Abstract
We assess the importance of land misallocation for productivity in agriculture using a quantitative model and detailed household-level data from Malawi. The land market is largely underdeveloped in Malawi as the vast majority of land is transmitted by inheritance, almost none of the land is purchased with a title, and a very small portion of operated land is rented. Moreover, the land distribution in Malawi is concentrated at extremely small operational scales, with more than 70 percent of all farmers operating less than 2 acres of land. Using the micro data, we estimate the distribution of farmer’s productivity and the wedges that rationalize the actual distribution of farm sizes in the context of a standard model of farm size with endogenous farm-level productivity. The micro data indicates that while the distribution of farmers productivity approximates a log normal with substantial dispersion, land is spread evenly among farmers at very low operational scales. The striking fact is that land size is unrelated to farm productivity. Under the counterfactual that land is efficiently allocated across existing productive farms, we find an increase in agricultural productivity of a factor of 4-fold, which is orders of magnitude larger than found in the related misallocation literature. This result emphasizes the importance of land market efficiency for agricultural productivity and suggest that land misallocation may explain the large productivity differences in agriculture across countries.

Keywords: misallocation, land, productivity, agriculture, Malawi, micro data.
JEL codes: O1, O4.
1 Introduction

A critical question in the field of economic growth and development is why some countries are rich and others are poor. The literature has offered many useful perspectives but in this paper we focus on two. First, the literature has argued that agriculture plays an important role in accounting for productivity differences between rich and poor countries (e.g., Restuccia, Yang, and Zhu, 2008). This is because poor countries are much less productive and allocate most of their labor in agriculture than rich countries. This literature has also shown that low productivity in agriculture with a subsistence constraint on food can explain the large fraction of employment in agriculture in poor countries. As a result, the key question is why labor productivity is so low in poor countries. Second, the literature has argued that the (mis)allocation of factors of production across heterogeneous production units can explain differences in measured productivity across countries (e.g., Banerjee and Duflo, 2005; Restuccia and Rogerson, 2008; and Hsieh and Klenow, 2009). Moreover, Adamopolous and Restuccia (forthcoming) show that misallocation in the agricultural sector may explain low productivity in poor countries. In particular, in understanding low productivity in agriculture in poor countries, we argue in this paper that misallocation of land may be important. The evidence suggests this is plausible in poor countries: most labor and land operate in subsistence rural agriculture, farms (family farms) operate at extremely low operational scale, and market for land restricted by institutional and policy features.

We assess the quantitative importance of land misallocation using detailed household-level micro data from Malawi. Malawi represents an interesting case since the land market is largely underdeveloped: Land is typically transmitted by inheritance (73%), granted by local leaders (11.7%), or received as bride price (1.9%); only 1.1% of land is purchased with a title; and only 6.9% of land is rented-in. Moreover, the land distribution in Malawi is fairly concentrated in extremely small operational scales, with more than 70 percent of all households operating less than 2 acres of land (less than an hectare). Moreover, using the detailed micro data we
estimate farm productivity—which are approximately log normally distributed with substantial
dispersion—and show that essentially land size is unrelated to farm productivity. This finding
is indicative of land misallocation across heterogeneous productive farmers.

To assess the role of land misallocation on agricultural productivity in Malawi, we develop a
model of farm size along the lines of Lucas (1978), Adamopoulos and Restuccia (forthcoming),
and Bello, et al. (2011). Individuals are heterogeneous in their ability to operate a farm. A farm
is a technology that requires the input of the farmer and land. We abstract from capital, hired
labor and labor supply differences across farmers, however, we do control for those differences in
our empirical counterpart of the model. Farmers have the option to invest in the productivity of
the farm operation at a cost. Hence, in this framework, land misallocation affect the expected
value of farms and may discourage productivity investments. To put it differently, the inability
to obtain the land required to grow the operational scale associated with higher productivity
prevents productive investments. We calibrate the model to macro and micro data for Malawi.
In particular, we are able to estimate and identify with the model both total factor productivity
at the farm level and the output wedges that are required to rationalize the observed land
distribution in a well-functioning land market. The counterfactual where output wedges are
removed—that is where land misallocation is eliminated so land is allocated efficiently across
farmers—produces a 4-fold increase in agricultural productivity. This is a large productivity
impact of misallocation compared to the related literature on misallocation and productivity.
Moreover, this increase in productivity results from reallocating land (and other factors) across
existing farmers. It is well known in the literature that higher productivity in agriculture from
better allocation of factors will result in fewer farms, increasing farm size and potentially further
increasing labor productivity in agriculture.

Our paper relates to the broad literatures on misallocation and productivity and misallocation
in agriculture discussed earlier. In addition, our paper is closely related to a set of papers em-
phasizing the misallocation of land as an important factor in understanding low productivity
in agriculture in poor countries, for instance Adamopoulos and Restuccia (forthcoming), Chen (2014), and Adamopoulos and Restuccia (2014). Chen (2014) extends a standard model of farm size allowing for untitled land to study the aggregate impact of differences in the extent of untitled land across countries. Adamopoulos and Restuccia (2014) study the misallocation and productivity impact of a specific land policy: land reforms. We differ from these papers in studying land misallocation in an economy with essentially no land market to assess the aggregate impact of land misallocation. While our paper has a weaker policy prescription relative to the previous papers, the aggregate impact of land misallocation we find is several orders of magnitude larger than previously found, suggesting the large role of market imperfection in land markets for the low productivity in poor countries. Like Adamopoulos and Restuccia (2014), our paper uses micro data to study macro development, an approach that relates to a growing literature in macroeconomics (e.g. Hsieh and Klenow, 2009; Buera, Kaboski, and Shin, 2014; Gollin, Lagakos, and Waugh, 2014; and Koh and Santaeulalia-Llopis 2014 among others).

The paper proceeds as follows. In the next section, we describe the important elements of the micro data for our analysis and in Section 3 we report our main findings from the micro data and other aggregate data. In Section 4 we describe the model. Section 5 we calibrate and estimate the model to macro and micro data of Malawi, perform a series of quantitative experiments to assess the productivity impact of land misallocation, and discuss the results. We conclude in Section 6.

2 Data

We use a new and unique nationally-representative household data from the Malawi 2010/11 Integrated Survey on Agriculture (ISA) collected by the World Bank. The original sample in-

\footnote{For a comprehensive discussion of the ISA surveys for different Sub-Saharan countries see de Magalhaes and Santaeulalia-Llopis (2014)}
cludes 12,271 households of which about 81% live in rural areas. The survey follows a stratified 2-stage sample design. First, 768 enumeration areas (EAs) were selected with probability proportional to size (PPS) within each district; Malawi is divided by 27 districts. Second, random systematic sampling was used to select 16 primary households and 5 replacement households from the household listing for each sample EA. ISA provides excruciating detail on agricultural production, land and its quality, agricultural capital equipment and structures, and individual labor supply of all household members (including hired labor or free/exchanged labor). This detailed information on inputs and outputs makes this dataset ideal for our exercise. Importantly, the sample is rolled over 12 months from March 2010 to March 2011, which implies we can take good care of seasonal behavior of agricultural product, nonagricultural income, consumption, and labor supply. This control over seasonality is impossible to do in previous LSMS data for all these economic components at once. Overall, ISA represents a substantial improvement with respect to previous LSMS questionnaires.

**Agricultural Production** Agricultural production represents 70% of all income in rural Malawi. Agricultural production in Malawi has very well defined seasonal patterns with the major harvest for the rainy season arriving in March. To compute agricultural production we use information collected for each household separately per crop and plot that includes all that was harvested through the previous year: from the rainy season (93%), dry (Dimba) season (3%) and from permanent crops/trees (4%). In this bundle, maize represents (78%) of total production and tobacco (19%). That is, we are clearly facing an economy where subsistence on food consumption is the norm (an issue we come back below). This means that this information is recovered retrospectively. To evaluate potential recall and telescoping error

---

2The sampling frame is based on the 2008 Population and Housing Census (PHC) that includes three major regions of Malawi, namely North, Center and South; and is stratified into rural and urban strata. The urban strata include the four major urban areas: Lilongwe City, Blantyre City, Mzuzu City, and the Municipality of Zomba.

3To control for monthly variation we need a second wave of data (that is, we need at least two observations for the same month, but different years). We use the Malawi IHS collected between March 2004 and March 2005 for that purpose, see de Magalhaes and Santaaulalia-Llopis 2014

4In our computations, we do not remove losses (i.e., harvest crop that was lost in the post-harvest period because of incidents, rodents, thefts, etc.).
we conduct robustness of our results by only using households interviewed within 3 months of the last rainy season. The quantity produced of a same crop can be reported in different units by households (kg, bags, heaps, pails, etc.) To deal with this we use a 'price conversion' unit procedure to convert all crop-units pairs into the same units (say Kg) (see de Magalhaes et. and Santaeulalia-Llopis 2014). Further, an important aspect of agricultural production is that a large part is not sold. To estimating the value of unsold agricultural production we use the median consumption prices of that crop in region-season pairs.

**Land** Our preferred measure of household land is the sum of the size of each cultivated household plot. Then, rented-in land (6.9% of all plots) adds to household land size. The total number of plots cultivated per household are 1.8. Plot size is recorded in acres using GPS (with precision of 1% of an acre) for 98% of plots (for the remaining 2% of plots, size is estimated). Further, we have built a quality-land index (see details in Appendix) including the following plot characteristics: soil quality (good, fair, or poor), extent of erosion (none, low, moderate, high), slope (flat, slight slope, moderate slope, or steep), and irrigation system (divert stream, bucket, hand pump, treadle pump, motor pump, gravity, no irrigation, and other), and the method of obtaining water for the plot (watering can, hose pipe, sprinkler, drip irrigation, flooding, and other (spec.)). Our index and the self-reported price are highly correlated, .46. While in our benchmark analysis we do not adjust for quality, we find that

---

5The units of each crop bought (or sold) are first converted to a modal unit used for reporting. Unit price for the same crop in a given region and season are used to generate household specific conversion rates. We pick the median conversion rate (if there are at least 7). All items are converted into Kg.

6An example of this strategy described in de Magalhaes and Santaeulalia-Llopis goes as follows: Farmers sell maize in grain. That has an average price at the gate of 40 MWK per Kg. However, the average price of ufa (maize flour) which is how most maize is consumed is 60 MWK per kg. The rationale for using consumption prices is that they represent the opportunity cost of having not produced the maize and, hence, needing to buy it in the market. Instead, the World Bank uses the sale price at the gate to estimate income (this is true for all previous LSMS datasets). We find that whether we use consumption or at-the-gate prices does not alter the quantitative implications of misallocation.

7The area harvested might be less than the area planted (drought, rains, fire, insects, etc).

8Land represents a large part of household wealth portfolio, 46%. House is 27%, livestock is 11%, agricultural equipment and structures (tools, barns) is 3%, see de Magalhaes and Santaeulalia-Llopis 2014

9An alternative measure: self-reported value: “If you were to sell this [PLOT] today, how much could you sell it for?” No land markets might introduce measurement errors in self-reported land values. However, (logged) land size and land value are positively correlated .37 and the variance of logged land size is 60% the variance of logged
our results on the implications of misallocation will not subject to quality adjustments.

**Agricultural Capital Equipment and Structures** Agricultural capital equipment includes implements (hand hoe, slasher, axe, sprayer, panga knife, sickle, treadle pump, watering can, and so on) and machinery (e.g., ox cart, ox plough, tractor, tractor plough, ridger, cultivator, generator, motorized pump, grain mill, and so on). Agricultural capital structures includes chicken houses, livestock kraals, poultry kraals, storage houses, granaries, barns, pig sties, etc. We use the estimated current selling price of capital items (if you wanted to sell one of this [ITEM] today, how much would you receive?) as proxy for capital services after conditioning on its use (did your household use the [ITEM] during the last 12 months?). We aggregate across items; hence, no asset index is needed. This selling price (not available in previous LSMS data) takes care of capital quality and depreciation (some items might be of lesser quality or broken or deteriorated). It also avoids the cumbersome use of the age of capital to impute current value from the value at the time of purchase (which requires recalling and depreciation assumptions by asset’s age).

**Individual and Household Labor Supply** In Malawi, a large proportion of the households members, beyond the household head, contribute to agricultural work. In Malawi, household size is 4.57 with extended families in which several generations live together in a single household. We define household members as individuals that have lived in the household at least 9 months in the last 12 months. Individual information about each household member’s extensive and intensive margins of labor supply is collected: (i) weeks worked, (ii) days per week, and (iii) hours per day. This information is provided retrospectively (as agricultural production) by plot, by agricultural activity covering the entire agricultural production: land preparation/

---

10 The household head is identified as the person who makes economic decisions in the household.
11 Household members potentially include family (e.g., children, spouses, siblings, and parents) and also non-relatives (e.g., lodgers and servants). Individual characteristics of each household member are collected. They include sex, age, language, religion, educational attainment, health behavior and status, marital status, societal system, geographical variables, migration characteristics, and so on.
planting, weeding/fertilizing, and harvesting and by season (rainy, dry and permanent). The hired labor and free/exchange labor (e.g. exchange labourers or assistance for nothing in return) number of days worked by men, women and children are also collected by plot, activity and season. Our preferred measure of household labor supply is aggregate hours of all individuals (household members and nonmembers) supplied in all plots cultivated by the household in the rainy season. Sensitivity on these measures is done (e.g. head hours, only members hours, head days or weeks, and so on). This detailed information on individual agricultural hours through the entire year jumps over any seasonal component of labor supply; that is, we do not rely on data on labor supply related to 'last week/month' behavior.  

**Measurement Error** There are very few missing observations, a blessing for a survey in poor countries. Our understanding from the World Bank field managers in charge of the data collection is that this is due to the fact that respondents took the survey as 'official'. Further, the collection strategy (survey questions) in many instances check for internal consistency reliability (e.g., individuals are asked total sales, and also sales by crop; the interviewer checks that the sums coincide). This helps decrease any type of measurement error. We exclude outliers: Trimming the top and bottom 1% (increasing the trimming does not change much our productivity results). While not in our benchmark, to deal with potential recall and telescopic error in agriculture production and activities we reconduct our exercise using only the households that were interviewed within three months of the harvest period; these imples one-fourth of our total observations.  

---

12It is well known that in questionnaires that are not rolled-over the entire year, labor supply/wages that refer to 'last week/month' have a seasonal bias; even with a questionnaire that is rolled over the entire year, the distribution of hours will be distorted unless data are de-seasonalized (e.g. by using a second rolling survey). Our ISA data do not suffer from this bias.

13We need to redo it for rainy and dry seasons separately as well.
3 Facts

Malawi is an extremely poor country in Africa. According to international comparable measures of output across countries, Malawi GDP per capita is 1/35 that of the United States. Moreover, growth in GDP per capita has not been able to keep pace to that of developed countries so that relative GDP per capita has fallen from 1/29 in 1950 to around 1/50 in the mid 2000. The difference in income reflects mostly a difference in labor productivity so GDP per worker or per labor hour is similar to that of GDP per capita. But labor productivity in the agricultural sector is much lower in Malawi, relative to the aggregate, for instance, agricultural GDP per worker relative to that of the United States in 1990 is 0.96 percent (a 104-fold difference)\textsuperscript{14} The distinction between aggregate productivity and agricultural productivity is important since in Malawi most of the employment is in the agricultural sector: the share of employment in agriculture in Malawi for 1990 is 62% whereas this share is only 2.5% in the United States. As emphasized in Restuccia, Yang, and Zhu (2008) this suggests that differences in labor productivity in the non-agricultural sector are much smaller than the aggregate and that understanding low productivity in Malawi hinges critically in understanding low productivity in the agricultural sector. Before we move on to analyze the micro data, we note that the operational scale of farms in Malawi is extremely small, with an average farm size of 0.7 hectares, whereas average farm size is 187 hectares in the United States and 16.1 hectares in Belgium. Belgium is a good reference for a developed country since the land endowment (measured as land per capita) is similar to that of Malawi (land per capita is 0.56 ha in Malawi, 0.5 ha in Belgium, and 1.51 in the United States). We also note that 82% of land in Malawi is suitable for cultivation of which 16% held by estates (export crops) and 84% by smallholders (mostly customary). See Table 1 for summary statistics on farm size between Malawi, Belgium and the United States.

Using the micro data for Malawi from ISA described in the previous section, we emphasize the

\textsuperscript{14}Data for 1990 from Adamopoulos and Restuccia, (forthcoming). For the aggregate GDP per worker (relative to the US) is 2.6% in 1990 (39-fold difference).
Table 1: Size Distribution of Farms (% of Farms by Size)

<table>
<thead>
<tr>
<th>Hectares</th>
<th>Malawi (cum)</th>
<th>Belgium (cum)</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 1 Ha</td>
<td>77.7</td>
<td>14.6</td>
<td>–</td>
</tr>
<tr>
<td>1 – 2 Ha</td>
<td>17.3</td>
<td>8.5</td>
<td>–</td>
</tr>
<tr>
<td>2 – 5 Ha</td>
<td>5.0 (100)</td>
<td>15.5 (38.6)</td>
<td>10.6</td>
</tr>
<tr>
<td>5 – 10 Ha</td>
<td>0.0</td>
<td>14.8</td>
<td>7.5</td>
</tr>
<tr>
<td>10+ Ha</td>
<td>0.0</td>
<td>46.6</td>
<td>81.9</td>
</tr>
<tr>
<td>Average Farm Size (Ha)</td>
<td>0.7</td>
<td>187.0</td>
<td>16.1</td>
</tr>
</tbody>
</table>

Note: From the World Census of Agriculture 1990 reported in Adamopoulos and Restuccia (forthcoming).

following observations.

**Operational Scale**  The operational scale of farms is extremely small. For each household, we have information of the amount of land used for agricultural production regardless of the land status (whether land is owned, rented, etc.). We find that more than 70 percent of farms operate under 2 acres (less than 1 hectare). Table 2 reports the size distribution of farms in the household data.

Table 2: Size Distribution of Farms

<table>
<thead>
<tr>
<th>Acres</th>
<th>% of Farms</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 1</td>
<td>40.0</td>
</tr>
<tr>
<td>≤ 2</td>
<td>73.0</td>
</tr>
<tr>
<td>≤ 3</td>
<td>90.0</td>
</tr>
<tr>
<td>≤ 4</td>
<td>95.0</td>
</tr>
</tbody>
</table>

**Land Use**  Most farms use the land they have (almost no sales and rental of land).

**Subsistence Consumption**  In the micro data, most farmers produce for own consumption. To illustrate this finding, we report in Table 3 some standard proxies for subsistence
consumption for each decile of the distribution of agricultural production. The first proxy, Food Insecurity in the second column, reports the response to the following question: “Over the past 12 months, has the household faced a situation in which there was not enough food?” The answer is recorded as yes or no and we report in this column the fraction of households in each decile that responds yes. More than 80 percent of households in the lowest decile faced a situation of not enough food. Even for the households in the highest decile, almost 30 percent faced some shortage of food. The second proxy we report, food consumption over non-durable consumption (third column), indicates the high degree of reliance on food of household’s consumption budget, in average 67 percent of household non-durable consumption is food. The last proxy we report is food consumption relative to agricultural production for each household. Most household consume more food than they produce.

Table 3: Subsistence Consumption, Malawi LSMS-ISA 2010

<table>
<thead>
<tr>
<th>Ag. Production Deciles</th>
<th>Food Insecurity (last 12m)</th>
<th>Food Consumption/ Nondurable Cons.</th>
<th>Food Consumption/ Ag. Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom 10%</td>
<td>80.7</td>
<td>0.56</td>
<td>3.09</td>
</tr>
<tr>
<td>10-20%</td>
<td>67.8</td>
<td>0.64</td>
<td>2.80</td>
</tr>
<tr>
<td>20-30%</td>
<td>65.7</td>
<td>0.65</td>
<td>2.34</td>
</tr>
<tr>
<td>30-40%</td>
<td>60.6</td>
<td>0.68</td>
<td>2.00</td>
</tr>
<tr>
<td>40-50%</td>
<td>57.4</td>
<td>0.68</td>
<td>1.87</td>
</tr>
<tr>
<td>50-60%</td>
<td>50.4</td>
<td>0.67</td>
<td>1.44</td>
</tr>
<tr>
<td>60-70%</td>
<td>47.4</td>
<td>0.68</td>
<td>1.22</td>
</tr>
<tr>
<td>70-80%</td>
<td>43.4</td>
<td>0.68</td>
<td>1.02</td>
</tr>
<tr>
<td>80-90%</td>
<td>35.2</td>
<td>0.69</td>
<td>0.80</td>
</tr>
<tr>
<td>Top 10%</td>
<td>28.0</td>
<td>0.68</td>
<td>0.53</td>
</tr>
<tr>
<td>Total</td>
<td>50.6</td>
<td>0.67</td>
<td>1.54</td>
</tr>
</tbody>
</table>

Figure 1 reports the histograms of the share of food in total consumption and the share of food in agricultural production. In each case, the vertical lines represent the median in each distribution.
Farms by Size  We first characterize farms by land size and summarize this in Figure 2. In Panel A and B we observe that larger farms use more capital and put more hours than smaller farms, with the capital to hour ratio being constant across land size as indicated in Panel C. This implies that capital and hours are tightly connected to the amount of land in the farm (household). Panels E and F document the output per unit of land and output per hour. Both of these are roughly constant across farm size. This is consistent with capital and hours tightly related to land and then output. To understand these patterns, note that in a standard framework of farm size (e.g., Adamopoulos and Restuccia, forthcoming), with no frictions and distortions, more productive farms are larger and demand more capital and hours than less productive farms. And the larger demand for factors tend to offset differences in productivity so as to the value of marginal products equalize across farms. Indeed, under standard Cobb-Douglas production function for farms, this process will tend to equalize the yield (output per unit of land) and labor productivity across farmers. Hence, this characterization seems indicative of efficiency in factor markets in Malawi. But is it the case? The answer hinges critically whether larger farms are indeed more productive than smaller farms. Using the micro data we find no connection productivity and size among farms in Malawi, we document this fact next.
Figure 2: Capital, Labor, Land and Output by Farm Size

Panel A: Capital

Panel B: Hours

Panel C: Capital per Hour

Panel D: Capital to Land Ratio

Panel E: Output per Land

Panel F: Output per Hour
Farms by Productivity  To investigate the connection between productivity and size across farms in Malawi, we estimate farm productivity by assuming a simple production function. We estimate farm-level total factor productivity as $z_i = y_i/l_i^\gamma$. The production function at the farm level associated with this equation is the same we use in the model in the next section. Nevertheless, we have verified that our finding are robust to other more elaborate specifications of the production function. Figure 3 reports variables by farm-level TFP. The patterns are striking. The data show that capital and hours in farms are unrelated to farm productivity (Panels A and B) so unlike in the previous findings where capital and hours were tightly related to land size, capital and hours are not which indicates that the size of farms is not related to productivity. Indeed, in Panel C we corroborate that land is unrelated to farm productivity (perhaps even slightly negatively related!) and this finding is consistent with our characterization of the land market in Malawi where the amount of land in farms is related to inheritance, norms, and redistribution and that access to land is severely restricted in rental and sale markets so more productive farmers cannot grow their farm size. These observations indicate that output per unit of land (Panel E) and output per hour (Panel F) are positively related to farm productivity. Roughly speaking, with a constant amount of land per farm, output is tightly related to farm productivity because more productive farmers produce more output given inputs but cannot adjust the amount of land, whereas output per hour is also positively related to productivity but less so than land productivity. But note that in the context of the framework described previously, output per unit of land and labor productivity should be constant across farm productivity, so this is evidence of land misallocation.

How important is land misallocation for productivity? We assess the quantitative contribution of factor misallocation documented in this section for productivity by considering an otherwise standard model of farm size estimated using the micro data for Malawi. We pursue this in the next two sections.
Figure 3: Capital, Labor, Land and Output by Farm TFP

Panel A: Capital

Panel B: Hours

Panel C: Land

Panel D: Capital to Land Ratio

Panel E: Output per Land

Panel F: Output per Hour
4 Model

We develop an industry model of farm size with endogenous farm productivity investment. The model builds from the framework of Lucas (1978), Adamopoulos and Restuccia (forthcoming), and Bello, et al. (2011). We focus on the determinants of productivity in the agricultural sector and hence abstract from the sectoral composition of the economy. See Admopoulos and Restuccia (forthcoming) for a treatment of farm size and productivity in a general equilibrium two-sector model.

Technology There is a single good produced in each period. The unit of production in agriculture is a farm (household farms) that differ on total factor productivity. We assume there is a large number of farms indexed by \( i \) with access to the following decreasing returns to scale production function

\[
y_i = A\kappa(zs_i)^{1-\gamma}x_i^\gamma, \quad 0 < \gamma < 1,
\]

where \( x \) is a composite factor input which may include land, capital and other factors. In what follows we assume the only input is land so \( x = l \) but we discuss below alternative specifications. The parameter \( A \) is economy-wide productivity, \( \kappa \) is an agriculture-specific productivity parameter, and \( zs \) is idiosyncratic productivity. There are two components of farm productivity: \( z \) is determined by endogenous investment (once-and-for-all) at a cost \( c(z) \) which will be the same across all farmers and \( s \) is idiosyncratic across farmers and is drawn from a known distribution \( F(s) \). This distribution will be estimated using the micro data from Malawi. Both \( z \) and \( s \) are constants and known before the demand decision for land. Note that from the specification of the farm technology for output, data on farm output \( y_i \) and land \( l_i \), \( s_i \) can be estimated up to a constant. Total factor productivity at the farm level (up to the constants \( A\kappa z \)) is then given by \( s_i^{1-\gamma} \) as reported in Section 3.
Static Farm’s Problem  Since our objective is to evaluate the extent and productivity impact of land misallocation using the micro data, it will be convenient to characterize land misallocation as arising from output distortions (either taxes or subsidies on output) in an otherwise standard market clearing condition for land. The problem for the farmer of maximizing profits is given by:

\[ \pi(s_i, \tau_i; z) = \max_{l_i} \left\{ A \kappa(z s_i)^{1-\gamma} l_i^\gamma - q l_i \right\}, \]  

implying the following first order condition for the demand for land \( l_i \):

\[ (1 - \tau_i) A \kappa(z s_i)^{1-\gamma} x_i^{\gamma-1} = q. \]  

This condition implies the following land demand function:

\[ \bar{l}(z, s_i, \tau_i) = \left( \frac{\gamma (1 - \tau_i) A \kappa}{q} \right)^{\frac{1}{\gamma}} z s_i. \]  

We can use FOC and data on \( l_i \) and estimated \( s_i \) to get \( \tau_i \) (up to a constant), i.e., the \( \tau \)'s that rationalize the observed land sizes given the estimated productivities.

Productivity Investment  Farmers can invest in the productivity of the farm \( z \) at a cost \( c(z) \). This decision occurs before the realization of \( s \). (We can explore an alternative assumption where this decision occurs after the realization of \( s \) so different farmers will invest differently in \( z \).) In order to characterize the optimal choice of \( z \) for the farmer, it is convenient to note that the land demand in equation (4) is linear in \( z \). Substituting \( \bar{l} \) into \( \pi \) implies that profits are also linear in \( z \). We use this property of profits to write \( \pi \) as \( \pi = z \hat{\pi} \). Then the problem of a farmer is choosing \( z \) to maximize the value of the farm \( W \) by solving:

\[ W = \max_z \left\{ z \sum_{s, \tau} \hat{\pi}(s, \tau; q) g(s, \tau) - c(z) \right\}. \]
We denote $\bar{z}$ the optimal productivity choice. In our quantitative work, we assume the following cost function $c(z) = Bz^\phi$ with $\phi > 1$. Hence, the first order condition from this problem implies that

$$
\bar{z} = \left( \frac{\sum_{s,\tau} \pi(s, \tau)g(s, \tau)}{B\phi} \right)^{1/(\phi-1)}.
$$

Note also that $\bar{l}$ is strictly decreasing in $q$ so in this framework $q$ is such that it equates the given supply of land with the aggregate demand from farms.

**Definition of Equilibrium**  A competitive equilibrium is a price for land $q$, decision rules and value of farms $\bar{l}(s, \tau)$, $\bar{z}$, $\pi(s, \tau)$, and $W$, such that: (i) Given $q$, $W$ and $\pi(s, \tau)$ are maximized in 2 and 5 and $\bar{l}$ and $\bar{z}$ are optimal decision rules from these problems given by equations (4) and (6), and (ii) the land market clears, $L = \sum_{s,\tau} \bar{l}(s, \tau)g(s, \tau)$.

## 5 Quantitative Analysis

In this section, we calibrate and estimate the model to micro and aggregate data for Malawi. We then conduct a series of experiments to assess quantitatively the extent to which land is misallocated in Malawi and the associated productivity losses of that misallocation. We then discuss these results.

### 5.1 Calibration

We calibrate and estimate the economy to data for Malawi. The parameters to be calibrated are: $\gamma$, $A$, $\kappa$, $B$, $\phi$, $L$, and the variables $s_i$ and $\tau_i$ across farmers. Since our focus is on comparing the outcomes of the current distribution of land in Malawi to a counterfactual alternative where land is distributed according to the productivity of farmers, we normalize each of $A$, $\kappa$ and $B$ to 1. The parameter $\gamma$ determines the returns to scale at the farm level and the profits of
farmers. We select its value based on the micro evidence, $\gamma = 0.85$, which is similar to the values used in the literature (see for instance Restuccia and Rogerson, 2008). There is not a lot of estimates of $\phi$ we can rely on, we set it to 2 based on estimates of innovation in the trade literature but will conduct sensitivity analysis for a reasonable range of values of this parameter. We estimate $s_i$ based on the micro data for individual household farmers described in the data section. We emphasize that the detailed data allows to construct real measures of productivity (where outputs of different crops are aggregated using a common set of prices across households). We construct measures of income that control for many elements including capital use, rain season, geography, type of crop, and total hours of work. We associate the adjusted output per hour measure of each individual household in the data to output by the farm in the model $y_i$ and land per hour to $l_i$. Then given our normalization for $A$ and $\kappa$, the production function of the farm in equation (1) can be used to estimate $zs_i$ as:

$$\bar{z}s_i = \left( \frac{y_i}{A\kappa l_i^\gamma} \right)^{1/(1-\gamma)}.$$

We note that $s_i$ and $\bar{z}$ (and $A$ and $\kappa$) cannot be separated by the data but it is possible in the model. Nevertheless, given the assumptions that imply a constant $\bar{z}$ across households, the data pins down the range of differences in idiosyncratic productivity since the ratio of farm productivity among any given farmers depend only on the idiosyncratic component (i.e., $A$, $\kappa$, and $\bar{z}$ are constant across farmers). Figure 4 documents the histogram of relative farm-level TFP, $(s_i/s_j)^{1-\gamma}$.

We note that the distribution is well approximated by a log normal with a significant range, for instance the log difference of the highest and lowest productivity is 5 log points which translates into a 150-fold difference. The log difference between the 90 to 10 percentiles is 2.1, the 75 to 25 percentiles is 1.1, and the standard deviation of the log of farm TFP is 0.86. This finding is striking given that most farmers operate near subsistence as discussed earlier. To place the dispersion in productivity in perspective we note that this distribution of productivity is similar
to that of the U.S. manufacturing and slightly smaller dispersion that in China and India for the manufacturing sector, as reported by Hsieh and Klenow (2009) and reproduced in Table 4.

The last remaining item to estimate is the implied output wedges needed to reconcile the land allocation in Malawi across farmers. In the context of the model with well-functioning market arrangement for land, output wedges are required for the actual allocation of land to be unrelated to the productivity of the farmers as discussed in Figure 4. Using the first order condition for land in 3, we calculate relative wedges as:

$$\frac{(1 - \tau_i)}{(1 - \tau_j)} = \frac{s_i}{s_j} \left( \frac{l_i}{l_j} \right)^{(\gamma - 1)},$$

where the relative $s_i$'s are calculated as described earlier and $l_i$ is given by the data on farm size. The distribution of these wedges across farmers is reported in Figure 5.
Table 4: Dispersion of Farm and Plant Level Productivity

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Malawi 2010</th>
<th>USA 1977</th>
<th>China 1998</th>
<th>India 1987</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>0.86</td>
<td>0.85</td>
<td>1.06</td>
<td>1.16</td>
</tr>
<tr>
<td>75-25</td>
<td>1.08</td>
<td>1.22</td>
<td>1.41</td>
<td>11.55</td>
</tr>
<tr>
<td>90-10</td>
<td>2.14</td>
<td>2.22</td>
<td>2.72</td>
<td>2.77</td>
</tr>
<tr>
<td>N</td>
<td>10,000</td>
<td>164,971</td>
<td>95,980</td>
<td>31,602</td>
</tr>
</tbody>
</table>

Note: Data for USA, China, and India refer to manufacturing establishments from Hsieh and Klenow (2009). SD is the standard deviation of log productivity; 75-25 is the log difference between the 75 and 25 percentile and 90-10 the 90 to 10 percentile difference in productivity. N is the number of observations in each dataset.

5.2 Counterfactual Experiments

We solve for the equilibrium of the model by first guessing a price of land $q$, solving for the decision rules of the farmers, aggregating the land demands, and then iterating on $q$ until the land market clears. Our economy is identified up to a constant so in what follows we consider a reasonable range for that constant. The constant relates to the average tax/subsidy implied by the land allocation in the data. We consider a situation that most households are taxed (i.e. the allocation of land is such that absent distortions most households would demand more land than what they currently have which is what we think characterizes the situation in Malawi, see for example US Aid and others). This corresponds to a relative factor of 1.0 in which we take the relative output wedges in Figure 5. Another possibility is that half the households are taxed and half subsidized, in this case the relative factor is set to the median of the relative output wedge which in this case is 0.06. We also consider a range of relative factors in between 0.06 and 1.

We compare all the cases discussed above with the case of no distortions, that is a situation where there are no output wedges so land is allocated efficiently on the basis of farm produc-
We note that the increase in aggregate output per capita, TFP, and land productivity from eliminating land misallocation is large in all cases of the relative factor. For instance, the increase in TFP is a factor of 3.3-fold for the case in which half of the farmers are taxed, whereas the factor is 4.2-fold in the case that most farmers are taxed. This is a very large increase in productivity as a result of a reduction in misallocation, at least compared to the misallocation and productivity literature when evaluating specific policies (increases on the
order of 5 to 30 percent) and even from eliminating wedges in manufacturing in China and India relative to the wedges in the United States from Hsieh and Klenow (2009), increases that range between 30-60 percent. Another finding to note from Table 5 is that misallocation of land in a narrow sense accounts for most of the productivity gains from reducing misallocation since the increase in once-and-for-all increase in average productivity $\bar{z}$ only contributes to roughly 6 percent of the productivity gains.

5.3 Discussion

We note that a drawback of the current setup is that productivity gains are not associated with a moment of labor away from agriculture, a factor that we have abstracted from. For this reason, average farm size remains constant in all the experiments we conducted. As is well known from standard sectoral models, an increase in productivity in agriculture is associated with a reduction in the share of employment in agriculture which will increase farm size, see for example Adamopoulos and Restuccia (forthcoming). Moreover, if there is selection in the movement of labor away from agriculture, for instance as emphasized in Lagakos and Waugh (2013), the increase in agricultural productivity can be larger, in the calibration of Lagakos and Waugh up to 2 times larger. We plan to incorporate the general equilibrium effect by extending the model to a two sector setting. We will also explore the selection channel.

Our analysis has abstracted from capital and other inputs besides land. This abstraction is motivated by the empirical evidence in Section 3 that capital and hours are tightly linked to land. As a result, the key misallocation is land and the model emphasizes the productivity impact of this misallocation assuming all other factors will move in tandem.

The biggest limiting factor in the data is the fact that we have a single cross section. While the data is able to isolate seasonal patterns, it may be the case that the estimated productivity does not reflect a permanent measure of productivity for each farmer. We plan to address this in
two forms. First, by evaluating the robustness of results under certain averaging and by using other pieces of data that are correlated to weather shocks for instance rain measures. Second, by exploiting a new round of data collected in a recent time period that will allow us to have a panel dimension.

6 Conclusions

We assessed the importance of land misallocation for productivity in agriculture using a quantitative model and detailed household-level data from Malawi. The micro evidence indicated substantial misallocation of land in Malawi as farm size is unrelated to farm productivity, partly the result of a largely underdeveloped land market. We considered an industry model of agriculture featuring a non-degenerate distribution of farm sizes. We calibrated and estimated the model using aggregate and micro data for Malawi. Under the counterfactual that land is efficiently allocated across existing productive farmers, we find an increase in agricultural productivity of a factor of 4-fold. This increase in productivity is orders of magnitude larger than the ones found in the related literature on misallocation and productivity (see for instance Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013; and Hopenhayn, 2014). This result emphasizes the importance of land market efficiency for agricultural productivity and suggest that land misallocation may explain the large productivity differences in agriculture across countries.

Our analysis takes the distribution of farmlands across productive farmers as given and asked about the efficiency gains of reallocation. Understanding the institutional and policy elements leading to the misallocation of land is useful in connecting the results with policy and development strategies. It may also be of interest to study the dynamic implications of misallocation for productivity (see for instance Restuccia and Rogerson, 2013 and Restuccia, 2013) whereby a reduction in misallocation encourages the more productive farmers to grow, utilize modern
inputs (mechanization, chemical seeds, etc.), and invest in the quality of land. Further, in a
consumption insurance exercise using panel data for Malawi, Uganda and Tanzania, de Magal-
haes and Santaulealia-Llopis (2014) show the ability of farmers to insure against income risk is
rather low, in particular for those with lower land holdings independently of their agricultural
productivity, implying that the missallocation of land potentially alters the ability to insure
consumption. We leave these interesting and important extensions of our analysis for future
research.
References


