ISSN 2451-5728

Serie Documentos de Trabajo del IIEP

Nº 55 - Septiembre de 2020

LAND MARKET DISTORTIONS AND AGGREGATE AGRICULTURAL PRODUCTIVITY: EVIDENCE FROM GUATEMALA

Braulio **Britos** - Manuel A. **Hernandez**
Miguel **Robles** - Danilo **Trupkin**

Britos, B. et al. (2020). Land market distortions and aggregate agricultural productivity: evidence from Guatemala. *Serie Documentos de Trabajo del IIEP, 55*, 1-44. http://iiep-baires.econ.uba.ar/documentos_de_trabajo

CONICET UBA FACULTAD DE CIENCIAS ECONÓMICAS IIEP - BAIRES

Instituto Interdisciplinario de Economía Política de Buenos Aires IIEP-BAIRES

http://iiep-baires.econ.uba.ar/

Instituto Interdisciplinario de
Economía Política de Buenos Aires
[IIEP-BAIRES]

Universidad de Buenos Aires, Facultad de Ciencias Económicas
Instituto Interdisciplinario de Economía Política de Buenos Aires
Av. Córdoba 2122 - 2º piso (C1120 AAQ)
Ciudad Autónoma de Buenos Aires, Argentina
Tel +54 11 5285-6578

http://iiep-baires.econ.uba.ar/

Consejo Nacional de Investigaciones Científicas y Técnicas
Ministerio de Ciencia, Tecnología e Innovación Productiva
Av. Rivadavia 1917 (C1033AAJ)
Ciudad Autónoma de Buenos Aires, Argentina
Tel +54 11 5983-1420

http://www.conicet.gov.ar/

ISSN 2451-5728

Los Documentos de Trabajo del IIEP reflejan avances de investigaciones realizadas en el Instituto y se publican con acuerdo de la Comisión de Publicaciones. L@s autor@s son responsables de las opiniones expresadas en los documentos.
Desarrollo editorial: Ed. Hebe Dato

LAND MARKET DISTORTIONS AND AGGREGATE AGRICULTURAL PRODUCTIVITY: EVIDENCE FROM GUATEMALA

Braulio Britos
UNIVERSITY OF MINNESOTA. ESTADOS UNIDOS.
brito032@umn.edu

Manuel A. Hernandez
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE. ESTADOS UNIDOS.
m.a.hernandez@cgiar.org

Miguel Robles
UNIVERSIDAD DEL PACÍFICO. PERÚ.
lmroblesf@up.edu.pe

Danilo Trupkin
UNIVERSIDAD DE BUENOS AIRES. FACULTAD DE CIENCIAS ECONÓMICAS. DEPARTAMENTO DE ECONOMÍA. BUENOS AIRES, ARGENTINA.
CONICET-UNIVERSIDAD DE BUENOS AIRES. INSTITUTO INTERDISCIPLINARIO DE ECONOMÍA POLÍTICA DE BUENOS AIRES (IIEP). BUENOS AIRES, ARGENTINA.
danilo.trupkin@fce.uba.ar

Distorsiones en los mercados de tierra y productividad agrícola agregada: evidencia de Guatemala

ABSTRACT

Farm size and land allocation are important factors in explaining lagging agricultural productivity in developing countries. This paper examines the effect of land market imperfections on land allocation across farmers and aggregate agricultural productivity. We develop a theoretical framework to model the optimal size distribution of farms and assess to what extent market imperfections can explain non-optimal land allocation and output inefficiency. We measure these distortions for the case of Guatemala using agricultural census microdata. We find that due to land market imperfections aggregate output is 19% below its efficient level for both maize and beans and 31% below for coffee, which are three major crops produced nationwide. The regions with higher distortions show a higher dispersion in land prices and less active rental markets. We also find that the degree of land market distortions across locations co-variate with road accessibility and ethnicity and, in a lower extent, with education.

RESUMEN

El tamaño del establecimiento y la asignación de tierras son factores importantes para la explicación del rezago de la productividad agrícola en los países en desarrollo. Este artículo examina formalmente el efecto de las distorsiones en los mercados de tierras sobre la asignación de este factor entre productores, y sobre la productividad agrícola agregada. Desarrollamos un marco teórico para modelar la distribución óptima del tamaño de los establecimientos y argumentamos hasta qué punto las distorsiones de mercado pueden explicar tanto la asignación subóptima de tierras como la (potencial) ineficiencia en la producción. Medimos estas distorsiones para el caso de Guatemala, utilizando microdatos del censo agropecuario. Encontramos que, debido a las imperfecciones en el mercado de tierras, el producto agregado es un 19% menor que su nivel eficiente tanto para maíz como para frijoles y un 31% menor para café, los tres principales cultivos producidos a nivel nacional. Las regiones con mayores distorsiones muestran una dispersión más alta en los precios de las tierras y mercados de arrendamiento menos activos. También hallamos que el grado de distorsión entre áreas covaría con la accesibilidad vial y la etnia y, en un menor grado, con la educación.

Keywords: Land market distortions - Output inefficiency - Agricultural productivity - Guatemala
Palabras claves: Distorsiones en mercados de tierra - Ineficiencia - Productividad agrícola - Guatemala

JEL Codes: O13, Q15 y O40

We thank the valuable comments of seminar participants at the European Economic Association Annual Meeting, Asian Meetings of the Econometric Society, Latin American and Caribbean Economic Association Annual Meeting, International Association of Agricultural Economists Triennial Congress, World Bank Conference on Land and Poverty, IFPRI and UNSAM. The study also benefited from helpful comments from Francisco Ceballos, Wenli Li, Francisco Olivet, Heitor Pellegrina, Maximo Torero and Rob Vos. We thank the Instituto Nacional de Estadística (INE) and the Secretaria de Seguridad Alimentaria y Nutricional (SESAN) of Guatemala for their help in accessing and collecting part of the data used in the analysis. All errors are our own.

1 Introduction

It is well established that agriculture plays a key role in explaining the large disparities in aggregate productivity between developing and developed countries (Caselli, 2005; Restuccia et al., 2008; Lagakos & Waugh, 2013). Poor countries employ most of their workers in agriculture and are much more unproductive than rich countries. As noted by Adamopolous & Restuccia (2014), farm size and land allocation are important factors in explaining this lagging agricultural productivity in poor countries. In particular, there are important differences in the size distribution of farms between rich and poor countries, where farms in poor countries have a much smaller operational scale and large farms have a significantly higher labor productivity than smaller ones. Further understanding farm size patterns, land productivity and allocation and the drivers of these processes is critical to reduce the agricultural productivity gap in developing countries.

The objective of this study is twofold. First, we formally assess the impact of land market imperfections on the allocation of land across farmers and on agricultural productivity, holding other factors constant. We develop a model with endogenous distributions of both land size and location to characterize land (mis)allocation in the presence of land market frictions in the agricultural sector. Second, we quantify the magnitude of these distortions on output efficiency using the case of Guatemala as an example. We also examine potential factors associated with these distortions by exploiting efficiency differences across locations, which can help to elaborate policies to improve efficiency in land markets.

We focus on white maize, black beans and coffee, which are three major crops produced nationwide and make up a large share of agricultural employment in Guatemala. The estimation results show that due to land market imperfections aggregate output is, on average, 81% of the efficient output for both maize and beans and 69% for coffee. Several robustness checks support these findings, suggesting that land market distortions could play a larger (negative) role among high-value cash crops relative to staple crops. We also find that areas with higher distortions (inefficiencies) exhibit a higher dispersion in land prices and less dynamic rental markets. Similarly, the extent of land market distortions across locations seem to co-variate with accessibility (road connectivity) and cultural aspects (ethnicity) and, in a lower extent, with the level of education.

The study ties into the general literature on factor misallocation across heterogeneous production units and productivity. Hopenhayn & Rogerson (1993) use the equilibrium model developed in Hopenhayn (1992) and show that dismissal taxes can distort labor allocation across firms and have important welfare losses through a decrease in average labor productivity (of over 2%). Similarly, Restuccia & Rogerson (2008) focus on capital misallocation and show that policies creating distortions on prices faced by producers can lead to large distortions on total factor productivity (TFP) and aggregate output (in the range of 30-50%). Across countries, Gollin et al. (2014) find evidence of labor misallocation between agricultural and non-agricultural sectors. The authors use microdata for 151 countries and show that output per worker in the agricultural sector is roughly half of the value in the non-agriculture sector, and the differences are more pronounced among developing countries.[1]

Closer to our study, Restuccia & Santaeulalia-Llopis (2017) study misallocation across household farms in Malawi. The authors estimate farms' TFP using detailed household-level survey data and find that input allocation is relatively constant across farms despite large differences in TFP. Their results indicate that agricultural productivity would increase by a factor of 3.6 if inputs were reallocated efficiently. Factor misallocation in their study is linked to restricted land markets as most of the land is directly assigned by village chiefs (i.e. are not marketed); as a result, the potential gains from reallocation are 2.6 times larger for farms with no marketed land relative to farms with marketed land. However, more recently, Gollin & Udry (2019) cast doubt on the role of land misallocation on agricultural productivity using survey panel data from farms in Tanzania and Uganda. The authors develop a framework that distinguishes between measurement error, unobserved heterogeneity and potential misallocation and find that measurement error and heterogeneity together account for nearly 90% of the dispersion in measured productivity. These findings suggest that the potential efficiency gains through land reallocation across farmers

[1] Other studies that analyze the link between factor misallocation, aggregate productivity and output include Hsieh & Klenow (2009), Bartelsman et al. (2013), David et al. (2016) and Bento & Restuccia (2017).

may be lower than previous findings. While we implement a different approach using (less detailed) nationwide census microdata, our results for Guatemala are highly robust across regions and closer to the estimates of Gollin & Udry (2019), particularly for maize and beans.

Our paper also contributes to the literature examining potential factors correlated with misallocation. On this matter, Restuccia & Rogerson (2017) review the literature on the effect of misallocation on productivity and conclude that there is no dominant source of misallocation as multiple factors seem to contribute to the total effect (e.g., taxes and regulations, preferential market access, subsidies and market imperfections such as market power and frictions). Adamopoulos & Restuccia (2017) evaluate the role of land quality and geography on agricultural productivity differences, and find that the rich-poor agricultural yield gap is not due to land quality differences but to a lower efficiency in crop production. Chen (2017) models the effect of untitled lands, which creates misallocation, on agricultural productivity and finds that land titling can increase productivity across countries by up to 82.5% (where about half of the increase results directly from eliminating land misallocation). Chen et al. (2019) assess the role of land markets on factor misallocation in Ethiopia, where the state owns the land, and show that land rentals significantly reduce misallocation and increase agricultural productivity. Similarly, Chamberlain & Ricker-Gilbert (2016) find evidence that rental markets contribute to efficiency gains in Malawi and Zambia by facilitating the transfer of land from less-able to more-able smallholder farmers.[2]

The remainder of the paper is structured as follows. Section 2 presents the theoretical model and its implications. Section 3 describes the data and the context of land distribution and productivity among the selected crops in Guatemala. Section 4 quantifies and discusses the distortions and resulting inefficiencies. Section 5 concludes and provides some policy recommendations.

[2]The list of papers is certainly more extensive, including the broader literature on the link between land tenure, institutions and agricultural productivity. For some recent studies, see Goldstein & Udry (2008), Besley et al. (2012), De Janvry et al. (2015), Jayne et al. (2016), Foster & Rosenzweig (2017) and Henderson & Isaac (2017).

2 Model

We develop a model featuring endogenous distributions of both land size and location to characterize agricultural land (mis)allocation. In each period the economy produces an agricultural good with both land and labor, while other factors are held constant across farmers. Land misallocation occurs because farmers who operate lands across locations (areas) face a transaction cost, which increases with distance.[3]

We abstract in the theoretical framework from other production factors as we are interested in characterizing the effect of land market imperfections on land allocation and output efficiency. If we explicitly include other factors in our modeled setup (such as improved inputs and machinery and equipment), the resulting distortions would naturally change due to the *span of control* (of adding more factors).[4] Yet, the magnitude of these changes ultimately depend on the assumed income shares of these other factors, which are typically small in developing countries compared to land and labor (see, e.g., Chen et al., 2019). In the empirical section, we still control for these other factors when modeling farm productivity and assess the sensitivity of our results to different income shares of land (i.e. importance of land relative to other factors in the production technology).

2.1 Set Up

The agricultural good is produced by a farmer with managerial skills s. We express land and production per unit of labor. In particular, we assume that a farmer of type i has the following simplified production function

$$y_i = s_i l_i^{\alpha}$$

where y_i is the agricultural output and l_i is land size, both normalized by the amount of labor employed at the farm level. The choice variable is the ratio land per unit of labor

[3] A location should be viewed as a delimited area of land such as a municipality, community, village or, in the limit, a plot. In the empirical section we define the geographic boundaries considered to perform the quantitative analysis.

[4] Further details are available upon request.

so the technology is characterized by decreasing returns to scale on that ratio.[5] Parameter $\alpha \in (0,1)$ captures land elasticity.

The farmer's managerial ability s follows a known time-invariant distribution with cumulative distribution function $F(s)$ and probability density function $f(s)$ with support $S = [\underline{s}, \overline{s}]$. Consider a discrete number of farmers with different managerial skills denoted by s_i, distributed across N locations. For simplicity, we assume that there is only one type of farmer per location, which implies that i hereafter identifies the location and farmer type.

Finally, we assume that there is a transaction cost for farmers who demand land across locations. Let $\tau_i(l_i^f)$ be the transaction cost that farmer i has to pay to operate in different locations, where $l_i^f \equiv l_i - \bar{l}_i$ and $\bar{l}_i$ is the (fixed) supply of land at the local market. We assume that $\tau_i(\cdot)$ increases both with distance (or lower accessibility) to other locations and with land size; i.e. $\tau_i'(\cdot) > 0$.[6]

These costs can be justified in several ways. They can be interpreted as the difficulties faced by farmers who demand land in other locations where information is more scarce and the lack of information increases with distance or lower accessibility; the existence of asymmetries in the form of certain (market) power from insiders; transportation costs for implementing effective managerial control; among other factors. These transaction costs (imperfections) result in misallocations in the land market. Below we quantify the role that the resulting land market distortions play in terms of welfare losses (output inefficiencies).

2.2 Farmer's Problem

A farmer with managerial ability s_i in location i demands land in order to maximize profits, taking the rental price q as given and subject to the non-negative constraints $l_i \geq 0$ and $l_i^f \geq 0$.

[5] Efficiency in our setting thereby involves a set of reallocations of land per unit of labor across farmers.
[6] Note that τ_i can also be interpreted as a tax rate, on the margin, for initially large farms.

The farmer's problem is defined as follows

$$\max_{l_i} \pi(s_i) = \left\{ s_i l_i^{\alpha} - q l_i - \tau_i(l_i^f) \mathbb{1}(l_i > \bar{l}) \right\}$$

where $\mathbb{1}(l_i > \bar{l})$ is an indicator function for the farmer's demand across locations.

The optimal condition for the i^{th} farmer is given by

$$\alpha s_i l_i^{\alpha-1} = q + \tau_i'(\cdot) \mathbb{1}(l_i > \bar{l}). \tag{1}$$

Without loss of generality, assume

$$\tau_i(l_i - \bar{l}) = \frac{\tau_i}{2}(l_i - \bar{l})^2.$$

Condition (1) becomes

$$\alpha s_i l_i^{\alpha-1} = q + \tau_i(l_i - \bar{l}) \mathbb{1}(l_i > \bar{l}). \tag{2}$$

Lastly, for every pair of farmers i and j, we have the following optimal relative allocation of land

$$\frac{l_i}{l_j} = \left(\frac{s_i}{s_j}\right)^{\xi} \left(\frac{q_i}{q_j}\right)^{-\xi} \tag{3}$$

where $q_i \equiv q + \tau_i(l_i - \bar{l})\mathbb{1}(l_i > \bar{l})$ denotes the total renting cost of land for farmer i and $\xi \equiv \frac{1}{1-\alpha}$.

2.3 Market Equilibrium

To solve for market equilibrium, we proceed as follows. The market clearing condition for the (aggregate) economy amount of land L is given by

$$L = \sum_{i=1}^{N} l_i. \tag{4}$$

Using conditions (3) and (4), we get the following expression for the individual land allocation of equilibrium

$$l_i = \left(\frac{s_i}{q_i}\right)^{\xi} \widetilde{L} \tag{5}$$

where $\widetilde{L} \equiv S_1^{-1} L$ and $S_1 \equiv \sum_{i=1}^{N} \left(\frac{s_i}{q_i}\right)^{\xi}$.

Then, individual output y_i becomes

$$y_i = \left(\frac{s_i}{q_i^{\alpha}}\right)^{\xi} \widetilde{L}^{\alpha}.$$

The aggregate (economy-level) output results in

$$Y = S_2 \widetilde{L}^{\alpha} \tag{6}$$

where $S_2 \equiv \sum_{i=1}^{N} \left(\frac{s_i}{q_i^{\alpha}}\right)^{\xi}$.

2.4 The Efficient Allocation

Expression (6) shows the aggregate output resulting from potential inefficiencies arising from the (mis)allocation of land (what we call *actual output*). This aggregate output can be compared with a (theoretical) aggregate output that would result from a social planner who solves a simple land-allocation problem given the overall distribution of farmers' productivity (what we call *efficient output*).

A market equilibrium without distortions would result in the same aggregate, efficient output as the output from the social-planner allocation. Consider, for instance, the special case in which there are no transaction costs; i.e. $\tau_i = 0$ for all i. In this case, the total rental cost of land becomes q for all farmers as land-market imperfections disappear.

Formally, we can define farmer i's efficient land allocation as l_i^*. Then, provided that $q_i = q$ for all i, we obtain from (5) the following expression for the efficient individual land size

$$l_i^* = s_i^{\xi} \widetilde{L}^* \tag{7}$$

where $\widetilde{L}^* \equiv S^{-1}L$ and $S \equiv \sum_{i=1}^{N} s_i^{\xi}$.

From expression (7), each farmer's efficient output is given by

$$y_i^* = s_i^{\xi}\widetilde{L}^{*\alpha}.$$

Summing up over the distribution of farmers, we obtain an expression for the aggregate (economy-level) efficient output equal to

$$Y^* = S^{1-\alpha}L^{\alpha}. \tag{8}$$

Finally, by comparing the actual output Y defined in (6) with the efficient output Y^* defined in (8), the ratio Y/Y^* offers a measure of the degree of output inefficiencies resulting from land misallocation. Figure 1 illustrates these inefficiencies considering a simple case with two farmer types (locations).

Using (6) and (8) we can explicitly derive the following output *efficiency ratio*

$$\begin{aligned} \frac{Y}{Y^*} &= \frac{S_2 S_1^{-\alpha}}{S^{1-\alpha}} \\ &= \frac{\left[\sum_{i=1}^{N}\left(\frac{s_i}{q_i^{\alpha}}\right)^{\xi}\right]\left[\sum_{i=1}^{N}\left(\frac{s_i}{q_i}\right)^{\xi}\right]^{-\alpha}}{\left[\sum_{i=1}^{N} s_i^{\xi}\right]^{1-\alpha}}. \end{aligned} \tag{9}$$

If $\tau_i = 0$ for all i (i.e. no land market imperfections), it can easily be shown that $S_1 = S_2 = S$ and $Y = Y^*$, which results in an efficiency ratio equal to 1.

We return to this discussion in Section 4 where we calculate the above ratio to quantify the magnitude of output inefficiencies for selected crops in Guatemala.

2.5 Model Implications

A characterization of the equilibrium results, both under the distorted (actual) and efficient (theoretical) land allocation, allows us to find several powerful model implications.

First, recall from equation (3) that the relative (distorted) land allocation between farmers i and j results in

$$\frac{l_i}{l_j} = \left(\frac{s_i}{s_j}\right)^{\xi} \left(\frac{q_j}{q_i}\right)^{\xi}.$$

This expression indicates that the relative land size of farmer i increases with managerial ability (s_i/s_j) and decreases with the renting cost (q_i/q_j), which is a function of the relative distortions (i.e. the transaction costs faced by each farmer).

Likewise, the relative efficient land allocation (without distortions) between farmers i and j is given by

$$\frac{l_i^*}{l_j^*} = \left(\frac{s_i}{s_j}\right)^{\xi}. \tag{10}$$

This expression shows that the relative efficient land size between farmers i and j solely depends on the managerial abilities (s_i/s_j).

Combining equations (3) and (10), we can derive a relationship between the two land allocations from which we can make some inference about the size distribution of farms. Let $z_{ij} \equiv l_i/l_j$ and $z_{ij}^* \equiv l_i^*/l_j^*$. Then,

$$\frac{z_{ij}}{z_{ij}^*} = \left(\frac{q_j}{q_i}\right)^{\xi}. \tag{11}$$

In general, we can characterize the following model implications from equations (3), (10) and (11):

1. The efficient land size of less productive farmers will be lower than the efficient land size of more productive ones.

2. The actual (likely distorted) land size of less productive farmers will not necessarily be lower than the actual land size of more productive ones.

3. The higher the transaction cost τ for farmer i, the lower her relative land size in equilibrium (z_{ij}) compared to the efficient ("ought-to-be") land size (z_{ij}^*).

4. The higher the land market distortions (the higher the τ's across farmers), the higher the price dispersion in the market (i.e. the higher the differences between any q_i and q_j) and the less efficient the allocation of land.

3 The Case of Guatemala

Guatemala is an interesting case study as it exhibits a large degree of heterogeneity in terms of climate, geography, ethnic composition and rural development. There is also a wide variation of agricultural activities; from large/medium- to low-scale farming and from high-value export crops such as coffee and sugar cane to food staple crops such as maize and beans. For the analysis below, we group the 22 departments of the country into six major geographic regions as shown in the map in Figure 2.[7] The departments within each region share similar socioeconomic, accessibility and agro-climatic conditions.

3.1 Data

The dataset used in the analysis is the microdata from the last census of agriculture in Guatemala, 'IV Censo Nacional Agropecuario 2003', corresponding to the crop year 2002-03 collected by the National Statistical Institute (INE). The census includes information on land size and use (for crops, cattle farming and other activities), production, labor and input use, machinery and equipment ownership, farmers' socioeconomic characteristics and geographic location. We focus on white maize, black beans and coffee, which results in a working sample of 396,317, 113,133 and 147,353 producers for each crop.[8]

Maize, beans and coffee are three major crops produced nationwide in Guatemala and together generate 62% of the agricultural employment (MAGA, 2011; MAGA, 2013). Maize, specially white maize, is by far the most common and extended food crop produced in the country with a annual harvested area of 841,094 hectares (Ha) and total production of 1,672,527 metric tonnes (MT) as of 2011/12; the major producer regions are Peten-Izabal (where maize production is combined with cattle farming activities), Western Highlands ('Altiplano Occidental') and Verapaces. Beans is the second major staple crop, which is

[7]We exclude the department of Guatemala from the Central region since the capital city is located in this department and there is a much lower presence of agricultural activities relative to other activities, as opposed to the other departments. The estimation results, however, are not sensitive to including this department.

[8]Around 2.5% of producers from the raw census data are excluded from the analysis due to missing observations, likely typos and extreme values for key variables of interest. Additional details are available upon request.

mainly produced for self-consumption and local markets across the country, with an annual harvested area of 238,140 Ha and production of 199,946 MT; the major producer regions are the Dry Corridor ('Corredor Seco'), Peten-Izabal and Western Highlands. Coffee is the second major export crop and is produced in multiple regions with an annual harvested area of 252,415 Ha and production of 245,752 MT; the major producer regions include Western Highlands, Pacífico-Bocacosta and Verapaces.[9] Focusing on these crops allows as to make comparisons across regions as well as to assess whether the inefficiencies resulting from potential land misallocation are more acute for certain types of crops, i.e. crops that involve small versus small/medium scale production, staple crops for subsistence and local markets versus cash crops for external markets.

Table 1 provides general descriptive statistics of the land size distribution in Guatemala. The top panel of the table shows the size distribution of landholdings dedicated to agricultural activities in the whole country and disaggregated by region. Farms under one hectare, considered as "infra-subsistence" farms by the Ministry of Agriculture (INE, 2006), comprise close to 65% of the landholdings in Guatemala whereas very large farms (over 20 hectares) comprise less than 2% of the landholdings.[10] The small size of landholdings is a regular pattern in developing countries as opposed to developed countries, where a large share of farms operate under much larger scales.[11]

The table further shows a large variation in the size distribution of landholdings across regions. For instance, 84% and 80% of the agricultural landholdings in the Central region and Western Highlands are smaller than one hectare; the case of the former is explained by the lower agricultural development in the central part of the country, while in the case of the latter this region is the poorest in terms of economic and rural development and

[9] Sugar cane is the main export crop in Guatemala but its production, which is basically large-scale farming (with a total harvested area of 239,261 MT), is concentrated in a specific region (Pacífico-Bocacosta), reason why we exclude it from the present study.

[10] As noted by Durr (2016), while the vast majority of farms in Guatemala are small, close to two thirds of the agricultural land in the country correspond to large-scale farms.

[11] See Lowder et al. (2016) and Lowder et al. (2019) for an extensive comparison of farm size and distribution across low-, lower-middle, upper-middle and high-income countries. Restuccia & Santaeulalia-Llopis (2017) report, for instance, that more than 81% and 46% of farms in the United States (US) and Belgium have more than 10 hectares, and only 15% of farms in Belgium have less than one hectare and none in the US.

there is a large presence of smallholder, subsistence agriculture. In contrast, the Petén-Izabal region combines agricultural with cattle farming activities and thus exhibits larger landholdings than the rest of the country.

The lower panels of the table present the land size distribution for the crops of interest. We generally observe a larger prevalence of smaller-scale farming in beans production across regions, relative to maize and coffee. The average landholding size dedicated to beans is 0.71 hectares versus 1.02 hectares for maize and 1.47 hectares for coffee.

Table 2 presents summary statistics of yields per worker by land size (for less than and more than one hectare) for each crop and region. Yields are a standard measure of agricultural land productivity defined as production (in quintals) per hectare.[12] Two interesting patterns emerge from the table. First, small farms mostly exhibit higher (and more dispersed) yields than large farms, which is indicative of decreasing returns to scale. This inverse relationship between land size and yields is commonly found in the literature (Barrett, 1996; Place, 2009; Barrett et al., 2010). Second, there are large differences in both the average and dispersion of yields across regions by crop, where Pacífico-Bocacosta seems to be the region with the highest yields and Verapaces the region with the lowest yields.

In Figure 3, we order the departments by quintiles of per capita volume of production and report the corresponding shares of landholdings of less than one hectare dedicated to each crop by quintile (left figures) and the shares of landholdings with more than 20 hectares (right figures). Q1 or Quintile 1 represents 20% of the departments with the lowest (per capita) production volumes, while Q5 or Quintile 5 represents 20% of the departments with the highest production volumes. We observe that among departments with lower production volumes, the share of "infra-subsistence" landholdings is in most cases larger than among departments with higher production volumes. The opposite is true for very large landholdings of more than 20 hectares, which have a larger relative presence among departments with higher production volumes. These aggregate production and land size patterns at the department level are in line with the cross-country evidence presented

[12] One quintal is equivalent to 100 pounds (46 kilograms).

in Adamopolous & Restuccia (2014) and indicative of potential land market distortions (inefficiencies).

3.2 A Measure of Farmer Productivity

We now turn to the calculation of our productivity measure at the farmer level for white maize, black beans and coffee using a regression-based approach. Following the theoretical setup, we are interested in deriving farmer-level productivity (ability) measures that will permit to compare hypothetical (efficient) land allocations with actual allocations to quantify the resulting output inefficiencies.

First, we derive a measure of managerial ability s for each farmer i based on the production function defined in Section 2

$$y_i = s_i l_i^{\alpha}.$$

We take natural logs of both sides of the equation and use the census microdata on output and land size (per unit of labor). We assume $\alpha = 0.3$, which is an intermediate value between the land income share estimated by Valentinyi & Herrendorf (2008) for the US ($\alpha = 0.18$) and the value estimated by Restuccia & Santaeulalia-Llopis (2017) for Malawi ($\alpha = 0.39$). The larger α value in Malawi is explained by the lower level of mechanization in the agriculture sector relative to, for example, the US. While a benchmark value of 0.30 is reasonable based on the level of agricultural development in Guatemala, which may also vary by crop, we additionally consider alternative values of 0.2 and 0.4 for sensitivity analysis.

Second, for each crop, we regress the derived s_i series on a set of control variables that include farmers' characteristics, input and machinery use, and agricultural practices. In particular, we account for the farmer's gender, age and years of education, share of family labor force, use of machinery and equipment, use of enhanced seeds, fertilizer and pesticides, if farm has an irrigation system and number of crops cultivated (as a proxy of specialization) plus administrative area (municipality) fixed effects.[13] The regression

[13] We also include an indicator variable that takes the value of one when the farmer's characteristics

results for the full sample are presented in Table A1 in the Appendix.[14]

The residual of this regression is our measure of farmer productivity.[15] Figure A1 in the Appendix plots the corresponding regional distributions (kernel densities) of the estimated productivities for each crop. It is clear that the distributions are very similar across regions, which is indicative of alike farmer-level heterogeneities across locations.[16] These comparable distribution patterns permit to infer that potential differences in output efficiencies across regions, discussed in the next section, are not necessarily driven by heterogeneity differences across locations.

Table 3 reports, in turn, summary statistics of the derived farmer productivities (per worker) by land size for each crop and region. Opposed to the yields presented in Table 2, farmer productivities are on average higher among farms of one or more hectares than among farms of less than one hectare. These opposite relationships between yields and productivity with land size are in line with the results of Aragon et al. (2019) for Uganda.[17] Overall, the positive correlation between productivity and land size suggests that land allocation is related in some degree to a farmer's ability. Next, we evaluate to what extent the current allocation of land is efficient.

(gender, age and years of education) are missing, and zero otherwise. For these cases, we assign the median value at the municipality level. The estimated distortions are though not sensitive to not imputing missing values for these variables (which results in fewer observations for the analysis) or excluding these variables from the regressions. Further details are available upon request.

[14]The coefficients of the control variables generally have the expected signs. We find, for example, a positive correlation across all crops between our s measure and education, use of equipment, presence of irrigation system and level of specialization (i.e. smaller number of different crops produced). Female producers, in contrast, are associated with a lower productivity as well as younger farmers (except for beans).

[15]In the estimations below we perform separate regressions by department in order to further account for potential heterogeneity in the partial correlations of the control variables across areas.

[16]The only exception is Peten-Izabal for coffee, which is precisely the crop-region where we have fewer observations; there are only 782 coffee producers in Peten-Izabal compared to 4,847-152,524 producers in the other crop-region pairs, as reported in Table 2.

[17]As noted by the authors, yields may capture farm productivity combined with decreasing returns to scale and market imperfections.

4 Quantitative Analysis

In this section, we quantify the magnitude of land misallocation for the three selected crops in Guatemala. We assess how the actual land allocation for each crop compares with a benchmark, efficient allocation chosen by a hypothetical social planner based on the estimated productivities. We also examine whether the resulting inefficiencies are in line with some of the model implications outlined in Section 2. Finally, we evaluate the potential channels that may explain the observed distortions across locations.

4.1 Output Efficiency Ratios

Our approach is built on Restuccia & Santaeulalia-Llopis (2017). First, we solve a simple optimization problem for a hypothetical social planner intending to maximize aggregate output by allocating land according to the distribution of farmers' productivity. The solution to this problem gives as a result the same aggregate output as the one described in expression (8) under market equilibrium without distortions. Second, we compare this efficient aggregate output with the output that results from the land size distribution found in the data, which is characterized in expression (6).

The efficient land allocation in a given location can be obtained by solving the following social planner problem

$$Y^* = \max_{\{l_i\}} \sum_{i=1}^{N} s_i l_i^{\alpha}, \quad \text{subject to} \quad L = \sum_{i=1}^{N} l_i$$

where Y^* denotes the efficient output.

The solution to the optimization problem is straightforward as the marginal product of land must be equal across farmers. The following is an expression for the efficient land allocation of an individual farmer

$$l_i^* = \frac{s_i^{1/(1-\alpha)}}{\sum_i s_i^{1/(1-\alpha)}} L.$$

Hence, the optimal land size of each farmer depends on her productivity relative to the whole distribution of productivities. Letting $S_i \equiv s_i^{1/(1-\alpha)} / \sum_i s_i^{1/(1-\alpha)}$, it follows that

$$Y^* = \sum_{i=1}^{N} s_i \left(S_i L\right)^{\alpha}.$$

Lastly, we compare the efficient output with the output under the current land allocation, defined as

$$Y = \sum_{i=1}^{N} s_i l_i^{\alpha}$$

where l_i is directly observed in the census microdata.

Table 4 presents the estimated efficiency ratios Y/Y^* by crop and region. The reported ratios are the corresponding averages of the departmental ratios in a region weighted by the number of producers in each department.[18] A higher efficiency ratio indicates that the current land allocation is closer to the optimal allocation from a social planner's perspective. We provide results for α values of 0.2, 0.3 and 0.4, where a larger value of α implies a lower level of mechanization in the production process.

The table shows varying degrees of inefficiencies among the selected crops that can be attributed to land misallocation (all else equal). For $\alpha = 0.3$, our benchmark value, the efficiency ratio ranges between 79% and 83.5% for white maize across regions, between 79.8% and 83.4% for black beans and between 60.3% and 71% for coffee. Interestingly, the regions with the largest efficiency ratios are generally the regions where a significant share of the production of each crop concentrates (except Peten-Izabal for coffee), while the region with the highest yields (Pacífico-Bocacosta) shows one of the lowest efficiency ratio.

Overall, we find a larger output efficiency for maize and beans, which are staple crops with a higher prevalence of small-scale and subsistence agriculture, relative to coffee, which is a high-value export crop characterized by small- and medium-scale production. The average efficiency ratio for both maize and beans is around 81%, which implies an aggregate

[18]Deriving an initial measure at the department level assumes that all farmers in a given department could eventually trade land between them. In general, the larger the geographic area, the larger the potential distortions and vice versa.

output gap between the current land allocation and the theoretically efficient allocation of one fifth; the average efficiency ratio for coffee is roughly 69%, which implies an output gap of about one third. In the hypothetical case that all land market distortions were removed, total output would increase to a lower extent for white maize and black beans than for coffee.[19] Although not reported, we find a positive correlation at the farmer level between the recovered q prices (based on optimal condition (2)) and our s measure of farmer productivity for all three crops, which further suggests that the estimated distortions affect relatively more productive farmers.

Note that all efficiency ratios increase when considering a lower α value (0.2) than the benchmark value of 0.3 and decrease when considering a higher α value (0.4), while the differences across crops and regions remain with these alternative values. This is explained by the fact that a larger α value implies a lower level of mechanization or, equivalently, a larger importance of land relative to other factors in the production technology and thereby results, ceteris paribus, in higher distortions from misallocating this factor across farmers. Considering the relatively lower level of mechanization in maize and beans production compared to coffee, the efficiency ratios for $\alpha = 0.4$ could be viewed as an estimated lower bound for these crops (76% on average), whereas the efficiency ratios for $\alpha = 0.2$ could be viewed as an estimated upper bound for coffee (77%).

Similarly, Table A2 in the Appendix reports efficiency ratios using land size instead of number of producers in each department as weights. The estimated output gaps and differences across crops and regions are not sensitive to this alternative weighting. The average efficiency ratio is around 82% for maize and beans and 65% for coffee (for $\alpha = 0.3$). Peten-Izabal is similarly the region with the largest efficiency ratios for each crop and Pacífico-Bocacosta exhibits one of the lowest ratios.

Lastly, in light of the recent discussion in Gollin & Udry (2019), we recognize that our results could be affected by the presence of other potential unobserved factors, such as measurement error and input quality, not accounted for in the estimations. The authors show, for example, that cross-plot measurement error within farms in Tanzania and Uganda

[19] Yellow maize shows a similar efficiency ratio than white maize, while sugar cane (produced in specific regions) shows an even lower efficiency ratio than coffee. Additional details are available upon request.

appears to be an important source of unobserved variation in productivity that can affect the estimated efficiency gains through land reallocation. While the lack of specific input-output plot-level data prevents us from implementing a similar approach as these authors, we can still assess the sensitivity of our results to excluding multi-plot farmers (52% of our working sample).[20] As shown in Table A3 in the Appendix, the resulting inefficiencies when only considering the subsample of farmers with one plot are very similar to when considering the full sample. The average efficiency ratio is 81% for maize and beans and 66% for coffee. These findings provide additional support to our base results.

4.2 Output Efficiency, Price Dispersion and Rental Markets

One of the implications from the theoretical framework described in Section 2 is that we should expect a higher dispersion in land prices among areas with larger market distortions (output inefficiencies). In particular, higher transaction costs τ in an area should result in more dispersed land prices and in a sub-optimal share of land transactions, which prevents the most productive farmers to work at their (larger) optimal scale. To explore this model implication, we rely on a complementary dataset from a three-year panel survey of households collected between 2012 and 2014 over half of the municipalities in the country.[21] The survey included a module on agricultural land markets that inquired about land prices and transactions. In one of the questions, households (smallholders) were asked to provide the price per hectare of what they would consider to be the most productive agricultural land in their municipality (i.e. in their immediate geographic area), which permits us to derive a measure of price dispersion at the municipality level.

Figure 4 plots the efficiency ratios (Y/Y^*) by municipality from the census microdata (for $\alpha = 0.3$) against the corresponding quartile coefficient of dispersion (QCD) of the reported land prices for the municipalities covered in the supplementary survey.[22] We ob-

[20]Hence, we assume that the amount of land in a given department is determined by the sum of land of all farmers reporting one plot in the department, which can trade land among them.

[21]The survey was part of the evaluation of a large-scale program executed by the Government of Guatemala against food insecurity and malnutrition and covered 176 of the 340 municipalities in the country, particularly the poorest and with the highest stunting rates.

[22]For ease of exposition, hereafter we present efficiency results for $\alpha = 0.3$. The estimated efficiency

serve an inverse efficiency-price dispersion relationship among all three crops and available regions, in line with the theoretical implication that more efficient areas (municipalities) should show a lower dispersion in land prices.[23]

Figure 5 presents, in turn, scatterplots of efficiency ratios and the share of agricultural land that is reported rented in a municipality. We observe in this case a positive correlation between these two variables indicating that those municipalities with a higher prevalence of land rentals are also those with seemingly lower land market distortions. Overall, we find that more efficient areas exhibit both a lower dispersion in land prices and more active rental markets.

4.3 Potential Channels of Distortions

We now turn to assess the potential channels or factors correlated with the estimated distortions and output inefficiencies by crop. For this purpose, we regress the efficiency ratios (Y/Y^*) at the municipality level on a set of indicators related to education, ethnicity, road and service accessibility and social conflict in the area, obtained from multiple data sources and different available years. These indicators include illiteracy rate, rate of indigenous population, road connectivity index, share of households with electricity and cellphones, and rate of extortions.[24] Table 5 presents the estimation results. The variables were first standardized for comparability purposes and the reported standard errors are clustered at the department level.[25] Columns (2), (4) and (6) include regional fixed effects.

ratios at the municipality level assumes that all farmers in a given municipality can only trade land between them. The QCD is equal to the difference between the 75-th and 25-th percentile of prices divided by the sum of both percentiles. In the plots we only include municipalities with at least ten price observations in the supplementary survey.

[23] The same pattern holds if we use instead the coefficient of price variation or if we use the price per hectare that the farmer valued her own land, after controlling for self-reported land quality (on a scale 1-10).

[24] The illiteracy rate is obtained from the Comision Nacional de Alfabetizacion (CONALFA) for 2014; rate of indigenous population from the Population Census (INE) for 2002; road connectivity index (weighted sum of paved and unpaved road kilometers normalized by extension area and population) from the Ministry of Agriculture (MAGA) for 2008; share of households with electricity and cellphones from the National Survey of Living Conditions (ENCOVI) for 2006; and rate of extortions from INFOSEGURA-Guatemala database for 2017.

[25] The number of observations differs across crops as not all crops are produced in all municipalities and the set of available indicators differs across locations.

We find that road accessibility and ethnicity appear to be positively correlated with the estimated efficiency ratios across all crops. The more connected the municipality in terms of paved (and unpaved) roads, normalized by extension area and population, the higher the efficiency ratio. Hence, higher transaction costs resulting from lower accessibility could be playing some role in explaining land market (output) inefficiencies in an area. Similarly, the larger the share of indigenous population in a municipality, the higher the efficiency ratio, which suggests that cultural aspects may also be explaining part of the observed distortions. This is correlated with the fact that within rural areas dominated by indigenous populations in Latin America, we typically observe more social cohesion (CEPAL, 2007); there is more trust and less information asymmetries. In rural areas with a lower prevalence of indigenous populations, there are probably more cultural barriers and information asymmetries between neighboring populations.

Regarding other indicators, illiteracy rate is negatively correlated with output efficiency, at least for white maize and black beans. All else equal, we expect a higher market dynamism and subsequent better land allocation among locations with more educated people. The rate of extortions, which is a proxy of social conflict in an area, is likewise negatively associated with output efficiency, although the correlations are only statistically significant at conventional levels in very few cases (for maize and coffee). Access to services such as electricity and cellphones, which act as proxies of easiness of information flow and development in an area, do not seem to be correlated with output efficiency.

5 Concluding Remarks

Farm size and land allocation play an important role in explaining lagging agricultural productivity in developing countries. This paper assesses the impact of land market distortions on land allocation and aggregate agricultural productivity. We develop a theoretical model to examine to what extent market distortions can explain non-optimal land allocation and output efficiency. We then quantify these distortions using census microdata from Guatemala. The estimation results show that due to land market imperfections aggregate output is roughly 81% of the efficient output for both white maize and black beans and 69% for coffee, which are the three major crops produced across the country. More efficient areas seem to exhibit a lower dispersion in land prices and more dynamic rental markets, in line with our theoretical discussion. We also find that the degree of land market distortions across locations appear to co-variate with road accessibility and ethnicity and, in a lower extent, with the level of education.

While there are some variations in efficiencies across regions, the overall findings indicate the presence of larger distortions for high-value export crops such as coffee (with an estimated output gap of one third), relative to staple crops such as maize and beans (with an output gap of one fifth). This suggests that the latter, with a higher share of small-scale and subsistence agriculture, may already be operating close to its maximum production potential such that eliminating land market distortions will have a smaller effect on reaching their optimal output level. In contrast, the elimination of these distortions for coffee could have a larger effect on its aggregate productivity and expansion of the agricultural sector towards more high-value cash crops.

The analysis examining potential factors associated with output inefficiencies suggest, for example, the importance of continuing improving accessibility and education as well as further recognizing and addressing likely cultural barriers. Certainly, policies in this regard, such as investment in road infrastructure and education, will require some time to become effective, while overcoming cultural differences may even take more time. For now, areas with a higher prevalence of indigenous population already seem to be operating more efficiently. Considering that the mobile penetration rate in rural Guatemala is over 90%, market information systems exploiting new technologies of information could also help, at least in the short-term, to develop or expand rental land markets across the country, maybe within areas that share similar cultural (ethnic) characteristics, and reallocate land from less to more productive producers. Pilot programs to assess farmers' willingness to rent in/out land and whether providing market information effectively contributes to the generation of rental markets, are an avenue of future work along these lines.

Finally, while several robustness checks support our main findings, we acknowledge that we cannot fully discard the presence of potential measurement error and other unobservables in our estimations. In this regard, our results should be interpreted with caution given the nature of our data. Similarly, our analysis is based on data from the 2003 agricultural census, while a more recent census in the country is still lacking. If, for instance, production technologies have generally improved in Guatemala over the past years, our estimation approach would imply lower gains from an efficient allocation of land, holding constant other factors.

References

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

Adamopoulos, T. & Restuccia, D. (2017). Geography and agricultural productivity: Cross-country evidence from micro plot-level data. Society for Economic Dynamics, 2017 Meeting Papers No. 1180.

Aragon, F., Restuccia, D., & Rud, J. P. (2019). Are small farms really more productive than large farms? NBER Working Paper No. 26331.

Barrett, C. B. (1996). On price risk and the inverse farm size-productivity relationship. *Journal of Development Economics*, 51(2), 193–215.

Barrett, C. B., Bellemare, M. F., & Hou, J. Y. (2010). Reconsidering conventional explanations of the inverse productivity–size relationship. *World Development*, 38(1), 88–97.

Bartelsman, E., Haltiwanger, J., & Scarpetta, S. (2013). Cross-country differences in productivity: The role of allocation and selection. *American Economic Review*, 103(1), 305–334.

Bento, P. & Restuccia, D. (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics*, 9(3), 267–303.

Besley, T. J., Leight, J., Pande, R., & Rao, V. (2012). The regulation of land markets: Evidence from tenancy reform in India. C.E.P.R. Discussion Paper No. 8736.

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of economic growth*, 1, 679–741.

CEPAL (2007). Cohesion social: Inclusion y sentido de pertenencia en America Latina y el Caribe. Naciones Unidas. Chile: Santiago de Chile [In Spanish].

Chamberlain, J. & Ricker-Gilbert, J. (2016). Participation in rural land rental markets in Sub-Saharan Africa: Who benefits and by how much? Evidence from Malawi and Zambia. *American Journal of Agricultural Economics*, 98(5), 1507–1528.

Chen, C. (2017). Untitled land, occupational choice, and agricultural productivity. *American Economic Journal: Macroeconomics*, 9(4), 91–121.

Chen, C., Restuccia, D., & Santaeulalia-Llopis, R. (2019). The effects of land markets on resource allocation and agricultural productivity. NBER Working Paper No. 24034.

David, J. M., Hopenhayn, H. A., & Venkateswaran, V. (2016). Information, misallocation, and aggregate productivity. *The Quarterly Journal of Economics*, 131(2), 943–1005.

De Janvry, A., Emerick, K., Gonzalez-Navarro, M., & Sadoulet, E. (2015). Delinking land rights from land use: Certification and migration in Mexico. *American Economic Review*, 105(10), 3125–3149.

Durr, J. (2016). The political economy of agriculture for development today: The "small versus large" scale debate revisited. *Agricultural Economics*, 47, 671–681.

Foster, A. D. & Rosenzweig, M. R. (2017). Are there too many farms in the world? Labor-market transaction costs, machine capacities and optimal farm size. NBER Working Paper No. 23909.

Goldstein, M. & Udry, C. (2008). The profits of power: Land rights and agricultural investment in Ghana. *Journal of Political Economy*, 116(6), 981–1022.

Gollin, D., Lagakos, D., & Waugh, M. E. (2014). The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2), 939–993.

Gollin, D. & Udry, C. (2019). Heterogeneity, measurement error and misallocation: Evidence from African agriculture. NBER Working Paper No. 25440.

Henderson, H. & Isaac, A. G. (2017). Modern value chains and the organization of agrarian production. *American Journal of Agricultural Economics*, 99(2), 379–400.

Hopenhayn, H. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60(5), 1127–1150.

Hopenhayn, H. & Rogerson, R. (1993). Job turnover and policy evaluation: A general equilibrium analysis. *Journal of Political Economy*, 101(5), 915–938.

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

INE (2006). Encuesta Nacional Agropecuaria (ENA) 2006. Instituto Nacional de Estadistica. Guatemala: Ciudad de Guatemala [In Spanish].

Jayne, T. S., Chamberlin, J., Traub, L., Sitko, N., Muyanga, M., Yeboah, F. K., Anseeuw, W., Chapoto, A., Wineman, A., Nkonde, C., et al. (2016). Africa's changing farm size distribution patterns: The rise of medium-scale farms. *Agricultural Economics*, 47(S1), 197–214.

Lagakos, D. & Waugh, M. E. (2013). Selection, agriculture and cross-country productivity differences. *American Economic Review*, 103(2), 948–980.

Lowder, S. K., Sanchez, M. V., & Bertini, R. (2019). Farms, family farms, farmland distribution and farm labour: What do we know today? FAO Agricultural Development Economics Working Paper 19-08. Rome, FAO.

Lowder, S. K., Skoet, J., & Raney, T. (2016). The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Development*, 87, 16–29.

MAGA (2011). El Agro en Cifras 2011. Ministerio de Agricultura, Ganaderia y Alimentacion. Guatemala: Ciudad de Guatemala [In Spanish].

MAGA (2013). El Agro en Cifras 2013. Ministerio de Agricultura, Ganaderia y Alimentacion. Guatemala: Ciudad de Guatemala [In Spanish].

Place, F. (2009). Land tenure and agricultural productivity in Africa: A comparative analysis of the economics literature and recent policy strategies and reforms. *World Development*, 37(8), 1326–1336.

Restuccia, D. & Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous plants. *Review of Economic Dynamics*, 11(4), 707–720.

Restuccia, D. & Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3), 151–174.

Restuccia, D. & Santaeulalia-Llopis, R. (2017). Land misallocation and productivity. NBER Working Paper No. 23128.

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

Valentinyi, A. & Herrendorf, B. (2008). Measuring factor income shares at the sectoral level. *Review of Economic Dynamics*, 11(4), 820–835.

Figure 1: Equilibrium in simple model with two farmer types (locations)

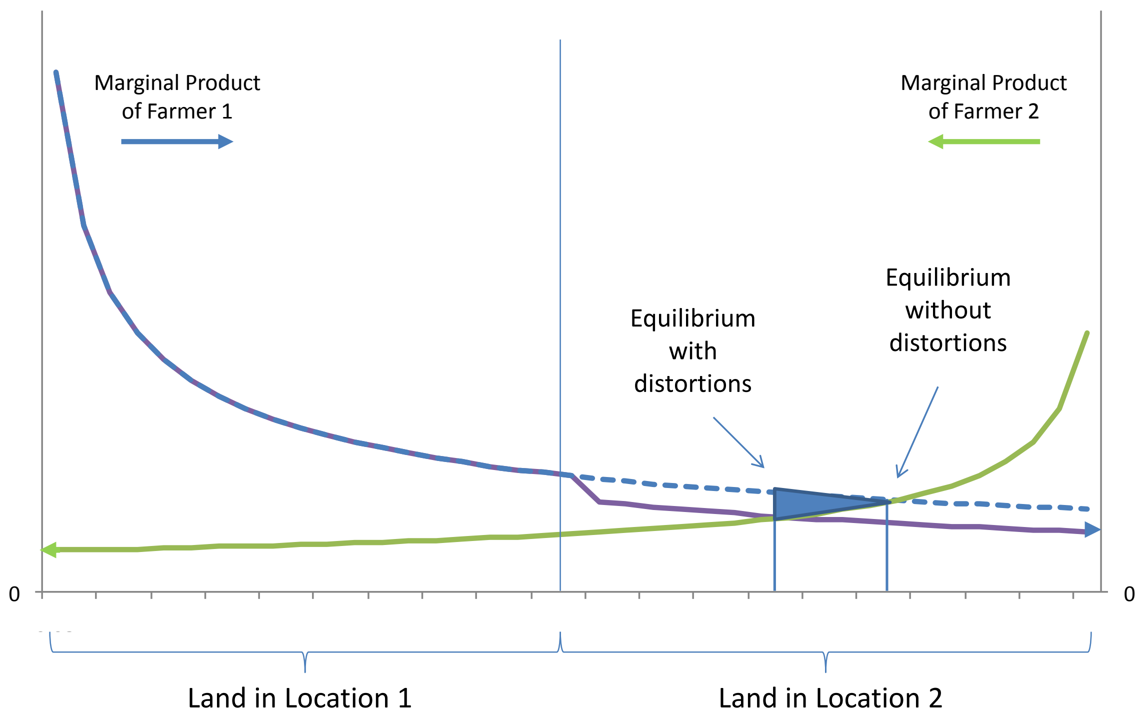


Figure 2: Map of Guatemala and regions considered

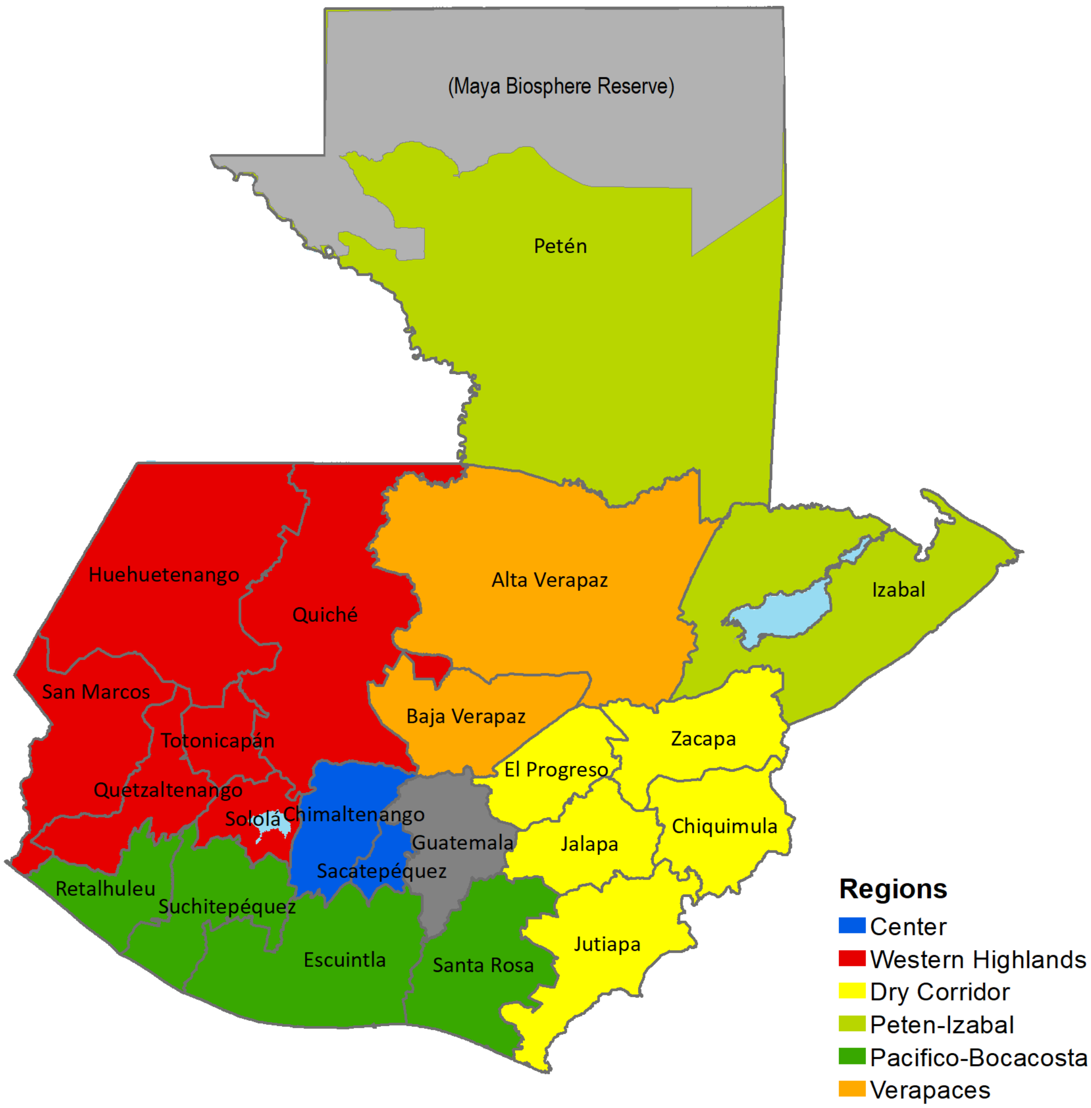
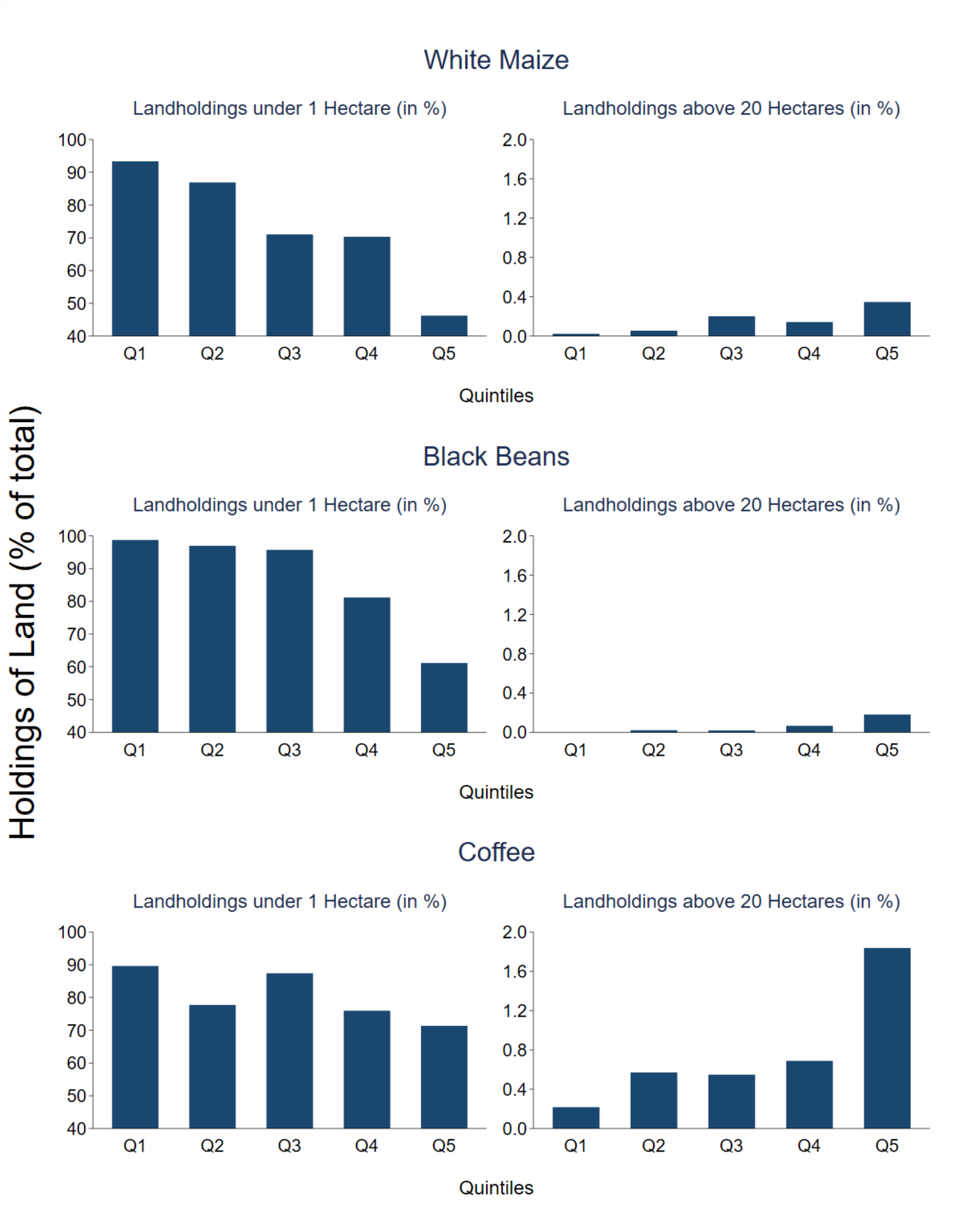
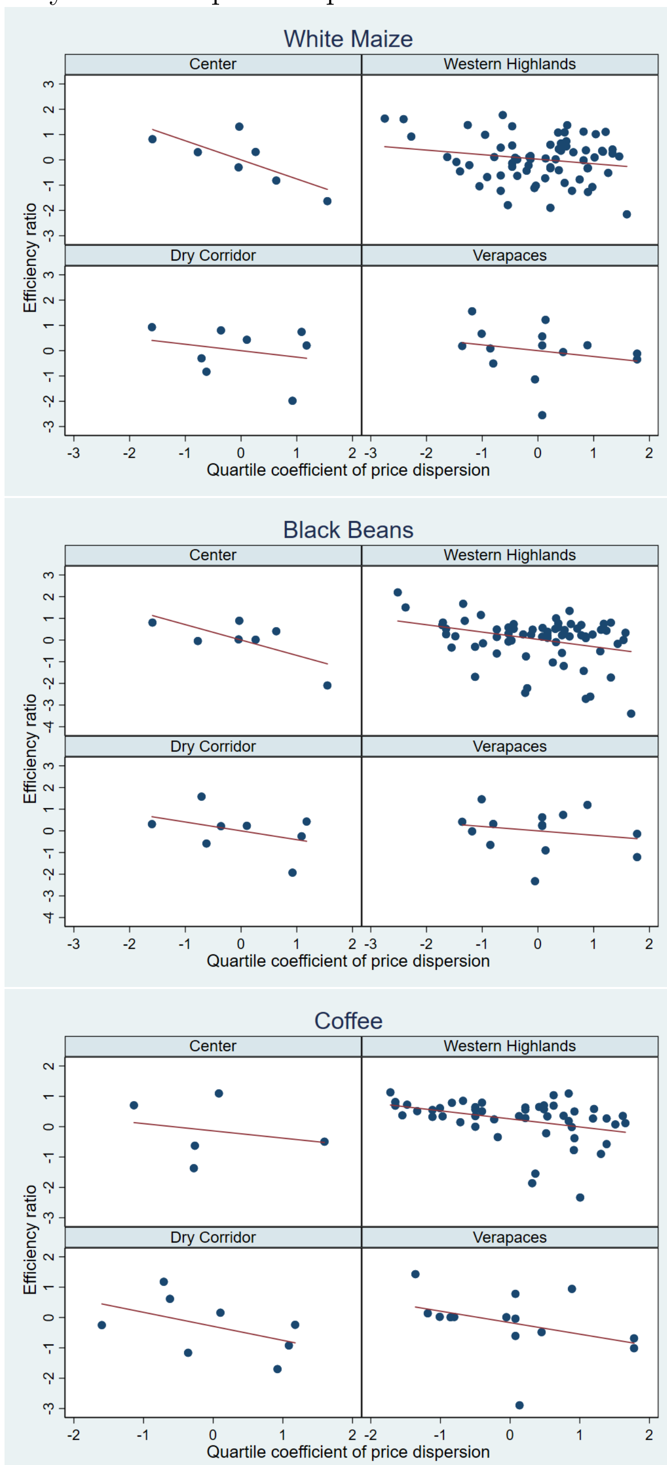


Figure 3: Share of small and very large landholdings across departments by quintile of per capita production volumes



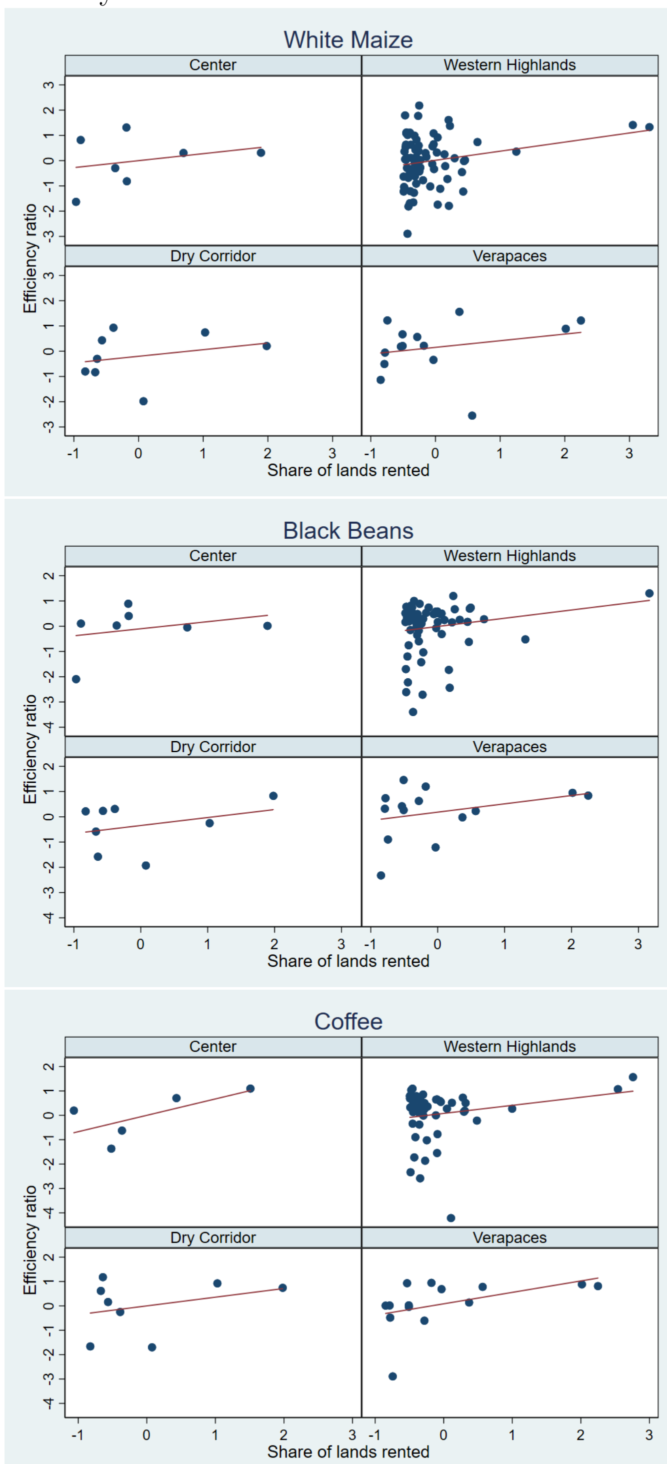
Note: The departments were ordered by quintiles of per capita volume of production.

Figure 4: Efficiency ratio and price dispersion of best land in each municipality



Note: The quartile coefficient of price dispersion is equal to the difference between the 75-th and 25-th percentile of land prices divided by the sum of both percentiles, where price is the price per hectare that surveyed farmers considered to be the most productive agricultural land in their municipality. The efficiency ratios are derived assuming $\alpha = 0.3$. Variables are standardized for comparability purposes across regions within each crop.

Figure 5: Efficiency ratio and share of lands rented in each municipality



Note: The efficiency ratios are derived assuming $\alpha = 0.3$. Variables are standardized for comparability purposes across regions within each crop.

Table 1: Size distribution of landholdings devoted to agriculture and by crop

Landholding size	All regions	Center	Western Highlands	Dry Corridor	Peten-Izabal	Pacifico-Bocacosta	Verapaces
Less than 1 Ha	64.7%	83.9%	79.9%	53.2%	17.7%	60.3%	39.4%
1-2 Ha	17.2%	11.9%	12.0%	27.1%	17.5%	18.2%	26.2%
2-5 Ha	11.9%	3.3%	5.8%	15.3%	29.0%	15.2%	23.6%
5-10 Ha	2.9%	0.4%	1.4%	2.5%	11.1%	2.9%	6.2%
10-20 Ha	1.6%	0.2%	0.6%	1.0%	9.4%	1.4%	2.9%
More than 20 Ha	1.7%	0.3%	0.3%	0.8%	15.3%	2.1%	1.6%
White maize							
Less than 1 Ha	73.3%	95.5%	90.5%	70.8%	20.9%	66.3%	65.1%
1-2 Ha	15.1%	3.6%	7.0%	20.2%	25.1%	18.0%	25.4%
2-5 Ha	9.3%	0.7%	2.2%	7.8%	40.4%	12.8%	8.8%
5-10 Ha	1.6%	0.1%	0.2%	0.9%	10.2%	1.9%	0.6%
10-20 Ha	0.5%	0.0%	0.1%	0.3%	2.7%	0.7%	0.1%
More than 20 Ha	0.2%	0.0%	0.0%	0.1%	0.7%	0.4%	0.1%
Black Beans							
Less than 1 Ha	82.1%	97.4%	97.0%	79.1%	43.7%	92.3%	94.7%
1-2 Ha	10.2%	2.1%	2.3%	14.4%	25.0%	5.4%	3.9%
2-5 Ha	6.5%	0.5%	0.6%	5.5%	25.8%	1.7%	1.2%
5-10 Ha	1.0%	0.0%	0.1%	0.7%	4.2%	0.4%	0.1%
10-20 Ha	0.2%	0.0%	0.0%	0.2%	1.0%	0.1%	0.0%
More than 20 Ha	0.1%	0.0%	0.0%	0.1%	0.3%	0.1%	0.0%
Coffee							
Less than 1 Ha	81.0%	78.7%	82.4%	75.4%	74.8%	62.8%	92.6%
1-2 Ha	10.6%	13.4%	11.1%	12.1%	13.2%	16.9%	4.6%
2-5 Ha	5.8%	4.0%	4.7%	9.1%	9.4%	13.6%	1.9%
5-10 Ha	1.2%	1.3%	0.8%	2.1%	1.6%	3.0%	0.4%
10-20 Ha	0.5%	0.8%	0.3%	0.8%	0.3%	1.3%	0.2%
More than 20 Ha	0.8%	1.8%	0.7%	0.5%	0.7%	2.5%	0.3%

Note: Calculations based on size of landholdings dedicated to agriculture in the top panel and to the production of white maize, black beans and coffee in the other panels. The Center region includes the departments of Sacatepequez and Chimaltenango; Western Highlands includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Dry Corridor includes Chiquimula, Jutiapa, Jalapa, El Progreso and Zacapa; Peten-Izabal includes Peten and Izabal; Pacifico-Bocacosta includes Retalhuleu, Suchitepequez, Escuintla and Santa Rosa; and Verapaces includes Alta Verapaz and Baja Verapaz.

Table 2: Average and dispersion of yields per worker by land size

	All regions		Center		Western Highlands		Dry Corridor		Peten-Izabal		Pacifico-Bocacosta		Verapaces	
	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha
							White Maize							
Mean	16.0	12.0	18.2	10.4	15.7	11.3	13.3	7.9	18.4	14.0	28.0	19.0	11.3	9.1
St. Dev.	15.6	12.4	16.6	10.5	14.4	11.9	13.6	9.0	15.4	12.5	22.5	16.5	11.4	10.2
IQR (P75-P25)	16.2	12.3	15.2	9.8	17.2	11.1	12.6	7.0	17.2	14.2	25.8	15.7	11.7	9.4
Observations	289,338	106,979	30,669	1,460	138,173	14,351	39,841	16,448	9,091	35,653	24,181	12,836	47,383	26,231
							Black Beans							
Mean	3.9	3.5	4.0	2.9	4.1	3.0	4.0	2.5	4.8	4.4	5.1	2.6	3.2	2.4
St. Dev.	3.8	3.4	3.7	3.0	4.0	3.7	3.8	2.7	3.8	3.7	4.7	2.8	3.3	2.6
IQR (P75-P25)	3.9	3.6	3.0	3.0	4.3	2.8	3.7	2.3	5.2	4.4	5.2	2.8	3.0	2.4
Observations	92,650	20,483	8,121	221	21,673	629	26,774	7,078	8,397	10,868	4,468	379	23,217	1,308
							Coffee							
Mean	16.0	9.6	17.3	11.3	16.7	8.2	16.5	9.7	9.3	11.3	24.5	12.5	11.1	6.4
St. Dev.	21.9	15.0	21.0	14.9	21.9	12.3	23.0	16.3	13.8	19.4	30.1	18.3	16.0	10.1
IQR (P75-P25)	15.5	8.9	17.6	10.4	18.7	7.7	16.1	8.8	10.0	10.8	22.9	11.2	10.5	5.8
Observations	119,306	28,047	5,702	1,543	53,388	11,367	16,539	5,481	599	183	11,745	6,987	31,333	2,486

Note: The Center region includes the departments of Sacatepequez and Chimaltenango; Western Highlands includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Dry Corridor includes Chiquimula, Jutiapa, Jalapa, El Progreso and Zacapa; Peten-Izabal includes Peten and Izabal; Pacifico-Bocacosta includes Retalhuleu, Suchitepequez, Escuintla and Santa Rosa; and Verapaces includes Alta Verapaz and Baja Verapaz. The Interquartile Range (IQR) is equal to the difference between the 75-th percentile and 25-th percentile values. Yields are expressed in quintals per hectare (per worker).

Table 3: Average and dispersion of productivity per worker by land size

	All regions		Center		Western Highlands		Dry Corridor		Peten-Izabal		Pacifico-Bocacosta		Verapaces	
	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha	< 1 Ha	>= 1 Ha
							White Maize							
Mean	-0.11	0.31	-0.03	0.73	-0.06	0.54	-0.15	0.37	-0.45	0.11	-0.23	0.44	-0.18	0.33
St. Dev.	0.66	0.68	0.63	0.66	0.66	0.70	0.62	0.65	0.63	0.65	0.68	0.65	0.65	0.68
IQR (P75-P25)	0.87	0.89	0.83	0.87	0.88	0.90	0.82	0.84	0.81	0.86	0.89	0.85	0.87	0.88
Observations	289,338	106,979	30,669	1,460	138,173	14,351	39,841	16,448	9,091	35,653	24,181	12,836	47,383	26,231
							Black Beans							
Mean	-0.11	0.48	-0.03	1.00	-0.04	1.25	-0.12	0.45	-0.47	0.37	-0.06	0.68	-0.06	1.02
St. Dev.	0.76	0.74	0.68	0.69	0.77	0.81	0.70	0.70	0.71	0.68	0.72	0.77	0.83	0.85
IQR (P75-P25)	1.00	0.95	0.92	0.96	1.01	1.23	0.93	0.91	0.91	0.9	0.98	0.95	1.11	1.13
Observations	92,650	20,483	8,121	0,221	21,673	629	26,774	7,078	8,397	10,868	4,468	379	23,217	1,308
							Coffee							
Mean	-0.17	0.71	-0.17	0.61	-0.15	0.69	-0.24	0.74	-0.25	0.80	-0.36	0.61	-0.08	1.05
St. Dev.	0.99	1.08	0.97	1.02	0.94	1.03	1.10	1.11	1.20	1.82	1.01	1.06	1.00	1.21
IQR (P75-P25)	1.28	1.31	1.25	1.24	1.20	1.24	1.47	1.42	1.71	2.90	1.28	1.28	1.31	1.46
Observations	119,306	28,047	5,702	1,543	53,388	11,367	16,539	5,481	599	183	11,745	6,987	31,333	2,486

Note: The Center region includes the departments of Sacatepequez and Chimaltenango; Western Highlands includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Dry Corridor includes Chiquimula, Jutiapa, Jalapa, El Progreso and Zacapa; Peten-Izabal includes Peten and Izabal; Pacifico-Bocacosta includes Retalhuleu, Suchitepequez, Escuintla and Santa Rosa; and Verapaces includes Alta Verapaz and Baja Verapaz. The Interquartile Range (IQR) is equal to the difference between the 75-th percentile and 25-th percentile values.

Table 4: Efficiency ratios (Y/Y^*)

Region	White Maize			Black Beans			Coffee		
	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$
All regions	86.7%	81.0%	75.8%	86.5%	80.8%	75.5%	77.3%	68.7%	61.3%
Center	87.4%	81.9%	76.8%	87.7%	82.4%	77.4%	70.0%	60.3%	52.9%
Western Highlands	86.1%	80.4%	75.2%	86.0%	80.1%	74.6%	78.0%	69.6%	62.3%
Dry Corridor	86.8%	81.1%	75.6%	86.1%	80.2%	74.8%	78.3%	69.8%	62.2%
Peten-Izabal	88.6%	83.5%	78.6%	88.5%	83.4%	78.7%	79.6%	71.0%	63.0%
Pacifico-Bocacosta	84.9%	79.0%	73.8%	85.8%	79.8%	74.4%	74.0%	64.8%	57.4%
Verapaces	87.1%	81.4%	75.9%	85.8%	79.8%	74.2%	78.7%	70.1%	62.6%

Note: The regional ratios reported are the corresponding averages of the departmental ratios in a region weighted by the number of producers in each department. The ratio for All regions is the weighted average across regions. The Center region includes the departments of Sacatepequez and Chimaltenango; Western Highlands includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Dry Corridor includes Chiquimula, Jutiapa, Jalapa, El Progreso and Zacapa; Peten-Izabal includes Peten and Izabal; Pacifico-Bocacosta includes Retalhuleu, Suchitepequez, Escuintla and Santa Rosa; and Verapaces includes Alta Verapaz and Baja Verapaz.

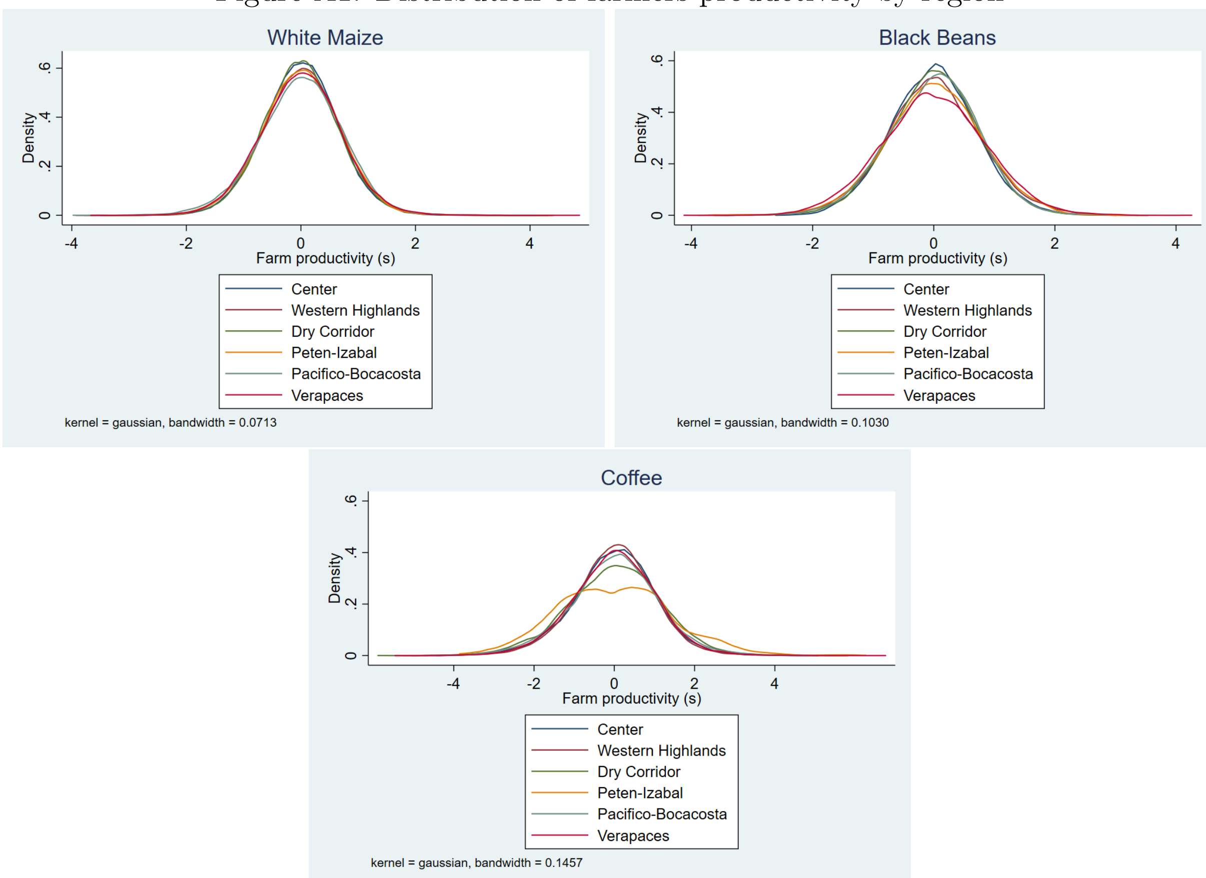
Table 5: Regression of efficiency ratio (Y/Y^*) on indicators at municipality level

Coefficient	(1)	(2)	(3)	(4)	(5)	(6)
	White Maize		Black Beans		Coffee	
Iliteracy rate	-0.143**	-0.137**	-0.125***	-0.092*	-0.072	-0.168
	(0.068)	(0.057)	(0.039)	(0.044)	(0.121)	(0.122)
Share of indigenous population	0.197*	0.276*	0.066	0.151*	0.201*	0.330**
	(0.098)	(0.145)	(0.076)	(0.079)	(0.105)	(0.151)
Road connectivity index	0.133***	0.133**	0.165**	0.180***	0.145*	0.052
	(0.045)	(0.057)	(0.061)	(0.059)	0.087	(0.105)
Share of households with electricity	-0.020	-0.011	-0.084	-0.071	-0.071	-0.084
	(0.064)	(0.056)	(0.049)	(0.052)	(0.097)	(0.087)
Share of households with cell phones	0.067	0.077	0.041	0.051	-0.029	-0.021
	(0.058)	(0.057)	(0.054)	(0.050)	(0.084)	(0.066)
Rate of extortions	-0.337*	-0.291	-0.039	-0.012	-0.350**	-0.102
	(0.192)	(0.246)	(0.096)	(0.114)	(0.141)	(0.199)
Constant	0.011	-0.068	-0.064	-0.128	-0.201	-0.451**
	(0.125)	(0.139)	(0.084)	(0.127)	(0.192)	(0.176)
Regional fixed effects	No	Yes	No	Yes	No	Yes
Observations	139	139	126	126	108	108
R-squared	0.187	0.211	0.090	0.151	0.130	0.286

Note: Each observation corresponds to a municipality. The efficiency ratios correspond to $\alpha = 0.3$. Variables standardized prior to the regression. Standard errors reported in parentheses clustered at the department level. ***, ** and * denotes significance at 1%, 5% and 10% level.

Appendix

Figure A1: Distribution of farmers productivity by region



Note: Farmers productivity derived based on full-sample estimations depicted in Table A1.

Table A1: Regression of Ln s measure on set of characteristics at farmer level, full sample

Coefficient	(1) White Maize	(2) Black Beans	(3) Coffee
Age	0.001*	-0.002***	0.004***
	(0.000)	(0.000)	(0.001)
Years of schooling	0.003*	0.007**	0.039***
	(0.001)	(0.003)	(0.006)
If female	-0.195***	-0.173***	-0.267***
	(0.009)	(0.015)	(0.023)
Household Labor / Total Labor	0.627***	0.641***	0.148
	(0.072)	(0.046)	(0.133)
If has machinery	0.127***	0.008	0.178***
	(0.028)	(0.017)	(0.036)
If has equipment	0.074***	0.050***	0.126***
	(0.015)	(0.009)	(0.028)
If uses high-performance seeds	0.069***	0.074***	-0.004
	(0.011)	(0.024)	(0.045)
If uses organic fertilizer	0.005	0.064***	0.128***
	(0.010)	(0.022)	(0.027)
If uses chemical fertilizer	0.062**	0.045	0.080*
	(0.029)	(0.047)	(0.043)
If uses pesticide	0.101***	0.019	-0.019
	(0.015)	(0.028)	(0.026)
If has irrigation system	0.072**	0.089*	0.138*
	(0.034)	(0.050)	(0.075)
Number of different crops	-0.308***	-0.249***	-0.393***
	(0.014)	(0.015)	(0.017)
Constant	2.241***	1.456***	1.425***
	(0.038)	(0.083)	(0.089)
Observations	396,317	113,133	147,353
R^2	0.498	0.493	0.333

Note: The regressions include municipality fixed effects and an indicator variable that takes the value of one for farmers with missing gender, age and years of education, and zero otherwise. Standard errors reported in parentheses clustered at the department level. ***, ** and * denotes significance at 1%, 5% and 10% level.

Table A2: Efficiency ratios (Y/Y^*), weighted by land size

Region	White Maize			Black Beans			Coffee		
	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$
All regions	87.1%	81.6%	76.4%	87.2%	81.7%	76.5%	74.5%	65.3%	57.7%
Center	87.4%	81.9%	76.8%	87.7%	82.4%	77.4%	70.0%	60.2%	52.9%
Western Highlands	86.0%	80.2%	75.0%	86.0%	80.1%	74.6%	73.8%	64.5%	56.9%
Dry Corridor	86.7%	81.0%	75.5%	86.1%	80.2%	74.8%	78.0%	69.3%	61.8%
Peten-Izabal	88.7%	83.7%	79.0%	88.5%	83.5%	78.8%	79.6%	71.0%	63.0%
Pacifico-Bocacosta	84.6%	78.6%	73.4%	86.0%	80.0%	74.6%	73.3%	63.9%	56.4%
Verapaces	87.1%	81.4%	75.9%	85.8%	79.7%	74.1%	78.6%	70.0%	62.5%

Note: The regional ratios reported are the corresponding averages of the departmental ratios in a region weighted by land size in each department. The ratio for All regions is the weighted average across regions. The Center region includes the departments of Sacatepequez and Chimaltenango; Western Highlands includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Dry Corridor includes Chiquimula, Jutiapa, Jalapa, El Progreso and Zacapa; Peten-Izabal includes Peten and Izabal; Pacifico-Bocacosta includes Retalhuleu, Suchitepequez, Escuintla and Santa Rosa; and Verapaces includes Alta Verapaz and Baja Verapaz.

Table A3: Efficiency ratios (Y/Y^*), considering only farmers that report one plot

Region	White Maize			Black Beans			Coffee		
	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$
All regions	86.9%	81.1%	75.9%	86.6%	80.9%	75.6%	75.4%	65.8%	58.4%
Center	88.4%	83.4%	78.5%	88.3%	83.5%	78.6%	69.7%	53.2%	45.8%
Western Highlands	86.3%	80.3%	75.2%	86.0%	79.8%	74.3%	77.2%	68.7%	61.6%
Dry Corridor	87.2%	81.7%	76.4%	86.1%	80.7%	75.2%	75.8%	67.9%	60.4%
Peten-Izabal	88.7%	83.7%	78.8%	88.7%	83.7%	78.9%	76.1%	65.2%	56.3%
Pacifico-Bocacosta	84.4%	78.1%	72.9%	83.8%	77.1%	71.5%	70.2%	58.2%	50.5%
Verapaces	87.2%	81.3%	75.8%	86.0%	79.9%	74.4%	75.7%	66.0%	58.3%

Note: The regional ratios reported are the corresponding averages of the departmental ratios in a region weighted by the number of producers in each department. The ratio for All regions is the weighted average across regions. The Center region includes the departments of Sacatepequez and Chimaltenango; Western Highlands includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Dry Corridor includes Chiquimula, Jutiapa, Jalapa, El Progreso and Zacapa; Peten-Izabal includes Peten and Izabal; Pacifico-Bocacosta includes Retalhuleu, Suchitepequez, Escuintla and Santa Rosa; and Verapaces includes Alta Verapaz and Baja Verapaz.