Land market distortions: Theory and evidence from Guatemala

M. Hernandez[1]; B. Britos[2]; M. Robles[3]; D. Trupkin[4]

1: International Food Policy Research Institute, Markets, Trade and Institutions Division, United States of America, 2: Universidad Carlos III Madrid, Department of Economics, Spain, 3: The World Bank, , United States of America, 4: Ministerio de Tra

Corresponding author email: m.a.hernandez@cgiar.org

Abstract:

Farm size and land allocation are important factors in explaining lagging agricultural productivity in developing countries. This paper formally examines the effect of land market distortions on the allocation of land across farmers and overall agricultural productivity. We first develop a theoretical framework to model the optimal size distribution of farms and assess to what extent market distortions can explain non-optimal land allocation and output inefficiency. We then calibrate the model to the case of Guatemala and evaluate potential drivers of the distortions across locations. We find that aggregate agricultural productivity across regions is over the range of 54-95% of the efficient output for different major crops considered. We evaluate alternative factors correlated with these distortions and provide some policy recommendations to improve efficiency.

Acknowledgement: We thank the Instituto Nacional de Estadistica (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. Francisco Ceballos provided excellent research assistance. All errors are our own.

JEL Codes: Q15, O47

#678

Land market distortions:
Theory and evidence from Guatemala

January 2018

Abstract

Farm size and land allocation are important factors in explaining lagging agricultural productivity in developing countries. This paper formally examines the effect of land market distortions on the allocation of land across farmers and overall agricultural productivity. We first develop a theoretical framework to model the optimal size distribution of farms and assess to what extent market distortions can explain non-optimal land allocation and output inefficiency. We then calibrate the model to the case of Guatemala and evaluate potential drivers of the distortions across locations. We find that aggregate agricultural productivity across regions is over the range of 54-95% of the efficient output for different major crops considered. We evaluate alternative factors correlated with these distortions and provide some policy recommendations to improve efficiency.

Keywords: Land markets, distortions, agricultural productivity, Guatemala

JEL Classifications: O13, Q15, O40

1 Introduction

It is well established that agriculture plays a key role in explaining the large disparities in labor 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; there are important differences in the size distribution of farms between rich and poor countries with a considerably smaller operational scale of farms in poor countries, and with large farms having a significantly higher labor productivity than smaller ones. Further understanding farm size patterns, the allocation of land through markets 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, formally assess the impact of land market distortions on the allocation of land across farmers and on agricultural productivity. Second, quantify the magnitude of these distortions using as an example the case of Guatemala and examine potential drivers of these distortions exploiting differences across locations. Ultimately, the study intends to discuss alternative policies to improve efficiency in land markets through the reduction of market frictions.

We focus on maize, sugar cane and coffee, which are the three major crops produced in Guatemala and generate most of the agricultural employment. The estimation results show that aggregate agricultural productivity across regions is over the range of 54-95% of the efficient output for the three crops considered. Maize generally shows a larger output efficiency compared to coffee and sugar cane, which is indicative that land market distortions play a more important (negative) role among high-value, export crops. Additional assessments across locations show that accessibility is a major factor explaining these distortions followed by education, as opposed to cultural factors.

The study ties into the general literature on factor misallocation across heterogeneous production units and productivity. Hopenhayn & Rogerson (1993), using the equilibrium model developed in Hopenhayn (1992), show that labor allocation across firms is distorted by dismissal taxes and can have important welfare losses through a decrease in average labor productivity (of up to 2%). In a similar line, Restuccia & Rogerson (2008) 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%). More recent studies that analyze the link between factor misallocation, aggregate productivity and output include Bartelsman et al. (2013), Bento & Restuccia (2017), David et al. (2016), Hsieh & Klenow (2009).

Closer to our study, Gollin et al. (2014) find evidence of labor misallocation across sectors using micro data for 80 countries. In particular, they show that the output per worker in the agricultural sector is roughly half the value in the non-agriculture sector, and the differences are more pronounced among developing countries. Using household-level data for Malawi, Restuccia & Santaeulalia-Llopis (2017) estimate the TFP for farms and find no relationship between TFP and operated land size and capital. The authors estimate that agricultural productivity would increase by a factor 3.6-fold if factors were reallocated to their efficient use. Factor misallocation in their study though is directly linked to restricted land markets, as most of the land is assigned by village chiefs and not marketed.

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., tax code and regulations, preferential market access, subsides, 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 create 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 is directly from eliminating land misallocation). Chen et al. (2017) assess the role of land markets on factor misallocation in Ethiopia, where the state owns the land, and show that land rentals significantly bring down misallocation and increase agricultural productivity. Similarly, Chamberlain & Ricker-Gilbert (2016) find evidence that rental markets contribute to efficiency gains within smallholder farmers in Malawi and Zambia by facilitating the transfer of land from less-able to more-able producers.[1]

The remainder of the paper is structured as follows. Section 2 presents the theoretical model and its implications. Section 3 describes the data and a set of empirical facts. Section 4 calibrates the model to the case of Guatemala and discusses the quantitative results. Section 5 concludes.

2 Model

We develop a model featuring endogenous distributions of both farm size and location to characterize land (mis)allocation in the agricultural sector across locations (regions). In each period the economy produces an agricultural good with land as the only factor of production.[2] Land misallocation occurs because farmers who operate lands across regional borders face a transaction cost, which increases with distance.

2.1 Farmers

The agricultural good is produced with land by a farmer with managerial skills s. The farm technology is characterized by decreasing returns to scale. In particular, a farmer of type i has the following production function

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

where y_i is the agricultural output of farm i and l is the amount of land. The parameter α captures the land elasticity.

[1]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) and De Janvry et al. (2015).

[2]We abstract from the capital decision, which is not the focus of this study.

The economy is endowed with a fixed amount of total land L. The farmer's managerial ability 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 N types of farmers with different managerial skills denoted by s_i, for $i = 1, ..., N$, distributed in N regions located across the country. For simplicity, we assume that there is only one type of farmer per region and rank the managerial ability such that $s_1 > s_2 > ... > s_N$. Hence, regions differ in terms of technology.

Without loss of generality, we further assume that the regions (denoted hereafter as r) are located linearly in terms of proximity, such that r_i is the adjacent region (the neighbor) of r_{i+1}, and the subindex i identifies both the region and the managerial ability. Based on the assumption above, r_1 is the most productive region; its neighboring, next closest location r_2 is the second most productive; and so forth. This assumption is not necessary for our purposes, but simplifies the analysis.

Finally, we assume that there is a transaction cost τ for a farmer who demands land across the border, and this cost is increasing with distance. That is, $\tau_{ij} = \tau(d_{i,j})$, with $\tau_{ii} = 0$ and $\tau'(\cdot) > 0$, where $d_{i,j}$ is defined as the distance between region i and j. These costs can be justified in several ways. They can be interpreted as the difficulties faced by farmers who demand land in other markets where information is more scarce for newcomers (and the lack of information increases with distance), the existence of asymmetries in the form of certain (market) power for insiders, and transportation costs for implementing effective managerial control, among other factors. These imperfections result in misallocations in the land market. Below we quantify the role that these failures play in terms of welfare losses.

2.2 Land Market Equilibrium

The firm's problem in the agricultural sector is defined as follows. A farmer with managerial ability s_i located in region i maximizes profits by demanding land and taking the rental prices of land (q) as given,

$$\max_{\{l_{ij}\}_{j=1}^{N}} \pi(s_i) = \left\{ s_i l_i^{\alpha} - \sum_{j=1}^{N} (q_j + \tau_{ij}) l_{ij} \right\}$$

where l_{ij} is the demand for land in market j of a farmer in region i, and $l_i = \sum_{j=1}^{N} l_{ij}$ is the total demand for land of farmer i.

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

$$\alpha s_i {l_i}^{\alpha-1} = q_j + \tau_{ij} \quad \text{in all markets } j.$$

We abstract from differences in land endowments and assume that the supply of land is the same across regions, i.e., $l_j^s = l^s$ for all $j = 1, ..., N$. The market clearing condition for land in each region becomes,

$$\sum_{i=1}^{N} l_{ij} = l_j^s = l^s.$$

Further, the aggregate (economy level) land market clearing condition is $L = \sum_{i=1}^{N} l_i$.

To solve for the equilibrium in the land market, we make the following additional assumption.

Assumption: Let the equilibrium rental price of region i in autarky be equal to q_i^A (i.e. when no trade is allowed among regions). We assume that $\tau_{ij} < |q_i^A - q_j^A|$ for all i and j.

This assumption implies that there is always some trade between regions (farmers purchasing lands across the border), even under the presence of transaction costs.

Result: Given that $s_1 > s_2 > ... > s_N$, then $l_{ij} = 0$ for $j < i$ (i.e. $l_{21} = l_{32} = ... = 0$).

The optimal farm size of less productive farmers will be lower than the optimal farm size of more productive ones. Hence, there is no equilibrium in which farmers in less productive regions rent land in more productive markets; otherwise, it would not be an optimal allocation.

We have then a system of $(N-1)$ equations for each land market clearing condition, and

$(N-1)$ conditions equating the marginal product of land for all farmers, net of transaction costs,

$$
\begin{aligned}
l_{11} &= l \\
l_{12} + l_{22} &= l \\
&\vdots \\
l_{(N-1)N} + l_{NN} &= l \\
f_{l_1}(l_{11} + l_{12}) &= f_{l_2}(l_{22} + l_{23}) + \tau_{12} \\
f_{l_2}(l_{22} + l_{23}) &= f_{l_3}(l_{33} + l_{34}) + \tau_{23} \\
&\vdots \\
f_{l_{N-1}}(l_{(N-1)(N-1)} + l_{(N-1)N}) &= f_{l_N}(l_{NN}) + \tau_{(N-1)N}
\end{aligned}
$$

where $f_{l_i}(\cdot)$ is farmer i's marginal productivity of land and the function argument identifies farmer $i's$ demand for land in all markets, $l_i = \sum_{j=1}^{N} l_{ij}$. Substituting the set of $(N-1)$ market clearing conditions for land into the $(N-1)$ conditions that equate the marginal productivity of land for all farmers, net of transaction costs, we obtain the following set of equilibrium conditions:

$$
\begin{aligned}
f_{l_1}(2l - l_{22}) &= f_{l_2}(l + l_{22} - l_{33}) + \tau_{12} \qquad (1) \\
f_{l_2}(l + l_{22} - l_{33}) &= f_{l_3}(l + l_{33} - l_{44}) + \tau_{23} \qquad (2) \\
&\vdots \\
f_{l_{N-1}}(l + l_{(N-2)(N-2)} - l_{(N-1)(N-1)}) &= f_{l_N}(l_{NN}) + \tau_{(N-1)N}.
\end{aligned}
$$

We have a system of $(N-1)$ equations to solve for the $(N-1)$ unknowns l_{ii} for $i = 2, ..., N$. We use the set of $(N-1)$ land market clearing conditions to solve for the remaining (cross-region) demands $l_{i(i+1)}$ for $i = 1, ..., N-1$.

Notice that $l_{11} = l > 0$, i.e., there is a positive equilibrium land allocation for a typical farmer in its own region. This is because of the assumption of a Cobb-Douglas production function for all farmers, which enables us to avoid the cases in which "foreign" farmers take-over all the land in a region and, at the same time, restricts farmers to only demand land in neighboring regions (i.e. $l_{ij} = 0$ for $|j - i| > 1$).

The whole system can be easily solved as follows. First, from equation (1), we solve for l_{33} as a function of l_{22}. In equation (2), we plug l_{33} in terms of l_{22} and obtain an expression for l_{44} as a function of l_{22}. We proceed sequentially to obtain an expression for l_{NN} in terms of l_{22}, i.e. $l_{NN}(l_{22})$. Since $f_{l_N}[l_{NN}(l_{22})] = f_{l_1}(2l - l_{22}) - \sum_{j=1}^{N-1} \tau_{jN}$, we obtain a solution for l_{22} and, then, for the remaining unknowns.

Finally, the set of N equilibrium land rental prices is given by $q_1 = f_{l_1}(l + l_{12})$ and $q_j = q_{j-1} - \tau_{j-1,j}$ for all $j > 1$.

2.3 A Simple Example

Consider two types of farmers, two regions, and the technology for producing the agricultural good defined above as $y_i = s_i l_i^\alpha$.

Let $s_1 > s_2$. Then, the two equilibrium conditions to solve for l_{12} and l_{22} are the following,[3]

$$\begin{aligned} \alpha s_1 (l + l_{12})^{\alpha - 1} &= \alpha s_2 (l_{22})^{\alpha-1} + \tau_{12} \\ l_{12} + l_{22} &= l, \end{aligned}$$

which implies a unique equation to solve for l_{22}^*,

$$\alpha s_1 (2l - l_{22})^{\alpha-1} = \alpha s_2 (l_{22})^{\alpha-1} + \tau_{12}.$$

Once obtained l_{22}, we solve for $l_{12} = l - l_{22}$.

Consider the following parameter values:

α	s_1	s_2	l	τ_{12}
0.5	2	1	1	0.2

[3] Remember that an optimal allocation implies $l_{11}^* = l$ and l_{21}^*=0.

Then, the optimal allocation is equal to,

l_{11}	l_{12}	l_{21}	l_{22}	q_1	q_2
1	0.4	0	0.6	0.85	0.65

We observe that the typical farmer in Region 1 (the more productive one) captures the whole land in its market, plus the 40% of land in Region 2 where the less productive farmer is located.

This result allows us to characterize both farm size and location, and quantify the distortions. For instance, if $\tau_{12} = 0$ (i.e. there are no transaction costs/distorsions), we obtain the frictionless land allocation (which we denote as l_{ij}^* for all i, j):

l_{11}^*	l_{12}^*	l_{21}^*	l_{22}^*	q_1^*	q_2^*
1	0.6	0	0.4	0.79	0.79

Thus, the typical farmer in Region 1 captures the 60% of land in Region 2 in addition to the whole land in her market. Finally, if we compare the aggregate agricultural ouput produced by the distorted allocation ($y = y_1 + y_2$) with the one produced by the frictionless allocation ($y^* = y_1^* + y_2^*$), with this simple example we find an inefficiency rate of 1%. See Figure 1 for further details.

3 The Case of Guatemala

This section describes the context of land markets in Guatemala. We rely on a number of alternative data sources that allow us to explore some empirical regularities across the country. We first present the data sources used.

3.1 Data

We rely on two different data sources. We use micro-data from the "IV Censo Nacional Agropecuario 2003" corresponding to the crop year 2002-03. This is the last census of agriculture in Guatemala, which is nationally representative and includes information on land use (for crops, cattle farming and other activities), production, input use and land quality variables, farmers' socioeconomic characteristics, among others. We complement the census data with a dataset from a three-year panel survey of households collected between 2012 and 2014 as part of the monitoring and evaluation of the Zero Hunger Pact, a large-scale program executed by the Government of Guatemala against malnutrition. The surveys were administered to urban and rural households in more than half of the administrative areas (known as municipalities) in the country with the highest rates of child malnutrition.[4] These surveys included a module on agricultural land markets that inquired about land prices, past transactions and general perceptions on the development of local land markets.

3.2 Land Markets in Guatemala

Guatemala is an interesting case study for analysis 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 commercial to subsistence farming and from relatively large-scale farming of export crops such as sugar cane and bananas or cattle farming to medium- and small-scale farming of high-value crops such as coffee and vegetables or basic grains such as maize and beans. For the analysis, we divide the country into six major geographic regions as shown in the map in Figure 2.

We begin by providing general descriptive statistics on the size of landholdings using data from the agricultural census. The top panel of Table 1 shows the size distribution of farms in the country as a whole and disaggregated by region. Overall, small farms (under 1 hectare) represent the majority of farms in Guatemala (almost 70%), while very large farms (over 20 hectares) represent only 0.5% of the farms.[5] 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. For instance, Restuccia & Santaeulalia-

[4]Guatemala is divided into 22 departments and 340 municipalities. The survey covered 176 municipalities.

[5]We observe a similar distribution if we consider landholding size used for temporary or permanent crops alone.

Llopis (2017) report more than 81% and 46% of farms in the United States (US) and Belgium having more than 10 hectares, and only 15% of farms in Belgium having less than one hectare and none in the case of the US.

The table also shows a large variation in the distribution of landholding size across regions. For instance, more than 85% of the agricultural landholdings in the Central region ('Centro') and the Western Highlands ('Altiplano Occidental') are smaller than one hectare; the case of the former is explained by the low 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 human development and there is a large presence of smallholder, subsistence agriculture. In contrast, the Petén-Izabal region concentrates most of the cattle farming activities in the country and thus exhibits much larger landholdings than the rest of the regions.

The bottom panels of Table 1 show, in turn, some variations in landholding size for white maize, sugar cane and coffee. We focus on these crops because they are the three major crops produced in Guatemala and generate more than 63% of the agricultural employment. Maize, specially white maize, is by far the most common and extended crop produced in the country with a total planted area of 841,094 hectares (Ha) and production of 1,672,527 metric tonnes (MT) as of 2011/12, and is basically oriented for the local market; the major producer regions are Peten-Izabal (where maize production is combined with cattle farming activities, reason why the landholdings are larger), Western Highlands ('Altiplano Occidental') and Verapaces. Coffee is also produced in multiple regions across the country with a total planted area of 252,415 Ha and annual production of 245,752 MT, and is the second major crop exported; the major producer regions include Western Highlands, Pacífico-Bocacosta and Verapaces. Sugar cane production is more concentrated in certain regions, particularly in Pacífico-Bocacosta, with a total planted area of 239,261 Ha and annual production of 2,019,622 MT, and is the major crop exported (MAGA, 2011). Focusing on these crops further permits to assess whether the (mis)allocation of land across agents with varying levels of productivity is more acute for certain type of crops, i.e. crops that involve small versus medium/large scale production, crops for subsistence/internal versus commercial/external markets.

To introduce the concept of productivity and provide an overview of its variability across the country, Table 2 presents summary statistics for white maize, sugar cane and coffee yields, a commonly used agricultural productivity measure. We observe again large differences in both the average levels and the variability of yields across regions by crop. Yet, this level of variability is not necessarily indicative of the presence of land market inefficiencies as it could result from an optimal allocation of production factors. We return to this point in Section 4 below.

Figure 3 shows, in turn, some aggregate patterns for the size of landholdings dedicated to maize, sugar cane or coffee relative to the level of agricultural development of the different departments. As a proxy of agricultural development we use the value of total agricultural production per capita constructed from the census data.[6] The figure shows the average share of very small (under 1 hectare) and very large (over 20 hectares) farms by income index quintiles (where Q1 represents the 1/5 of departments with the lowest levels of income). We observe that at lower income levels, the share of very small farms tends to be much higher than at higher income levels. The opposite is true for very large farms, which gain relative importance in departments with higher levels of income. This is in line with the international evidence presented in Adamopolous & Restuccia (2014) and is indicative of the presence of inefficiencies in land markets.[7]

Figure 4 provides additional empirical motivation on the discussion of land market inefficiencies described in Section 2. In the panel collected in more than half of the country between 2012-2014, households were asked to provide the price per hectare of what they would consider to be the most productive agricultural land in their corresponding administrative area (municipality). The purpose of the question was to evaluate the variability in farmers' perceptions about land prices in their immediate geographic area. We thus calculate the coefficient of variation for this price across all farmer responses within each administrative area and use it as a proxy measure for land market imperfections. Figure 4 shows a scatterplot of the concentration ratio in landholdings (calculated from the

[6]The production is valued using local crop prices at the municipality level.
[7]We generally find a similar pattern when separately considering each crop.

census data and defined as the fraction of landholdings held by the largest 10% farmers) and our proxy measure of market imperfection for each administrative area covered in the survey. Recall from the theoretical model in Section 2 that higher transaction costs τ result in a sub-optimal limited degree of transactions in land markets, which impedes the most productive farmers to work at their (larger) optimal scale. The inverse relationship observed in the figure is indicative of a negative correlation between potential land market imperfections and allocative efficiency.[8]

3.3 A Measure of Farm Productivity

We now calculate a measure of total productivity at the farm level for white maize, sugar cane and coffee. We begin by taking the residual from the production function defined in Section 2 for each particular farmer i in region j,

$$y_{ij} = s_{ij}(l_{ij})^{\alpha}.$$

We obtain the measure for total farm productivity s using the census micro-data on land and crop yields and assuming $\alpha = 0.18$, which is the value for land income share estimated in Valentinyi & Herrendorf (2008).[9] To obtain a more accurate measure of total farm productivity we account for a set of control variables, including farmer's age and years of education, whether he/she uses enhanced seeds, type and quantity of input used, use of irrigation systems and number of crops cultivated (to capture the level of specialization). In particular, we estimate the calculated series of s_{ij} for each crop as a function of the above set of control variables plus administrative area fixed effects via ordinary-least squares. The residual of this regression is our measure of farm productivity.

The regression results are presented in Table A1 in the Appendix. The coefficients of the

[8]This relationship holds if we use the price per hectare that the farmer valued her own land, after controlling for land quality. The relationship also holds when using 5% and 20% concentration ratios.

[9]We perform below a sensitivity analysis with $\alpha = 0.39$, which is the value estimated by Restuccia & Santaeulalia-Llopis (2017) for the case of Malawi. The larger value of α in Malawi is explained by the lower level of mechanization in the agriculture sector relative to, for example, the US. Based on the development of Guatemala, a range between $\alpha = 0.18$ and $\alpha = 0.39$ seems to be appropriate for this country.

control variables generally have the expected signs. For all crops, total farm productivity is positively correlated with farmer's age, years of schooling, use of high-performance seeds and fertilizer, and use of irrigation systems. The level of specialization (defined as number of different crops produced), in contrast, appears to affect total farm productivity differently depending on the type crop. Specialization is positively associated with farm productivity in the case of sugar cane and coffee but negatively associated in the case of maize.

For the country as a whole, the correlation between farm productivity and landholding size results in the range of 0.2-0.3, depending on the crop considered. These correlations are statistically significant at conventional levels, which suggests that the allocation of land is related in some degree to the farmers' productivity. Next, we discuss to what extent this allocation is efficient. To do this, we evaluate how the actual allocation compares with a benchmark, efficient allocation chosen by a hypothetical social planner.

4 Quantitative Analysis

In this section we quantify the magnitude of land misallocation in Guatemala at the country level. We then repeat the analysis at the regional level to evaluate geographic heterogeneity. Finally, based on the model developed in Section 2, we assess the potential channels that may explain the nature of the distortions, focusing on land market imperfections.

Our approach is built on Restuccia & Santaeulalia-Llopis (2017). First, we solve a simple optimization problem of a hypothetical social planner intending to maximize aggregate output by allocating land according to the overall distribution of farmers' productivity. Second, we compare the aggregate output that results from this efficient allocation by the social planner with the actual aggregate output that would result from the land size distribution found in the data.

An efficient allocation for region j can be obtained by solving the following social planner problem,

$$Y_j^* = \max_{\{l_{ij}\}_i} \sum_{i=1}^{N} s_{ij}(l_{ij})^\alpha,$$

subject to

$$L_j = \sum_{i=1}^{N} l_{ij}$$

where Y_j^* denotes the efficient output.[10]

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,

$$l_{ij}^* = \frac{s_{ij}^{1/(1-\alpha)}}{\sum_i s_{ij}^{1/(1-\alpha)}} L.$$

The theory suggests that each farmer's land size depends on her productivity relative to the whole distribution of farm productivity. Letting $S_{ij} \equiv s_{ij}^{1/(1-\alpha)} / \sum_i s_{ij}^{1/(1-\alpha)}$, it follows that,

$$Y_j^* = \sum_{i=1}^{N} s_{ij} \left(S_{ij} L \right)^{\alpha}.$$

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

$$Y^c = \sum_{i=1}^{N} s_{ij} (l_{ij}^c)^{\alpha}$$

where l_{ij}^c denotes the actual land extension of an individual farmer observed in the census data.

Table 3 presents the results of this exercise in terms of the efficiency ratio Y^c/Y^*. The results are reported by crop and region. A higher (lower) efficiency ratio indicates that the current land allocation across economic agents is closer to (farther from) the optimal allocation from a social planner's perspective for a given crop in a corresponding location.[11] In general, we observe an important degree of inefficiency arising from land misallocation.

[10] Since we focus on the land-market distortions channel, we assume that the set of control variables is exogenously given for the planner.

[11] We first calculate the farm productivity measures by department in order to account for potential heterogeneity in the response to the control variables used to clean the measure. We then calculate the efficiency ratios by department and report the average efficiency ratio across departments in a given region.

The ratio varies between 54% and 95% across regions and crops. On average, the gap of aggregate output between the current land allocation and the theoretically efficient allocation is around one fifth; hence, if market imperfections were removed, total output would increase by roughly 25%.

Interestingly, we find a larger output efficiency for a staple crop such as maize (with a larger prevalence of small-scale agriculture) as opposed to high-value, export crops such as coffee and, in particular, sugar cane.[12] The regions with the largest maize efficiency ratio (over 95.5%) are Peten-Izabal, Verapaces and Western Highlands ('Altiplano Occidental'), which are also the regions where most maize is produced but are not necessarily the most productive ones (see Table 2). This also suggests that maize production may already be operating close to its maximum production potential in several locations such that further eliminating land market distortions will not increase by much maize output. In contrast, the Pacífico-Bocacosta region, where most sugar cane production concentrates and with a high productivity, shows the smallest sugar cane efficiency ratio (54%); removing land market imperfections in this case could almost double the output in this region. Coffee efficiency ratios (71-87%) are more mixed across regions with high or low production (and productivity) levels.

Table A2 in the Appendix shows the efficiency ratios when alternatively using $\alpha = 0.39$. The magnitude of land misallocation is higher compared to the benchmark case ($\alpha = 0.18$) because larger values of α imply a larger share of land in the production technology and, thus, higher costs from misallocating this factor across farmers. We observe again a large heterogeneity across regions and crops, with an efficiency ratio ranging in this case from 27% to 89%. Overall, the results are in line with the patterns observed in the benchmark case in terms of the regions showing relatively higher (or lower) output efficiencies by crop.

Lastly, as an additional robustness check, we report the efficiency ratios calculated without including farmers' socioeconomic characteristics such as age and education as control variables in the first-stage regressions (in Table A1). These variables show an important number of missing values (close to 20%) in the census data, and (as expected)

[12] Although not reported, we find similar large efficiency ratios for yellow maize and beans, which are also major staple crops grown in Guatemala.

these characteristics are mainly missing for larger farmers, which could generate some bias in the results. Table A3 in the Appendix shows the corresponding results. We find somewhat larger efficiency ratios, particularly for sugar cane that has a higher prevalence of large-scale agriculture, which suggests a potential negative selection bias when excluding these observations. Still, the differences in efficiencies across crops and regions generally remain; e.g., Pacífico-Bocacosta is the region with the lowest output efficiency ratio (for sugar cane) and Peten-Izabal is the region with the largest efficiency ratio (for white maize).

4.1 Potential channels of distortions

We now turn to assess the potential channels or factors correlated with the estimated inefficiencies by crop type. For this purpose, we recover efficiency ratios (Y^c/Y^*) by municipality and regress them on a set of indicators related to education, ethnicity, socioeconomic characteristics, land ownership, and accessibility in the area. In particular, these include illiteracy rate, rate of indigenous population, rate of Spanish-speaking households, average age of household head, average dependency ratio (population 0-14 years old and 65 and over relative to 15-64 years old), proportion of lands owned, average distance (in hours) to other municipalities in the same department, and population density (population per square kilometer).[13]

Table 4 presents the estimation results.[14] The variables were first standardized for comparability purposes and the reported standard errors are clustered at the department level. Two interesting patterns emerge from the table. First, the observed efficiency ratios seem negatively correlated with the two measures of accessibility or proximity included in the regression analysis. The more distant (in terms of time traveled) a municipality is from

[13]The average distance between municipalities is calculated through an accessibility model that relies on spatial analysis using a geographic information system (GIS) and a combination of spatial variables that influence the movement of people. The analysis assumes that people travel via highways, major roads, or walkways (when these exist) and around facilities near their homes (e.g., schools, healthcare centers), and takes into account the presence of natural barriers (e.g., rivers, lakes) as well as the traveling speed (slope). The model simulates the time (hours) it takes a person to reach the nearest location using the fastest available method and route of travel. Additional details are available upon request.

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

the other municipalities in the department, the lower the efficiency ratio; a one standard deviation increase in this distance is associated with 0.161 standard deviations decrease in the efficiency ratio for white maize and 0.225 standard deviations decrease for sugar cane. Similarly, the less dispersed the population in a municipality the smaller the inefficiencies; a one standard deviation increase in population density is associated with 0.256, 0.555 and 0.378 standard deviations increases in the efficiency ratio for white maize, sugar cane and coffee. Hence, larger transaction costs in terms of required distances traveled and feasibility of information flow play some role in explaining output inefficiencies.

Second, the level of education in the municipality is also positively correlated with output efficiency, at least for white maize and sugar cane. A one standard deviation increase in the illiteracy rate is associated with 0.173 and 0.158 standard deviations decrease in the efficiency ratio for these two crops. A larger dynamism in land markets, and subsequent better allocation of land, is more likely to happen among locations with more educated people. Ethnicity, captured through the rate of indigenous population, appear negatively correlated with output efficiency for all three crops although the correlations are not statistically significant. In the same vein, an apparent puzzling result is the negative correlation between the rate of Spanish-speaking households and output efficiency for maize, which suggests that in locations with less Spanish speakers there are less distortions; note though that in many locations the production of maize is mainly dominated by smallholder farmers (in several cases for subsistence purposes) with important cultural (language) barriers, such that there could be less information asymmetries or a larger social cohesion in (smaller) communities dominated by non-Spanish speakers.

5 Conclusion

Farm size and land allocation play an important role in explaining lagging agricultural productivity in developing countries. This paper evaluates the impact of land market distortions on land allocation and agricultural productivity. We first 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 the case of Guatemala as an example. The results show that aggregate agricultural productivity across regions is over the range of 54-95% for maize, sugar cane and coffee, which are the three most important crops produced in the country. We also find that output efficiency is positively correlated with accessibility and, to a lower extent, education, but not with potential cultural factors.

These findings indicate the presence of larger distortions for high-value, export crops such as coffee and, in particular, sugar cane, as opposed to staple crops such as maize. It may be the case that the latter, with a larger prevalence of small-scale, subsistence agriculture, may already be operating close to its maximum production potential such that further eliminating land market distortions will not have a major effect on reaching the (already close to) optimal output level. Yet, for coffee and sugar cane the elimination of these distortions could have a significant effect on the further production expansion of these crops and, consequently, the development of the agricultural sector.

The analysis examining the potential factors associated with output inefficiencies indicate the importance of improving accessibility as well as education. Certainly, policies in this regard, such as investment in roads, infrastructure and education, will require some time to become effective. In the short-term, however, market information systems exploiting new technologies of information (e.g., the mobile penetration rate in rural Guatemala is over 85%) could help to develop or expand land markets across the country and reallocate land from less to more productive producers. Pilot programs to assess whether providing market information contributes in the generation of rental markets and to assess the willingness among farmers to rent in/out land, particularly among locations with larger distortions, are an avenue of future work along these lines.

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.

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.

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. (2017). 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.

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.

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.

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

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

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., Tao Yang, D., & Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, 55, 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 the simple 2-region model

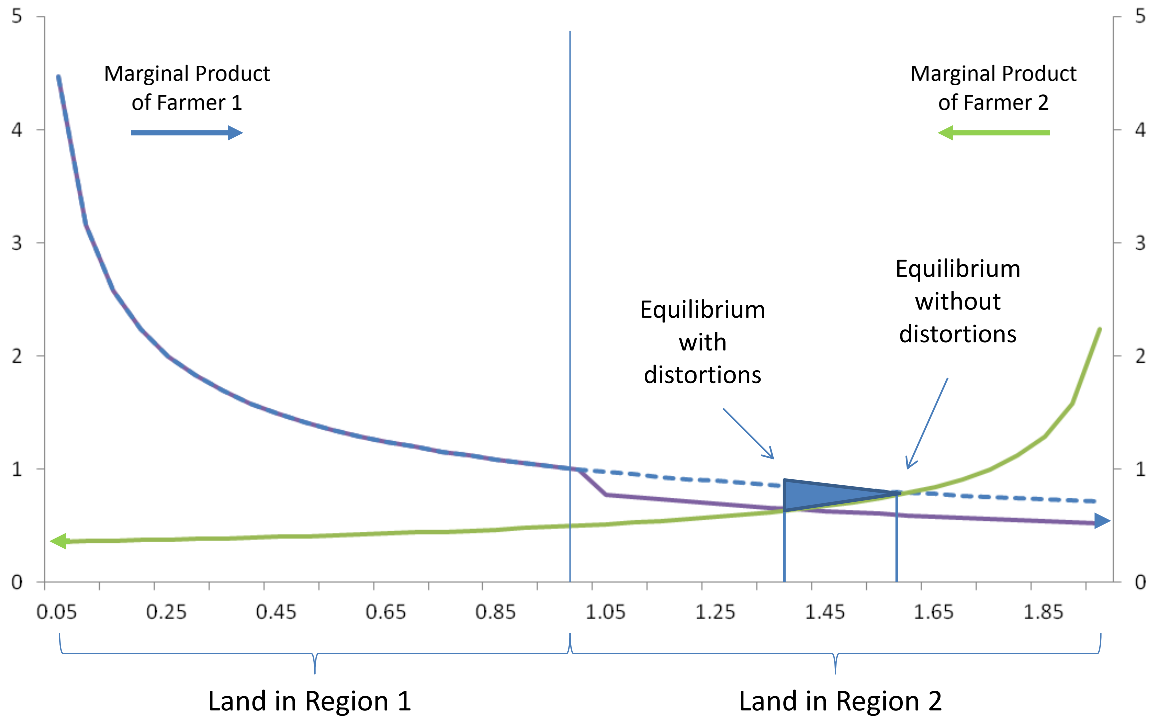


Figure 2: Map of Guatemala and regions considered

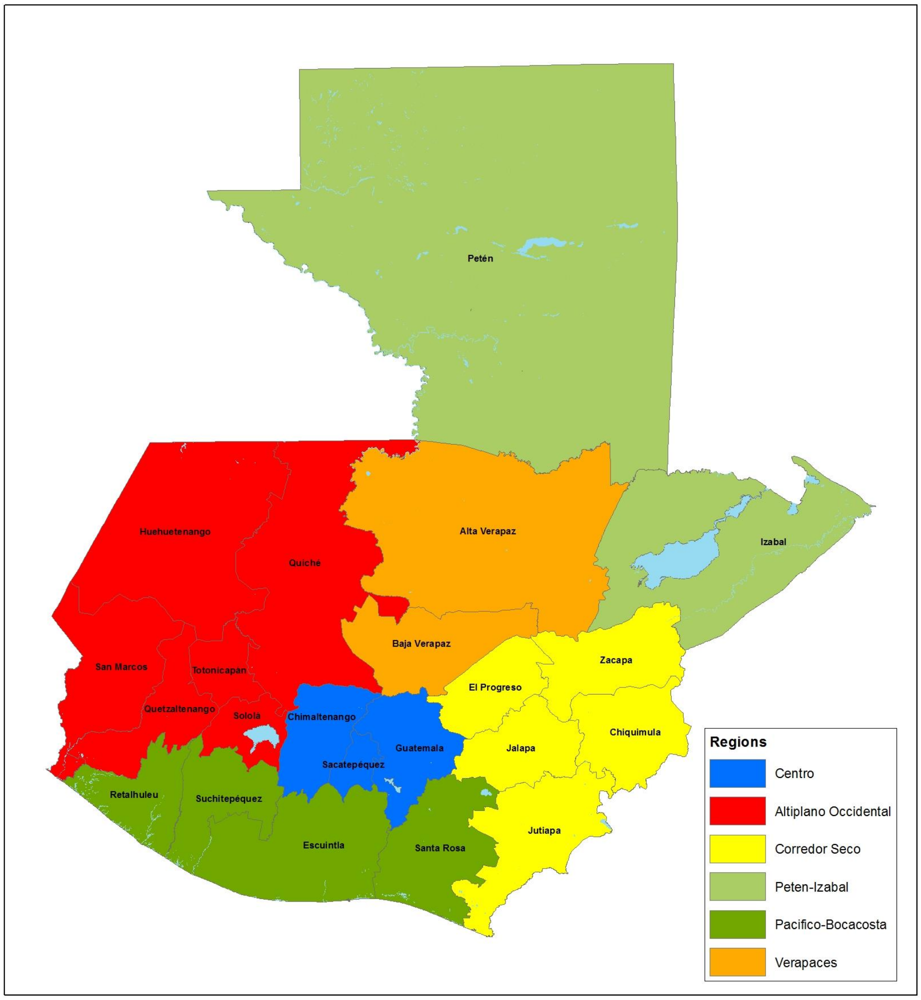
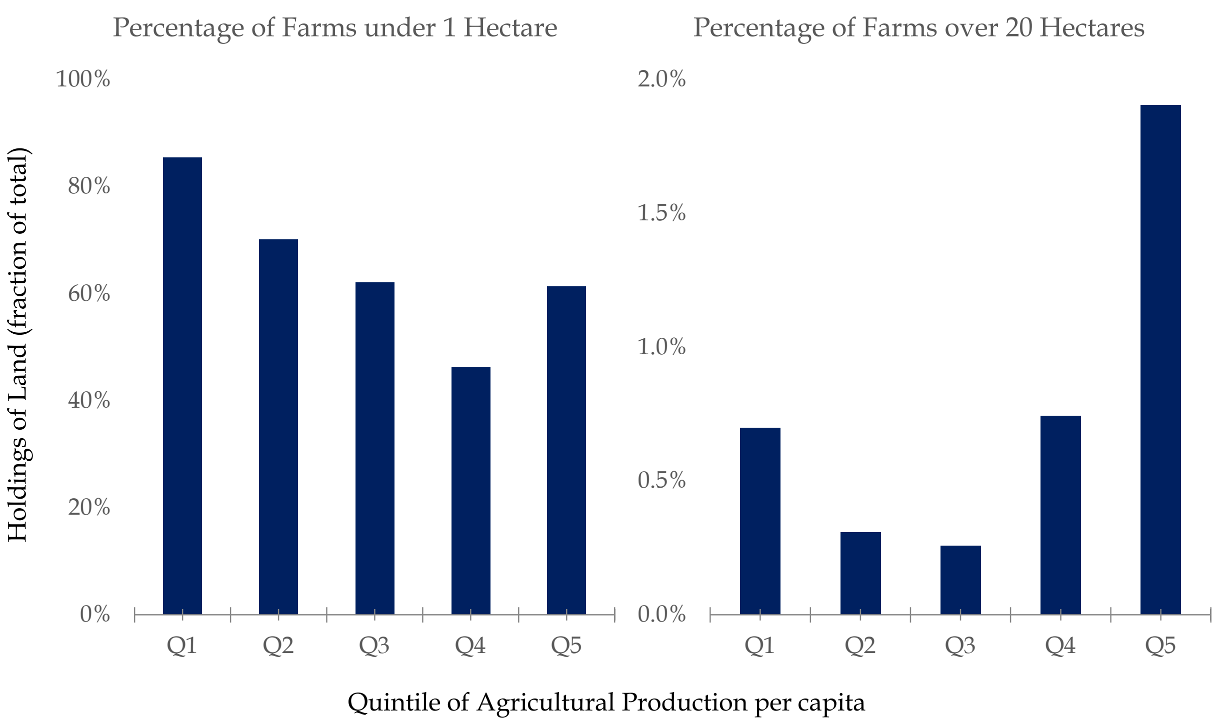
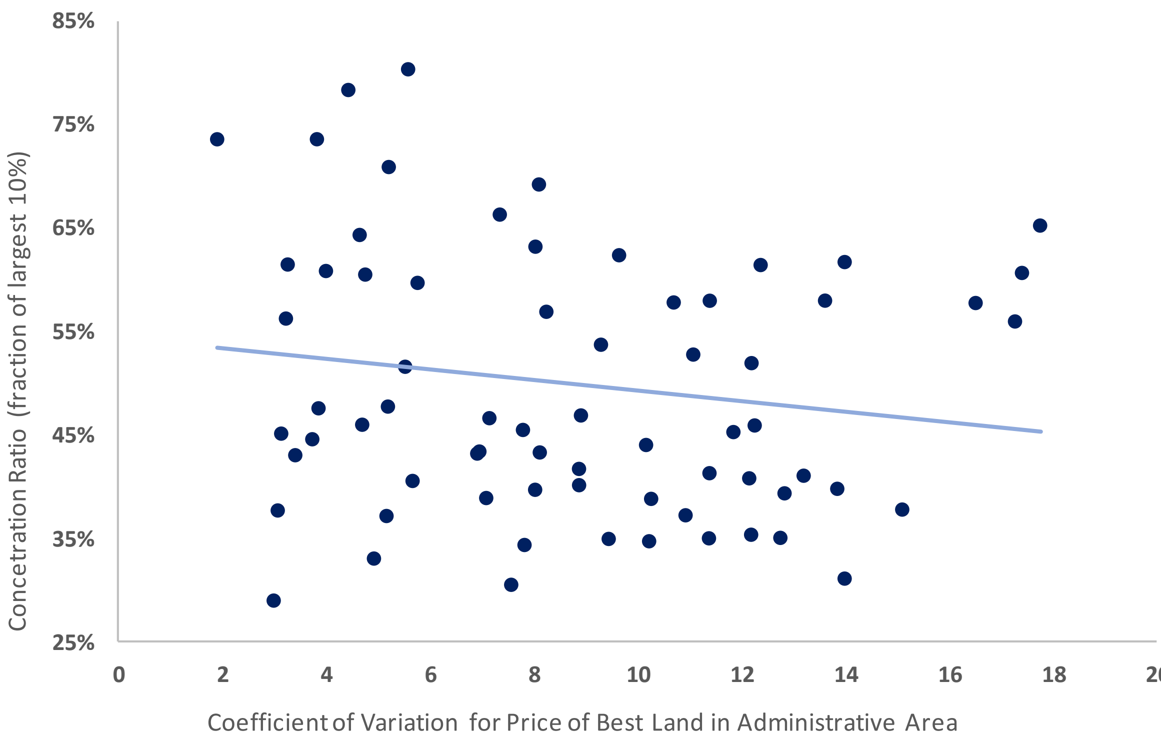


Figure 3: Share of small and large farms across departments by income quintile



Note: Based on landholdings dedicated to production of white maize, sugar cane and coffee.

Figure 4: Landholdings Concentration ratio and land price dispersion



Note: The vertical axis measures the share of landholdings held by the 10% largest farmers in the municipality. The horizontal axis measures the coefficient of variation of the (self-reported) price of the most productive land in the municipality.

Table 1: Size distribution of farms (% of farms by size)

Landholding size	All regions	Centro	Altiplano Occidental	Corredor Seco	Petén-Izabal	Pacífico-Bocacosta	Verapaces
Less than 1 Ha	69.2%	85.2%	85.3%	54.8%	17.8%	60.7%	57.1%
1 - 2 Ha	17.1%	9.6%	10.3%	25.5%	21.2%	19.5%	29.6%
2 - 5 Ha	10.4%	3.4%	3.7%	16.1%	40.0%	14.5%	11.9%
5 - 10 Ha	2.1%	0.6%	0.4%	2.5%	15.0%	2.5%	1.0%
10 - 20 Ha	0.7%	0.4%	0.1%	0.8%	4.7%	1.0%	0.2%
More than 20 Ha	0.5%	0.8%	0.2%	0.3%	1.2%	1.8%	0.2%
White maize							
Less than 1 Ha	73.5%	93.5%	90.5%	70.8%	20.9%	66.3%	65.1%
1 - 2 Ha	15%	4.8%	7%	20.2%	25.1%	18%	25.4%
2 - 5 Ha	9.3%	1.4%	2.2%	7.8%	40.4%	12.8%	8.8%
5 - 10 Ha	1.6%	0.2%	0.2%	0.9%	10.2%	1.9%	0.6%
10 - 20 Ha	0.5%	0.1%	0.1%	0.3%	2.7%	0.7%	0.1%
More than 20 Ha	0.2%	0.1%	0%	0.1%	0.7%	0.4%	0.1%
Sugar Cane							
Less than 1 Ha	89.8%	81.2%	93.1%	90.5%	97%	46%	98.3%
1 - 2 Ha	3.7%	4.8%	4.4%	4.8%	1.5%	10.4%	1%
2 - 5 Ha	2.3%	2.5%	2.2%	3.4%	0.9%	8%	0.6%
5 - 10 Ha	0.7%	0.7%	0.2%	1.2%	0.3%	3.8%	0.1%
10 - 20 Ha	0.4%	0.7%	0%	0.1%	0.3%	3.8%	0%
More than 20 Ha	3.1%	10.1%	0%	0.1%	0%	27.9%	0%
Coffee							
Less than 1 Ha	82.6%	80.6%	83.8%	77.4%	83.8%	63.8%	93.6%
1 - 2 Ha	9.6%	9.8%	10.2%	11.2%	8.4%	16.4%	4%
2 - 5 Ha	5.3%	4.4%	4.3%	8.3%	6%	13.2%	1.7%
5 - 10 Ha	1.1%	1.5%	0.7%	2%	1%	2.9%	0.3%
10 - 20 Ha	0.5%	1.1%	0.3%	0.7%	0.2%	1.3%	0.1%
More than 20 Ha	0.9%	2.6%	0.6%	0.5%	0.4%	2.4%	0.3%

Note: Region Centro includes the departments of Guatemala, Sacatepequez and Chimaltenango; Altiplano Occidental includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Corredor Seco 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: Dispersion of productivity across farms

White Maize	All regions	Centro	Altiplano Occidental	Corredor Seco	Petén-Izabal	Pacífico-Bocacosta	Verapaces
Mean	26.9	32.4	28.4	23.8	22.6	37.9	20.4
St. Dev.	13.8	15.0	13.3	12.8	10.1	14.8	10.5
Coefficient of Variation	0.5	0.5	0.5	0.5	0.4	0.4	0.5
Ratio P75/P25	2.0	2.1	2.1	2.1	1.8	2.0	2.0
Observations	403,242	38,561	152,524	56,289	44,745	37,017	73,614
Sugar Cane	All regions	Centro	Altiplano Occidental	Corredor Seco	Petén-Izabal	Pacífico-Bocacosta	Verapaces
Mean	213.7	351.0	102.8	150.4	169.4	735.0	142.0
St. Dev.	337.3	466.5	164.0	180.4	199.8	531.5	181.5
Coefficient of Variation	1.6	1.3	1.6	1.2	1.2	0.7	1.3
Ratio P75/P25	17.2	24.0	5.0	14.1	13.5	14.5	11.5
Observations	8,211	223	3,432	866	242	1,111	2,293
Coffee	All regions	Centro	Altiplano Occidental	Corredor Seco	Petén-Izabal	Pacífico-Bocacosta	Verapaces
Mean	34.3	40.4	30.6	41.3	27.2	43.3	30.6
St. Dev.	29.2	31.5	26.9	33.2	26.7	32.5	25.4
Coefficient of Variation	0.8	0.8	0.9	0.8	1.0	0.8	0.8
Ratio P75/P25	3.6	3.6	3.6	3.6	8.3	3.1	3.0
Observations	150,111	9,497	64,813	22,020	783	18,732	33,819

Note: Region Centro includes the departments of Guatemala, Sacatepequez and Chimaltenango; Altiplano Occidental includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Corredor Seco 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 3: Efficiency ratio (Y^c/Y^*)

Crop	All regions	Centro	Altiplano Occidental	Corredor Seco	Petén-Izabal	Pacífico-Bocacosta	Verapaces
White Maize	92.5%	90.4%	92.6%	92.2%	95.4%	90.1%	94.5%
Sugar Cane	74.3%	61.8%	77.9%	81.4%	91.4%	53.7%	79.6%
Coffee	80.4%	71.1%	77.1%	86.7%	85.3%	76.6%	85.3%

Note: All regions shows the simple average across regions; Region Centro includes the departments of Guatemala, Sacatepequez and Chimaltenango; Altiplano Occidental includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Corredor Seco 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 4: Regression of efficiency ratio (Y^c/Y^*) on indicators at municipality level

Coefficient	(1) White Maize	(2) Sugar Cane	(3) Coffee
Illiteracy rate	-0.173**	-0.158*	0.041
	(0.074)	(0.076)	(0.043)
Rate of indigenous population	-0.088	-0.153	-0.017
	(0.136)	(0.151)	(0.117)
Rate spanish-speaking households	-0.285***	-0.070	0.038
	(0.094)	(0.119)	(0.103)
Average age household head	0.168	0.049	0.051
	(0.104)	(0.087)	(0.116)
Average dependency ratio	0.046	0.190	-0.020
	(0.073)	(0.109)	(0.084)
Proportion of lands owned	0.031	0.065	-0.083
	(0.115)	(0.051)	(0.077)
Average distance (hours) to other municipalities	-0.161*	-0.225***	-0.042
	(0.087)	(0.064)	(0.109)
Population density	0.256**	0.555*	0.378*
	(0.115)	(0.265)	(0.187)
Constant	-0.036	-0.024	0.046
	(0.088)	(0.097)	(0.083)
Observations	141	98	128
R-squared	0.167	0.236	0.066

Note: Each observation corresponds to a municipality. 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

Table A1: Regression of Ln farm productivity (s_{ij}) on set of characteristics at farmer level

Coefficient	(1) White Maize	(2) Sugar Cane	(3) Coffee
Age	0.005***	0.015***	0.011***
	(0.000)	(0.003)	(0.001)
Years of schooling	0.015***	0.109***	0.087***
	(0.002)	(0.020)	(0.009)
If uses high-performance seeds	0.096***	0.136*	0.103*
	(0.015)	(0.070)	(0.052)
If uses organic fertilzer	0.033***	0.184**	0.206***
	(0.009)	(0.074)	(0.027)
If uses chemical fertilizer	0.129***	0.163	0.225***
	(0.032)	(0.111)	(0.047)
If uses pesticide	0.168***	0.023	0.086**
	(0.016)	(0.081)	(0.033)
If irrigation system	0.142***	0.324***	0.194***
	(0.038)	(0.075)	(0.049)
Number of different crops produced	0.052***	-0.100**	-0.042**
	(0.017)	(0.040)	(0.019)
Constant	2.502***	6.112***	2.283***
	(0.053)	(0.445)	(0.148)
Observations	325,407	6,532	120,140
R-squared	0.438	0.503	0.297

Note: Standard errors reported in parentheses clustered at the department level. ***, ** and * denotes significance at 1%, 5% and 10% level.

Table A2: Efficiency ratio (Y^c/Y^*) with $\alpha = 0.39$

Crop	All regions	Centro	Altiplano Occidental	Corredor Seco	Petén-Izabal	Pacífico-Bocacosta	Verapaces
White Maize	83.9%	79.4%	84.6%	82.9%	89.4%	80.1%	87.3%
Sugar Cane	54.5%	41.3%	58.7%	62.5%	81.2%	26.7%	56.5%
Coffee	62.3%	48.2%	57.9%	73.0%	69.9%	56.0%	69.1%

Note: All regions shows the simple average across regions; Region Centro includes the departments of Guatemala, Sacatepequez and Chimaltenango; Altiplano Occidental includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Corredor Seco 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 ratio (Y^c/Y^*) excluding farmers' characteristics as control variables

Crop	All regions	Centro	Altiplano Occidental	Corredor Seco	Petén-Izabal	Pacífico-Bocacosta	Verapaces
White Maize	96.6%	97.0%	96.2%	96.3%	97.5%	96.1%	96.6%
Sugar Cane	79.1%	66.6%	82.0%	85.6%	92.3%	67.2%	80.9%
Coffee	89.8%	90.8%	86.1%	92.8%	85.4%	90.3%	93.3%

Note: All regions shows the simple average across regions; Region Centro includes the departments of Guatemala, Sacatepequez and Chimaltenango; Altiplano Occidental includes Huehuetenango, Quiche, San Marcos, Quetzaltenango, Totonicapan and Solola; Corredor Seco 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.