A. D. and Rosenzweig, M. R. (2011). Are indian farms too small? mechanization, agency costs, and farm efficiency. *Econojmic Growth Center, Yale University New Haven CT*.

Gollin, D., Lagakos, D., and Waugh, M. E. (2013). The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2):939–993.

Karlan, D., Osei, R., Osei-Akoto, I., and Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints *. *The Quarterly Journal of Economics*, 129(2):597–652.

Kehoe, T. J., Pujolas, P. S., and Ruhl, K. J. (2016). The opportunity costs of entrepreneurs in international trade. *Economics Letters*, 146:1–3.

Krishna, K. and Sheveleva, Y. (2017). Wheat or strawberries? intermediated trade with limited contracting. *American Economic Journal: Microeconomics*, 9(3):28–62.

Lagakos, D. and Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, 103(2):948–80.

Levchenko, A. A. (2007). Institutional quality and international trade. *The Review of Economic Studies*, 74(3):791–819.

Lucas Jr, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, pages 508–523.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Melo, J. O. (1995). *Historia de Colombia El Establecimiento de la Dominacion Espanola*. Biblioteca de la Presidencia de la Republica, Bogota, Colombia.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4):683–701.

Morten, M. and Oliveira, J. (2016). The effects of roads on trade and migration: Evidence from a planned capital city. *NBER Working Paper*, 22158.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics*, 122(2):569–600.

Parson, J. (1961). *La Colonizacion Antioquena en el Occidente Colombiano*. Archivo de la Economia Nacional, Banco de la Republica, Bogota, Colombia.

Pavcnik, N. (2002). Trade liberalization, exit, and productivity improvements: Evidence from chilean plants. *The Review of Economic Studies*, 69(1):245–276.

Pellegrina, H. S. (2017). The causes and consequences of the spatial organization of agriculture: Evidence from brazil. *Working Paper*.

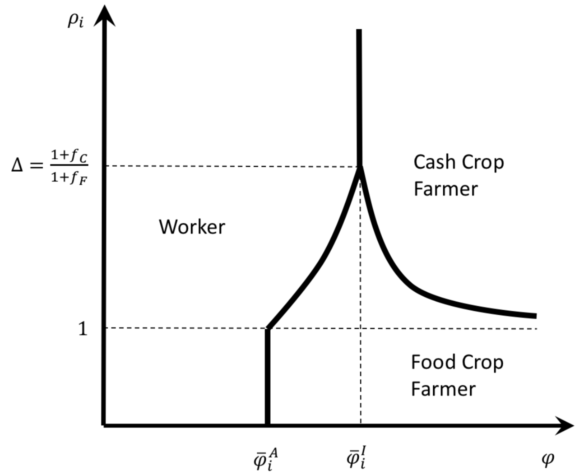
Rajan, R. G. and Zingales, L. (1998). Financial dependence and growth. *The American Economic Review*, 88(3):559–586.

Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, 55(2):pp. 234–250.

Sotelo, S. (2016). Domestic trade frictions and agriculture. *Working Paper*.

Tombe, T. (2015). The missing food problem: Trade, agriculture, and international productivity differences. *American Economic Journal: Macroeconomics*, 7(3):226–58.

Figure 1: Types of Local Equilibria, Occupational Choices and Selection into Trade

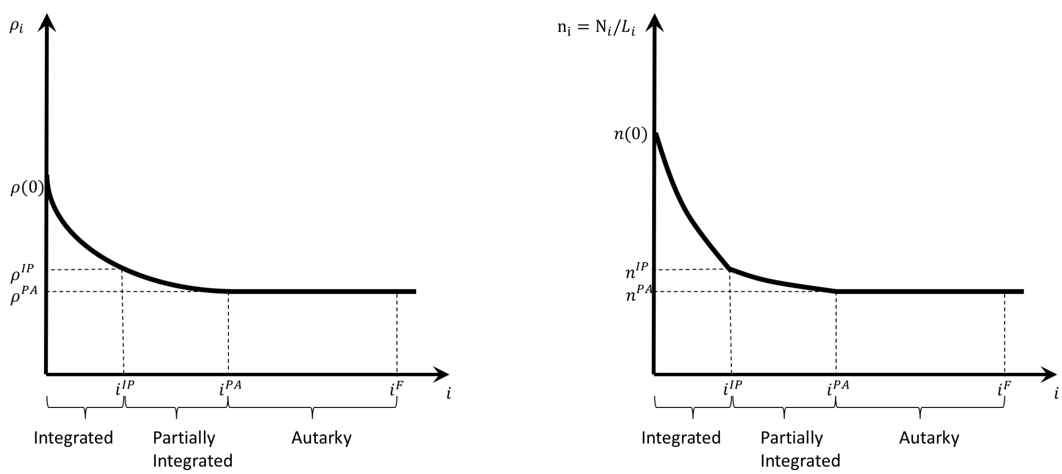


Notes: This figure illustrates the relationship between the cash crop premium and the occupational choices of individuals in the model. The y-axis is the cash crop premium and the x-axis the farming skill of an individual.

Figure 2: Relative Prices, Population Density and the Patterns of Commercialization

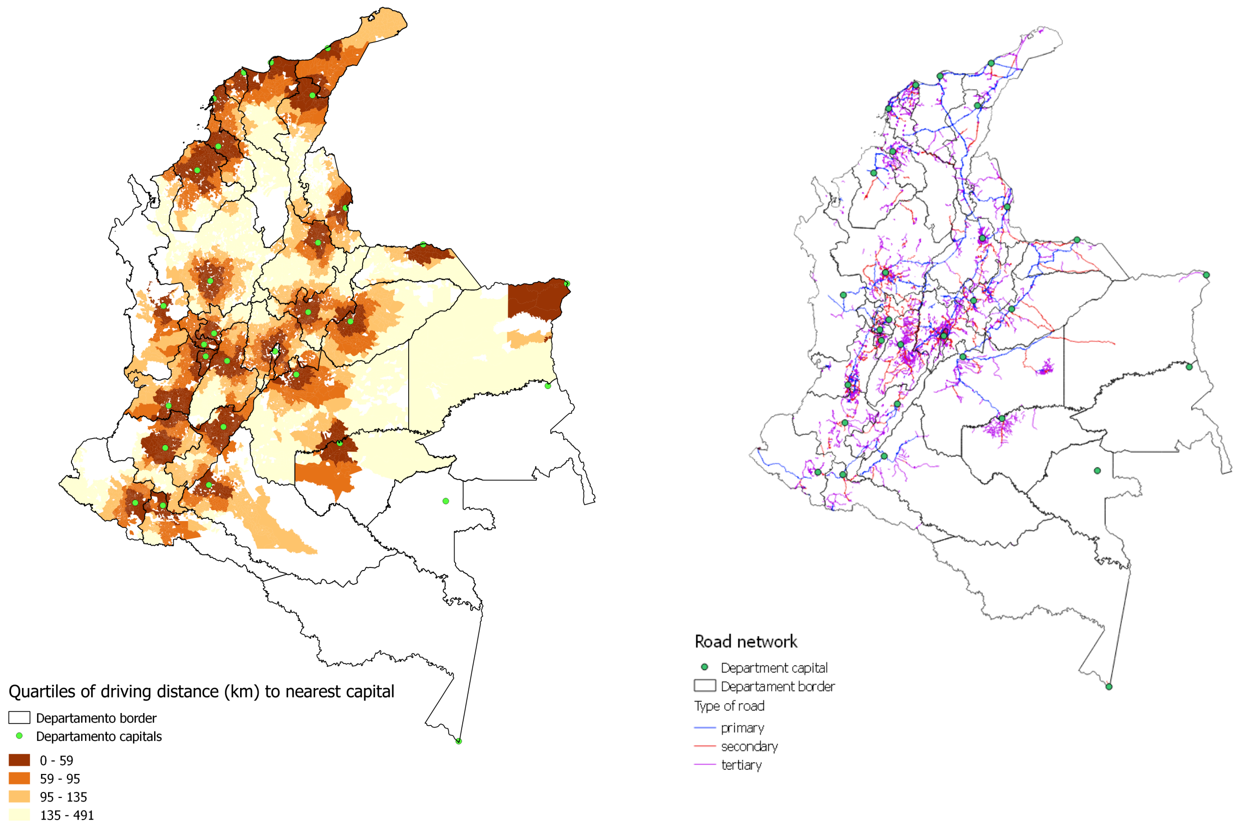
Panel A: Relative price

Panel B: Population density



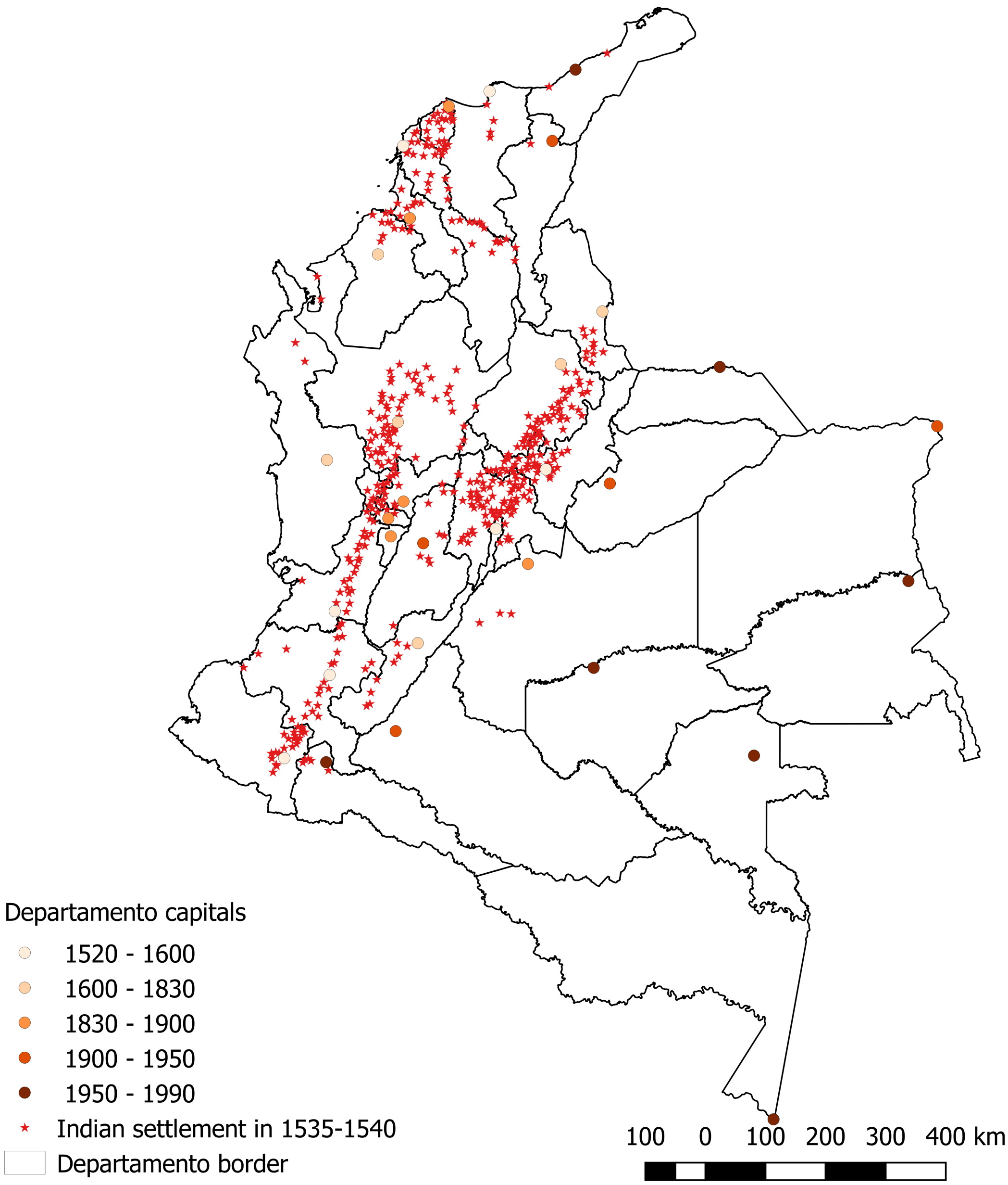
Notes: This figure illustrates the spatial equilibrium in the model. Panel A shows the relative price of the cash crop relative to the food crop the distance to urban centers. Panel B shows the relationship between population density and the distance to urban centers.

Figure 3: Road Network and Location of Urban Centers



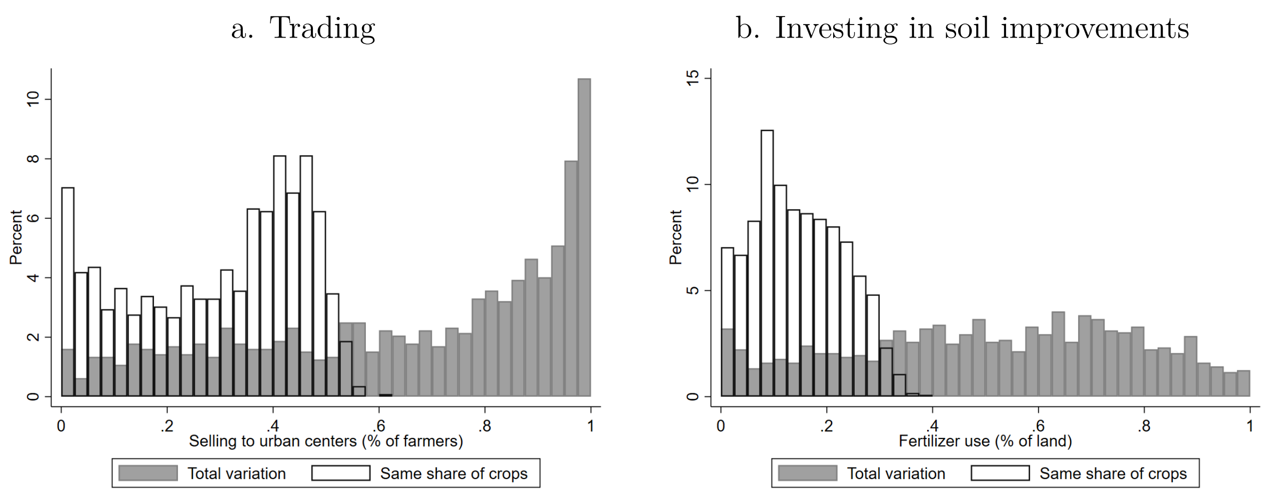
Notes: The figure on the left shows the travel distance by road from municipios to the nearest urban center. For most municipios, distances were calculated using Google Maps. For a set of municipios located closer to the Amazon forest region we used the road network presented on the right hand side.

Figure 4: Indian Settlements 1510-1561 and Current Departamento Capitals



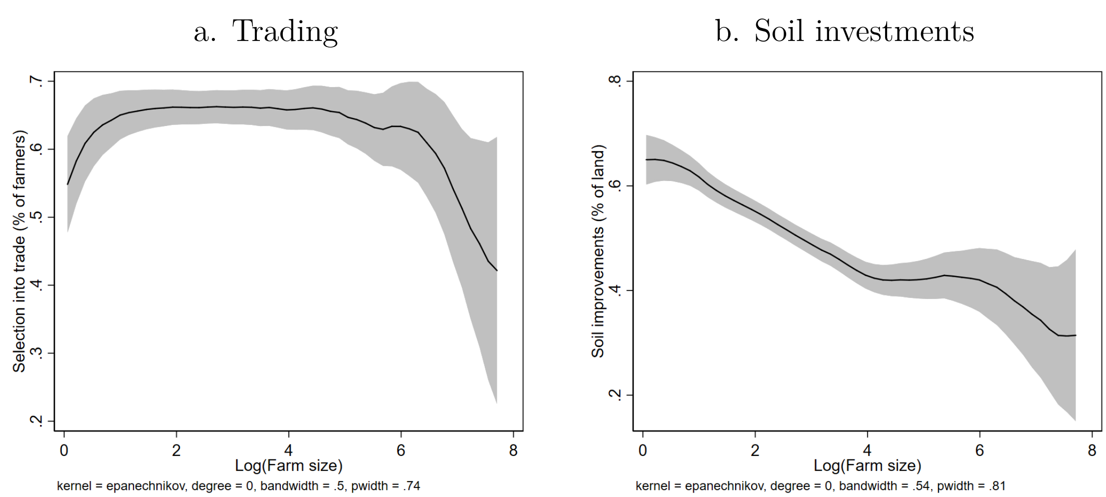
Notes: This figure shows the location of indigeous settlements and the location of current departamento capitals. Indigenous settlements come from the compilation of historical records presented in Melo (1995).

Figure 5: The Share of Farmers Trading and The Share of Land using Soil Improvements across Municipalities in Colombia



Notes: This figure shows the share of farmers selecting into trading and the share of land with soil improvements across municipalities. It also presents the distribution when there is no variation across municipalities in the share of land in each crop.

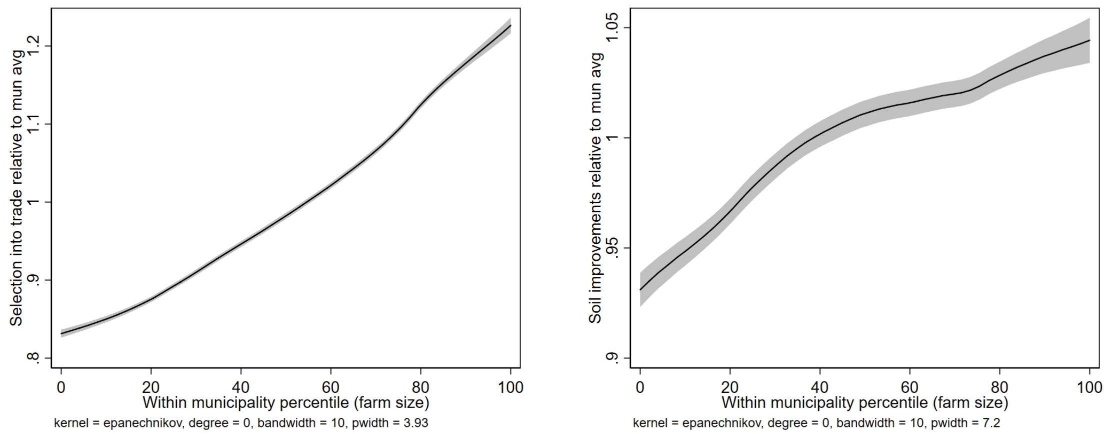
Figure 6: Farmers' Choices according to Farm Size between Municipalities



Notes: Panel a in this figure presents a non-parametric regression of the share of land with soil improvements in a county against the share of farmers selling their output to urban centers or international markets. Panel b presents the a non-parametric regression of the share of farmers trading and the average farm size in the Municipality

Figure 7: Farmers' Choices according to Farm Size within Municipality

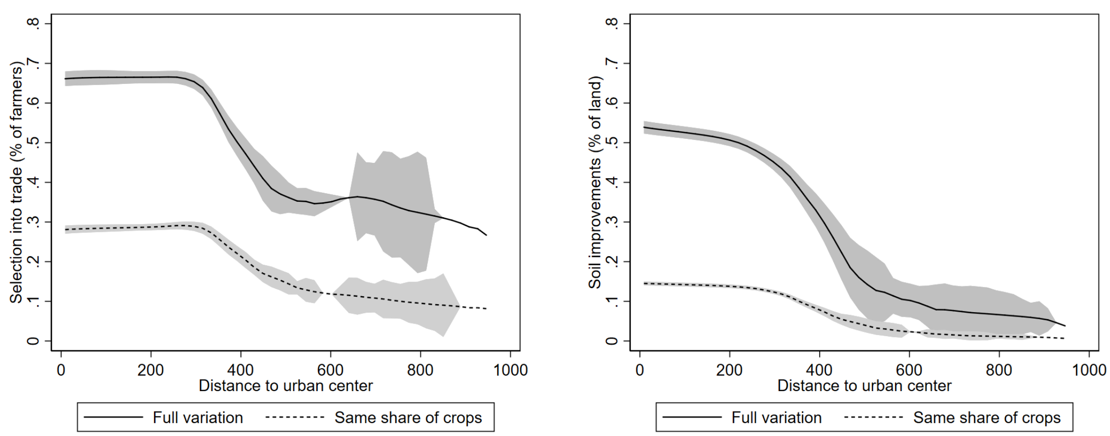
a. Trading b. Soil investments



Notes: This figure shows in the x-axis the percentile of a farm in terms of hectares relative to other farms within a municipality.

Figure 8: Farmers' Choices and The Distance to Urban Centers

a. Trade and market access b. Soil investments and market access



Notes: This figure shows a non-parametric regression of the share of farmers selling to urban centers on distance to urban centers as well as the share of farmers investing in soil improvements. In both cases, the figure includes non-parametric regression based on a decomposition exercise where we assume that all municipalities have the same share of farmers producing each crop.

Table 1: Summary Statistics for Selected Crops

	(1) Soil improvements (% of land)	(2) Distance to urban center (km)	(3) Own-consumption (% of farms)	(4) Cash destination (% of farms)
African palm (1%)	56.6	139.5	35.6	95.2
Banana (2%)	41.4	124.3	67.4	83.6
Cocoa (4%)	48.8	136.0	36.9	91.2
Coffee (35%)	75.9	101.6	34.5	89.2
Kidney bean (1%)	68.4	102.2	49.8	51.5
Pineapple (1%)	47.5	122.0	71.1	82.2
Plantain (17%)	49.3	141.8	64.2	53.1
Potato (3%)	71.8	72.4	92.3	25.2
Rice (3%)	54.0	131.3	40.1	78.9
Sugar cane (panela) (10%)	51.2	118.7	33.4	84.9
Sugar cane (sugar) (1%)	84.1	56.4	37.2	84.2
White maize (3%)	49.4	108.6	71.2	40.2
Yam (1%)	32.0	119.8	80.4	28.9
Yellow maize (5%)	50.2	112.0	71.1	39.4
Yucca (8%)	34.8	160.9	87.7	40.2
Total (100%)	58.9	118.7	51.6	70.2

Notes: This table shows summary statistics for the 15 largest crops in terms of planted area. Our sample accounts for 82% of the planted area and 90% of farmers in Colombia. Parenthesis next to the name of the crop indicate the percentage of plots producing the respective crop. Source: CNA-2014 DANE.

Table 2: The Impact of Distance to Urban Centers

	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) IV	(6) IV	(7) OLS	(8) IV
Panel A: Dep var is the % of farmers trading								
Dist to UC	-0.040**	-0.061***	-0.057**	-0.071***	-0.078***	-0.077***	-0.023***	-0.036***
	(0.018)	(0.022)	(0.022)	(0.014)	(0.022)	(0.023)	(0.008)	(0.010)
R^2 or F	0.018	0.412	0.423	51.563	73.922	82.689	0.468	82.689
Panel B: Dep var is the % of land with soil improvements								
Dist to UC	-0.085***	-0.063***	-0.055***	-0.103***	-0.077***	-0.070***	-0.016***	-0.021***
	(0.012)	(0.009)	(0.007)	(0.019)	(0.007)	(0.010)	(0.003)	(0.004)
R^2 or F	0.107	0.355	0.371	51.563	73.922	82.689	0.419	82.689
Panel C: First stage regression								
Dist to Pre-Col Cities				0.791***	0.927***	0.891***		0.891***
				(0.110)	(0.110)	(0.100)		(0.100)
R^2				0.598	0.726	0.735		0.735
Obs	1117	1117	1117	1117	1117	1117	1117	1117
Departamento FE		✓	✓		✓	✓		✓
Geography		✓	✓		✓	✓		✓
Institutions and Conflict			✓			✓		✓

Notes: Robust standard errors clustered at the nearest urban center level in parenthesis. *** denotes significance at 1% level, ** significance at 5% and * at 10%. Distance to urban centers is measured in 100 kilometers. Geographic variables are: altitude, ruggedness, distance to the closest river, total rainfall recorded in the closet gauging station, and the share of the area in the municipality in six different climatological zones. Institutions and Conflict variables include: years of presence of non-state armed actors in the municipality, number of hectares planted with coca, and the homicide rate in the municipality in 2013, the year of foundation of the municipality, two measures calculated by the central government that capture effectiveness and managerial capacity of the local government, investments in sewage and piped water, and the presence of the public agricultural bank in the municipality.There are 33 *departamento* fixed effects. OLS columns report R^2 and IV columns report F from Kleibergen-Paap weak instrument test.

Table 3: The Impact of Market Access on the Patterns of Agricultural Specialization

	Dependent variable is log of					
	Farms	Farms	Sales	Sales	Area	Area
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Interaction with share of farmers trading (OLS)						
Share × Dist to UC	-0.037	-0.153***	-0.129***	-0.339***	-0.105***	-0.319***
	(0.025)	(0.034)	(0.037)	(0.050)	(0.034)	(0.046)
R^2	0.343	0.382	0.382	0.413	0.357	0.392
Panel B: Interaction with share of farmers with soil improvements (OLS)						
Share × Dist to UC	-0.276***	-0.190***	-0.418***	-0.351***	-0.407***	-0.344***
	(0.040)	(0.037)	(0.052)	(0.054)	(0.050)	(0.049)
R^2	0.350	0.416	0.389	0.442	0.364	0.416
Panel C: Interaction with share of farmers trading (IV)						
Share × Dist to UC	-0.636***	-1.763***	-0.847***	-2.450***	-0.806***	-2.306***
	(0.131)	(0.391)	(0.179)	(0.560)	(0.169)	(0.498)
F	88	37	88	37	88	37
Panel D: Interaction with share of farmers with soil improvements (IV)						
Share × Dist to UC	-0.448***	-0.340***	-0.598***	-0.450***	-0.568***	-0.468***
	(0.073)	(0.076)	(0.095)	(0.100)	(0.086)	(0.096)
F	300	126	300	126	300	126
Obs	8000	8000	8000	8000	8000	8000
Municipalities	1095	1095	1095	1095	1095	1095
Crops	15	15	15	15	15	15
Fixed effects	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓

Notes: Robust standard errors clustered at the municipality level in brackets. *** denotes significance at 1% level, ** significance at 5% and * at 10%. Explanatory variables are standardized. For controls, we include: geographic variables such as altitude, ruggedness, distance to the closest river, total rainfall recoded in the closet gauging station, and the share of the area in the municipality in six different climatological zones; interactions of distance to urban centers with the share of farmers with high school in each crop and the share of farmers with machinery in each crop; interactions of the factor endowments in terms of machinery and farmers with high school with the share of farmers with high school in each crop and the share of farmers with machinery in each crop.

Table 4: Calibrated Parameters

		Definition for Selection into Trade					
		Narrow		Baseline		Broad	
Parameter	Description	Estimate	Std	Estimate	Std	Estimate	Std
γ	Land intensity	0.30	-	0.30	-	0.30	-
δ	Decay in cash crop premium	0.06	-	0.08	-	0.21	-
ρ_0	Cash crop premium	1.02	0.02	1.11	0.03	1.42	0.13
Δ	Fixed costs for cash crop	1.03	0.03	1.18	0.05	1.55	0.17
σ^2	Variance of skills	0.44	0.13	0.93	0.09	1.18	0.13

Notes: This table shows the estimates of the structural parameters of the model. Parameters without std. errors were calibrated. Decay for trade cost is given by 1000 *km*. Standard errors are obtained via bootstrapping using 100 random samples. We present results using three definitions of selection into trade: a narrow one that considers mostly international markets, a baseline one that considers retail regional markets and a broader one that also considers farm-gate sales.

Table 5: Mechanisms driving the Gains from Trade

		Def for Selection into Trade		
Discount functions	Mechanism	Narrow	Baseline	Broad
$L(\delta, \Delta)$	Selection into trade	0.99	0.95	0.91
$\Omega(\delta)$	Decay in cash crop premium	0.99	0.99	0.97

Notes: This table shows the estimates of the functions in the model that govern the gains from moving from an economy where all rural localities are in autarky to the current trading economy.

Table 6: Productivity Gains from Removing Barriers to Trade

		Discount on friction					
Barrier	Parameter	1	0.8	0.6	0.4	0.2	0
Panel A: Impact on productivity							
Fixed cost	$1+(\Delta-1)\times(\text{discount})$	1	1.005	1.012	1.021	1.031	1.041
Cash crop decay	$\delta\times(\text{discount})$	1	1.001	1.003	1.005	1.008	1.010
Panel B: Impact on share of farmers producing cash crop							
Fixed cost	$1+(\Delta-1)\times(\text{discount})$	0.47	0.63	0.87	0.98	0.99	1.00
Cash crop decay	$\delta\times(\text{discount})$	0.47	0.50	0.53	0.56	0.59	0.61

Notes: This table shows the impact of gradually removing the trade frictions of the economy.

Appendix

A Additional Details about the Data

A.1 Data

The main dataset in our empirical analysis is the 3rd Colombian Agricultural Census. This census was conducted by the statistical bureau of Colombia (DANE) between November of 2013 and December of 2014. It covers the entire Colombian territory and collects data on legal agricultural activities in private farms and collective land owned by minorities. We brought the micro-data and cleaned it for our analysis. The information is originally available at the plot-level. We aggregate this information into farm and municipality level for the purposes of this paper. Since farms can produce more that one crop, when aggregating at the farm level we focus on the crop with the largest share of area in each farm. Also, our main analysis focuses on the 15 most important crops in terms of area planted in country. There were a couple of limitations when we include all crops. One is that some of them tend to be highly concentrated in specific municipalities and provide no variation in export-orientation or use of soil improvements across regions. This limits our analysis of the crop-specialization by region. Second, we do not have price data for many of these crops, which keeps us from measuring sales. Third, some crops had no information on production but positive information on quantity when production was small, which indicates problems of measurement error.

The geographic variables that we use are the altitude of the municipality, a measure of the ruggedness of the terrain, total rainfall in the year of the survey, proximity to rivers, and 6 variables capturing the share of the area in a municipality under 6 different climatological zones. To compute these variables we performed the geometric intersection between the municipality boundaries and the raster and shapefiles containing information on each of these variables. For elevation and terrain roughness we use ASTER GDEM rasters and computed the weighted average of all pixel values on each variable within a municipality. For climatological zones we use estimations from the Caldas Lang model for climatological zoning. This model classifies geographic areas according to temperature, annual precipitation and elevation, ranging from warm and dry to extremely cold and humid. We use the shapefiles with this information provided by the National Institute of Hydrology and Meteorology (IDEAM). In addition, we use maps on the location of rivers to compute the distance from the centroid of each municipality to the closest river and measures of annual rainfall provided by IDEAM to compute a measures of water availability in the municipality the year of the survey.

A.2 Overview of the crops included in the data

Coffee, banana and African palm oil are the most important exportable crops in the country. Their production are not only important for the local economy, but also for international markets. Colombia is among the world's top four countries in coffee, African palm oil and bananas exports. Thanks to the diversity of elevations, climates and soil compositions, several regions of the country produce coffees of diverse qualities. Most varieties are suitable for hilly terrains, and production usually take place in small-scale farms. Cooperatives play a key role in the distribution of coffee. Minimum coffee prices are determined by the producers association, which works together with cooperatives of producers to guarantee unlimited purchases at fixed prices. Some cooperatives also make contracts with international buyers to sell special coffees at higher prices. Indeed, as shown in table A1, a large share of farmers sell their products to cooperatives in the case of coffee.

The production of bananas and African oil palm takes place in large landholdings. Due to its ecological conditions of humidity and temperature, banana cultivation concentrates in the Caribbean region. While the cultivation of African oil palm is widely spread across several regions. Palm oil mills are located in the producing regions and buy the raw product directly from farmers. African palm is typically not used as food and farmers sell the bulk of their products to wholesale industries (see A1). Exports of cocoa have recently gained importance in Colombia. Like coffee, cocoa is grown in small farms. There are two large processing companies that buy the raw product from agricultural intermediaries in the local market. Cooperatives of producers have emerged recently seeking to reduce intermediation costs and facilitate contracts with international buyers. Similar to African palm, cooperatives play a smaller role in the distribution of cocoa relative to coffee.

Sugarcane, rice and corn are other important cash crops in the country. There are two different varieties of corn and sugarcane. The scale of production and the destination markets vary across these varieties. In particular, medium to large-scale landholdings specializes in the production of the type of sugarcane that industrial plants use for the production of sugar. While small-family farms grow sugarcane for the production of *panela*[26] in small rustic plants. The *departamento* of Valle del Cauca in the south west of the country concentrates most of the sugarcane for sugar. While panela sugarcane is widely spread across the country. Similarly, production of white corn takes place in small farms, geographically spread across the country, and human consumption is its main destination. The production of yellow corn

[26] Unrefined sucrose product that results from boiling sugarcane juice

takes place in large mechanized farms and its destination is livestock feeding. Rice is planted in large land-holdings, located in the easter plains and in the mid-west of the country.

Tubes, including potatoes, yam and yucca, and plantains are the most important food crops in the country. Production of these crops usually takes place in small farms. The soil and climate of the mountains are suitable for growing potatoes. While plantains, yam and yucca are suited for lower elevation and warmer temperatures. We can see that these crops a large part of the farmers use the production of their farm for their own subsistence A1.

B Details about the Empirical Analysis

In the paper, we refer to robustness checks using alternative definitions of dependent variables. Below, we provide a summary of the additional results that we include in this appendix. Our results hold using a series of alternative specifications.

1. Table A2 shows the impat of market acess using a dependent variable of selection into trade that is constructed from a dataset that includes all the crops in the data.

2. Table A3 shows the impact of market access using as a dependent variable a narrow definition of selection into trade based only on farmers who sell to cooperatives and to international markets.

3. Table A3 shows the impact of market access using as a dependent variable a broader definition of selection into trade that also includes farm gate selling and municipality markets.

4. Table A3 also shows the impact of market access when we use distance to ports and major airports according to customs data as the explanatory variable.

5. Table A4 shows the impact of market access on the specialization of a county using the nature of the crops in the data to construct the measure of selection into trade.

6. Table A5 shows the first stage of the regression of the impact of market access on the patterns of crop specialization, which was omitted from the main text.

7. Table A7 shows the impact of market access on the specialization of municipalities in crops that are commercially-oriented using the extensive margin of specialization. In other words, a dummy indicating whether the municipality produces a specific crop.

8. Table A8 shows the impact of the log of market access on the selection of farmers into trade.

B.1 Farm size distribution

Figure A5 shows that the farm size distribution in terms of sales has a log-normal shape. This is in line with the skill distribution used in Adamopoulos and Restuccia (2014). This motivate us to estimate the model based on a log-normal distribution.

C Details of the Model

This section describes the equations of the model in detail. In the paper, we provide a full description of the set up of the model.

C.1 Profit Maximization

Crops are indexed by k, rural localities by i and individuals are referred according to their farming skill φ. Individuals can pay a fixed cost f_k to set up a farm. If they do so, they maximize the following profit function

$$\pi^*_{ki}(\varphi) = \max_{l_{ki}(\varphi)} p_{ki}a_{ki}\varphi^{1-\gamma}l_{ki}(\varphi)^{\gamma} - r_i l_{ki}(\varphi) - w_i f_k,$$

where p_{ki} is the price, a_{ki} is the productivity, $l_{ki}(\varphi)$ is the land use and r_i is the land rent.

Solving this maximization problem gives

$$l^*_{ki}(\varphi) = \varphi \left(\frac{p_{ki}a_{ki}\gamma}{r_i}\right)^{\frac{1}{1-\gamma}},$$

and

$$\pi^*_{ki}(\varphi) = \varphi\Gamma \left(\frac{p_{ki}a_{ki}}{r_i^{\gamma}}\right)^{\frac{1}{1-\gamma}} - w_i f_k.$$

For the cash crop, we will use the following equations

$$l^*_{Ci}(\varphi) = \varphi\rho_i \left(\frac{p_{Fi}a_{Fi}\gamma}{r_i}\right)^{\frac{1}{1-\gamma}} \tag{A1}$$

and

$$\pi^*_{Ci}(\varphi) = \varphi\rho_i\Gamma \left(\frac{p_{Fi}a_{Fi}}{r_i^{\gamma}}\right)^{\frac{1}{1-\gamma}} - w_i f_C, \tag{A2}$$

where we defined $\rho_i \equiv (\frac{p_{Ci}a_{Ci}}{p_{Fi}a_{Fi}})^{\frac{1}{1-\gamma}}$ as in the article. In what follows, when convenient, we will also use $\Delta \equiv (1+f_C)/(1+f_F)$ to derive the model.

C.2 Local equilibria

C.2.1 Local Equilbrium 1: Integrated Economy (I)

The cutoff $\bar{\varphi}_i^I$ is given by $\pi_{Ci}^*(\bar{\varphi}_i^I) = w_i$, which gives

$$\bar{\varphi}_i^I = w_i \Delta(1+f_F) \Big/ \Gamma \rho_i \left(\frac{p_{Fi} a_{Fi}}{r_i^\gamma} \right)^{\frac{1}{1-\gamma}}. \tag{A3}$$

From the full employment, we get

$$N_{Ci}(1+f_C) = N_i,$$

since $N_{Ci} = N_i(1-G(\bar{\varphi}_i^I))$, this condition can be written as

$$(1-G(\bar{\varphi}_i^I)) = \frac{1}{1+f_C}.$$

Note that the full employment equation provides implicitly the value of $\bar{\varphi}_i^I$ as a function of f_C. Using this result, we can work with the market clearing condition

$$N_{Ci} \int_{\bar{\varphi}_i^I}^{\infty} l_{ki}^*(\varphi) g(\varphi | \varphi > \bar{\varphi}_i^I) d\varphi = L_i.$$

Substitute the demand for land in equation (A1) and use $N_{Ci} = N_i(1-G(\bar{\varphi}_i^I))$ to get

$$N_i(1-G(\bar{\varphi}_i^I))\rho_i \left(\frac{p_{Fi} a_{Fi} \gamma}{r_i} \right)^{\frac{1}{1-\gamma}} E(\varphi | \varphi > \bar{\varphi}_i^I) = L_i.$$

Now, use the cutoff equation to obtain

$$N_i \frac{w_i \Delta(1+f_F)}{\Gamma \rho_i \bar{\varphi}_i^I} \gamma^{\frac{1}{1-\gamma}} (1-G(\bar{\varphi}_i^I)) E(\varphi | \varphi > \bar{\varphi}_i^I) = r_i L_i,$$

which can be written as

$$\frac{w_i N_i}{(1-\gamma)} \frac{\Delta(1+f_F)}{\rho_i \bar{\varphi}_i^I} \rho_i (1-G(\bar{\varphi}_i^I)) E(\varphi | \varphi > \bar{\varphi}_i^I) = \frac{r_i L_i}{\gamma}.$$

Here, we define

$$\alpha_i^I = \frac{\bar{\varphi}_i^I}{E(\varphi | \varphi > \bar{\varphi}_i^I)},$$

which is the share of payments to wages and the opportunity cost of farmers relative to profits.

The payments to land must equal the payments to labor adjusted by a term that is related to the efficiency of skill use in the economy. Isolate r_i and we get

$$r_i = w_i \frac{\gamma}{(1-\gamma)\alpha_i^I} \frac{N_i}{L_i}.$$

Substitute this equation back into the cutoff in equation (A3)

$$\bar{\varphi}_i^I = w_i \Delta(1+f_F) \Big/ \Gamma \rho_i \left(\frac{p_{Fi} a_{Fi}}{\left(w_i \frac{\gamma}{(1-\gamma)\alpha_i^I} \frac{N_i}{L_i} \right)^{\gamma}} \right)^{\frac{1}{1-\gamma}}.$$

Isolating w_i gives

$$w_i = (1-\gamma) p_{Fi} a_{Fi} \left(\frac{\bar{\varphi}_i^I}{1+f_F} \frac{\rho_i}{\Delta} \right)^{1-\gamma} \left(\alpha_i^I \frac{L_i}{N_i} \right)^{\gamma},$$

which can also be written as

$$w_i = (1-\gamma)\alpha_i^I p_{Fi} a_{Fi} \left(h_i^I \right)^{1-\gamma} \left(\frac{L_i}{N_i} \right)^{\gamma}.$$

where

$$h_i^I \equiv \frac{\rho_i E(\varphi | \varphi > \bar{\varphi}_i^I)}{1+f_C}.$$

The GDP is given by $Y_i = r_i L_i / \gamma$. As such, the expected real income of individuals i before observing their farming skill φ is

$$u_i^I = (1-\gamma) \frac{r_i L_i}{\gamma} \frac{1}{p_{Fi}} \frac{1}{N_i}.$$

Using the formula for land rents gives

$$u_i^I = \frac{w_i}{\alpha_i^I} \frac{1}{p_{Fi}}.$$

Now, using the formula for wages we get

$$u_i^I = (1-\gamma) a_{Fi} \left(h_i^I \right)^{1-\gamma} \left(\frac{L_i}{N_i} \right)^{\gamma}. \tag{A4}$$

C.2.2 Local Equilbrium 2: Partially Integrated Economy (P)

For the partially integrated economy, we have the following cutoffs

$$\bar{\varphi}_i^P = w_i(1+f_F)\Big/\Gamma\left(\frac{p_{Fi}a_{Fi}}{r_i^{\gamma}}\right)^{\frac{1}{1-\gamma}} \tag{A5}$$

and

$$\tilde{\varphi}_i^P = \bar{\varphi}_i^P\left(\frac{\Delta-1}{\rho_i-1}\right).$$

For the full employment condition, we can get

$$(1-G(\tilde{\varphi}_i^P)(\Delta-1)+(1-G(\bar{\varphi}_i^P)) = \frac{1}{1+f_F}. \tag{A6}$$

The solution for $\tilde{\varphi}_i^P$ is implicitly given by the equations above and we can write $\bar{\varphi}_i^P$ and $\bar{\varphi}_i^P$ as a function of f_C, f_F and ρ_i.

The land market clearing is given by

$$\left(\frac{p_{Fi}a_{Fi}\gamma}{r_i}\right)^{\frac{1}{1-\gamma}}\left(N_{Ci}\int_{\tilde{\varphi}_i^P}^{\infty}\rho_i g(\varphi|\varphi>\tilde{\varphi}_i^P)d\varphi + N_{Fi}\int_{\bar{\varphi}_i^P}^{\tilde{\varphi}_i^P}\varphi g(\varphi|\tilde{\varphi}_i^P>\varphi>\bar{\varphi}_i^P)d\varphi\right) = L_i.$$

using the properties of conditional probabilities and $N_{Ci} = N_i(1-G(\tilde{\varphi}_i^P))$ and $N_{Fi} = N_i(G(\tilde{\varphi}_i^P) - G(\bar{\varphi}_i^P))$ we get

$$\left(\frac{p_{Fi}a_{Fi}\gamma}{r_i}\right)^{\frac{1}{1-\gamma}}\left(N_i\rho_i\int_{\tilde{\varphi}_i^P}^{\infty}g(\varphi)d\varphi + N_i\int_{\bar{\varphi}_i^P}^{\tilde{\varphi}_i^P}\varphi g(\varphi)d\varphi\right) = L_i.$$

Add and substracting $N_i\int_{\tilde{\varphi}_i^P}^{\infty}\varphi g(\varphi)d\varphi$, and use the properties of the conditional probabilities again to obtain

$$N_i\left(\frac{p_{Fi}a_{Fi}\gamma}{r_i}\right)^{\frac{1}{1-\gamma}}\Big((1-G(\tilde{\varphi}_i^P))(\rho_i-1)E(\varphi|\varphi>\tilde{\varphi}_i^P)+(1-G(\bar{\varphi}_i^P))E(\varphi|\varphi>\bar{\varphi}_i^P)\Big) = L_i.$$

Substitute the cutoff condition here to obtain

$$N_i\frac{w_i(1+f_F)}{\Gamma\bar{\varphi}_i^P}\Big((1-G(\tilde{\varphi}_i^P))(\rho_i-1)E(\varphi|\varphi>\tilde{\varphi}_i^P)+(1-G(\bar{\varphi}_i^P))E(\varphi|\varphi>\bar{\varphi}_i^P)\Big)\gamma^{\frac{1}{1-\gamma}} = r_iL_i.$$

Using the full employment condition to substitute $(1+f_F)$, define the share of payments to wages and opportunity cost of farmers

$$\alpha_i^P \equiv \frac{\bar{\varphi}_i^P((1-G(\tilde{\varphi}_i^P)(\Delta-1)+(1-G(\bar{\varphi}_i^P)))}{\Big((1-G(\tilde{\varphi}_i^P))(\rho_i-1)E(\varphi|\varphi>\tilde{\varphi}_i^P)+(1-G(\bar{\varphi}_i^P))E(\varphi|\varphi>\bar{\varphi}_i^P)\Big)}.$$

Using this equation, we obtain

$$\frac{w_iN_i}{(1-\gamma)\alpha_i^P} = \frac{r_iL_i}{\gamma}.$$

Isolate r_i

$$r_i = w_i\frac{\gamma}{(1-\gamma)\alpha_i^P}\frac{N_i}{L_i}.$$

Substitute back into the cutoff equation (A5)

$$\bar{\varphi}_i^P = w_i(1+f_F)\Big/\Gamma\left(\frac{p_{Fi}a_{Fi}}{\left(w_i\frac{\gamma}{(1-\gamma)\alpha_i^P}\frac{N_i}{L_i}\right)^{\gamma}}\right)^{\frac{1}{1-\gamma}}.$$

Isolating wages and simplifying gives

$$w_i = (1-\gamma)\alpha_i^P p_{Fi}a_{Fi}\left(h_i^P\right)^{1-\gamma}\left(\frac{L_i}{N_i}\right)^{\gamma},$$

where

$$h_i^P \equiv \frac{\Big((1-G(\tilde{\varphi}_i^P))(\rho_i-1)E(\varphi|\varphi>\tilde{\varphi}_i^P)+(1-G(\bar{\varphi}_i^P))E(\varphi|\varphi>\bar{\varphi}_i^P)\Big)}{(1+f_F)((1-G(\tilde{\varphi}_i^P)(\Delta-1)+(1-G(\bar{\varphi}_i^P))}.$$

Using this equation for nominal wages, we get the following expression for expected real income

$$u_i^P = (1-\gamma)a_{Fi}\left(h_i^P\right)^{1-\gamma}\left(\frac{L_i}{N_i}\right)^{\gamma}. \tag{A7}$$

C.2.3 Local Equilbrium 3: Autarky (A)

For the economy in autarky, we can solve it as in the fully integrated equilibrium and get the following real income

$$u_i^A = (1-\gamma)a_{Fi}\left(h_i^A\right)^{1-\gamma}\left(\frac{L_i}{N_i}\right)^{\gamma}, \tag{A8}$$

where

$$h_i^A \equiv \frac{E(\varphi|\varphi>\bar{\varphi}_i^A)}{1+f_F}.$$

C.2.4 Local GDP

The product per worker is given by

$$y_i = (1-\gamma)a_{Fi}\left(h_i^L\right)^{1-\gamma}\left(\frac{L_i}{N_i}\right)^{\gamma},$$

where the subscript L denotes the type of local equilibrium in i ($L \in \{A, P, I\}$). Multiplying by the mass of worker N_i gives

$$Y_i = (1-\gamma)a_{Fi}\left(h_i^L\right)^{1-\gamma}(N_i)^{1-\gamma}(L_i)^{\gamma}.$$

Since markets are perfectly competitive and there are no market failures, the solution to the problem of agents is the one that maximizes the local product of the economy. Therefore, we can rewrite the GDP in i as

$$Y_i = h_i(N_i)^{1-\gamma}(L_i)^{\gamma}, \tag{A9}$$

where we defined

$$h_i = (1-\gamma)a_{Fi}\max\{h_i^A, h_i^P, h_i^I\}^{1-\gamma}, \tag{A10}$$

which is a term capturing the aggregate TFP of county i.

C.3 Spatial equilibrium

Let u^* be the welfare in spatial equilibrium where agents are indifferent between localities. Then, the national population employment condition is given by

$$N = \left(\int_0^{i^F} L_i\left(\frac{h_i}{u^*}\right)^{\frac{1}{\gamma}} d_i\right)$$

Isolating u^* gives

$$u^* = N^{-\gamma}\left(\int_0^{i^{i^F}} L_i(h_i)^{\frac{1}{\gamma}} d_i\right)^{\gamma}.$$

C.4 Gains from trade

To obtain the welfare gains described in the article, we define the following terms

$$u^A = N^{-\gamma} \left(\int_0^{i^F} L_i (h^A)^{\frac{1}{\gamma}} d_i \right)^{\gamma},$$

which can be written as

$$u^A = h^A \left(\frac{A}{N} \right)^{\gamma},$$

where we defined $A \equiv \int_0^{i^F} L_i (a_{Fi})^{\frac{1}{\gamma}} d_i$ and $h^A \equiv (1-\gamma)(\frac{(1-G(\bar{\varphi}^A))E(\varphi|\varphi>\bar{\varphi}^A)}{1+f_F})^{1-\gamma}$. For the full trading economy with no selection, we have

$$\bar{u} = \bar{h} \left(\frac{A}{N} \right)^{\gamma},$$

where $\bar{h} \equiv h^A \rho_0$. Now, the welfare when $f_C \to f_F$ is

$$u^S = \left(\int_0^{i^{i^F}} L_i (\rho_i h^A)^{\frac{1}{\gamma}} d_i \right)^{\gamma},$$

where

$$u^S = h^A \left(\int_0^{i^{i^F}} L_i (\rho_i)^{\frac{1}{\gamma}} d_i \right)^{\gamma},$$

Now, assuming that every rural location has the same endowment of land $L_i = L$, we can write the gains from trade as

$$\frac{u^*}{u^A} = L(\delta, \Delta) \Omega(\delta) \frac{\bar{u}}{u^A},$$

where we defined the loss function related to selection

$$L(\delta, \Delta) = \left(\frac{\int_0^{i^F} \left(h_i \right)^{\frac{1}{\gamma}} d_i}{\int_0^{i^F} \left(\rho_i h^A \right)^{\frac{1}{\gamma}} d_i} \right)^{\gamma},$$

and the pure loss function related to trade costs

$$\Omega(\delta) = \left(\int_0^{i^{i^F}} \left(\frac{\rho_i}{\rho_0} \right)^{\frac{1}{\gamma}} d_i \right)^{\gamma}.$$

D Proof of Propositions

D.0.1 Proof of proposition 1

We define the integrated economy as the case where $w_i > \pi^*_{Fi}(\bar{\varphi}^I_i)$, which is the maximum profit that the farmer with the lowest skill φ would obtain if she became a worker rather than a producer of F. In other words, any individual with a skill φ lower than $\bar{\varphi}^I_i$ would choose to become a worker instead of becoming a food producer. Using the cutoff equation for $\bar{\varphi}^I_i$ in (A3) , one can show that

$$w_i > \pi^*_{Fi}(\bar{\varphi}^I_i) \iff \rho_i > \Delta,$$

which completes the proof.

D.0.2 Proof of proposition 2

We define a partially integrated economy as the case where $\Phi_{Ci} \neq \emptyset$ and $\Phi_{Fi} \neq \emptyset$. Therefore, our proposition can be re-stated as

$$\Phi_{Ci} \neq \emptyset \wedge \Phi_{Fi} \neq \emptyset \iff \Delta > \rho_i > 1$$

First, we prove the only if statement. Assume that $\Delta > \rho_i > 1$ is true. Note that since profits are increasing in φ, there are always agents who become farmers. Therefore, we can exclude the case where $\Phi_{Ci} = \emptyset \wedge \Phi_{Fi} = \emptyset$. Due to $\rho_i > 1$, there exists an agent with a sufficiently high skill φ^* that satisfies $\pi^*_{Ci}(\varphi^*) > \pi^*_{Fi}(\varphi^*)$ and chooses to become a producer of C, so that $\Phi_{Ci} \neq \emptyset$. Due to $\Delta > \rho_i$, we know that $\Phi_{Fi} = \emptyset \wedge \Phi_{Ci} \neq \emptyset$ does not hold. Therefore, we must have $\Phi_{Ci} \neq \emptyset \wedge \Phi_{Fi} \neq \emptyset$. For the if part, assume that $\Phi_{Ci} \neq \emptyset \wedge \Phi_{Fi} \neq \emptyset$. If $\rho_i > \Delta$, then we have $\Phi_{Ci} \neq \emptyset \wedge \Phi_{Fi} = \emptyset$. If $1 > \rho_i$, then $\pi^*_{Ci}(\varphi*) < \pi^*_{Fi}(\varphi*)$ for any $\varphi*$ and all farmers produce food. Therefore, if $\Phi_{Ci} \neq \emptyset \wedge \Phi_{Fi} \neq \emptyset$ then $\Delta > \rho_i > 1$.

Regarding the partition part of the proposition. Define the profit difference for the farmer producing C relative to F as $\Delta\pi(\varphi)_{FCi} \equiv \pi^*_{Ci}(\varphi) - \pi^*_{Fi}(\varphi)$. One can show that $\partial\Delta\pi(\varphi)_{FCi}/\partial\varphi > 0$, that is, an increase in φ increases the incentives to produce C. Therefore, relative to the agent φ^{CF}_i who is indifferent between the production of these two crops, all agents with $\varphi > \varphi^{CF}_i$ will produce C. This defines the cutoff $\tilde{\varphi}^P_i$. Now, define the difference between the profit of farmers producing F and wages as $\Delta\pi(\varphi)_{FWi} \equiv \pi^*_{Fi}(\varphi) - w_i$. One can also show that $\partial\Delta\pi(\varphi)_{FCi}/\partial\varphi > 0$. By the previous part of the proof, we are in the equilibrium where the set of farmers who choose to become a food crop is non empty. Therefore, we know that there must be a $\bar{\varphi}^P_i$ such that agents above this cutoff produce the food crop. In addition, since the set of farmers producing the cash crop is also non empty and that profits are increasing in φ, we know that there are some agents above $\bar{\varphi}_i^P$ that produce the cash crop. With this, we complete the proof.

D.0.3 Proof of proposition 3

To prove this result, we first note that the ratio of the cutoffs is given by

$$\frac{\tilde{\varphi}_i^P}{\bar{\varphi}_i^P} = \frac{\Delta - 1}{\rho_i - 1}.$$

With an increase in ρ_i, the ratio of the cutoffs must decrease. As a consequence, we also know that the share of farmers becoming cash crop producers conditional on becoming farmers increases $(1 - G(\tilde{\varphi}_i^P))/(1 - G(\bar{\varphi}_i^P))$. Now, manipulate full employment condition to obtain

$$\frac{1 - G(\tilde{\varphi}_i^P)}{1 - G(\bar{\varphi}_i^P)}(\Delta - 1) + 1 = \frac{1}{(1 + f_F)(1 - G(\bar{\varphi}_i^P))}.$$

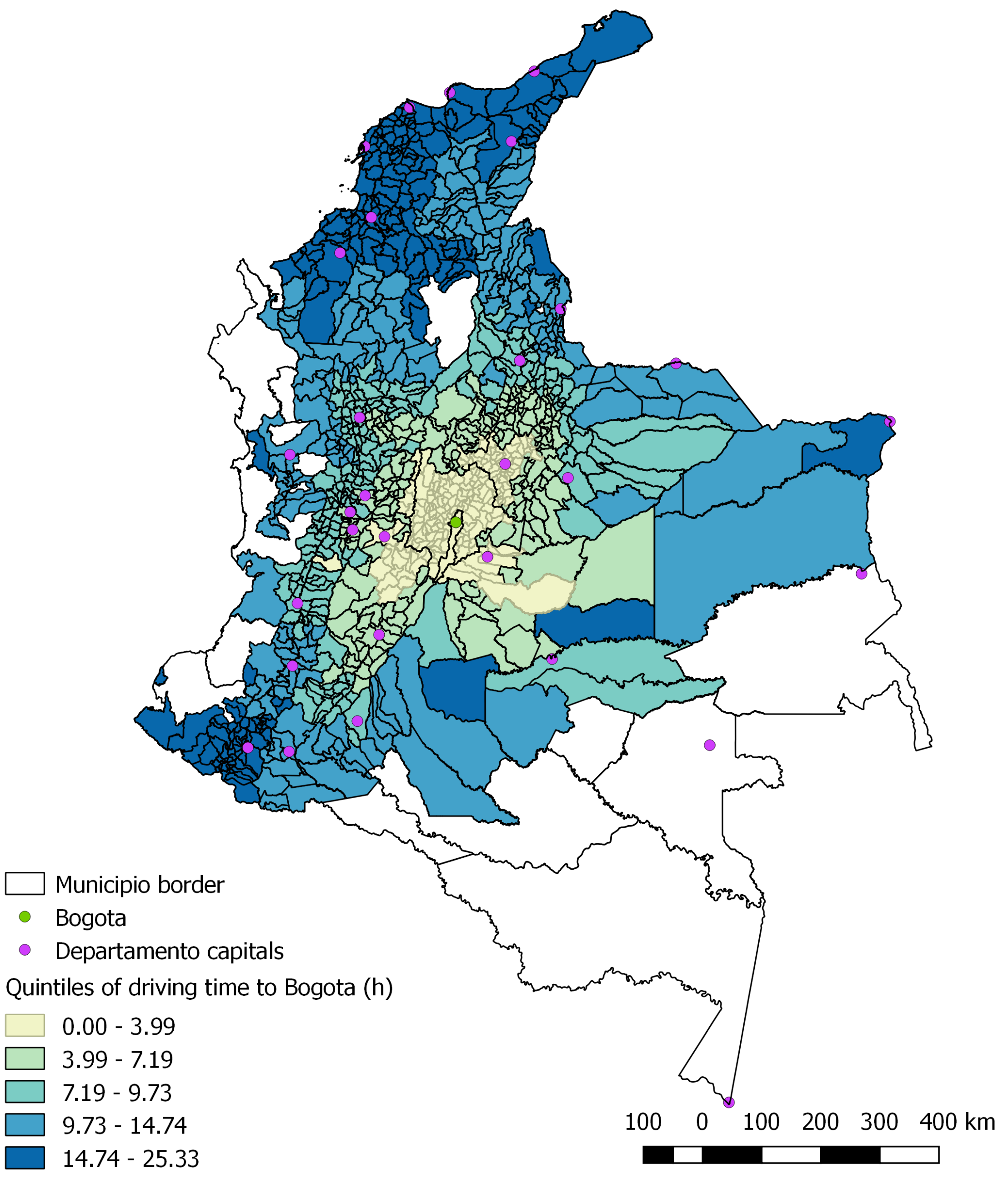
The left hand side in the equation above increases, therefore, the term $(1 - G(\bar{\varphi}_i^P))$ must decrease and $\bar{\varphi}_i^P$ increase. In summary, with an increase in ρ_i, the share of cash crop producers increase relative to food producers, but the share of individuals becoming farmers decrease.

D.0.4 Proof of proposition 4

If a location $i = 0$ is an integrated economy where $\Phi_{C0} \neq \emptyset$ and $\Phi_{F0} = \emptyset$, then we have $\rho_0 > \Delta > 1$. If the location i^F is in autarky, then we must have $p^*_{Ci^F} a_{Ci^F} < p^*_{Fi^F} a_{Fi^F}$, where $p^*_{Ci^F}$ is the price that a trader could obtain by purchasing C from i^F and selling it in $i = 0$ and $p^*_{Fi^F}$ the price of purchasing food in i^0 and selling it in i^F. In this case, $\rho i^F < 1$. Using the non-arbitrage condition, define a cash crop premium function $\rho(i) = (a(p_{C0}/p_{F0})exp(-2\delta d_{0i}))^{1/(1-\gamma)}$, where a is the relative productivity of the economy. This function is continuous and decreases with i. Since $i = 0$ is an integrated economy, then $\rho(0) \geq \Delta > 1$, and for $i = i^F$,$\rho(i^F) < 1 < (p^*_{Ci^F}/p^*_{Fi^F})a$. Due to the continuity of $\rho()$, there must be a i^{IP} such that $\rho(i^{IP}) = \Delta > 1$. For all $i \leq i^{IP}$ we have $\rho(i \geq \Delta$ and a full trading equilibrium holds. In addition, there must be a i^{PA} such that $\rho(i^{IP}) > \rho(i^{PA}) \geq 1$. In the region between i^{PA} and i^{IP}, the conditions for the dual equilibrium holds. Finally, for i^{PA} until i^F, we know that $\rho(i^{PA}) < 1$ and an autarky equilibrium exists.

E Additional Figures and Tables

Figure A1: Driving time to Bogota



Notes: The figure shows driving time from urban areas of municipios to Bogota. Driving time was calculated using Google Maps. Municipios in white are not connected to the road network. It takes about 17 hours to drive from Bogota to Cartagena, in the northern coast. The driving distance between these two cities is 1,042 kilometers (647.5 miles).

Figure A2: Distance to current urban centers and distance to historical cities

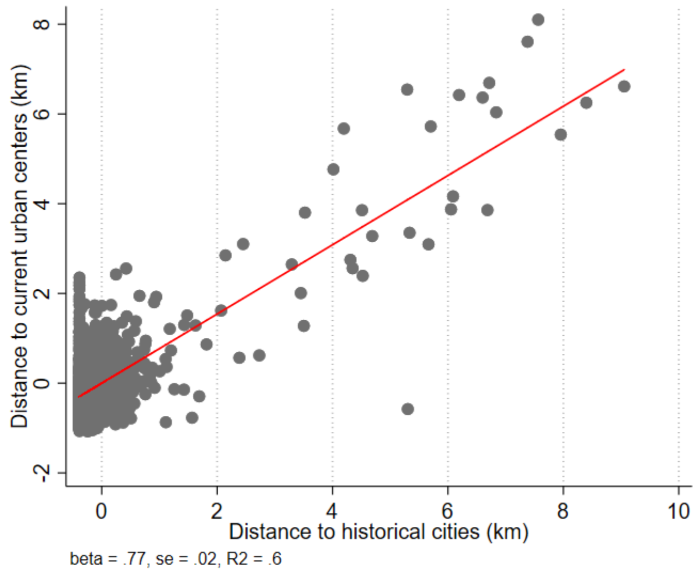
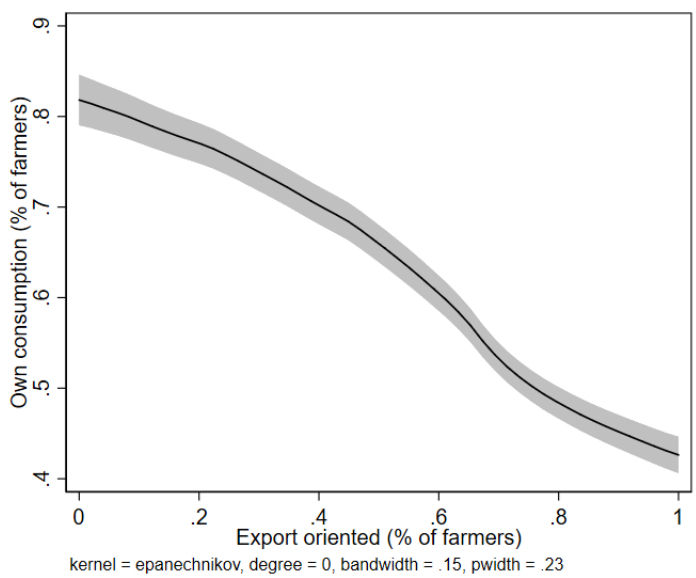


Figure A3: Link between Urban Destination and Own-Consumption



Notes: This figure shows a non-parametric regression of the share of farmers selling to urban centers and the share of farmers consuming their own products.

Figure A4: Revenues per Hectare relative to Municipality Average

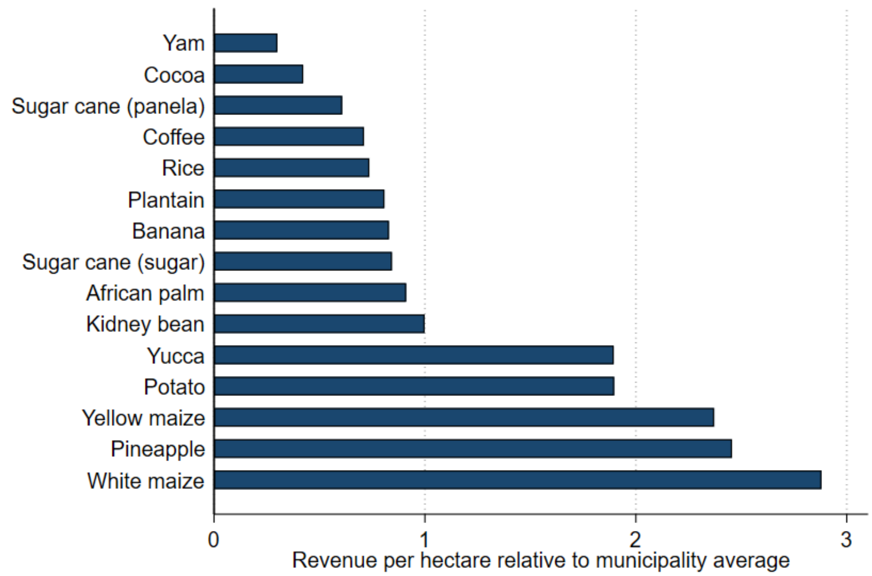
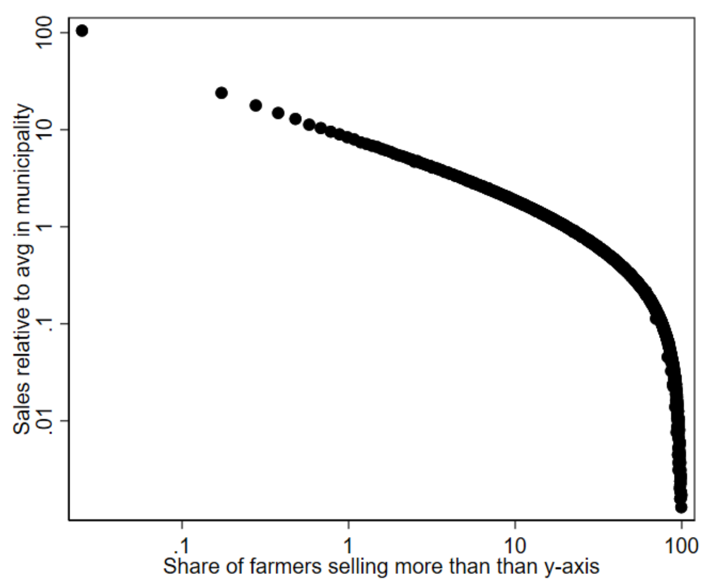


Figure A5: Sales Distribution of Farmers within Municipalities



Notes: To construct this figure, we first calculate the percentile in the distribution of sales for each farm within a municipality and divide the total sales by the average sales in the municipality. We then collapse the data into 1000 groups according to the distribution in the percentile distribution of sales.

Table A1: Complete Summary Statistics for Selected Crops

	(1) Machinery in farm (% of farms)	(2) High education managers (% of managers)	(3) Fertilizer use (% of land)	(4) Distance to urban center (km)	(5) Own-consumption (% of farms)	(6) Cooperatives (% of farms)	(7) Wholesale industry (% of farms)
African palm (1%)	36.1	36.4	56.6	139.5	35.6	5.1	93.7
Banana (2%)	26.1	30.7	41.4	124.3	67.4	34.7	18.3
Cocoa (4%)	35.2	25.3	48.8	136.0	36.9	22.4	86.0
Coffee (35%)	30.4	23.0	75.9	101.6	34.5	85.9	24.1
Kidney bean (1%)	23.5	24.9	68.4	102.2	49.8	28.3	16.6
Pineapple (1%)	32.2	35.7	47.5	122.0	71.1	16.7	22.1
Plantain (17%)	32.5	31.4	49.3	141.8	64.2	28.6	22.0
Potato (3%)	15.4	22.2	71.8	72.4	92.3	3.0	9.1
Rice (3%)	26.3	34.2	54.0	131.3	40.1	9.3	74.9
Sugar cane (panela) (10%)	23.3	23.0	51.2	118.7	33.4	25.9	81.7
Sugar cane (sugar) (1%)	23.0	31.6	84.1	56.4	37.2	24.8	83.5
White maize (3%)	18.3	29.1	49.4	108.6	71.2	15.0	19.0
Yam (1%)	14.9	37.6	32.0	119.8	80.4	3.3	19.0
Yellow maize (5%)	21.7	30.8	50.2	112.0	71.1	12.9	17.7
Yucca (8%)	26.0	31.8	34.8	160.9	87.7	12.4	21.7
Total (100%)	27.9	27.2	58.9	118.7	51.6	43.3	34.1

Notes: This table shows summary statistics for the 15 largest crops in terms of total sales. Our sample accounts for 82% of the total sales in Colombia. Parenthesis next to the name of the crop indicate the percentage of plots producing the respective crop. Revenues are given in Colombian pesos. Source: CNA-2014 DANE.

Table A2: The Impact of Distance to Urban Centers - all crops

	(1)	(2)	(3)	(4)
	OLS	OLS	IV	IV
Panel A: Dep var is the % of farmers trading				
Dist to UC	-0.054***	-0.067***	-0.087***	-0.088***
	(0.015)	(0.022)	(0.014)	(0.019)
R^2 or F	0.043	0.433	47.263	72.978
Panel B: Dep var is the % of land with soil improvements				
Dist to UC	-0.076***	-0.062***	-0.089***	-0.074***
	(0.010)	(0.009)	(0.017)	(0.008)
R^2 or F	0.105	0.306	47.263	72.978
Panel C: First stage regression				
Dist to Pre-Columbian Cities			0.798***	0.937***
			(0.116)	(0.111)
R^2			0.600	0.729
Obs	1118	1118	1118	1118
Departamento FE		✓		✓
Geography		✓		✓

Notes: Robust standard errors clustered at the nearest urban center level in parenthesis. Distance to urban centers is measured in 100 kilometers. Geographic variables are: altitude, ruggedness, distance to the closest river, total rainfall recoded in the closest gaugin station in 2013, and the share of the area in the municipality in six different climatological zones. There are 33 *departamento* fixed effects. OLS columns report R^2 and IV columns report F from Kleibergen-Paap weak instrument test.

Table A3: The Impact of Distance to Urban Centers using Alternative Variables

	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) OLS	(6) IV
Panel A: Dep var is the % of farmers exporting using narrow definition						
Dist to UC	-0.038***	-0.033**	-0.066***	-0.058***	-0.020***	-0.053***
	(0.011)	(0.014)	(0.010)	(0.006)	(0.007)	(0.010)
R^2 or F	0.059	0.373	51.563	73.922	0.461	73.922
Panel B: Dep var is the % of farmers exporting using narrow definition						
Dist to Port	-0.024***	-0.032**	-0.041***	-0.052***	-0.012	-0.047***
	(0.008)	(0.012)	(0.004)	(0.007)	(0.007)	(0.010)
R^2 or F	0.053	0.375	485.925	290.709	0.460	290.709
Panel C: Dep var is the % of farmers exporting according to the broad definition						
Dist to UC	-0.065***	-0.039***	-0.098***	-0.042**	-0.017***	-0.022***
	(0.018)	(0.013)	(0.026)	(0.019)	(0.005)	(0.006)
R^2 or F	0.037	0.480	51.563	73.922	0.510	73.922
Panel D: Dep var is the % of farmers exporting according to the broad definition						
Dist to Port	-0.063***	-0.046***	-0.061***	-0.037**	-0.013*	-0.020***
	(0.014)	(0.017)	(0.012)	(0.016)	(0.006)	(0.005)
R^2 or F	0.081	0.485	485.925	290.709	0.509	290.709
Panel E: Dep var is the % of land with soil improvements						
Dist to UC	-0.085***	-0.063***	-0.103***	-0.077***	-0.019***	-0.022***
	(0.012)	(0.009)	(0.019)	(0.007)	(0.003)	(0.003)
R^2 or F	0.107	0.355	51.563	73.922	0.402	73.922
Panel F: Dep var is the % of land with soil improvements						
Dist to Port	-0.055***	-0.058***	-0.064***	-0.068***	-0.020***	-0.020***
	(0.009)	(0.011)	(0.007)	(0.006)	(0.003)	(0.003)
R^2 or F	0.106	0.355	485.925	290.709	0.412	290.709

Notes: Robust standard errors in parenthesis clustered at the nearest urban center group. Distance to urban centers is measured in 100 kilometers. OLS columns report R^2 and IV columns report F from Kleibergen-Paap weak instrument test. Sample size is 1117. See Table 2 for explanations on each column.

Table A4: The Impact of Distance to Urban Centers and The Choice of Cash Crops - alternative measure

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	IV	IV	OLS	IV
Panel A: Dep var is the % of farmers producing cash crops						
Dist to UC	-0.045***	-0.032	-0.070***	-0.055**	-0.016**	-0.030***
	(0.016)	(0.023)	(0.020)	(0.022)	(0.006)	(0.007)
R^2 or F	0.024	0.280	51.562	73.918	0.500	73.918
Panel B: First stage regression						
Dist to Pre-Columbian Cities			0.791***	0.927***		0.927***
			(0.110)	(0.110)		(0.110)
R^2			0.598	0.726		0.726
Obs	1117	1117	1117	1117	1117	1117
Departamento FE		✓		✓	✓	✓
Geography		✓		✓	✓	✓

Notes: Robust standard errors clustered at the nearest urban center level in parenthesis. Cash crops include: rice, exportable banana, cocoa, coffee, sugarcane, yellow corn, African palm oil, pineapple, roses and carnations. Distance to urban centers is measured in 100 kilometers. Geographic variables are: altitude, ruggedness and the share of the area in the municipality in six different climatological zones. There are 33 *departamento* fixed effects. OLS columns report R^2 and IV columns report F from Kleibergen-Paap weak instrument test.

Table A5: First Stage Impact of Instrument Interacted with Trade Intensity

	Dep var is interaction of dist to UC with			
	Export Intensity	Export Intensity	Fertilizer Intensity	Fertilizer Intensity
	(1)	(2)	(3)	(4)
Share × Dist to Pre-Columbian Cities	0.776***	0.590***	0.547***	0.243***
	(0.048)	(0.057)	(0.063)	(0.043)
R^2	0.674	0.731	0.290	0.623
Obs	8000	8000	8000	8000
Municipalities	1095	1095	1095	1095
Crops	15	15	15	15
Fixed effects	✓	✓	✓	✓
Controls		✓		✓

Notes: Robust standard errors clustered at the municipality level in parenthesis. See table the table in the main paper for a description of controls.

Table A6: The Impact of Market Access on the Patterns of Agricultural Specialization - alternative definition of food and cash crops

	Dependent variable is log of					
	Farms	Farms	Sales	Sales	Area	Area
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Interaction with share of farmers producing cash crops (OLS)						
Share × Dist to UC	-0.062***	-0.038*	-0.150***	-0.166***	-0.121***	-0.147***
	(0.019)	(0.022)	(0.028)	(0.032)	(0.026)	(0.031)
R^2	0.337	0.368	0.383	0.403	0.347	0.372
Panel B: Interaction with share of farmers producing cash crops (IV)						
Share × Dist to UC	-0.504***	-1.004***	-0.676***	-1.257***	-0.610***	-1.329***
	(0.098)	(0.226)	(0.148)	(0.314)	(0.143)	(0.302)
F	42	21	65	31	42	21
Obs	8364	8364	8000	8000	8364	8364
Municipalities	1095	1095	1095	1095	1095	1095
Crops	18	18	15	15	18	18
Fixed effects	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓

Notes: Robust standard errors clustered at the municipality level in brackets. Explanatory variables are standardized. Cash crops include: rice, exportable banana, cocoa, coffee, sugarcane, yellow corn, African palm oil, pineapple, roses and carnations. For controls, we include: geographic variables such as altitude, ruggedness and the share of the area in the municipality in six different climatological zones; interactions of distance to urban centers with the share of farmers with high school in each crop and the share of farmers with machinery in each crop; interactions of the factor endowments in terms of machinery and farmers with high school with the share of farmers with high school in each crop and the share of farmers with machinery in each crop.

Table A7: The Impact of Market Access on the Patterns of Agricultural Specialization (Extensive Margin)

	Dep var is indicator if positive	
	Farms	Farms
	(1)	(2)
Panel A: Inter with share of farmers exporting (OLS)		
Share × Dist to UC	0.005	-0.029***
	(0.004)	(0.005)
R^2	0.393	0.425
Panel B: Inter with share of farmers with soil improv (OLS)		
Share × Dist to UC	-0.246***	-0.171***
	(0.038)	(0.036)
R^2	0.343	0.412
Panel C: Inter with share of farmers exporting (IV)		
Share × Dist to UC	-0.183***	-0.199***
	(0.022)	(0.029)
F	335	221
Panel D: Inter with share of farmers with soil improv (IV)		
Share × Dist to UC	-0.045***	0.001
	(0.005)	(0.004)
F	335	171
Obs	16425	16425
Municipalities	1095	1095
Crops	15	15
Fixed effects	✓	✓
Controls		✓

Notes: Robust standard errors clustered at the municipality level in brackets. Explanatory variables are based on extensive margin. For controls, we include: geographic variables such as altitude, ruggedness and the share of the area in the municipality in six different climatological zones; interactions of distance to urban centers with the share of farmers with high school in each crop and the share of farmers with machinery in each crop; interactions of the factor endowments in terms of machinery and farmers with high school with the share of farmers with high school in each crop and the share of farmers with machinery in each crop.

Table A8: The Impact of Distance to Urban Centers - Explanatory variable is in logs

	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) OLS	(6) IV
Panel A: Dep var is the % of farmers trading						
Log of Dist to UC	-0.065	-0.139*	-0.275***	-0.369***	-0.041**	-0.201***
	(0.073)	(0.077)	(0.079)	(0.099)	(0.016)	(0.050)
R^2 or F	0.006	0.402	106.825	17.264	0.500	17.264
Panel B: First stage regression						
Log of Dist to Pre-Columbian Cities			0.565***	0.562***		0.562***
			(0.055)	(0.137)		(0.137)
R^2			0.346	0.514		0.514
Obs	1117	1117	1117	1117	1117	1117
Departamento FE		✓		✓	✓	✓
Geography		✓		✓	✓	✓

Notes: Robust standard errors clustered at the nearest urban center level in parenthesis. Distance to urban centers is measured in 100 kilometers. Geographic variables are: altitude, ruggedness, distance to the closest river, total rainfall recorded in the closet gauging station, and the share of the area in the municipality in six different climatological zones. There are 33 *departamento* fixed effects. OLS columns report R^2 and IV columns report F from Kleibergen-Paap weak instrument test.


BANCO DE LA REPÚBLICA
COLOMBIA