DISCUSSION PAPER SERIES

DP14553

PRODUCTIVITY DISPERSION AND PERSISTENCE AMONG THE WORLD'S MOST NUMEROUS FIRMS

Kyle Emerick, Marshall Burke and Casey Maue

DEVELOPMENT ECONOMICS

CEPR

PRODUCTIVITY DISPERSION AND PERSISTENCE AMONG THE WORLD'S MOST NUMEROUS FIRMS

Kyle Emerick, Marshall Burke and Casey Maue

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Development Economics

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

PRODUCTIVITY DISPERSION AND PERSISTENCE AMONG THE WORLD'S MOST NUMEROUS FIRMS

Abstract

A vast firm productivity literature finds that otherwise similar firms differ widely in their productivity and that these differences persist through time, with important implications for the broader macroeconomy. These stylized facts derive largely from studies of manufacturing firms in wealthy countries, and thus have unknown relevance for the world's most common firm type, the smallholder farm. We use detailed micro data from over 12,000 smallholder farms and nearly 100,000 agricultural plots across four countries in Africa to study the size, source, and persistence of productivity dispersion among smallholder farmers. Applying standard regression-based approaches to measuring productivity residuals, we find much larger dispersion but less persistence than benchmark estimates from manufacturing. We then show, using a novel framework that combines physical output measurement, estimates from satellites, and machine learning, that about half of this discrepancy can be accounted for by measurement error in output. After correcting for measurement error, productivity differences across firms and over time in our smallholder agricultural setting closely match benchmark estimates for non-agricultural firms. These results question some common implications of observed dispersion, such as the importance of misallocation of factors of production.

Kyle Emerick - kyle.emerick@tufts.edu
Tufts University and CEPR

Marshall Burke - mburke@stanford.edu
Stanford University

Casey Maue - cmaue@stanford.edu
Stanford University

Productivity dispersion and persistence among the world's most numerous firms

Casey Maue[1,2*], Marshall Burke[2,3,4], Kyle Emerick[5,6]

[1]Emmett Interdisciplinary Program on Environment and Resources, Stanford University,

[2]Center on Food Security and the Environment, Stanford University

[3]Department of Earth System Science, Stanford University

[4]National Bureau of Economic Research

[5]Department of Economics, Tufts University

[6]Centre for Economic Policy Research

Abstract   A vast firm productivity literature finds that otherwise similar firms differ widely in their productivity and that these differences persist through time, with important implications for the broader macroeconomy. These stylized facts derive largely from studies of manufacturing firms in wealthy countries, and thus have unknown relevance for the world's most common firm type, the smallholder farm. We use detailed micro data from over 12,000 smallholder farms and nearly 100,000 agricultural plots across four countries in Africa to study the size, source, and persistence of productivity dispersion among smallholder farmers. Applying standard regression-based approaches to measuring productivity residuals, we find much larger dispersion but less persistence than benchmark estimates from manufacturing. We then show, using a novel framework that combines physical output measurement, estimates from satellites, and machine learning, that about half of this discrepancy can be accounted for by measurement error in output. After correcting for measurement error, productivity differences across firms and over time in our smallholder agricultural setting closely match benchmark estimates for non-agricultural firms. These results question some common implications of observed dispersion, such as the importance of misallocation of factors of production.

1 Introduction

Why are some firms more productive than others, in terms of their ability to turn inputs into outputs? This question lies at the center of a vast 'firm productivity' literature which seeks to describe broad patterns of economic performance as a function of the productivity of individual firms. Over time, this literature has established a set of key empirical stylized facts, namely that productivity differences among firms are large, even within narrowly-defined industries, and that these differences persist through time (Syverson, 2011).

These stylized facts derive largely from studies of manufacturing firms in developed countries. Figure 1 plots estimates of productivity dispersion we compiled from more than 30 published articles against average per capita income. Over 80% of estimates are based on data from manufacturing firms, and over 40% come from firms in countries with per capita income greater than 10,000USD. While manufacturing is a key sector in many developed countries, the most common type of firm and majority employer in developing countries is typically the small family-owned farm. Data from the World Census of Agriculture indicate that there are ∼570 million individual farms in the world, a number roughly on par with the estimated total number of firms across all non-agricultural sectors combined (Stein, Goland, and Schiff, 2010). Of these, 84% are less than two hectares in size and nearly all are family-owned (FAO, 2014). Approximately 2.5 billion people are estimated to reside in these "smallholder" farm households, representing at least 60% of the world's poor (Christen and Anderson, 2013). In terms of employment, using nationally-representative surveys from 25 African countries from the years 2006-2012, McMillan and Harttgen (2014) find that 48% of adults over the age of 25 are employed in agriculture.

Are productivity dynamics of developing-country agricultural firms different from their better-studied non-agricultural counterparts in developed countries? Figure 1 suggests that productivity dispersion is higher in countries with lower income levels, where agriculture constitutes a much higher share of total value added and employment. Yet there are few existing works that use micro data on agricultural firms across several developing countries to answer the basic questions about productivity dispersion and persistence that have been part-and-parcel of the mainstream firm productivity literature for decades. Given the important structural role smallholder agricultural firms occupy, closing this gap represents an important step towards establishing a comprehensive understanding of how firm productivity affects development and growth outcomes.

In this paper we characterize the size and sources of productivity dispersion and persistence

among smallholder agricultural firms. We use detailed micro data from over 12,000 small-holder farms ($> 93,000$ agricultural plots) collected in household-panel surveys conducted in four countries in Sub-Saharan Africa (Tanzania, Uganda, Nigeria, and Ethiopia). We estimate productivity as a reduced-form residual in the log-log regression of agricultural output on factor inputs, plus additional covariates and fixed effects. These additional explanatory variables allow us to measure productivity in a manner consistent with the existing research on non-agricultural firms, as well as control for environmental factors that play a relatively stronger role in agricultural production.

We document large apparent dispersion in productivity across firms in each country. As measured by the ratio of the 90th to the 10th percentile of the estimated productivity distribution, we find that dispersion is a factor of 1.2-2.0 times greater than previous estimates based on data from manufacturing firms. Persistence, as measured by the annualized autocorrelation of household-level productivity, is lower by a factor of two. Put simply, while some firms appear to be many times more productive than others in a given year, they might only be weakly more productive in subsequent years. To our knowledge, these are the first estimates of both the dispersion and persistence of total factor productivity differences for agricultural firms.

A primary contribution of our paper is then to advance our understanding of how much of the large productivity dispersion we observe among these firms can be accounted for by measurement error in either inputs or outputs, or by misspecification of the mapping of inputs to output. The motivation here is straightforward. In the firm productivity literature, a variety of economic mechanisms have been hypothesized as root causes or important consequences of productivity dispersion. These include misallocation, insecure property rights, and unobserved heterogeneity in managerial talent. These findings, and the policy prescriptions that arise from them, depend critically on accurate measures of productivity. However, accurately measuring productivity from actual production data can be difficult for any type of firm, and smallholder farms are no exception. In our data setting, farmers grow multiple crops, harvest multiple times, frequently don't keep formal records, and can be surveyed weeks or months after harvesting their fields. In addition, land tenure systems are typically informal, and farmers often struggle to estimate the sizes of their plots. Yet only in the past few years have researchers begun to analyze whether measurement error is an important source of measured productivity dispersion across firms (ex. Bils, Klenow, and Ruane (2017)). Focusing on agriculture, Gollin and Udry (2019) show that when taken together, measurement error, unobserved heterogeneity, and idiosyncratic shocks account for much of
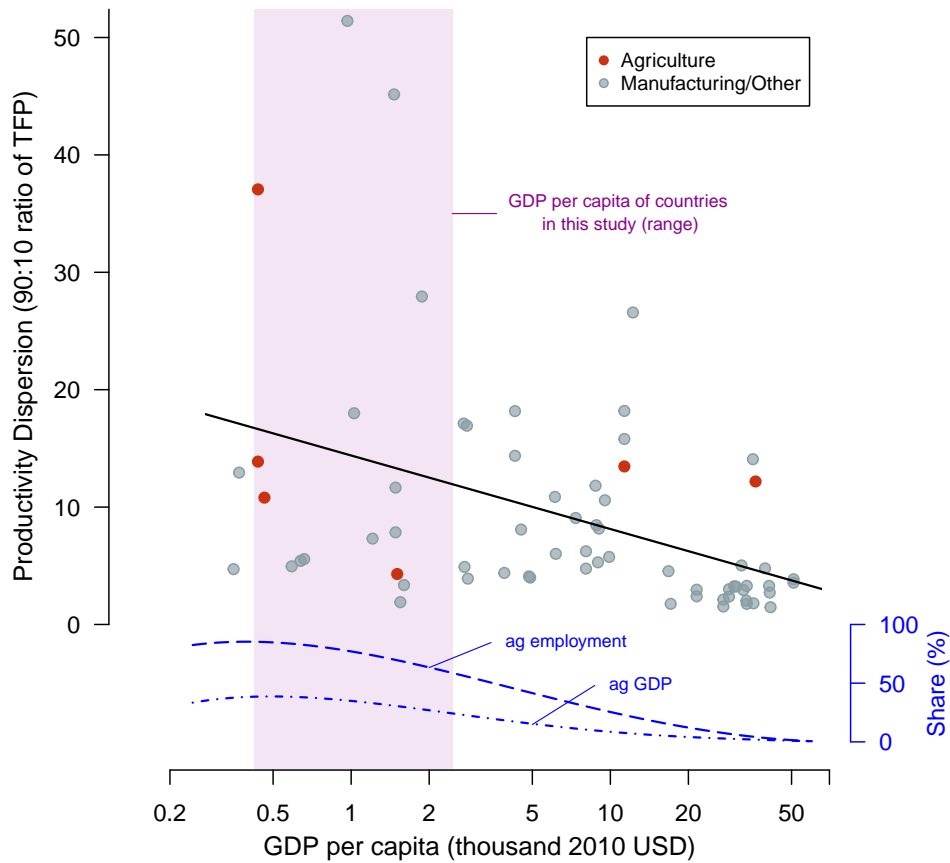
Figure 1: Estimates of productivity dispersion in previous studies of firm productivity. There are few studies of firm-level productivity dispersion in the agricultural sector (red dots), and the estimated relationship suggests dispersion may be higher at lower incomes where agriculture is a more important component of the overall economy. Each point represents an individual study's estimate of productivity dispersion across firms in a particular country-year, which we match to GDP per capita in the World Bank's World Development Indicators database (WDI). Data from 30 published economics papers and working papers representing 67 country-year estimates are shown. Grey indicates a study of firms in the non-agricultural sector (>90% manufacturing). Red indicates a study of agricultural firms. The shaded region indicates the range of per capita GDP values for the countries analyzed in this study. The dashed blue lines at the bottom show the distribution of 'share of employment in agriculture' and 'share of GDP from agriculture' for all countries in the WDI.

the productivity dispersion in Tanzania and Uganda.[2] But the literature has yet to quantify the size of measurement error, and how much of overall productivity dispersion it accounts for relative to other sources.

---

[2]Agricultural economists have also recently looked to measurement error as a potential explanation for why smaller farms appear to have higher land productivity (Carletto, Gourlay, and Winters (2015); Gourlay, Kilic, and Lobell (2017); Kilic et al. (2017); Bevis and Barrett (2017)). And in development economics, researchers have assessed whether data quality issues can account for large gaps in measured labor productivity between agriculture and non-agriculture in developing countries (Gollin, Lagakos, and Waugh (2014); Mccullough (2017)).

We offer novel empirical approaches for quantifying the effect of measurement error on estimates of productivity dispersion and persistence. We consider measurement error from two sources: misspecification of the production function, and measurement error in inputs and outputs. For the former, we adopt a machine learning approach that can flexibly account for any non-linearities or interactions in the production function that would not be picked up by a standard log-linear specification. Perhaps surprisingly, we find that simple log-linear production functions describe the data about as well as much more flexible machine learning approaches. At least in our setting, the Cobb-Douglas function predicts output well.

To quantify the role of measurement error in inputs and outputs, the core idea of our empirical approach is that multiple measures of inputs and output – all measured with noise for different reasons – can be used to construct bounds on the true variance in productivity across farms, and purge measurement error from estimates of persistence. In practice, we use estimates from surveys, satellites, and crop cutting (physical harvest measurement) when applying this technique to our data. Conservatively, our estimates indicate that measurement error in output accounts for 37-56% of the observed dispersion in productivity from surveys, and that measurement-error-corrected estimates of productivity dispersion are on par with benchmark estimates from Hsieh and Klenow (2009) for non-agricultural firms. Our results suggest that measurement error in output plays a large role relative to other proposed sources of heterogeneity such as managerial ability or unobserved temporal shocks. As our method only accounts for mismeasurement in some (not all) components of the production function, these estimates might even be a lower bound on the overall role of measurement error in cross-sectional estimates of productivity dispersion.

Correcting for measurement error in output also substantially increases our estimates of the persistence of productivity differences, relative to naive estimates. Although estimates vary to some degree by the approach used, our measurement-error-corrected persistence estimates are on par with benchmark estimates from manufacturing firms in developed countries (Foster, Haltiwanger, and Syverson, 2008).

Taken together, our findings have important implications for the understanding of productivity dynamics for agricultural firms in developing countries. They suggest that policies and economic theories premised upon observed patterns in conventional survey-based measures of productivity may be misguided, and highlight the importance of implementing at-scale multiple approaches to measuring output in agricultural surveys. Because, after accounting for measurement error, the productivity of the smallholder farmers in our data exhibits patterns consistent with the stylized facts of the firm productivity literature, our work also

suggests that economic mechanisms underlying firm productivity dynamics explored in previous studies, and their relationship to broader macroeconomic outcomes, may also apply in the developing-country agricultural context.

The rest of the paper is organized as follows. In Section 2, we provide an overview of our data and describe our empirical strategy for measuring productivity. Section 3 provides our initial estimates of the dispersion and persistence in productivity across agricultural firms in our survey data, with productivity appearing more dispersed and less persistent than is suggested by benchmark estimates in the non-agricultural firm productivity literature. To understand these results, in this section we also decompose the sources of productivity dispersion by implementing a range of different fixed-effects specifications, and provide evidence that unobserved factors which vary at small spatial scales, of which measurement error is one, are important determinants of productivity dispersion. In Section 4 we introduce our framework for quantifying the effect of measurement error and explain how it is built upon and relates to other previous studies. Section 5 presents our main results on the effect of measurement error and misspecification on our estimates of dispersion and persistence. In Section 6 we summarize our results and discuss their policy implications.

2   Data and Empirical Framework

2.1   Data

Our data come from the World Bank's Living Standards Measurement Study Integrated Surveys on Agriculture (LSMS-ISA), a set of nationally-representative household-level panel studies from multiple African countries. In particular, we use data from the LSMS-ISA surveys conducted in Tanzania, Uganda, Nigeria, and Ethiopia. Summaries of each of these country datasets are provided in Table 1. We define an 'agricultural firm' as a household that harvested crops from any land owned or rented/sharecropped by the household. Under this definition, we observe more than 12,000 unique agricultural firms in multiple survey waves and growing seasons over a period from 2008 to 2015. Farms in our dataset are small (median of 0.81 hectares[3]), rain-fed (<3% of plots are irrigated), family-owned (>80% of plots owned by household members), and located in rural areas (>85%).[4] Expenditures on capital inputs such as fertilizer, pesticide, or farm equipment are quite low (<\$30 for the

---

[3]For reference, the average farm-size in the U.S.in 2017 was 180 hectares

[4]A plot is generally defined as contiguous pieces of land on which a specific crop or mixture of crops is grown, and on which a single set of farm management practices are implemented.

median household), and labor is used much more intensively than capital in production.[5] More than 40% of plots cultivated by firms in our dataset are intercropped, and nearly 80% of households cultivate one or more of following key staple crops: maize, beans, cassava, sorghum, rice, wheat, teff. The estimated value of agricultural products produced during a growing season by the average household is $405, and average total annual household consumption is about $1450.[6]

Several key features of these data facilitate the analyses we undertake in this study. First, the LSMS took great effort to track and re-interview all households in the original sample during subsequent survey waves, with attrition rates generally less than 5%. Consequently, as shown in Table 1, the vast majority farm households (>85% across countries) in our data are observed in at least two survey waves. This panel dimension enables us to estimate productivity for the same households at multiple points in time, and thus allows us to generate the first-ever estimates of the persistence of total factor productivity for developing-country agricultural firms.

Second, the surveys provide extremely detailed information about agricultural production at the plot level, and, as shown in Table 1, we typically see multiple plots cultivated by the same household in a given growing-season-year, often with different plots managed by different individual farmers.[7] The granularity of these data allow us to construct measures of inputs, output, and productivity at the both the household- and plot-level, and this in turn allows us to assess how much fixed characteristics of households or farmers contribute to measured productivity dispersion. Few, if any, other surveys that have been administered at the scale of the LSMS-ISA contain similarly detailed agricultural data.

A third important feature of our data is the availability of multiple measures of land area and crop yield for a subset of plots. These measures can be used to construct multiple measures of productivity for the same production unit (plot or household) and, as described below, we exploit these multiple measures to quantify the importance of measurement error

---

[5]The median household used ∼150 person days in production across all cultivated plots per growing season. Even using a conservative back of the envelope wage estimate of $1 USD per person day, the typical household in our dataset spends approximately five times more on labor than capital inputs. By contrast, according to 2017 data on farm expenditure from the U.S. National Agricultural Statistics Service, the average farm in the U.S. spent twice as much on fertilizer, agricultural chemicals, and farm equipment alone as they did on labor.

[6]The average annual consumption per adult equivalent in our dataset is 378 USD, or slightly more than $1 per day. Compare this to the estimate of the international poverty line in 2019 of $1.90 per day

[7]A farmer is generally defined as the individual household member who is the primary decision maker regarding management practices on a particular plot.

Table 1: World Bank LSMS-ISA Surveys

	Tanzania (TZNPS)	Uganda (UNPS)	Nigeria (GHS)	Ethiopia (ERSS)
Years	2008 - 2012	2009 - 2012	2012 - 2015	2011 - 2015
Survey Waves	3	3	2	3
Growing Seasons	2	2	1	1
Farm Households	3503	2430	2985	3091
% HHs Obs. in multiple waves	0.67	0.95	0.93	0.89
HH-season-years	8863	10860	5051	7765
Plots/HH-season	1.8	2.9	1.8	4.9
Farmers/HH-season	1.2	1.1	1.1	1.1
Median Plot Size (ha)	0.4	0.2	0.3	0.1
Plot-season-years	15814	31408	9338	36906

in observed patterns of productivity dispersion in our data. More specifically, for between 60 and 90 percent of plots in each country, we observe both a farmer-estimate of plot area, as well as the area measured by survey enumerators using a GPS device. In Ethiopia, enumerators also conducted a "crop-cutting" exercise on $\sim 30\%$ of plots across all three survey waves.[8] During a crop-cut, enumerators randomly select a 2x2 meter section within a plot, harvest all the crops in the selected area, then weigh, dry, and weigh again the collected harvest. In conjunction with yields computed based on farmer-estimates of plot area and harvest quantity, the crop cuts provide us with a second measure of crop yields for the subset of plots on which they were conducted.

Finally, because we find in the Ethiopian LSMS data that measurement error in output contributes substantially to measured productivity dispersion, we also exploit two other datasets that allow us to further examine the importance of measurement error in output. In particular, we use survey data on maize farmers from Kenya and Uganda in which plot-level measures of output collected on the ground were matched to independent satellite-based estimates (Burke and Lobell, 2017; Lobell et al., 2018). While these data do not contain as much detail on farm inputs as our main LSMS datasets, they do allow us to assess whether dispersion in standard measures of land productivity (i.e., yield) are as affected as other productivity measures (i.e., total factor productivity) by measurement error in output in an alternative and independent data setting.

---

[8]For more detail on the co-occurrence of GPS and farmer-estimated area measures, and crop-cut-based vs. farmer-estimated yield in Ethiopia, see Tables A6 and A7 in Appendix A.5.3.

2.2 Measuring Productivity

The primary measure of productivity analyzed in the firm productivity literature is total factor productivity (TFP). As discussed in Syverson (2011), the literature takes two common approaches to estimating TFP. In the first, the researcher assumes firms' production technologies can be described using a known functional form (ex. Cobb-Douglas) and that the share of total costs firms allocate to each different factor input represents its output elasticity. Under these assumptions, TFP can be inferred directly by inverting the production function.[9] Reliable estimates of cost shares in smallholder African agriculture are unavailable, making this approach hard to implement in our setting.

The second approach, and the one we undertake in this paper, is to estimate the production function using data, and measure productivity as the Solow residual. Specifically, we measure productivity as the residual from a regression of log output on log inputs with additional controls and fixed effects. This approach has a number of advantages in our setting. First, it allows us to measure productivity in a manner consistent with previous studies in the firm productivity literature. For example, most studies of manufacturing-firm productivity focus on dispersion across firms within narrowly defined industries (ex. as defined by four-digit SIC product codes). We too focus on intra-sectoral variation in productivity by including crop-system fixed effects, thus isolating productivity comparisons to farms that produce similar goods. Second, our approach makes it straightforward to account for the fact that agricultural production, as compared to manufacturing, is likely more dependent on local environmental factors such as climate, weather, and soil characteristics. By including controls for these environmental factors and village-year fixed effects, we isolate productivity variation across farms that is not driven by these agronomic factors. Finally, this approach to measuring productivity also allows for a straightforward decomposition of the sources of productivity dispersion by altering the granularity of the fixed effects. This type of decomposition has been identified as an important part of the agenda for emerging research on firm productivity (Syverson, 2011), and provides another direct way to compare our results to previous research.

---

[9]Some additional assumptions, including whether or not firms' production technologies exhibit constant returns to scale, are also required. For applications of this approach see, for example, Foster, Haltiwanger, and Syverson (2008) or Midrigan and Xu (2014)

Our main specification is,

$$
\begin{aligned}
output_{imhvt} = {} & \beta_1 land_{it} + \beta_2 labor_{it} + \beta_3 capital_{it} \quad\quad\quad\quad\quad\quad (1)\\
& + \sum_k \beta_4^k geovar_{ht}^k + \textit{crop-system}_m + \textit{village-year}_{vt} + \epsilon_{imhvt},
\end{aligned}
$$

where i indicates the plot, t indicates the growing-season-year, h refers to the household that cultivates plot i, v indicates the village (cluster) in which household h is located, and m indicates the mixture of crops grown on plot i. The terms $\textit{crop-system}_m$ and $\textit{village-year}_{vt}$ indicate fixed-effects for the most commonly observed crop-mixtures in each country[10], and for each village-year, respectively. The term $geovar_{ht}^k$ denotes the household-level climate, weather, soil, and land quality variables.[11]

The remaining terms correspond to our plot-level measures of agricultural output and inputs, which are measured in logs.[12] Labor is measured as the total number of person-days spent on pre-harvest activities by either hired laborers or own household members.[13] Capital inputs are measured in value terms, as the sum of expenditures on variable inputs (seeds, fertilizer, pesticides and herbicides) and the farmer-reported value of the stock of owned and rented durable capital (tools, machinery, and structures).[14] Our primary measure of agricultural output is computed as the sum, across crops cultivated on each plot, of the product of the harvested quantity and a fixed national-level median crop price. To generate estimates of physical productivity we also measure output as the quantity (in kgs) of crops harvested for

---

[10]There are between 50 and 100 unique crops recorded in each country dataset, and thousands of plot-specific intercrop combinations. Rather than specifying a fixed effect for each of these, we classify each crop into one of 9 higher-order groups (cereals, vegetables and melons, fruits and nuts, oilseed crops, roots and tubers, beverage and spice crops, legumes, sugar crops, other crops) using the FAO's Indicative Crop Classification (ICC) system, maintaining separate groups for the 7 most common staple crops observed in the data (See Appendix A.2 for more on the definition of key crops). Crop mixtures are then defined based on the combination of crop-groups observed on each plot. This classification procedure results in 50-100 different crop-mixture classifications in each country dataset. By comparison, there are around 200 different four-digit product-category codes for the manufacturing sector in the International Standard Industrial Classification (ISIC) of all economic activities.

[11]Appendix A.3 provides a detailed summary of the geographic variables included in our main specification.

[12]See Appendix A.1 for summary statistics of these measures by country at the plot- and household-level

[13]For Uganda, labor inputs are not disaggregated by activity type, so person-days totals include time spent harvesting crops.

[14]In cases where farmers did not report expenditures, such as for own-produced organic fertilizer (e.g. animal manure) we use fixed national-level median prices to value capital inputs. Additionally, the value of tools, equipment and machinery is reported at the household level. To construct a plot-level measure, we attribute to each plot a share of household-level durable capital proportional to the plot's share of total household area.

a set of common staple crops in each country.[15] Our analysis necessarily requires a number of measurement choices. Appendix A.4 shows that alternative measurement choices lead to similar results.

We refer to our primary measure of productivity as revenue-based total factor productivity (TFPR). Unlike the firm productivity literature, the majority of agricultural goods produced by smallholder farms are homogeneous and undifferentiated, so differences in product quality are likely small. In addition, smallholder farmers take prices as given and a large share of output is allocated to home consumption. The re-weighting of crop outputs by national prices serves only to fix units when summing output of different crops. When focusing on the output of a single crop, we denote productivity as quantity TFP, or TFPQ. In some results we also consider crop yield, which we compute by residualizing crop-specific yields (harvested quantities over planted area) on observed geographic controls and the specified fixed effects. Across disciplines and fields of study, yields are the most commonly used metric of agricultural productivity, and can be computed in many settings which lack the necessary data to compute TFP. We include residual yield as a measure of land productivity for comparability with this broader academic context.

We estimate Equation (1) (separately for each country) using OLS. Log productivity is then estimated as the Solow residual, $\hat{\epsilon}_{imhvt}$. We acknowledge that this approach to estimation contrasts with other non-experimental approaches to estimating production functions with observational data. These include control-function approaches similar to those developed in Olley and Pakes (1996), Levinsohn and Petrin (2013) and Ackerberg, Caves, and Frazer (2015), as well as dynamic-panel techniques along the lines of Arellano and Bond (1991) and Blundell and Bond (2000). The primary advantage of these methods is that they provide a means of obtaining accurate estimates of the factor elasticities in the production function. However, neither are particularly well-suited to our empirical setting. In particular, control-function approaches leverage assumptions about the timing and relationship between productivity shocks and other types of production decisions, such as investment or purchase of intermediate inputs. In our data, intermediate inputs are not well observed, and the required assumptions about the timing of productivity shocks relative to input use decisions tend to be strong and untestable. With dynamic panel approaches, commonly used instruments include lags or (lagged) differences of output or input prices. Our panels are relatively

---

[15]Key crops are described in Appendix A.2. Valuing harvest quantity, versus relying on reported crop sales, accounts for the value of auto-consumed harvest. Aggregating harvest value across crops provides a comprehensive way to measure output produced by intercropped production systems. When measuring output as harvest value, we measure the land input as farmer-reported total plot area. For physical productivity, we use crop-specific planted area

short (2-3 periods) limiting the number of potential instruments of this type. Since our interest lies in analyzing patterns in the residuals, rather estimating factor elasticities, we instead just use OLS.

3   Size and Sources of Productivity Dispersion and Persistence

In this section, we implement the empirical framework described in Section 2 to generate baseline estimates of productivity at both the plot- and household-level. Using these estimates, we first quantify the magnitude of productivity dispersion across plots and the degree of persistence in household-level productivity. For both dispersion and persistence we benchmark our estimates against estimates from previous studies in manufacturing, finding greater dispersion and less persistence among firms in our dataset. To better understand the potential mechanisms underlying these patterns, we then decompose plot-level productivity dispersion into various components (ex. village, household, farmer, crop-system), finding substantial dispersion remaining at small spatial scales. This motivates our subsequent analysis of the effect of measurement error, which is introduced in the next section.

3.1   Plot-Level Productivity Dispersion

Figure 2 summarizes our estimates of the magnitude of plot-level dispersion in measured productivity. Each row of the figure pertains to one of our three key measures of productivity. The left column shows the kernel density of the distribution of log-productivity resulting from the estimation of Equation (1) for each country (colored lines). In each panel, these densities are plotted relative to an artificial distribution representing the TFPR of Indian manufacturing firms based on the values reported in Hsieh and Klenow (2009), hereafter referred to as HK (shaded grey area and bottom panel).[16] The right column of Figure 2 shows the estimated magnitude of productivity dispersion, as measured by the ratio of the 90th to the 10th percentile value of each distribution, across a range of model specifications. In this column, the solid points indicate the 90:10 ratio resulting from the estimation of

---

[16]In particular, we start with the 90:10 ratio of TFPR reported for Indian manufacturing firms in HK. The natural log of this ratio is equivalent to the difference between the 90th and 10th percentile values of the log TFPR distribution. Assuming this distribution is standard normal, $+/-$ half the log difference can be related to the z-score associated with these percentiles. From these z-scores, we estimate of the standard deviation, then generate the artificial distribution by taking draws from a standard normal distribution with this estimated variance. In general, we use HK as a benchmark for dispersion because it is one of the best-cited papers on productivity dispersion, is one of few studies that provide estimates of dispersion in developing countries, and because HK's estimate for dispersion among Indian manufacturing firms falls very close to the median of all the dispersion estimates we collected and presented in Figure 1.

productivity using our baseline specification of the production function (Equation (1)). This specification includes village-year fixed effects as well as crop-system fixed effects, such that we are comparing two farmers in the same village growing the same crop or set of crops in the same year. Lighter-colored points indicate alternative specifications containing fewer fixed effects and controls.[17] The estimates furthest to the right correspond to the 90:10 ratio when productivity is estimated without any controls or fixed effects.[18] The bottom-right panel contains a box-plot describing the (inter-quartile) range of estimates of dispersion in TFPR in the non-agricultural sector from across the published studies displayed in Figure 1. The dashed magenta lines in each panel are located at a value of 5.0, and indicate the magnitude of dispersion in TFPR observed among Indian manufacturing firms reported in HK.

Across countries and measures of productivity, the values shown in Figure 2 indicate that dispersion in productivity among developing-country agricultural firms is large, and substantially larger than benchmark estimates of dispersion in the non-agricultural sector. The dispersion estimates shown in the figure are relatively consistent across countries and measures of productivity The cross-country average 90:10 ratios of our baseline measures of TFPR, TFPQ, and residual yield are 8.86, 7.21, and 9.34, respectively.[19] These values are significantly larger than HK's estimate for Indian manufacturing firm TFPR and, except for TFPQ, exceed the 75th percentile value (8.15) of the box-plot in Figure 2. The large dispersion we observe across firms in our data is consistent with the findings of other recent studies which quantify dispersion among smallholder farmers using LSMS-ISA data.[20]

---

[17]Left-to-right, the controls and fixed effects included in the specifications represented by each row of dots in Figure 2 are: (i) village-year, crop-system, geovariables, (ii) village, year, crop-system, geovariables, (iii) level-2 administrative jurisdiction by year, crop-system, geovariables, (iv) level-2 administrative jurisdiction, year, crop-system, geovariables, (v) crop-system, year, geovariables, (vi) crop-system, year, (vii) crop-system, (viii) none.

[18]For TFPR and TFPQ, the interpretation of the dispersion estimate derived from this model is the variation in output (either harvest value or quantity) not attributable to variation in land, labor or capital input use. For yields, the dispersion estimate represents the magnitude of variation in crop-yields. For both TFPQ and yields, this specification does include a crop fixed-effect to facilitate pooling across key crops, and so the estimated residual dispersion is within each key crop.

[19]By country our estimates for TFPR are: 7.16, 8.71, 9.64, and 9.93, for Tanzania, Uganda, Nigeria, and Ethiopia, respectively. Our estimates for TFPQ, in the same order, are: 6.12, 8.16, 7.40, and 7.17. And for residual yield: 7.34, 10.63, 10.09, and 9.03

[20]Using an alternative empirical approach, Restuccia and Santaeulalia-Llopis (2015) estimate the 90:10 ratio of TFPQ to be 10.8 for Malawian smallholder farmers. The 90:10 ratios of plot-level yields reported in Gollin and Udry (2019) are 17.99 and 27.32 for Tanzania and Uganda, respectively
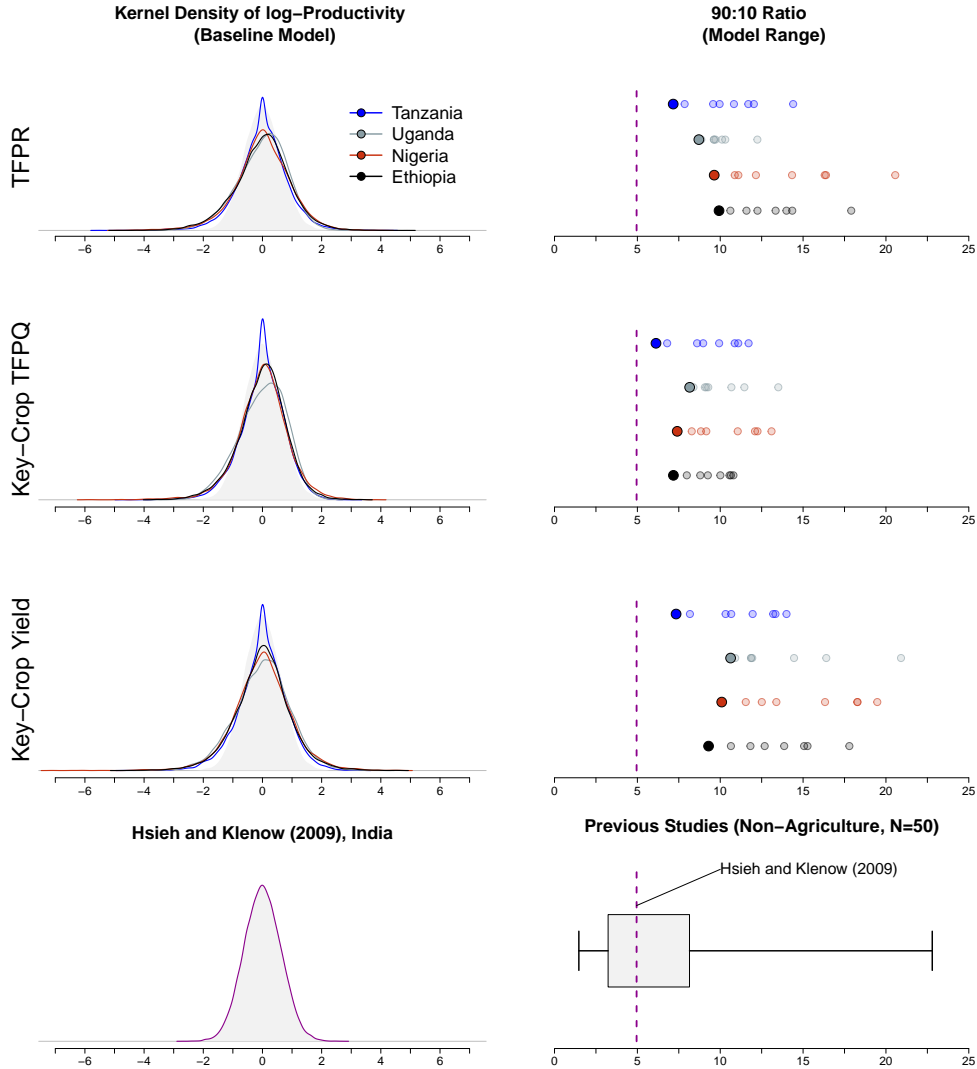
Figure 2: Dispersion in productivity among developing-country agricultural firms is larger than previous estimates from the non-agricultural sector. Colored lines in the left column show the kernel density of log productivity estimated using our baseline specification for each of our sample countries and productivity measures, relative to an artificial distribution representing the TFPR of Indian manufacturing firms (shaded grey areas) based on the dispersion in TFPR among these firms reported in Hsieh and Klenow (2009) (magenta dashed lines). Solid colored points in the right column show the 90:10 ratios associated with these baseline distributions. Lighter-colored points indicate the 90:10 ratios associated with alternative production function specifications that include less restrictive fixed effects. The box-plot in the bottom-right panel shows the (inter-quartile) range of estimates of dispersion in TFPR among non-agricultural firms from more than 30 published firm-productivity studies (50 country-year dispersion estimates in total).

3.2 Persistence in Household-Level Productivity

The above analysis suggests that some agricultural firms are dramatically more productive than others, with firms at the 90th percentile at least seven times more productive than firms at the 10th percentile. Do these differences in productivity persist over time? Answering this

question is key for understanding the underlying sources of productivity dispersion and how these patterns map to broader economy-wide productivity dynamics. To our knowledge, no estimates exist on the persistence of productivity differences over time among small developing-country agricultural firms.

To quantify whether large productivity differences persist over time, we focus only on our baseline measures of productivity, i.e. those resulting from the estimation of (variants of) Equation (1). Figure 3 summarizes our results. Again, each row corresponds to one of our three measures of productivity. In the left column (panels [a]-[c]) we show the linear auto-correlation of productivity among households observed in multiple survey waves, estimated from a pooled OLS regression. Because the number of survey waves and the time intervals between survey waves vary across countries in our sample, we estimate the autocorrelation between contemporaneous and lagged productivity for lags of one to four years. The shape and color of the plotted points indicate the country, and the error bars represent 95% confidence intervals. A linear fit to the point estimates is plotted as a dashed colored line in each panel, and the shaded colored regions represent the 95% confidence intervals on the predictions of this linear model.

To make our results comparable to estimates reported in the previous literature, in the right column of Figure 3 (panels [d]-[f]) we plot the annualized autocorrelation of productivity implied by each linear point estimate, where we annualize our linear estimates by raising them to a power of one over the length of the lag in years. The annualized values thus represent the year-on-year autocorrelation which, if applied over the number of years specified by each lag, would generate the linear estimates we observe in panels [a]-[c]. As a benchmark, in each of the right-column panels we also plot, in magenta, a central estimate of the annual autocorrelation (0.75) of productivity from Syverson's 2011 review of firm-productivity studies. Finally, colored dashed lines in panels [d]-[f] show the simple average of the depicted estimates.

Across measures of productivity, the values shown in Figure 3 indicate that the persistence of measured productivity among agricultural firms is low relative to estimates from non-agricultural firms. The cross-country, cross-lag average annualized autocorrelations for TFPR, TFPQ, and residual yield are 0.40, 0.44, and 0.37, respectively.[21] These values are nearly half the central estimates from the existing firm-productivity literature. For example,

---

[21]By country the cross-lag average annualized persistence of TFPR is: 0.49, 0.28, 0.38, and 0.41 for Tanzania, Uganda, Nigeria, and Ethiopia, respectively. Our estimates for TFPQ, in the same order, are: 0.57, 0.31, 0.34, and 0.44. And for residual yield: 0.53, 0.24, 0.27, and 0.33
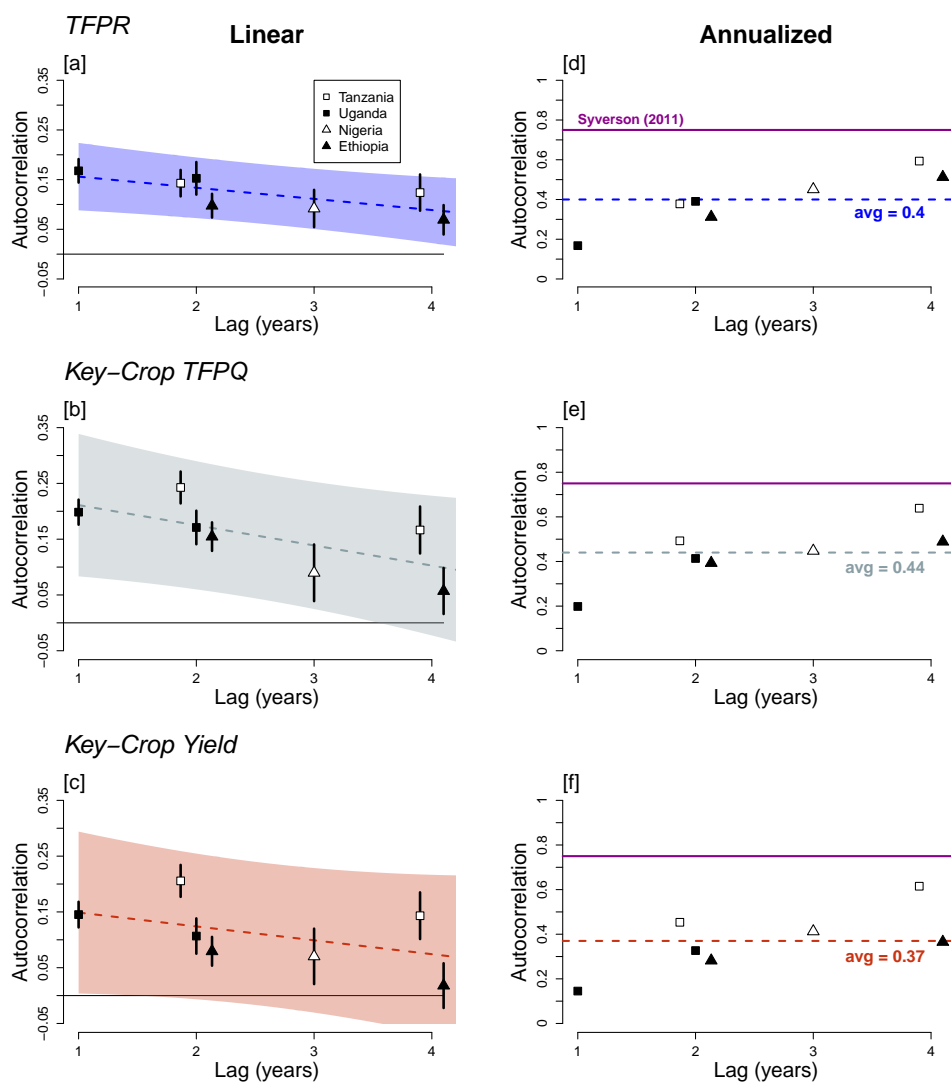
Figure 3: **The persistence of measured productivity among developing-country agricultural firms is low relative to previous estimates from non-agricultural firms.** In panels [a]-[c] plotted points represent estimates of the linear autocorrelation between contemporaneous household-level productivity and lagged productivity, across lags ranging from one to four years. Points' shape and color indicate the country, and the error bars represent 95% confidence intervals. A linear fit to the point estimates is plotted as a dashed colored line in each panel, and the shaded colored regions represent the 95% confidence intervals on the predictions of this linear model. Panels [d]-[f] show the year-on-year autocorrelation which, when applied over the number of years specified by each lag, generate the corresponding linear estimates in panels [a]-[c]. Solid magenta lines indicate a central estimate of the persistence of non-agricultural firms (0.75) from Syverson (2011), and colored dashed lines denote the average of the estimates depicted in each panel.

Foster, Haltiwanger, and Syverson (2008) find autocorrelation coefficients of approximately 0.75-0.8 among U.S. manufacturing firms. Consistent with our dispersion results, persistence is highest for TFPQ, and lowest for yields. However, we do see more heterogeneity across

countries. Across productivity measures and lags, the average annualized persistence by country is 0.53 for Tanzania, 0.28 for Uganda, 0.33 for Nigeria, and 0.39 for Ethiopia. The range of these estimates, expressed as a fraction of their mean, is 0.65, approximately twice the level of cross-country variation we observed in our measures of productivity dispersion.

For all productivity measures, as expected, the linear autocorrelations decay as the length of the lag increases. However, there is an increasing linear trend in our annualized autocorrelation estimates. This may be due to the presence of unobserved time-invariant determinants of productivity that vary at spatial scales below the village-level. Such factors would not be captured by the controls or fixed effects in Equation (1), and thus end up in our residual measure of productivity. Because they are time-invariant, these factors would be equally correlated across multiple observations of the same household, regardless of the number of years separating those observations. This time-invariant component of the measured linear correlation would then be up-weighted when we annualize our linear estimates. We present further evidence suggestive of small-scale unobserved heterogeneity in the next subsection.

3.3   Sources of Productivity Dispersion

Our results thus far indicate that there is more dispersion in productivity across plots cultivated by developing-country agricultural firms than among the non-agricultural firms typically studied in the firm productivity literature, and that productivity for smallholder farms is relatively less persistent. To better understand what mechanisms might explain these patterns, and to motivate our subsequent analysis of measurement error, we conduct a simple decomposition of the sources of productivity differences among farms. The spirit of this exercise is not to pinpoint specific explanations, as exploring all potential sources of productivity dispersion in smallholder agriculture in detail is beyond the scope of this study.[22] Rather, the decomposition we conduct serves to classify the spatial, temporal, and organizational scale at which important determinants of productivity vary.

Our decomposition proceeds as follows. First, we quantify dispersion when productivity is estimated using a model of the production function with minimal fixed effects or controls. We refer to the 90:10 ratio associated with these minimally-saturated models as the total variation or total dispersion in productivity. More specifically, for TFPR, total variation

---

[22]We do investigate whether alternative measures of output and factor inputs alter our baseline estimates of dispersion in TFPR in Appendix A.4. Some of these alternative measurement scenarios, such as using richer and more disaggregated measures of labor and capital inputs, or using local rather than national crop prices to value output, shed light on the importance of some specific sources of dispersion (ex. variation in labor quality or producer-specific prices). A wide variety of other hypotheses could be explored using our data.

is estimated as the 90:10 ratio of the residual in a model without any fixed effects or controls, i.e. just (log) harvest value on (log) land, labor, and capital inputs. In this case, the interpretation of total productivity dispersion is the magnitude of variation in all-crop harvest value not attributable to variation in input use. For TFPQ and yields, since we pool observations across multiple key crops in the estimation, we include a crop fixed effect. So for TFPQ, total dispersion is the amount of variation in harvest quantity not attributable to input use or differences between key crops in the average quantity harvested. And for yields, total productivity variation is simply the dispersion of observed crop yields controlling for average differences between key crops.

We then quantify the 90:10 ratio of productivity associated with a series of specifications of the production function which include additional and increasing granular fixed effects and controls. In particular, we start with a model which only contains crop-system fixed effects, then add survey-year fixed effects, household-level geovariable controls, and village fixed effects. We then substitute village for village-year fixed effects and drop the (now redundant) survey-year dummies, thus reproducing our baseline specification (Equation (1)). To this model we then add household fixed effects, and finally replace the household- with farmer-level fixed effects. For each of these increasingly saturated models, we calculate the difference between total dispersion and the 90:10 ratio of productivity estimated using the specified model, express this difference as a percentage of total dispersion, and interpret this percentage as the share of total variation explained by all the fixed effects and controls included in the model. Similarly, for each specification in this sequence, we also calculate the reduction in the 90:10 ratio relative to the previous model. This reduction can be interpreted as the amount of additional productivity variation that can be attributed to the newly-added fixed effects or controls. As with our cumulative measure, we express this additional variation explained as a percentage of total productivity dispersion.

Table 2 and Figure 4 summarize and contextualize our decomposition results. In the table, we report statistics that convey the magnitude of the dispersion we decompose, how much of it we are able to explain in aggregate, and the relative granularity of the different fixed-effects we include. In Figure 4, the left column (panels [a]-[c]) shows the 90:10 ratio of productivity associated with each of the model specifications described above – starting on the left with total productivity dispersion – for each of our sample countries and measures of productivity. In the right column, y-axis values indicate the share of total dispersion explained by all the fixed effects and controls in each model, and the grey percentage values at the top of each panel are the cross-country averages of the additional variation explained by moving from

18

Table 2: Dispersion Decomposition Factors

	Tanzania	Uganda	Nigeria	Ethiopia
Plot-level obs	15814	31408	9338	36906
Years	3	3	2	3
Crop-systems	67	86	68	54
Villages	381	576	343	589
Median HHs per village-year	8	9	10	10
Avg farmers per village-year	13.6	19.0	11.1	10.3
Total dispersion†	13.37	15.55	17.71	15.50
Max % Explained†	0.76	0.64	0.80	0.62
Max adj. R^2 †	0.71	0.53	0.70	0.63

Note: † indicates a cross-measure average

one model to the next.

Our first key observation is that, relative to developed-country non-agricultural firms, a larger share of productivity differences between firms in our data is not attributable to observable characteristics or fixed effects. As shown in Table 2, even with our most saturated regression model we can only explain between 62 and 80 percent of total productivity dispersion. Across countries and measures of productivity, the average percent of total variation explained by this model is 71%, and the average of the maximum adjusted R^2 obtained across models is 0.64. For our baseline model, the analogous values are 44% and 0.48, respectively. As a point of contrast, in their study of the firms in the Forbes 800, Bertrand and Schoar (2003) estimate a much sparser model of returns on a vector of time-varying firm characteristics plus firm and year fixed effects, and obtain an adjusted R^2 of 0.72.

Second, of the set of factors we consider explicitly, unobserved time-invariant characteristics of villages, households and individual farmers are the most important sources of productivity dispersion. On average across countries, unobserved time-invariant features of villages explain an additional 15% of total dispersion in TFPR, 18% of total dispersion in TFPQ and 19% of total dispersion in yields, even after controlling for differences driven by input use, fixed characteristics of different crop-systems, and observable agronomic conditions. Similarly, adding household fixed-effects to a model already containing geovariables plus crop-system and village-year fixed effects increases the percent of total variation explained by 14%, 16%, and 15% for TFPR, TFPQ, and yield, respectively. Farmer fixed effects explain an additional 11% for all three productivity measures. No other set of fixed effects or
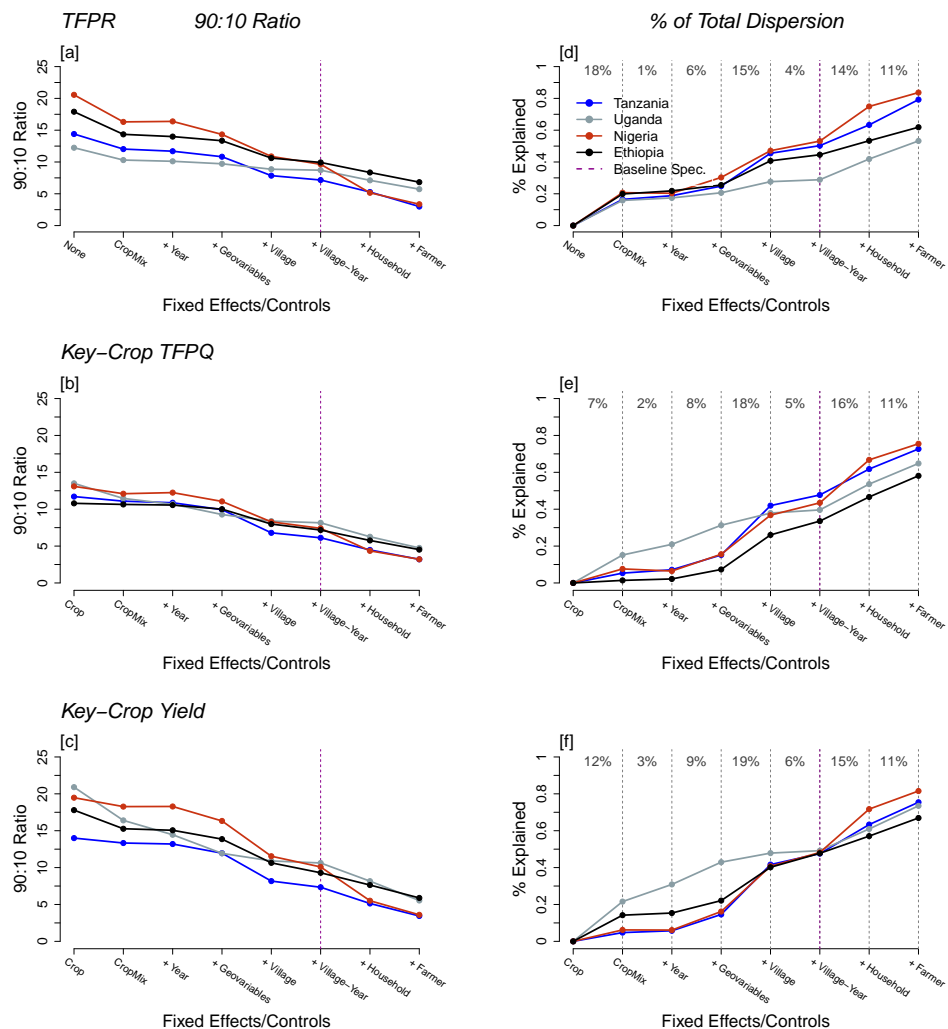
Figure 4: **Small-scale unobserved factors explain the largest shares of measured productivity dispersion, though a substantial portion of dispersion remains unexplained.** Panels [a]-[c] report the 90:10 ratio, by country, of each of our three key measures of productivity when estimated using a series of specifications of the production function which include additional and increasing granular fixed effects and controls. Points on the far left of each panel represent 'total productivity dispersion' and were estimated from a model including minimal fixed effects (either none for TFPR or crop-fixed effects for TFPQ and yield). Left-to-right, subsequent specifications add (i) crop-system and (ii) year fixed effects, (iii) geovariable controls, (iv) village, (v) village-year, (vi) household and (vii) farmer fixed effects. Panels [d]-[f] plot the share of total productivity dispersion explained by each specification. Grey percentage values at the top of each panel are the cross-country averages of the additional variation explained by moving from one model to the next. In all panels our baseline specification is highlighted with a magenta dashed line.

controls explains more than 10% of total variation on average.[23]

---

[23]Crop-system fixed effects do explain a large share of total productivity variation (18%) for TFPR, but considerably less for TFPQ (7%) and yields (12%). This is because for TFPQ and yields, we are only considering the subsample of plots where a key crop was cultivated. Thus, for these measures, crop-system

How do we interpret these decomposition results in light of our previous findings on dispersion and persistence? Our dispersion results indicate that differences in measured productivity across plots are large relative to dispersion in non-agricultural sector. The large share of total productivity dispersion attributable to village-, household-, and farmer-level fixed effects in our decomposition exercise suggest that important components of measured productivity are unobserved (or at least not captured by our observed measures of input use or agronomic conditions), and vary at small spatial and organizational scales (i.e. at or below the village-year level). Furthermore, the fact that we can only explain around 70 percent of total variation with the explanatory variables and fixed effects included in our most-saturated model indicates that factors which vary at even more granular spatial (ex. plot-level) and temporal (ex. intra-annual) scales may also play a significant role. On persistence, we find that households' measured productivity is only weakly correlated over time. Importantly, in our estimation of persistence, we measured productivity using our baseline model, which does not include household- or farmer-level fixed effects. Therefore, the productivity residuals we use to measure persistence retain the effects of any time-invariant characteristics of households or farmers. The observed lack of persistence, then, suggests such factors are less important drivers of changes in productivity over time than year-to-year fluctuations in other exogenous factors. Taken together, our results suggest that the most important sources of productivity dispersion are unobserved, vary over time and at small spatial scales (likely within plots and years), and are exogenous to time-invariant characteristics of farmers and households.

What do these features imply about the potential mechanisms underlying the productivity differences we observe among firms in our data? In general, residual measures of productivity are thought to include technology or management differences, differences in market power across firms, variation in external factors such as weather, and measurement errors. Given the small size of the firms we study, we can reasonably assume that variation in firms' market power minimally influence differences in measured productivity. Regarding differences in management practices, a number of studies of non-agricultural firms suggest firm performance is strongly determined by the management ability of individual managers or executives (ex. Bloom and Van Reenen (2007); Bushnell and Wolfram (2007); Bertrand and Schoar (2003)). Differences in managerial ability could explain the productivity dispersion

---

fixed effects only capture differences between the smaller set of production systems which include each key crop. The difference between the the amount of additional dispersion explained by crop-system fixed effects for TFPQ and yields is likely due to the fact that variation driven by capital and labor input end up in our residual yields measure. So differences in input use intensity across production systems will be captured by these fixed effects for yields, but not for TFPQ.

we see across — but not within — farmers. For instance, about half of the dispersion we estimate in Uganda and Ethiopia still remains when including fixed effects for individual farmers.[24]

Unmeasured small-scale time-varying heterogeneity in agronomic conditions would also conform with our descriptive results. Our baseline specification includes village-year fixed effects, which should account for most types of weather shocks.[25] But, highly local and plot-specific shocks, such as pest attacks or disease outbreaks could still explain part of the dispersion we estimate.

Measurement errors in inputs or outputs could also produce differences in measured productivity that exhibit patterns consistent with our descriptive results. Intuitively, adding random noise to measures of the components of the production function will increase the variance of the estimated residual. So if inputs and output are measured with more noise in our data than in the datasets used to study non-agricultural firms, this could explain why we observe greater dispersion in productivity. Additionally, measurement error occurs at a small spatial scale, as our measures of inputs and outputs are either plot- or household-specific. If measurement error is large, this could explain why the village, household, or farmer fixed effects in our decomposition exercise appear as the most important sources of cross-sectional productivity dispersion.[26] Finally, measurement error is time-varying and (if it is also idiosyncratic) uncorrelated with other time-invariant determinants of productivity. This could explain why we see low levels of persistence despite the large proportion of cross-sectional dispersion attributable to fixed characteristics of villages, households, and farmers.

Given that measurement error potentially rationalizes our initial descriptive findings and

---

[24]Farmer fixed effects limit our comparison to plots cultivated *by the same person*. It is well known that plot productivity may vary across members within a household. For example, Udry (1996) documents significant differences in productivity between plots managed by male versus female members of the same household.

[25]A large number of studies have documented a highly non-linear relationship between crop yields and temperature where below a threshold of about $34°C$ higher temperatures positively affect plant growth, but exposures to extreme temperatures above this threshold are damaging (ex. Schlenker and Roberts (2009); Schlenker and Lobell (2010)). These studies also indicate that the timing of exposure to extreme heat during the plant development cycle moderates the magnitude of the negative effect (Hatfield and Prueger (2015)).

[26]Note this is true for panel dimension as well. For example, imagine we observe a household in multiple survey waves that cultivated only a single plot. In theory, if measurement error is mean zero, then with enough observations the average effect of measurement error will go to zero, and the fixed-effect estimated for the household will not reflect the effect of measurement error. In practice, our panels are short (2-3 waves per survey). So if the variance of measurement error is large, it is very unlikely that the average effect of measurement error approximates zero. In this case, the household fixed effect will still capture the dispersion induced by measurement error.

decomposition results, in the next two sections we turn to investigating the significance of measurement error in our data explicitly. First, in the next section, we propose a framework for quantifying the effect of measurement error on measured productivity dispersion and persistence, and describe how we apply this framework to our data. Then, in Section 5, we present and discuss our measurement error results.

4   Quantifying the Effect of Measurement Error

Productivity is notoriously hard to measure. Wide dispersion and low persistence could reflect true differences in farmers' productivity, the stochastic nature of agricultural production, or it could simply reflect measurement error. A bevy of recent papers in agricultural economics on the farm size-productivity relationships highlight the potentially large role of measurement error in the agricultural context.[27]

In the firm productivity literature, measurement error has received somewhat less attention, though the topic has recently begun to attract more interest in the wake of Hsieh and Klenow (2009) and related studies documenting high levels of 'resource misallocation'. For instance, Rotemberg and White (2017) replicate Hsieh and Klenow's analysis using a modified version of the Indian Census of Manufactures dataset, which was cleaned in accordance with procedures typically implemented by the U.S. Census Bureau, and find that significant differences in the amount of dispersion among Indian versus U.S. manufacturers disappear. Relatedly, Bils, Klenow, and Ruane (2017) use panel data to show that approximately half of the dispersion in measured TFPR among Indian manufacturing firms is attributable to measurement error. In work closely related to ours, Gollin and Udry (2019) examine the impact of measurement error on productivity dispersion and the extent of resource misallocation among smallholder farms. They use an IV strategy to recover unbiased estimates of key factor elasticities, and show that accounting for measurement error and late-season productivity shocks reduces estimates of dispersion and misallocation substantially.[28]

Our paper builds on this work in several ways. First, we make use of the fact that agricultural surveys measure outputs and land inputs in different ways. Objective measures of plot area from GPS devices allow us to calculate measures of TFP dispersion that are not plagued

---

[27]These include studies on the role of measurement error in both farm size(Abay, Bevis, and Barrett, 2019; Carletto, Gourlay, and Winters, 2015; Carletto, Savastano, and Zezza, 2013; Cohen, 2019; Holden and Fisher, 2013; Kilic et al., 2017) and/or output (Barrett et al., 2017; Desiere and Jolliffe, 2018; Gourlay, Kilic, and Lobell, 2017; Lobell et al., 2018).

[28]Specifically, in Gollin and Udry (2019) measurement error in inputs, outputs, and the effect of late-season production shocks, account for between 66 and 90 percent of the observed variation in log productivity residuals.

by measurement error in survey-based measures of land area. For output, we show that multiple noisy measures of crop production for the same field can be used to derive bounds on the dispersion of real underlying productivity. In developing this 'multiple-measures' approach, we make progress towards understanding how much measurement error affects productivity estimates derived from survey-based measures that have conventionally been relied upon in the literature. Second, we study not only the potential error arising from mismeasurement of inputs and outputs, but also potential error arising from misspecification of the production function. To our knowledge, the literature has yet to recognize this latter source of measurement error. Finally, we also provide estimates of the effect of measurement error on the measured persistence of productivity, where the recent literature has focused exclusively on dispersion.

4.1 Bounding True Dispersion when Inputs or Outputs are Mismeasured

Our approach to understanding the influence of mismeasurement of inputs or outputs on estimated productivity dispersion is to use multiple measures of productivity to put bounds on the true variance of productivity across farmers. To do so, we make use of farmer-reported estimates of crop production and plot area, estimates of crop yields based on a procedure called "crop-cuting" where survey teams participate in partial plot harvests with farmers, estimates of maize yield from satellites (Burke and Lobell, 2017; Lobell et al., 2018), and measures of plot area recorded by survey enumerators using a handheld GPS device. These measures are described in more detail in Appendix A.5.

To see how we use these multiple measures to analyze the effect of measurement error on measured productivity dispersion, consider a plot i with "true" (unobserved) productivity ω_i, and define σ_ω^2 as the variance of ω across plots. This is the relevant parameter of interest because it represents the dispersion in true productivity — rather than any noise due to mismeasurement. Next, let ω_i^a and ω_i^b denote two noisy measures of productivity. For example, ω_i^a could represent the productivity estimate obtained from farmer-reported yield, and ω_i^b could be the estimate obtained from a crop cut. Formally, define these two measures of productivity as:

$$\omega_i^a = \omega_i + \epsilon_i^a \tag{2}$$
$$\omega_i^b = \omega_i + \epsilon_i^b,$$

24

where ϵ_i^a and ϵ_i^b are random sources of measurement error — both with a variance of σ_ϵ^2.[29] We assume that ϵ_i^a and ϵ_i^b are the only sources of measurement error, and that they are independent. In our example, this translates into an assumption that the errors from measuring yield with crop cutting (or satellites) are independent from those with farmer self-reported yield, and that there is no measurement error in other components of the production function (e.g. plot area). This assumption seems reasonable since errors from surveys are distinct from the types of errors that take place when the actual yield from a subset of the field is extrapolated to the entire field.

Next, we construct two composite measures of productivity using ω_i^a and ω_b^i – a "projected" and an "average" composite. The variances of these two measures put the bounds on the true variance, σ_ω^2. For the projected composite, first let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the estimated coefficients from the regression of ω^a on ω^b plus a constant.[30] Then define the projected composite as the predicted value from that regression:

$$\omega_i^{proj} = \hat{\beta}_0 + \hat{\beta}_1 \omega_i^b. \tag{3}$$

Second, let the average composite be defined as the simple average of the two productivity measures:

$$\omega_i^{avg} = \frac{\omega_i^a + \omega_i^b}{2}. \tag{4}$$

The variance of ω_i^{proj} establishes a lower bound on the true variance σ_ω^2. To see this, taking the variance of Equation (3) delivers

$$Var(\omega^{proj}) = \sigma_\omega^2 \left(\frac{\sigma_\omega^2}{\sigma_\omega^2 + \sigma_\epsilon^2} \right). \tag{5}$$

Equation (5) has intuitive properties. The term in the parentheses causes the variance of ω_i^{proj} to be biased downward. This bias becomes more severe as the observed measures of productivity get noisier, i.e. as σ_ϵ^2 increases. Conversely, the variance of the predicted value converges to the true variance as σ_ϵ^2 decreases. Equation (5) also shows that the gap between the true variance and the variance in observed productivity shrinks when ω_i^a and ω_i^b are more strongly correlated and ω_i^b is less noisy. Intuitively, if the second measure of productivity is strongly correlated with the first, but less variable, then more of the variation in the first

---

[29]An alternative form would allow for a different variance of the measurement error across the two productivity measures. Assuming a common variance has no meaningful effect on the bounds we calculate.

[30]That is, $\hat{\beta}_1 = \left(\omega^{b\prime} \omega^b \right)^{-1} \omega^{b\prime} \omega^a$, and $\hat{\beta}_0 = \bar{\omega}^a - \hat{\beta}_1 \bar{\omega}^b$, where $\bar{\omega}$ indicates a sample mean.

productivity measure can be attributed to actual dispersion rather than measurement error.

The upper bound on the true productivity variance comes from ω_i^{avg}. Returning to Equation (4), a straightforward derivation shows that $Var(\omega^{avg}) = \sigma_\omega^2 + \frac{\sigma_\epsilon^2}{2}$. The simple average of the two measures is "too noisy" because both measures are made up of the true variance and the random measurement error. As we would expect, the upper bound decreases when there is less random noise in the two measures of productivity. In combination, we can use these two measures to put bounds on the true variance in productivity. The approach does not require any distributional assumptions or assumptions about the determinants of measurement error. Instead, we need only for the measurement errors in the two measures to be uncorrelated.

4.2 Measurement Error and the Persistence of TFP

We next consider the question: how much does noise introduced by measurement error attenuate the measured persistence of productivity over time? Our method for answering this question again makes use of the fact that, for a subset of households, we observe the same set of alternative productivity measures in multiple survey waves. In particular, we use alternative measures of productivity across survey waves to estimate the persistence of true productivity (absent measurement error) using instrumental variables (IV), and compare these IV estimates to those obtained when using OLS regressions to estimate persistence, as is common in the firm productivity literature (Syverson, 2011). The difference between the two estimates thus provides a means of quantifying how much measured persistence is attenuated by measurement error.

For our IV regression, we estimate the autocorrelation of productivity across successive periods (generically, periods t and $t-1$) using one noisy productivity measure as an instrument for the other in the the earlier period $(t-1)$. In principle, using this IV approach will recover an unbiased estimate of the autocorrelation of true productivity, absent measurement error, across periods. The rationale is simple. Assuming, as above, that measurement errors in both noisy measures of productivity are random, then they are uncorrelated with themselves and each other both in cross-section and in time-series. In this setting, the separate measures of productivity – which are true productivity plus measurement error – satisfy the exclusion criteria for suitable instruments. First-stage predictions of one measure of productivity based on variation in the other in period $t-1$ will only reflect the common variation in true productivity, and so are "purged" of measurement error. Therefore, in the second-stage, the estimated correlation between first-stage predictions and noisily measured productivity in

period t will capture only the autocorrelation of true productivity.

Overall, this empirical approach follows Krueger and Lindahl (2001), who use multiple measures of education to correct for measurement error in cross-country estimates of the return to schooling. In applying this approach in our setting, we generate the first estimates of the effect of measurement error on the measured persistence of productivity for smallholder agricultural firms. Despite the importance of persistent productivity differences over time in the broader literature (Foster, Haltiwanger, and Syverson, 2008), the role of measurement error in these persistence estimates has yet to be considered.

4.3 Measurement Error from Production Function Misspecification

Analyses of productivity dispersion typically assume a Cobb-Douglas production function. In keeping with this approach, our main estimating equation (Equation (1)) is log-log in inputs and linear in household geovariable controls. However, if the relationship between log output and any of these factors is actually non-linear, then these non-linearities will end up in our productivity residuals. Interactions between components of the production function would operate similarly. There is substantial reason to think that agricultural production functions could have both important non-linearities and interactions: e.g returns to fertilizer are thought to be highly concave, and large amounts of fertilizer will have no (or limited) effect on output if the rains fail or if labor is unavailable at harvest.

To explore whether such non-linearities or interactions in the production function explain a significant portion of measured dispersion in TFPR, we compare dispersion under our baseline specification of the production function, to the dispersion of productivity residuals obtained using a machine learning algorithm to predict output based on inputs and geovariables. More specifically, we use random forests (Tin Kam, 1995; Amit and Geman, 1997) to predict log revenue using log inputs and geovariable controls after residualizing on crop-system and village-year fixed effects. See Appendix A.5.2 for an overview of the technical details of this prediction algorithm.[31]

The difference between observed log revenue (residualized on fixed effects) and the predictions of this residual generated by the random forest can be interpreted as an alternative measure of TFPR that accounts for non-linearities and interactions in the predictor variables (inputs

---

[31]Note, some of our geovariables are categorical, rather than continuous. These are not amenable to use as predictors in a random forest, and so are not included in the residual prediction stage.

and geovariables). Comparing the distribution of this alternative measure of TFPR to one we obtain using Equation (1) thus provides a means of assessing how much of the baseline dispersion in TFPR we measure is attributable to these non-linearities and interactions. To guard against overfitting, we also generate estimates of dispersion using cross-validated versions of both the log-linear and random forest models, where the models are trained on part of the sample and evaluated on held-out data.

5   Effect of Measurement Error on Measured Dispersion and Persistence

5.1   Impact of Production Function Misspecification

Turning now to results, we find little evidence that misspecification of the production function is an important source of measurement error in our estimates of cross-sectional productivity dispersion. Figure 5 displays the comparison between log-log and random forest estimates of dispersion for each of the countries in our sample. Each panel in the figure displays the (kernel density) distribution of two productivity residuals, one (in blue) generated using a Cobb-Douglas specification of the production function (Equation (1)) and the other generated using random forest predictions (in red).

The 90:10 ratios of the random-forest predictions are similar to those obtained using the Cobb-Douglas specification. Using a Cobb-Douglas production function, the 90:10 ratio of TFPR is 7.16 in Tanzania, 8.71 in Uganda, 9.93 in Nigeria, and 9.64 in Ethiopia. Using random forest to predict residualized output, the corresponding 90:10 ratios are 6.93, 8.14, 9.07, and 8.59. On average across countries, the 90:10 ratio of the random forest predictions are 7.3% lower than the associated predictions of the Cobb-Douglas model.[32] Comparing this result to the decomposition of productivity dispersion conducted in Section 3.3, the additional variation explained by accounting for nonlinearities and interactions between inputs and geovariables is comparable to the share of total variation explained by the linear effect of household geovariables on their own.

We also confirm using cross-validation that neither the linear nor random forest models are substantially overfitting the data on which they're trained (see Section A.5.2). Our overall interpretation from these results is that non-linearities and interactions between observed components of the production function do not explain a large share of variation in output.

---

[32]These percentages by country are 3.2, 6.5, 8.7, and 10.9% for Tanzania, Uganda, Nigeria, and Ethiopia, respectively
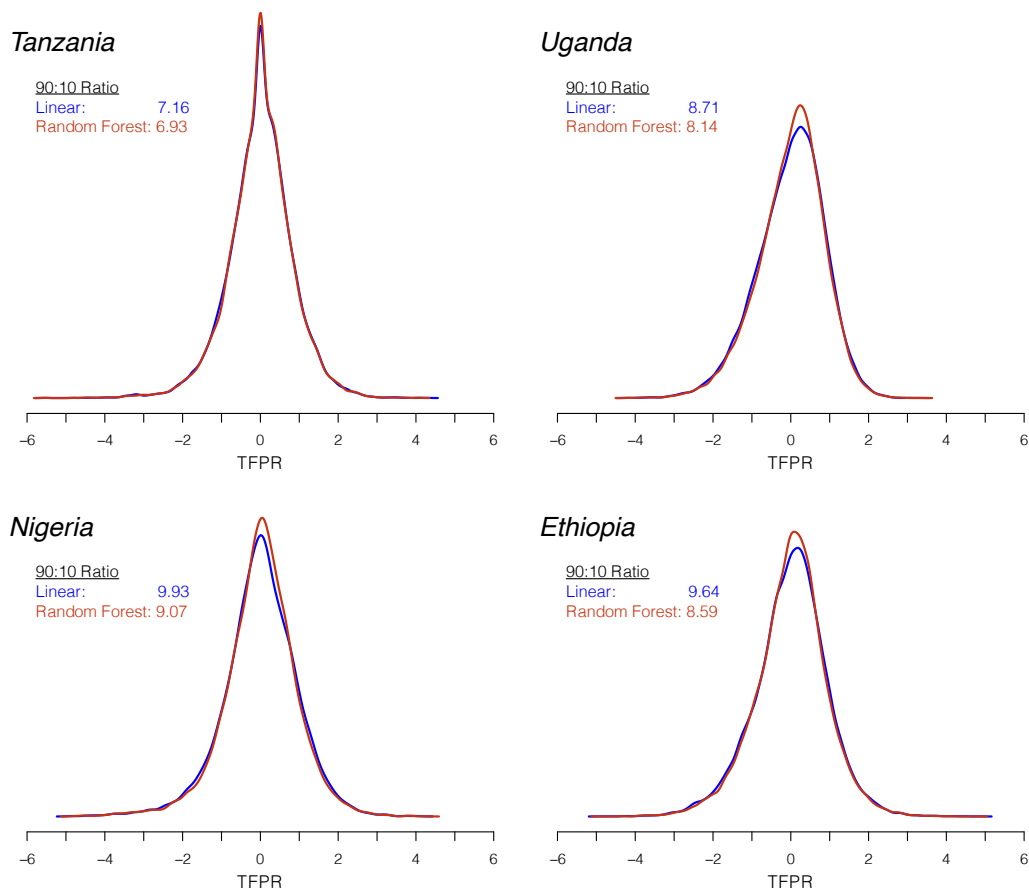
Figure 5: TFPR Estimated from Linear vs. Random Forest Models. In each panel, the kernel density of log TFPR estimated using our baseline linear specification (Equation (1)) is shown in blue, and the distribution resulting from random forest predictions is shown in red. 90:10 ratios associated with each set of productivity residuals are included as annotations.

Put simply, our results suggest that the production function for smallholder agricultural firms is essentially Cobb-Douglas.

5.2 Effect of Measurement Error in Inputs and Output on Productivity Dispersion

Figures 6 and 7 summarize our estimated bounds on productivity dispersion when accounting for measurement error in outputs. Figure 6 shows the bounds on productivity dispersion we derive using multiple measures of output from self reports and crop-cuts in Ethiopia, and Figure 7 shows bounds for yield (land productivity) dispersion using ground- and satellite-based maize yield estimates in our non-LSMS data, respectively. The base colors (grey and

red) in Figure 6 denote the two measures of physical productivity (TFPQ and residual yield) for which we estimate bounds. The shading of the rectangles indicates which measure of plot-size (farmer-reported versus GPS-measured) was used to generate the different output measures.[33] In Figure 7, the colors and x-axis labels denote country-years. In the bottom portion of each figure, we quantify the percentage of observed dispersion (based on farmer-estimated output/land) that is attributable to measurement error. More specifically, we compute the differences between the point estimate for 'farmer-estimated' dispersion (the ○'s) and the maximum (average composite) and minimum (projected composite) values of the bounded region, and express these as a percentage of the farmer-estimated magnitude of dispersion. The resulting tan bars in Figures 6 and 7 thus represent the percentage of dispersion in naively measured productivity (i.e. using farmer-estimated quantities without correcting for measurement error) that is attributable to measurement error. For consistency with our previous empirical results, the figures take the 90:10 ratio of the productivity distribution as the measure of dispersion. For consistency with the discussion in Section 4.1, we provide analogous results taking the variance of productivity to be the measure of dispersion in Figures A5-A8 in Appendix A.5.4.

In general, the bounds displayed in Figures 6 and 7 are very similar. In both cases, the percent of 'farmer-estimated' dispersion that is attributable to measurement error is large. Specifically, using the Ethiopia data, the bounds imply that at least 37.1% of baseline dispersion in TFPQ and 48.3% of baseline dispersion in yields can be attributed to measurement error in output, on average. Similarly, in our non-LSMS data, the estimated bounds imply at least 55.9% of variation in survey-based maize yields is due to measurement error.

The estimated upper bounds on productivity dispersion in both Figure 6 and Figure 7 lie below the benchmark level of dispersion observed in HK (90:10 ratio of 5.0). Using the Ethiopia crop-cut data, the mean 90:10 ratios in the bounded regions are 2.80 and 2.72 for TFPQ and residual yields, respectively. For the non-LSMS satellite-based bounds, the mean 90:10 ratio of yields in the bounded area is 2.31.[34] This suggests that, after accounting for

---

[33]Specifically, when evaluating dispersion in TFPQ, we require two measures of harvest quantities. One we observe directly in the form of farmer-estimates. For the second, we multiply crop-cut yields by the given plot area measure. When evaluating residual yields, we observe yields directly from the crop-cuts, and we compute farmer-estimated yields by dividing farmer-reported harvest quantity by the given area measure.

[34]The magnitude of these 90:10 ratios are similar to those reported by Gollin and Udry (2019). Specifically, after accounting for measurement error in factors of production and output and for late-season shocks, they find the 90:10 ratio of TFPQ drops from 14.8 to 4.4 in Tanzania and from 15.8 to 2.3 in Uganda. In Ethiopia, accounting for measurement error in output, we find that the 90:10 ratio of TFPQ drops from 7.17 (using our baseline model) to 2.8 (the mean of the bounded region), with the bounded region ranging from 1.44 to 4.18.
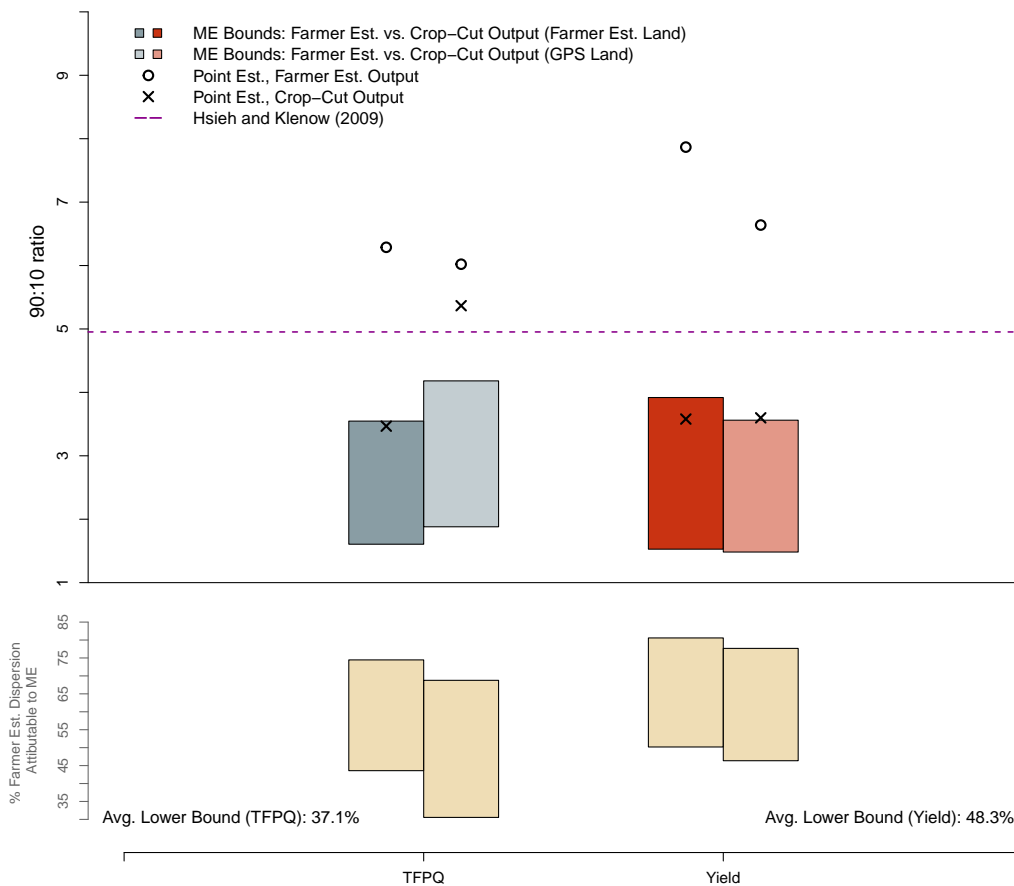
Figure 6: Bounds on cross-sectional productivity dispersion due to measurement error in output. Plotted points indicate the 90:10 ratios of the distributions of productivity residuals computed using farmer-estimated (○'s) and crop-cut based (×'s) measures of crop output individually. Colored polygons bound the range of 90:10 ratios between those associated with the distributions of the average (upper bound) and projected (lower bound) composites. For each measure of productivity, points and polygons on the left were calculated using output measures based on farmer-estimated plot size, and those on the right were calculated using output based on GPS-measured plot size. The tan polygons in lower panel indicate the percent of dispersion in productivity computed using farmer-estimated output that is attributable to measurement error implied by the estimates in the top panel. For example, the lower bound of each tan polygon is the difference between the ○'s and the upper value of the colored bounded area, expressed as percentage of the ○ values. The dashed magenta line represents the 90:10 ratio of the distribution of TFPR observed among Indian manufacturing firms in Hsieh and Klenow (2009).

measurement error in output, productivity dispersion among smallholder agricultural firms may actually be *lower* than dispersion among developing-country non-agricultural firms. However, this comparison is not completely fair, as measurement error is not accounted for in HK's dispersion estimate. More conservatively, these results do not substantiate the claim
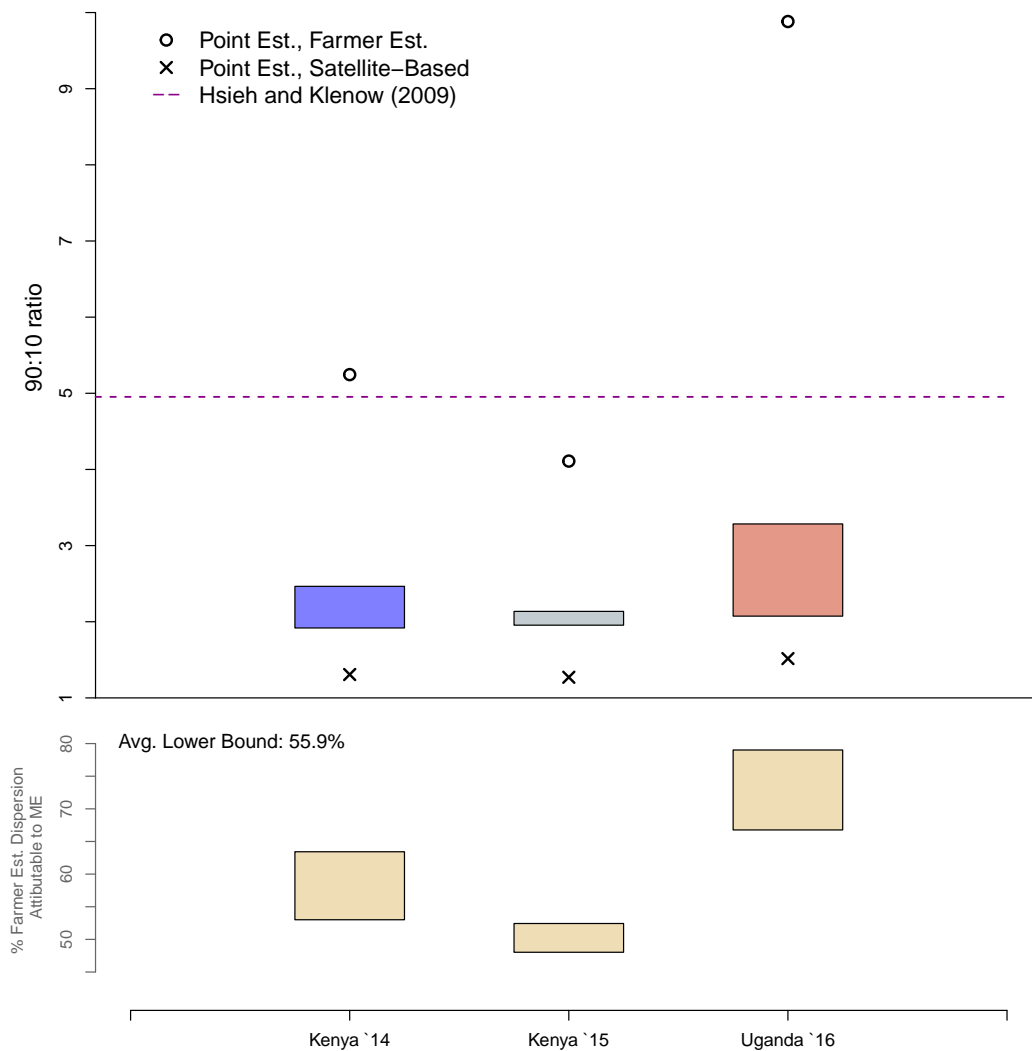
Figure 7: Bounds on dispersion due to measurement error in yield. Errors in measures of crop output in our non-LSMS data from Kenya and Uganda significantly magnify measured dispersion in maize yields, similar to the effect observed in Figure 7. Plotted points indicate the 90:10 ratios of the distributions of yields computed using farmer-estimated (○'s) and satellite-based (×'s) measures of crop output individually. Colored polygons bound the range of 90:10 ratios between those associated with the distributions of the average (upper bound) and projected (lower bound) composites. The tan polygons in lower panel indicate the percent of dispersion in productivity computed using farmer-estimated yield that is attributable to measurement error implied by the estimates in the top panel. For example, the lower bound of each tan polygon is the difference between the ○'s and the upper value of the colored bounded area, expressed as percentage of the ○ values. The dashed magenta line represents the 90:10 ratio of the distribution of TFPR observed among Indian manufacturing firms in Hsieh and Klenow (2009).

that there is more productivity dispersion in agriculture.

For measurement error in inputs, Figure A7 shows the bounds generated when we use farmer-

reported versus GPS-measured land to estimate productivity. The results indicate that the lower bound estimate of the percent of measured productivity dispersion that is due to measurement error in land is relatively small – less than 5% on average, with country-specific estimates ranging from 0-14%. Notably, the estimated bounds for residual yields in Figure A7 are substantially wider than those estimated for TFPR and TFPQ, in particular in the Uganda and Nigeria data. This likely reflects the multiplicative effect of measurement error in yield measures. In general, the narrowness of the estimated bounds for TFPR and TFPQ is consistent with the observation that the farmer-reported and GPS-based measures of plot size are relatively strongly correlated.[35] Consistent with this, the preceding analysis of measurement error in output shows that using objective GPS-based measures of plot sizes, rather than farmer estimates, does not meaningfully change our bounds. While plot sizes are clearly measured with error in agricultural surveys, this type of measurement error does not contribute significantly to our estimate of productivity dispersion across farms.

In contrast, the results of our bounding exercise indicate that measurement error in output accounts for large share of measured productivity dispersion, a finding borne out consistently using various measures of output. After accounting for measurement error in output, productivity dispersion among the small agricultural firms in our data no longer appears larger than benchmark estimates of dispersion for non-agricultural firms.

5.3   Impact of Measurement Errors in Inputs and Output on Persistence

Table 3 summarizes our results on the implications of measurement error for estimates of productivity persistence, using the IV approach described above. Panels in the table correspond to each of the three different key measures of productivity. The first row in each panel indicates the level of persistence observed when estimated using simple OLS, and the subsequent rows contain the results obtained using the different instruments described in Appendix A.5.1, where applicable. The values in Table 3 are the annualized autocorrelation coefficients derived from the linear estimates resulting from the IV. For each country and variant of the IV represented in the table, the associated linear regression results, including the estimated reduced-form, first-stage, and second-stage linear autocorrelations are documented in Tables A8-A15 in Appendix A.6. Note that because we find that measurement error in output accounts for a large amount of the variation in productivity, we do not report estimates of persistence derived using productivity computed using GPS-measured land area

---

[35]See Table A7 in Appendix A.5.3

as an instrument for productivity computed using farmer-estimated land area in Table 3, though these IV regressions are reported in Appendix A.6.[36]

Table 3: Annualized estimates of persistence, correcting for measurement error in either inputs or outputs. IV estimates labeled "2x-lag" instrument lagged productivity with twice-lagged productivity. Estimates labeled "FE vs crop cut" instrument productivity estimates using farmer-estimated output measures with output measures from crop cuts.

	Tanzania	Uganda	Nigeria	Ethiopia	Avg
TFPR					
Naive	0.38	0.17	0.45	0.31	0.33
IV: 2x-lag	1.06	0.90		1.09	1.02
Key-crop TFPQ					
Naive	0.49	0.20	0.45	0.39	0.38
IV: 2x-lag	0.84	0.92		0.93	0.90
IV: FE vs. Cropcut				0.62	0.62
Key-crop Yield					
Naive	0.45	0.15	0.41	0.28	0.32
IV: 2x-lag	0.83	0.69		0.82	0.78

Overall, results on persistence are consistent with the results from our dispersion bounding exercise. When we use crop-cut yield based TFPQ as an instrument for farmer-estimated harvest quantity based TFPQ in Ethiopia (the only location where that comparison is possible), the measured autocorrelation increases to 0.62, up from 0.39 when persistence is estimated using OLS. Across countries, when we use twice-lagged TFP to instrument for lagged TFP – which in principle should capture all sources of measurement error in TFP – we get estimates of persistence that are close to unity.

There are plausible threats to the identification of the linear autocorrelation coefficient for each of our IV approaches. For the IV approach that uses farmer-estimated versus crop-cut based measures of output in Ethiopia, the main concern is weak instruments. Encouragingly, for TFPQ we obtain a highly significant first-stage – the estimated linear autocorrelation is equal to 0.469 with a standard error of 0.025.[37] For our twice-lagged productivity IV strategy,

---

[36]More specifically, in this IV setting the exclusion restriction likely fails. This is because output is being measured in the same way for both alternative measures of productivity (i.e. the one computed using GPS-measured land and the one computed using farmer-estimated area). Thus, in the presence of measurement error in output, for which we find strong evidence, the two alternative measures of productivity will both incorporate the same measurement error in output. So they will be correlated with one another through something other than "true" productivity.

[37]When evaluating persistence in residual yields using this same approach, the estimated first-stage coefficient is 0.001 with a standard error of 0.032 (see Appendix Table A15). This weak first stage may result from the low raw correlation between farmer-estimated and crop-cut yields, and the fact that the autocorrelation of residual yields does not account for variation in the amount of labor and capital used by farmers in differ-

the biggest threat is serial correlation in measurement error, resulting in an upward bias in the IV persistence estimates. We perhaps see some evidence of such bias in our results, with estimates of the linear autocorrelation (at lags greater than or equal to 1 survey wave) using this IV strategy approaching unity in several cases (with values ranging from 0.674-1.184, see Appendix Tables A12-A14).

Despite these concerns, the general picture that emerges from Table 3 is consistent: measurement error appears to significantly attenuate measured persistence. Across countries and measures of productivity, the annualized autocorrelation of productivity estimated using OLS is 0.34. The comparable average across IV approaches using twice-lagged productivity or multiple measures of output is 0.66. One interpretation of these figures is that measurement error attenuates measured persistence by approximately 50%. Finally, despite the threats identified above, we note that these are the first reported estimates of the effect of measurement error on the persistence of productivity among small developing-country agricultural firms.

6   Discussion and Conclusion

We evaluate two key stylized facts which underlie much of the existing literature on firm productivity – that there is wide dispersion in productivity across firms, and that the productivity of individual firms is highly persistent over time – for an understudied and important type of firm, smallholder farms in developing countries. Our results indicate that, when using conventional survey-based approaches to estimating TFP, productivity dispersion among smallholder agricultural firms appears significantly larger than benchmark estimates for non-agricultural firms in developing countries, and persistence in productivity is nearly 50% lower. To better understand these results, we evaluate the sources of measured dispersion in productivity, finding that the largest shares of measured dispersion are attributable to fixed effects which capture unobserved factors that vary at the village-, farmer-, and household-levels.

Because measurement error can rationalize both these initial descriptive findings, we then evaluate the effect of measurement error on measured productivity dispersion and persistence explicitly. Making use of of the multiple measures of both inputs and outputs, we find that measurement error in output accounts for more than 37% of observed dispersion in TFPQ and yield, and that, accounting for measurement error, the dispersion in true productivity

---

ent survey waves. Because we do not obtain a significant first stage, we omit the corresponding second-stage estimates of measurement-error adjusted persistence in residual yields from Table 3, though these estimates are reported in Table A15 in Appendix A.6.

for smallholder agricultural firms in our data is similar to benchmark estimates for non-agricultural firms in developing countries.

As an extension of this analysis, we also investigate how much nonlinearities in the production function or interactions between production function components, which are not specified in our baseline model, contribute to measured dispersion. To do this, we compare dispersion estimates from our baseline linear model to those derived from machine learning predictions that capture the effects of these nonlinearities and interactions. We find that, on average across countries in our sample, only 7.3% of dispersion can be attributed to these factors and that this result is not driven by overfitting, suggesting that our results on dispersion are not dirven by incorrect specification of the functional form in the production function.

To evaluate the effect of measurement error on the persistence of productivity, we use an instrumental variables approach in which estimates based on multiple independent measures of inputs and output in earlier survey waves are used to instrument for lagged productivity in a persistence regression. Consistent with our dispersion results, we find that measurement error in output significantly attenuates measured persistence and that, accounting for measurement error, the persistence of productivity among smallholder farmers is on par with estimates of persistence for non-agricultural firms in developed countries.

Overall, we show that the stylized facts which underlie many analyses of the mainstream firm productivity literature – wide dispersion in productivity across firms, and high persistence in productivity over time – seem to apply in the developing country agricultural context. More significantly, we highlight the importance of accounting for measurement error in the evaluating patterns of firm productivity measured from agricultural survey data. Taking measurement error seriously alters the conclusions that can be derived from the observation that survey-based measures of productivity vary widely across farmers.

For instance, a number of recent papers have pointed to land misallocation as a key driver of low aggregate agricultural productivity in developing countries (Adamopoulos and Restuccia, 2014; Restuccia, 2016). The efficiency gains from policies targeted at land misallocation depend directly on the magnitude of productivity dispersion across farms. Our results indicate that a meaningful share of this dispersion estimated from agricultural surveys is due to measurement error in output. Not accounting for this measurement error is likely to overstate the benefits of reallocation across farmers.

Additionally, a large body of development scholarship argues that poor households (many of

which are engaged in agriculture) can be freed from "poverty traps" via discrete interventions that break them out of low-productivity, low-income equilibria (Kraay and Mckenzie, 2014). One-time interventions that make households more productive, such as training programs, will only have long-run impacts if the induced shocks to productivity are persistent. Our results show that not accounting for measurement error may cause one to understate the persistence of productivity differences across farmers over time.

Finally, our results suggest that improved methods for measuring productivity can help reduce the large amounts of measurement error in survey-based measures. The vast literature on productivity differences of manufacturing firms relies mostly on firm surveys for estimating productivity. Considering alternate measures of output, as we have done in the agricultural setting, may help researchers focusing on other sectors reduce the influence of measurement error on estimates of productivity dispersion.

References

Abay, Kibrom A., Leah Bevis, and Christopher B. Barrett. 2019. "Measurement Error Mechanisms Matter: Agricultural Intensification with Farmer Misperceptions and Misreporting." *NBER Working Paper 26066* .

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification properties of recent production function estimators." *Econometrica* 83 (6):2411–2451.

Adamopoulos, Tasso and Diego Restuccia. 2014. "The size distribution of farms and international productivity differences." *American Economic Review* .

Amit, Y and D Geman. 1997. "Shape Quantization and Recognition with Randomized Trees." *Neural Computation* 9 (7):1545–1588.

Arellano, Manuel and Stephen Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The Review of Economic Studies* 58 (2):277. URL http://restud.oxfordjournals.org/lookup/doi/10.2307/2297968.

Barrett, C., K. Abay, G. Abate, and T. Bernard. 2017. "Correlated Non-Classical Measurement Errors, ' Second Best ' Policy Inference and the Inverse Size-Productivity Relationship in Agriculture." *Working Paper* .

Bertrand, Marianne and Antoinette Schoar. 2003. "Managing With Style: The Effect of Managers on Firm Policies." *Quarterly Journal of Economics* CXVIII (4):1169–1208.

Bevis, Leah E M and Christopher B Barrett. 2017. "Close to the Edge : Do Behavioral Explanations Account for the Inverse Productivity Relationship?" *Working Paper* .

Bils, Mark, Peter J Klenow, and Cian Ruane. 2017. "Misallocation or Mismeasurement?" *Working Paper* (599):1–48. URL http://siepr.stanford.edu/sites/default/files/publications/599wp{_}1.pdf.

Bloom, Nicholas and John Van Reenen. 2007. "Measuring and Explaining Management Practices Across Firms and Countries." *Quarterly Journal of Economics* 72 (4):1329–1372.

Blundell, R. and S. Bond. 2000. "GMM Estimation with persistent panel data: an application to production functions." *Econmetric Reviews* 19 (3):321–340. URL http://eprints.ucl.ac.uk/15809/.

Burke, Marshall and David B. Lobell. 2017. "Satellite-based assessment of yield variation and its determinants in smallholder African systems." *Proceedings of the National Academy of Sciences* 114 (9):2189–2194. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1616919114.

Bushnell, James and Catherine Wolfram. 2007. "The Guy at the Controls: Labor Quality and Power Plant Efficiency." *NBER Working Paper 13215* .

Carletto, Calogero, Sydney Gourlay, and Paul Winters. 2015. "From Guesstimates to GP-Stimates: Land Area Measurement and Implications for Agricultural Analysis." *Journal of African Economies* 24 (5):593–628.

Carletto, Calogero, Sara Savastano, and Alberto Zezza. 2013. "Fact or artifact: The impact of measurement errors on the farm size-productivity relationship." *Journal of Development Economics* 103 (1):254–261. URL http://dx.doi.org/10.1016/j.jdeveco.2013.03.004.

Casley, Dennis J. and Krishna Kumar. 1988. *The Collection, Analysis, and Use of Monitoring and Evaluation Data*. Baltimore and London: Johns Hopkins University Press.

Christen, Robert and Jamie Anderson. 2013. "Segmentation of smallholder households: meeting the range of financial needs in agricultural families." Tech. rep., CGAP.

Cohen, Alex. 2019. "Estimating Farm Production Parameters with Measurement Error in Land Area." *Economic Development and Cultural Change* (June 2014):305–335.

Desiere, Sam and Dean Jolliffe. 2018. "Land productivity and plot size: Is measurement error driving the inverse relationship?" *Journal of Development Economics* 130 (October 2017):84–98. URL https://doi.org/10.1016/j.jdeveco.2017.10.002.

FAO. 2014. "The State of Food and Agriculture 2014: Innovation in Family Farming." Tech. rep., Food and Agricultural Organization of the United Nations.

Fermont, Anneke and Todd Benson. 2011. "Estimating Yield of Food Crops Grown by Smallholder Farmers: A Review in the Uganda Context." Tech. Rep. June, IFPRI Discussion Paper 01097, Washington, D.C.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability." *American Economic Review* 98 (1):394–425. URL http://www.jstor.org/stable/29729976{%}0Ahttp://about.jstor.org/terms.

Gollin, Douglas, David Lagakos, and Michael E Waugh. 2014. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129 (2):939–993.

Gollin, Douglas and Christopher Udry. 2019. "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture." *Working Paper 25440* .

Gourlay, Sydney, Talip Kilic, and David B Lobell. 2017. "Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda." *World Bank Policy Working Paper* (September).

Hatfield, Jerry L. and John H. Prueger. 2015. "Temperature extremes: Effect on plant growth and development." *Weather and Climate Extremes* 10:4–10.

Holden, Stein T and Monica Fisher. 2013. "Can Area Measurement Error Explain the Inverse Farm Size Productivity Relationship?" *Centre for Land Tenure Studies Working Paper* (October).

Hsieh, Chang-Tai and Peter J Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* CXXIII (November):1329–1372.

Kilic, Talip, Alberto Zezza, Calogero Carletto, and Sara Savastano. 2017. "Missing(ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements." *World Development* 92:143–157. URL http://dx.doi.org/10.1016/j.worlddev.2016.11.018.

Kraay, Aart and David Mckenzie. 2014. "Do Poverty Traps Exist? Assessing the Evidence." *Journal of Economic Perspectives* 28 (3):127–148.

Krueger, Alan B and Mikael Lindahl. 2001. "Education for growth: Why and for whom?" *Journal of Economic Literature* 39 (4):1101–1136.

Levinsohn, James and Amil Petrin. 2013. "Production Functions Estimating to Control for Using Inputs Unobservables." *The Review of Economic Studies* 70 (2):317–341.

Lobell, David B., George Azzari, Marshall Burke, Sydney Gourlay, Shenong Jin, Talip Kilic, and Siobhan Murray. 2018. "Eyes in the Sky, Boots on the Ground Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis in Uganda." URL http://documents.worldbank.org/curated/en/556261522069698373/pdf/WPS8374.pdf.

Mccullough, Ellen B. 2017. "Labor productivity and employment gaps in Sub-Saharan Africa." *Food Policy* 67:133–152. URL http://dx.doi.org/10.1016/j.foodpol.2016.09.013.

McMillan, Margaret S and Kenneth Harttgen. 2014. "What is driving the 'African Growth Miracle'?" *NBER working paper 20077* .

Midrigan, By Virgiliu and Daniel Yi Xu. 2014. "Finance and Misallocation : Evidence from Plant-Level Data †." *American Economic Review* 104 (2):422–458.

Olley, G Steven and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (6):1263–1297.

Poate, Derek. 1988. "A Review of Methods for Measuring Crop Production From Smallholder Producers." *Experimental Agriculture* 24:1–14.

Restuccia, Diego. 2016. "Resource Allocation and Productivity in Agriculture." *Working Paper* :2016.

Restuccia, Diego and Raul Santaeulalia-Llopis. 2015. "Land misallocation and productivity." *Available at SSRN 2607103* .

Rotemberg, Martin and T Kirk White. 2017. "Measuring Cross-Country Differences in Misallocation." *Working Paper* :1–27.

Schlenker, Wolfram and David B Lobell. 2010. "Robust negative impacts of climate change on African agriculture." *Environmental Research Letters* 5 (1):014010.

Schlenker, Wolfram and Michael J Roberts. 2009. "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change." *Proceedings of the National Academy of Sciences* 106 (37):15594–15598. URL http://www.pnas.org/content/106/37/15594.abstract.

Stein, Peer, Tony Goland, and Robert Schiff. 2010. "Two trillion and counting." *International Finance Corporation and McKinsey & Company* .

Syverson, Chad. 2011. "What determines productivity?" *Journal of Economic literature* 49 (2):326–365.

Tin Kam, Ho. 1995. "Random Decision Forests." In *Proceedings of the 3rd International Conference on Document Analysis and Recognition*. Montreal, Canada, 278–282.

Udry, Christopher. 1996. "Gender, agricultural production, and the theory of the household." *Journal of political Economy* :1010–1046.

A   Appendices

A.1   Plot and Household Summary Statistics

Table A1: Smallholder Farms in Our Dataset (plot-level)

Factor	Level	Country	Mean	Median	St.dev	Min	Max	N
Output (nominal USD)	Plot	Tanzania	149.56	75.44	184.22	0.13	714.98	15814
		Uganda	70.99	46.09	66.63	0.38	235.59	31408
		Nigeria	521.04	292.35	583.76	0.57	2130.36	9338
		Ethiopia	64.63	35.62	73.30	0.16	260.89	36906
		All	**126.74**	**53.04**	**248.17**	**0.13**	**2130.36**	**93466**
Land (hectares)	Plot	Tanzania	0.83	0.40	0.83	0.01	3.24	15814
		Uganda	0.26	0.20	0.21	0.01	0.81	31408
		Nigeria	0.43	0.30	0.42	0.01	1.66	9338
		Ethiopia	0.15	0.09	0.15	0.01	0.53	36906
		All	**0.33**	**0.20**	**0.47**	**0.01**	**3.24**	**93466**
Labor (person-days)	Plot	Tanzania	62.23	46.00	50.40	1.00	188.00	15814
		Uganda	39.94	35.00	25.80	1.00	118.00	31408
		Nigeria	152.18	96.00	153.46	1.00	572.00	9338
		Ethiopia	24.25	14.49	25.46	0.14	88.72	36906
		All	**48.73**	**30.00**	**68.00**	**0.14**	**572.00**	**93466**
Capital (nominal USD)	Plot	Tanzania	71.90	12.76	143.46	0.00	575.63	15814
		Uganda	17.65	7.63	24.59	0.00	86.49	31408
		Nigeria	45.79	22.78	56.68	0.01	204.14	9338
		Ethiopia	9.14	2.88	13.50	0.00	48.15	36906
		All	**26.28**	**6.31**	**67.89**	**0.00**	**575.63**	**93466**

Table A1: Table A1 shows summary statistics of inputs and output in our plot-level data. Plots in our dataset are small, employ low levels of capital inputs, and generate modest revenues. Output is measured as the dollar value of all crops harvested from each plot, valued using fixed national-average crop-specific prices (per kg) across survey-waves. Land is total plot area as estimated by farmers. Labor is measured in person-days worked by household members and hired laborers on each plot. Where permitted by the data (Tanzania, Nigeria, and Ethiopia), we exclude harvest labor in the calculation of the labor input. Capital is measured as the nominal dollar value of variable inputs (fertilizer, pesticide/herbicide) plus a share of the value of durable capital assets (farm equipment/machinery) owned by the household. The share of durable

capital attributed to each plot is given by the proportion of total household area represented by the plot. Output and capital are converted to nominal USD using the average country-specific exchange rate over the period spanned by all waves of each survey.

Table A2: Smallholder Farms in Our Dataset (household-level)

Factor	Level	Country	Mean	Median	St.dev	Min	Max	N
Output (nominal USD)	Household	Tanzania	290.21	148.63	351.48	0.14	1346.13	8863
		Uganda	223.31	160.98	195.80	0.97	700.91	10860
		Nigeria	989.99	643.41	976.03	1.54	3604.51	5051
		Ethiopia	410.75	279.88	392.13	0.25	1439.07	7765
		All	**405.27**	**214.72**	**546.56**	**0.14**	**3604.51**	**32539**
Land (hectares)	Household	Tanzania	1.57	1.01	1.49	0.02	5.67	8863
		Uganda	1.36	1.01	1.18	0.02	4.45	10860
		Nigeria	0.80	0.52	0.79	0.01	3.00	5051
		Ethiopia	1.03	0.71	0.97	0.01	3.63	7765
		All	**1.25**	**0.81**	**1.21**	**0.01**	**5.67**	**32539**
Labor (person-days)	Household	Tanzania	139.38	97.00	126.47	1.00	480.00	8863
		Uganda	317.91	249.00	255.08	1.00	1020.00	10860
		Nigeria	262.00	151.00	282.63	1.00	1063.00	5051
		Ethiopia	168.00	105.70	169.07	0.31	623.81	7765
		All	**224.83**	**146.00**	**226.18**	**0.31**	**1063.00**	**32539**
Capital (nominal USD)	Household	Tanzania	140.20	24.06	289.09	0.00	1164.63	8863
		Uganda	58.37	22.76	84.46	0.00	309.38	10860
		Nigeria	90.34	45.84	106.47	0.01	386.73	5051
		Ethiopia	61.66	30.83	71.76	0.00	270.25	7765
		All	**86.41**	**27.82**	**176.29**	**0.00**	**1164.63**	**32539**

Table A2: Table A2 shows summary statistics of inputs and output in our household-level panel data. Farms in our household dataset are small, employ low levels of capital inputs, and generate modest revenues. Output is measured as the dollar value of all crops harvested from all plots cultivated by each household, valued using fixed national average crop-specific prices (per kg) across survey-waves. Land is the total area of all plots cultivated by the household based on farmer estimates of plot area. Labor is the total number of person-days worked by household members and hired laborers on plots cultivated by the household. Where permitted by the data (Tanzania, Nigeria and Ethiopia) we exclude harvest labor in the calculation of the labor input. Capital is measured as the nominal dollar value of variable inputs (fertilizer, pesticide/herbicide) used on all plots cultivated by the household, plus the value durable capital assets (farm equipment/machinery) owned

by the household. Output and capital are converted to nominal USD using the average country-specific exchange rate over the period spanned by all waves of each survey.

A.2 Prevalence of Key Crops

Table A3: Key Crops Constitue a Significant Proportion of the Data

Country	Total Obs. (Plot-level)	Total Obs. (HH-years)	Crop	% of Plots (pure-stand)	% of Plots (intercrop)	% Panel Obs. (pure-stand)	% Panel Obs. (intercrop)
Tanzania	15814	8863	Maize	0.158	0.555	0.026	0.559
			Cassava	0.022	0.069	< 0.01	0.051
			Beans	0.02	0.184	< 0.01	0.197
			Sorghum	0.014	0.058	< 0.01	0.053
			Rice	0.094	0.126	0.019	0.137
			All Key Crops	0.318	0.789	0.049	0.712
Uganda	31408	10860	Maize	0.067	0.241	< 0.01	0.517
			Cassava	0.076	0.197	< 0.01	0.551
			Beans	0.07	0.251	< 0.01	0.569
			Sorghum	0.033	0.053	< 0.01	0.11
			All Key Crops	0.254	0.55	< 0.01	0.913
Nigeria	9338	5051	Maize	0.061	0.359	< 0.01	0.337
			Cassava	0.056	0.324	< 0.01	0.287
			Beans	0.018	0.224	< 0.01	0.213
			Sorghum	0.041	0.291	< 0.01	0.331
			Rice	0.046	0.063	< 0.01	0.059
			All Key Crops	0.222	0.851	< 0.01	0.759
Ethiopia	36906	7765	Maize	0.108	0.169	< 0.01	0.405
			Beans	0.053	0.074	< 0.01	0.177
			Sorghum	0.091	0.137	< 0.01	0.275
			Teff	0.124	0.131	< 0.01	0.297
			Wheat	0.073	0.079	< 0.01	0.188
			All Key Crops	0.451	0.561	0.018	0.735

Table A3: Table A3 shows that, in each of our sample countries, a select number of common crops constitute a significant portion of our plot- and household-panel samples. Column 3 is the total number of plot-level observations in each cross-sectional dataset. Column 4 shows the total number of household-year observations in the panel of households we observe in at least two years. Columns 5 is the proportion of plots that were only cultivated with each key crop. Column 6 is the proportion of plots where a key crop was cultivated, including intercropped plots. Column 7 is the proportion of household-years in the panel sample where households only cultivated a given crop. Column 8 shows the proportion of household-years in the panel sample where households cultivated a given crop, even if they also cultivated other crops. For the pure-stand columns (5 and 7) the 'All Key Crops' row indicates the sum across all crops listed in Table A3. For the 'intercrop' columns (6 and 8), the 'All Key Crops' row indicates the proportion of the total plot-level and panel sample in which plots/households cultivated any key crops. Note, crops that constitute less than 5% in all columns were excluded in each country subcomponent, but do figure in to the all key-crop totals.

A.3 Description of Household Geovariables

Table A4: Summary Statistics of Household Geovariables

Variable	Tanzania	Uganda	Nigeria	Ethiopia
Climate and Weather Variables				
Mean ann temp (°C * 10)	225.85, (25.64)	218.93, (18.06)	263.53, (10.07)	190.26, (31.70)
Avg temp of wettest qtr (°C * 10)	231.62, (26.75)	215.90, (16.42)	251.97, (11.24)	186.43, (32.43)
Avg ann precip (mm)	1114.85, (327.77)	1236.31, (177.07)	1331.35, (590.17)	1158.01, (379.34)
Tot precip in survey year (mm)	825.38, (298.80)	1201.84, (187.76)	1233.18, (414.40)	1157.88, (481.80)
Avg precip in wettest mo (mm)	230.53, (85.20)	177.41, (28.91)	275.27, (56.57)	234.99, (77.21)
Avg precip of wettest qtr (mm)	569.48, (186.81)	456.92, (72.91)	701.44, (181.83)	596.57, (204.55)
Precip in wettest qtr of survey year (mm)	488.43, (161.74)	459.99, (118.68)	646.49, (125.57)	578.68, (215.63)
Avg start of wettest qtr (dekad)	17.28, (16.67)	8.30, (4.31)	17.85, (1.04)	15.24, (3.04)
Start of wettest qtr in survey year (dekad)	15.45, (15.41)	8.68, (6.54)	17.93, (3.50)	19.07, (5.70)
Potential wetness index	13.44, (3.33)	1241.67, (253.84)	14.14, (2.96)	12.41, (1.68)
Geographic Variables				
Elevation (m)	928.99, (544.86)	1241.66, (253.82)	316.46, (217.13)	1918.44, (520.97)
Slope (%)	6.10, (6.00)	7.56, (7.34)	2.69, (2.73)	-
Agro-Ecological Zone (%)[†]				
Temperate/semiarid	53.55	0.00	0.00	0.00
Tropic-cool/humid	0.00	29.09	0.00	20.06
Tropic-warm/subhumid	0.00	17.17	51.91	45.52
Tropic-warm/humid	0.00	53.74	6.36	0.21
Terrain (%)[†]				
Mid-alt plains	6.12	0.00	32.79	1.60
High-alt plains	27.58	32.25	12.20	8.47
Mid-alt plateau	25.95	49.07	1.13	21.48
Mid-alt mountain	0.01	14.81	0.00	25.47
Soil Characteristics Variables				
Nutrient availability constraint (%)[†]				
None	30.62	14.24	38.86	64.08
Moderate	50.73	61.52	36.69	34.71
Severe	13.08	19.69	24.21	0.48
Nutrient retention constraint (%)[†]				
None	46.02	32.15	44.59	75.11
Moderate	46.64	61.87	50.07	24.13
Rooting conditions constraint (%)[†]				
None	70.10	73.13	65.29	52.32
Moderate	18.42	12.70	29.00	15.94
Root O_2 avail constraint (%)[†]				
None	81.19	69.53	84.14	86.16
Moderate	9.88	24.65	13.15	9.89
Excess salt constraint (%)[†]				
None	88.83	93.90	97.86	98.98
Soil toxicity constraint (%)[†]				
None	92.17	93.90	99.76	98.96
Land workability constraint (%)[†]				
None	48.73	66.89	61.27	17.00
Moderate	30.78	15.61	31.58	31.71
Severe	11.69	9.93	5.50	20.68
Observations	8863	10860	5051	7765

Table A4: Table A4 summarizes the household-level geovariables included in our baseline specification of the production function by country. For continuous variables, the table reports means and standard deviations (in parentheses) across all household-years in our household-level sample. For categorical variables, the table reports the % of household-year observations belonging to select categories. For variables denoted with a †, only a subset of categories are depicted. For weather and climate variables, long-run (i.e., beyond the period of each survey) climate averages are denoted with 'Avg'; the remaining variables indicate averages within each survey-year. Both the climate and weather variables are derived from external environmental datasets, which use statistical and process-models to spatially interpolate data.

A.4 Details of and Sensitivity to Measurement Choices

It is important to acknowledge that the richness of the LSMS-ISA data permit alternative choices about how to measure inputs and output in each country. Given the nascent state of the literature on developing-country agricultural total factor productivity, there are no well-established norms. In what follows, we describe our baseline measurement choices in detail, and conduct additional sensitivity analyses to evaluate the extent to which alternative measurement choices affect the distribution of estimated productivity. A description of the alternative measurement scenarios we evaluated is provided in Figure A1. Figure A2 summarizes of the sensitivity of our baseline measures of TFPR dispersion by country across these different scenarios.

Output: Our primary measure of agricultural output is computed as the sum, across crops cultivated within a given production unit (plot or household), of the product of the harvested quantity and a fixed national-level median crop price. By valuing harvest quantity, versus relying on reported crop sales, this measure accounts for the value of auto-consumed harvest. Aggregating harvest value across crops is also essential in the African agricultural context, where intercropped production systems are the norm. It is important to note that in using fixed national prices we abstract from variation in the value of output driven by differences in product quality or the effect of local demand conditions on crop prices. While this abstraction is likely unproblematic in our data setting[38], we cannot reject the possibility that ascribing the same prices to all firms actually creates systematic variation in measured output due to the fact that (relative) crop prices in some areas (or at some times) differed significantly from the national median levels. For this reason, we also measure agricultural output as the physical quantity (in kgs) of crops harvested from a given plot (or by a household) for a set of four to five key cereal crops in each country.[39] Obviously because physical harvest quantities do not depend on local prices, this measure of output provides a consistent means of comparing firms experiencing different local market conditions. The drawback is that we cannot aggregate quantities across the different crops produced by a particular production unit. Following the convention in the manufacturing firm-productivity literature, we present results using both value- and physical-based measures of output for all of our main analyses.

Land: In our plot-level analyses, we measure the land input as the total area of the plot as reported by the farmer. At the household-level, we use the sum of reported plot areas for all plots cultivated by the household. In the estimation of crop-specific physical productivity (i.e. when output is measured in kilograms of crop harvested, or as crop yield in mt/ha) we

---

[38]Because the majority of agricultural products produced by smallholder farmers are homogeneous, undifferentiated goods, differences in product quality across firms in our data are likely small. Regarding local prices and market conditions, our claim is that the notion of productivity most relevant for the comparison to previous studies of manufacturing firms is one of technical efficiency - that is how much physical 'stuff' can a firm produce with a given level of inputs. Using fixed national prices to value harvest simply provides a means of converting crop-specific physical output into a unit of measure (value) which can be consistently aggregated across crops.

[39]Specifically, we consider the following key crops: maize, beans, sorghum (all countries), cassava (Tanzania, Uganda, Nigeria), rice (Tanzania, Nigeria), wheat, and teff (Ethiopia). For more detail regarding the prevalence and selection of key crops, see Appendix A.2.

use crop-specific planted area at the plot-level, and the analogous sum for the household-level. As noted in the description of our data, we do observe GPS-based measures of plot area for a large fraction of plots. We choose to use farmer-estimated area because that is what is most commonly observed in other agricultural survey datasets. We explore the sensitivity of measured productivity to different choices of how to measure the land input (ex. as harvested area, or as estimated sales value of land) in Figure A2.

Labor: As discussed in Syverson (2011) the preferred measure of labor inputs is the total wage bill, as the wages paid to different workers should reflect their levels of human capital, and thus the wage bill is a quality-adjusted measure of labor inputs. However in our data, while the wage bill is reported for hired labor, no value is reported for the labor supplied by the members of farm households. Given that we cannot reliably estimate the implicit wages paid to household members, we measure the labor input as the total number of person-days supplied by either hired laborers or own household members. In most of our sample, the data also permit measuring the labor input in a more disaggregated way, i.e. with separate input quantities for labor supplied by own vs. hired workers, male vs. female workers, or labor allocated to different activities (ex. land preparation and planting vs. weeding and fertilizing). Additionally, the individual information supplied about household members, allow us to construct additional measures of labor quality such as the average age, sex, and education-level of own-farm workers.

Capital: We measure aggregate capital inputs in value terms, as the sum of expenditures on variable inputs (seeds, fertilizer, pesticides and herbicides) and the farmer-reported value of the stock of durable capital (tools, machinery, and structures) owned or rented in by the production unit. We also use imputed fixed national-level median prices to value the quantities of own-produced organic fertilizer (e.x. animal manure) used in production. Because durable capital is typically only reported at the household level, to construct a plot-level measure, we attribute to each plot a share of household-level durable capital proportional to the plot's share of total household area. As with our labor data, we evaluate how much measuring capital inputs in a more disaggregated fashion (ex. with separate measures of variable vs. durable capital, or for different types of variable and durable capital) affects our estimates of productivity in Figure A2.

A.5 Understanding Measurement Error

To bound the role of measurement error in output on estimated productivity dispersion, we use multiple measures of output: farmer self-reports, data from crop cuts, and estimates from satellites. "Crop-cutting" is a common measurement procedure in which a (random) subsection of a farmer's plot is harvested and the production from this section is processed (ex. dried) and weighed by enumerators. Because only a subset of the plot is harvested, yield measured from crop cuts is considered to be a noisy measure of yield from the entire field.[40]

---

[40]While some studies use crop-cuts as a benchmark for evaluating measurement error in farmer self-reports (ex. Barrett et al. (2017)), others evaluate the accuracy of crop-cuts relative to a gold standard benchmark where survey enumerators observe directly or participate in the harvest of an entire plot. These evaluations suggest that crop-cuts are also a noisy measure, where errors result from imperfect observance of the crop-cut

Figure A1: Description of Alternative Measurement Scenarios

	Baseline	Alternative Scenarios	N
Output	Tanzania All-crop harvest value. Quantities reported in kilograms, and valued using fixed national median crop prices (per kg). Prices computed using all paired observations of farmer reported (i) harvest quantity/estimated value or (ii) sales quantity and sales value, across all survey waves. Uganda, Nigeria, and Ethiopia All crop harvest value. Quantities converted from non-standard units using reported conversion factors (where available) and using fixed national average conversion factors for each crop-unit otherwise. Harvest valued using fixed national median crop prices (per kg) based on all paired quantity-value observations.	Tanzania 1. Harvest valued using fixed national median crop prices based only on reported crop *sales* 2. Harvest valued using fixed national median crop prices based only on farmer-reported *harvest value* 3. Harvest valued using farmer-reported values where available, and imputed using (baseline) fixed national median crop prices otherwise Uganda, Nigeria, and Ethiopia 1. Quantities converted to kilograms using estimated fixed national average conversion factors in all cases, rather than reported conversion factors where available (NGA) 2. Quantities converted to kilograms using spatially and temporally varying estimates of reported non-standard unit conversion factors, and valued using spatially/temporally varying median crop-prices (UG/ETH)	Tanzania: 3 Uganda: 1 Nigeria: 1 Ethiopia: 1
Land	All Countries Total plot/parcel/field area as reported by survey respondents.	All Countries 1. All-crop planted area, based on the reported percentage of each plot planted by crop and the farmer-estimated total plot area. 2. All-crop harvested area, based on the reported percentage of planted area that was harvested (by crop) and the farmer estimate of total plot area (TZ/NGA only). 3. Farmer estimated land value, or land value imputed using spatially and temporally varying estimates of land prices 4. Baseline measure of land input plus additional plot-level land quality controls (ex. wetness, steepness, elevation)	Tanzania: 4 Uganda: 3 Nigeria: 4 Ethiopia: 3
Labor	All Countries Total person-days worked on each plot	All Countries 1. Total person-hours based on reported hours worked, or average hours per day estimated from labor module 2. Person-days disaggregated by (i) own vs. (ii) hired labor 3. Person-days disaggregated by own vs. hired *and* by gender (male vs. female) 4. Total person hours for own labor and total hired-labor costs 5. Total person days plus additional plot- and household-level labor quality controls (ex. avg age, sex, marital status, education)	Tanzania: 5 Uganda: 5 Nigeria: 5 Ethiopia: 5
Capital	All Countries Value of all variable inputs used on plot, plus a share of the value of household durable capital. Variable and durable capital valued using national average prices by input-type within survey waves. Plot-share of household durable capital equal to the proportion of total household land area represented by the plot.	All Countries 1. Capital value disaggregated into variable (fertilizer, pesticide, seed) and durable (tools/machinery) capital input 2. Capital value disaggregated into variable vs. durable. Variable capital disaggregated by type (fertilizer, pesticide, and seed). Durable capital disaggregated by ownership-status (rented vs. owned)	Tanzania: 2 Uganda: 2 Nigeria: 2 Ethiopia: 2

Figure A1: Figure A1 Describes the alternative measurement scenarios evaluated in Figure A2.

Satellite-based estimates of crop yields are generated by extracting vegetation indices from satellite imagery within known field boundaries, and have been shown to be highly correlated with ground-based yield measures (Burke and Lobell, 2017; Lobell et al., 2018).[41] This innovative approach offers great promise for assessing dimensions of agricultural productivity at a larger scale and higher frequency than is possible using surveys alone, but improving and validating the accuracy of predictions generated using these methods is still limited by the availability of high-quality ground data.

---

protocols and from sampling error when yields are heterogeneous within plots (Casley and Kumar, 1988; Desiere and Jolliffe, 2018; Gourlay, Kilic, and Lobell, 2017; Fermont and Benson, 2011; Poate, 1988).

[41]The R^2's from the regressions of ground-measured crop yields on satellite-based predictions in these studies range from approximately 0.20-0.55.

Figure A2

Productivity Dispersion and Sensitivity to Measurement Choices
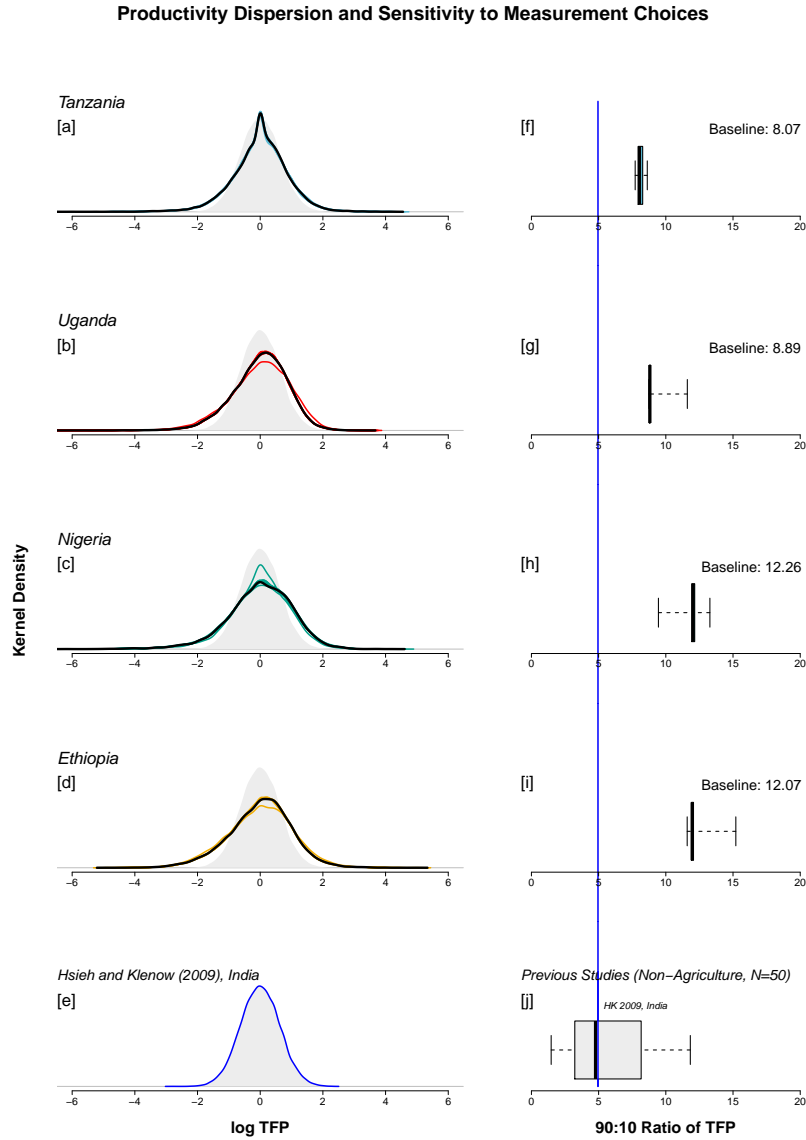
Figure A2: Figure A2 highlights the limited sensitivity of the estimated distribution of TFP to choices about how to measure inputs and output. Panels [a]-[d] on the left show the distribution of log TFP across plots in each of our sample countries. The bold black lines in these panels correspond to the distribution resulting from the estimation of our baseline regression specification, using our baseline measures of inputs and output. Each of the colored lines in panels [a]-[d] show the distribution resulting from a different set of measurement choices, using the same (baseline) regression specification. Appendix Table A4 describes each of the different scenarios evaluated for each country in detail. The box-plots in panels [f]-[i] on the right describe the range of 90:10 ratios across all the measurement-sensitivity scenarios shown on the left-side panel of the same row. In each box-plot, solid black lines are located at the median 90:10 ratio value, the colored box contains the inter-quartile range, and the whiskers extend to the extrema.

A.5.1 Empirical Implementation of Measurement Error Bounding Analysis

To implement our proposed method for bounding true productivity dispersion, we require two measures of productivity for the same plot. In practice, given that plot size and crop

yields are both multiply-measured in different subsets of our data, we can generate these alternative measures in two separate ways. The first and most broadly applicable way is to use the two different estimates of plot size – farmer-reported and GPS measured – in separate estimations of Equation (1) to generate two sets of productivity residuals. The subset of our data where we observe both farmer-estimates and GPS-measures of plot size is much larger than the analogous subset where we observe multiple measures of crop yields. In particular, we see multiple measures of plot size in all survey waves of all of our sample countries. By contrast, only in the Ethiopia LSMS data, and in the non-LSMS survey data from Kenya and Uganda do we observe multiple measures of crop yields. Furthermore, we can apply our bounding technique using multiple measures of plot size for all three of our key measures of productivity (TFPR, TFPQ, and yield). This is not the case for crop yields, where we can only estimate bounds on physical measures of productivity.

The second way to implement our bounding technique is to generate separate productivity residuals using multiple measures of crop yields. This, in turn, is done in different ways depending on the specific data being used. In the Ethiopia LSMS data, the multiple estimates of crop yields we observe are (1) the quotient of farmer-estimated harvest quantity and farmer-reported plot size and (2) the reported harvest quantity from crop cuts conducted by enumerators, divided by the crop cut area ($4m^2$). For computing bounds on the variance of residual yields, these two metrics can be used without further modification. When bounding dispersion in TFPQ, we multiply the computed crop-cut yields by either the farmer-reported or GPS-measured plot size to recover an estimate of total plot-level crop output.[42] This estimate, in conjunction with farmer-reported estimates of harvest quantity can then be used as the dependent variable on the left-hand side of Equation (1).

We can also generate bounds using the multiple estimates of crop yields contained in our supplementary datasets from Kenya and Uganda, which are described in (Burke and Lobell, 2017; Lobell et al., 2018). The most important feature of these data is that they contain survey-based estimates of crop yields that are matched to independent satellite-based yield estimates. Obtaining satellite estimates of yields for the plots in the LSMS data is not possible because plot-level geolocations are confidential in the publicly available versions of the data. Even the household geocoordinates recorded in the publicly-available LSMS are "jittered" for confidentiality reasons. The ability to compare the bounds on dispersion derived from survey- versus satellite-based yields, to those derived from farmer-reported versus crop-cut yields is a key advantage afforded by these data. However, the scope for applying our bounding technique in these data is more limited. In particular, they do not contain reliable measures of labor and capital input use. Consequently, we can only estimate bounds on the dispersion of crop yields, and not TFPQ. Additionally, farmers surveyed exclusively cultivated maize, so the measures of yield dispersion derived from these data differ slightly from those generated using the LSMS data, where we consider multiple key crops (including maize).

---

[42]Specifically, when using crop-cut TFPQ and TFPQ computed using farmer-estimated plot size, we multiply by the farmer-reported area. When using crop-cut TFPQ and TFPQ computed using GPS-measured plot size, we use the GPS-based measure.

Additionally, to maintain comparability with our descriptive analyses, we focus on the 90:10 ratio of the productivity distribution, rather than the variance, in our empirical bounding exercise. That is, for all measures of productivity and in all the aforementioned data settings, we use the 90:10 ratio of the average and projected productivity composites to bound the 90:10 ratio of true productivity. While the 90:10 ratio is a different measure of dispersion than the variance, for a broad set of probability distributions (including all symmetrical distributions) it is a simple (linear) function of the variance. Thus, for this set of distributions, the theoretical results established in Section 4.1 apply directly to the 90:10 ratio.

Next, consider the implementation of our instrumental variables approach to estimating the persistence of true productivity. In general, all the considerations related to applicability of our measurement error framework in different data settings described above also apply in our persistence analysis. As an exception, we note that the non-LSMS data from Kenya and Uganda is not constructed as a panel, so we do not generate any persistence estimates using these data. Additionally, because it is households, not plots, that are tracked across survey waves, we use household-level aggregates of the multiple measures of land area and crop yields to implement our persistence analysis. For the land input, our household-level measures are the sum of farmer-estimated and GPS-measured plot size across all plots cultivated by the household. For crop yields, we take the simple average yield across plots or crop-cuts. And for harvest quantities used to estimate TFPQ, we first back calculate plot-level harvest quantities from crop-cut yields using the approach described above then sum to the household level.[43] In summary, we generate IV estimates of persistence in TFPR, TFPQ, and residual yields using multiple household-level measures of land area in all countries in the LSMS data, but only evaluate TFPQ and yields using the alternative household-level measures of crop yields and harvest quantity in the Ethiopia data.

As an additional extension, in our empirical persistence analysis we also estimate a variant of the IV regression described in Section 4.2 in which we use twice-lagged productivity, rather than alternatively-measured productivity, as an instrument. In other words, if we let ω^a represent the noisy household-level measures of productivity that we observe using farmer-reported estimates of land area and output, then in this variant of our IV procedure, we use ω^a_{t-2} to instrument for ω^a_{t-1} in the panel regression of ω^a_t on ω^a_{t-1}. The motivation for implementing this alternative approach is to account for potential error in the measures of multiple production function components simultaneously. More specifically, while implementing our baseline IV strategy tells us how much measurement error in the multiply-measured quantity (either land area or output/yield) attenuates persistence, using twice-lagged productivity as an instrument controls for potential measurement error in all inputs and outputs used to generate a particular productivity measure. While this is more comprehensive, the assumptions required for validity are correspondingly stronger. In particular, we must assume that

---

[43]Note, we frequently observe households where at least one plot, but not all, were either GPS measured or crop-cut. When computing crop-cut based estimates of harvest quantity (for TFPQ) and GPS-measured land area at the household-level, we first substitute farmer-reported plot-level values for missing data prior to summing across plots. In this way, we retain the full set of households for where at least one plot was crop-cut/GPS measured, and differences in the household-level measures only reflect the differences between the alternative measures across plots that were multiply measured.

none of the measurement errors in any (rather than jut one) of the production function components are serially correlated. Practically, we implement this variant of our IV strategy for all three key measures of productivity in all countries except Nigeria, where we only have two survey waves of data.

A.5.2   Implementation of Random Forest

Random forest is a 'decision tree' algorithm in which predictions of outcomes are generated by identifying a sequence of binary breaks, termed 'branches', in the predictor data, and attributing the conditional mean of the observed outcome to observations within subsets delineated at each branch. In random forest, this process is implemented iteratively on subsets of the data (i.e. the algorithm 'grows' many decision trees), with branches chosen to minimize a loss function in each iteration. Final predictions are computed by aggregating across all the component decision-trees. Intuitively, random forest approximates fitting a step function to the observed outcomes, where each step is defined by a set of binary conditions across the predictor variables. Any effects of non-linearities and interactions between predictor variables in the determination of the outcome are thus captured in the conditional means associated with each branch. The algorithm is best suited for data settings similar to ours with many observations, and in which the only objective is to create the best possible predictions of the outcome variable (ex. in lieu of identifying the most important causal predictors).

Random forests are in general thought to be unlikely to overfit the data on which they're trained, as they sample only a subset of predictor variables at each node in a decision tree and then average over many trees. Because we wish to compare RF estimates to those from a linear model, one potential concern is that any differences between linear and random forest model estimates of productivity is because the linear model – but not the random forest – is overfitting the data. To evaluate this, we split our data into five disjoint train and test datasets and evaluate predictive performance only on held-out test data (i.e. 5-fold cross validation). We then assess overfitting by comparing (i) the 90:10 ratios, and (ii) the R^2 of the regression of observed residualized revenue on predicted residualized revenue for each type of model. If the linear model is overfitting, then this will lead to larger 90:10 ratios and lower R^2 values in the test data relative to the full sample.

The results from this exercise are displayed in Table A5 and suggest our linear model is not significantly overfit. Across countries and measures of model performance, we observe essentially no differences between the cross-validated models and those estimated on the full sample. Consistent with our baseline comparison, the random forest model performs modestly better than the linear model across both performance metrics. Our overall interpretation from these results is that non-linearities and interactions between observed components of the production function do not explain a large share of variation in output, nor are our linear models likely overfitting the data. Put simply, our results suggest that the production function for smallholder agricultural firms is essentially Cobb-Douglas.

Table A5: Cross-Validated Estimates of TFPR: Linear vs. Random-Forest

90:10 Ratio

	Linear (full sample)	Linear (cross-validated)	Random Forest (full sample)	Random Forest (cross-validated)
Tanzania	7.16	7.15	6.93	6.93
Uganda:	8.71	8.72	8.14	8.18
Nigeria:	9.93	9.92	9.07	9.07
Ethiopia:	9.64	9.66	8.59	8.61

R^2 of Observed on Predicted Residualized Revenues

	Linear (full sample)	Linear (cross-validated)	Random Forest (full sample)	Random Forest (cross-validated)
Tanzania	0.191	0.188	0.202	0.206
Uganda:	0.165	0.164	0.200	0.199
Nigeria:	0.081	0.077	0.138	0.136
Ethiopia:	0.296	0.295	0.331	0.328

Note: The second panel reports the adjusted R^2 from the regression of all-crop revenue, residualized on crop-system and village-year fixed effects, on the predictions of this residual generated by each model.

A.5.3 Descriptive Comparison of Multiple Land and Output Measures

The underlying relationship between different measures of land and yields is an integral to understanding the results we obtain in our empirical analysis of measurement error. For example, the strength of the correlation between each pair of alternative measure (and therefore the productivity residuals derived from them) determines the size of the estimated bounds on productivity dispersion, and the significance of the first stage in the IV persistence regression. How then, do the different measures of land area and output we observe in the different data settings described above compare?

We observe both farmer estimates and GPS measures of plot area for 72% of our plot-level observations. The correlation between the two measures is relatively strong. Across countries, the average R^2 of the regression of farmer-reported on GPS-measured plot size is 0.61. In general, the distributions of the two different measures are similar in all countries. However, the density of farmer-estimates exhibits bunching around integers or simple fraction values (see Figure A.5.2 below).[44] For example, in Tanzania, the density of farmer-estimates of plot size within 0.05 acre bins around values of 1, 2, and 3 acres are 12 to 17 times higher than the corresponding densities for GPS measured area. As a result of this bunching,

---

[44]For additional information on the proportion of plots that were GPS measured and the correlation between GPS and farmer-reported plot size by country, see Table A.5.1 in Appendix A.5. Additionally, we note that in Uganda, GPS measurements were taken at the parcel, rather than the plot-level, where parcels are defined as contiguous assemblages of plots. Consequently, we conduct our measurement error analysis in the Uganda data at the parcel-level.

the distribution of farmer-reported plot size exhibits slightly lower variance than the GPS measures. Across countries, the average standard deviation of farmer-estimated plot size is 0.59 ha, whereas the average standard deviation of GPS-measured area is 0.69 ha.

In the Ethiopia LSMS data, enumerators conducted a crop-cut on 31% of plots containing any key crop. In contrast to the multiple measures of plot size, the correlation between crop-cut yields and yields computed from farmer-reported harvest quantity and plot size is low. On average across key crops, the R^2 of the regression of farmer-estimated yields on crop-cut yields is 0.08.[45] The distribution of crop-cut yields is also significantly right-shifted relative to the farmer estimates (see Figure A4 in Appendix A.5). For example for maize, the mean farmer estimated yield is 3.9 mt/ha, whereas the mean crop-cut yield is 6.2 mt/ha. This may reflect a tendency for enumerators to randomly select areas for crop-cutting only in portions of plots where crops are being grown, which would result in higher measured yields relative to the yields computed based on the farmer estimates where total planted area (including areas with no harvestable crops) is the denominator. Crop-cut yields also exhibit lower variance than farmer estimates. Again for maize, the standard deviation of crop-cut yields is 5.2 mt/ha and 7.9 mt/ha for farmer-estimated yields. This likely results from the fact that, for farmer-estimated yields, both the numerator and the denominator are measured with error, whereas crop-cuts were conducted in a standardized $2 \times 2 \ m^2$ area.

In contrast to the crop-cuts in Ethiopia, the mean of satellite-based maize yields in Kenya and Uganda is similar to that of the survey-based measure, and the mean of both measures are similar in magnitude to mean yields we observe in the LSMS. For example, in Uganda, where both LSMS and non-LSMS surveys were conducted, the mean of farmer-estimated maize yields in the LSMS data is 1.19 mt/ha. In the non-LSMS data, mean maize yields are 0.81 mt/ha and 2.22 mt/ha for the survey- and satellite-based estimates, respectively. Relative to the crop-cuts, the satellite-based estimates of yield are also more strongly correlated with the survey-based estimates, with R^2 values ranging from 0.08-0.22. Finally, the satellite-based yield estimates also exhibit significantly lower variance than the survey-based measures, and this relative difference is greater than between the alternative yield measures in the Ethiopia LSMS data. On average across the countries and years of observation contained in the Kenya and Uganda data, the standard deviation of satellite-based maize yields is 0.32 mt/ha, whereas the standard deviation of survey-based yields is 1.01 mt/ha.[46]

---

[45]For additional information on the frequency of crop cuts for different key crops, and the correlation between crop-cut and farmer-estimated yields by crop, see Table A.5.3 in Appendix A.5.

[46]So in the Ethiopia data, crop-cut maize yields are approximately 34% less variable than the survey-based estimates, while in the non-LSMS Kenya and Uganda data, satellite-based yields are 68% less variable.

Table A6: Multiple Measures of Land Input in Most of the Sample

Country	Total Obs. (plot-level)	Farmer Estimated	GPS Measured	Both	R-Squared (FE $\sim$ GPS)
Tanzania	20881	20881 (100%)	13408 (64%)	13408 (64%)	0.69
Uganda (parcel-level)	34712	34712 (100%)	20064 (58%)	20064 (58%)	0.68
Nigeria	9338	9338 (100%)	8202 (88%)	8202 (88%)	0.43
Ethiopa	36932	36932 (100%)	31936 (86%)	31936 (86%)	0.65

Table A7: Multiple Measures of Output for Key Crops in Ethiopia

Crop	Total Obs (Plots)	Total Obs (HH-years)	Plots w. Crop-cut (%)	HH-years w. Crop-cut (%)	R^2: FE Yield $\sim$ CC Yield	R^2: GPS Yield $\sim$ CC Yield
Maize	5749	3488	1528 (27%)	1412 (40%)	0.04	0.03
Beans	2477	1770	1049 (42%)	950 (54%)	0.05	0.06
Sorghum	4786	2573	1309 (27%)	1177 (46%)	0.03	0.03
Teff	4814	2627	1357 (28%)	1229 (47%)	0.04	0.04
Wheat	2885	1707	937 (32%)	867 (51%)	0.04	0.06
All Key Crops	19900	6263	6180 (31%)	3815 (61%)	0.07	0.08

GPS−Measured vs. Farmer−Estimated Land Area

Tanzania
[a]

Uganda
[b]

Nigeria
[c]
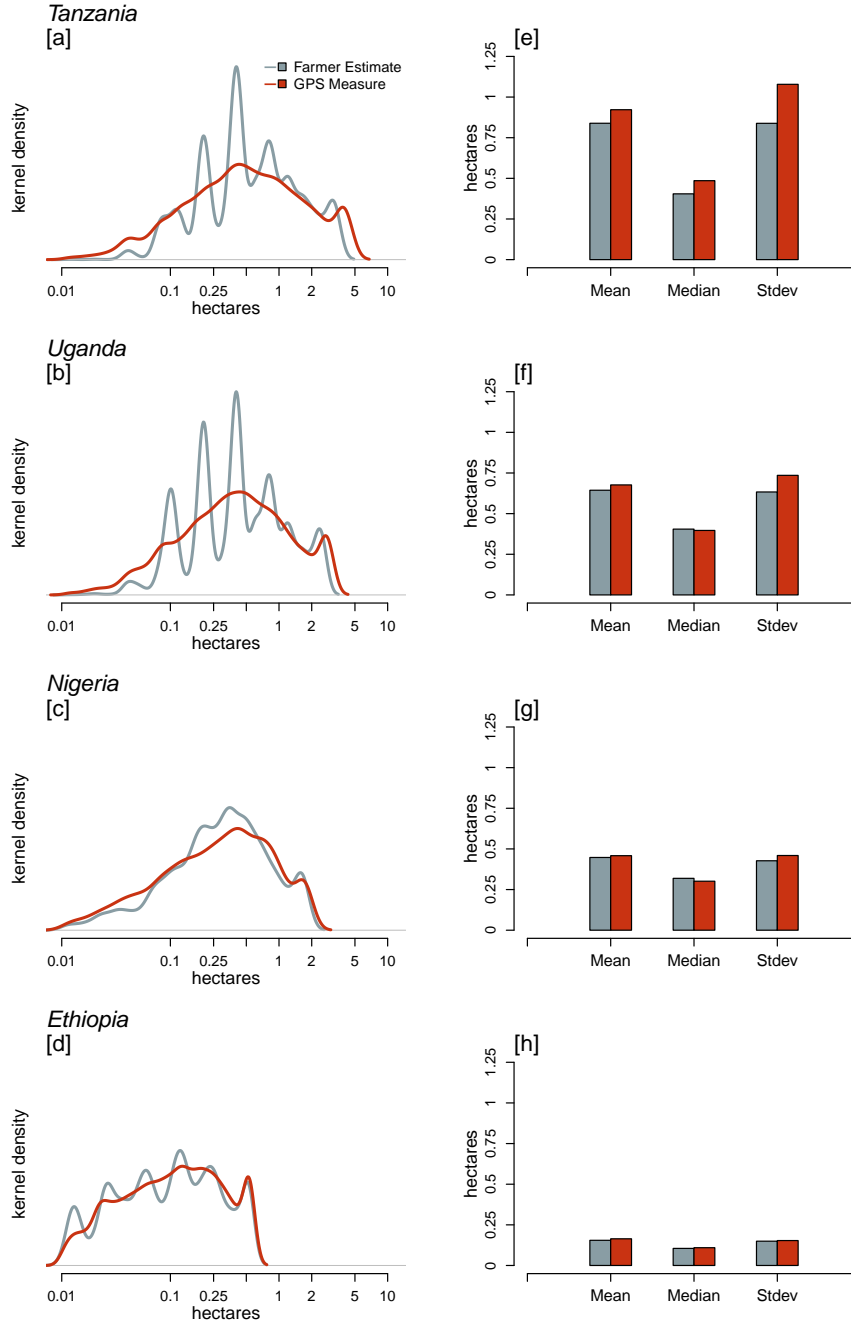
Ethiopia
[d]

Ethiopia: Crop–Cut vs. Computed Yields for Key Crops
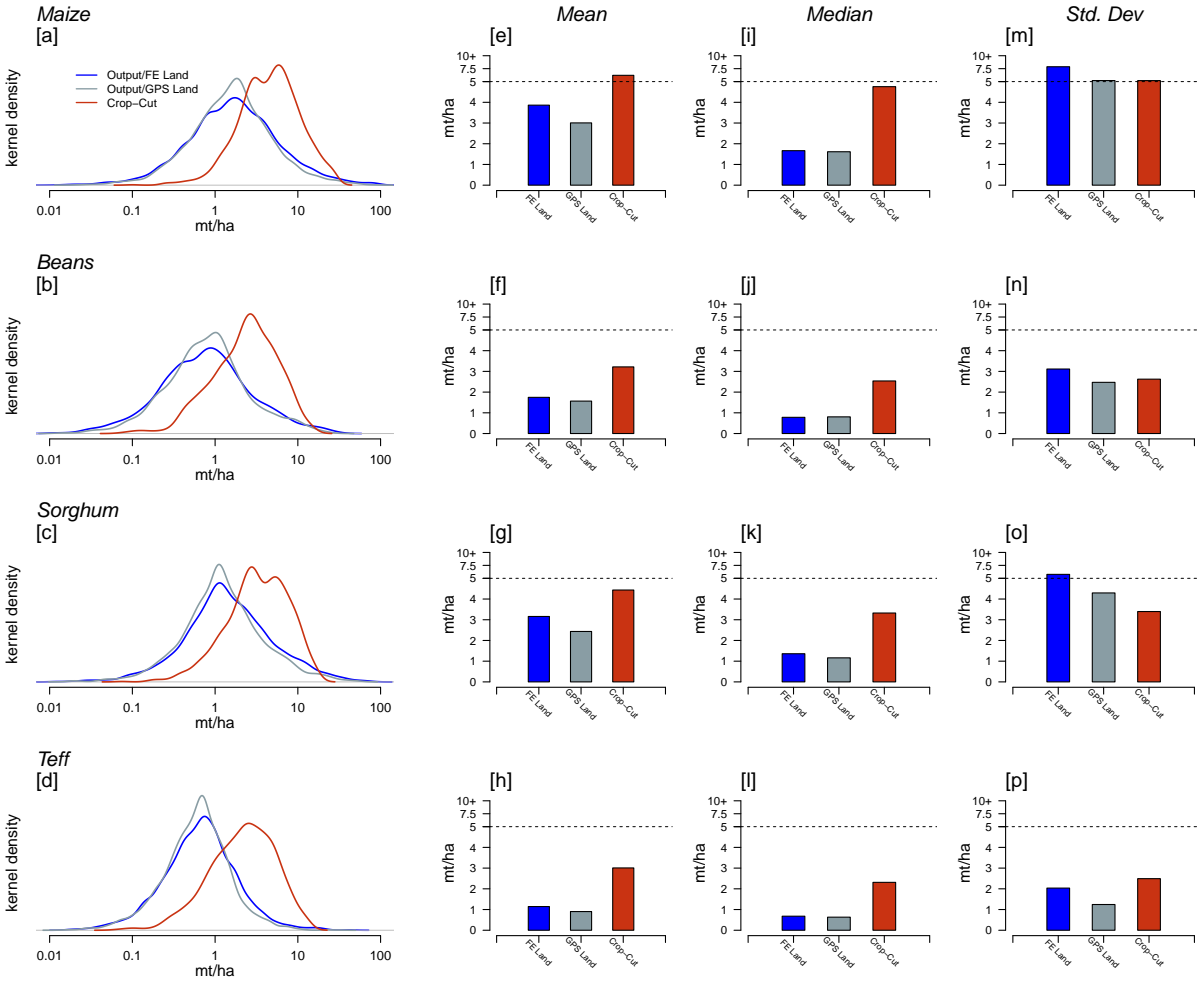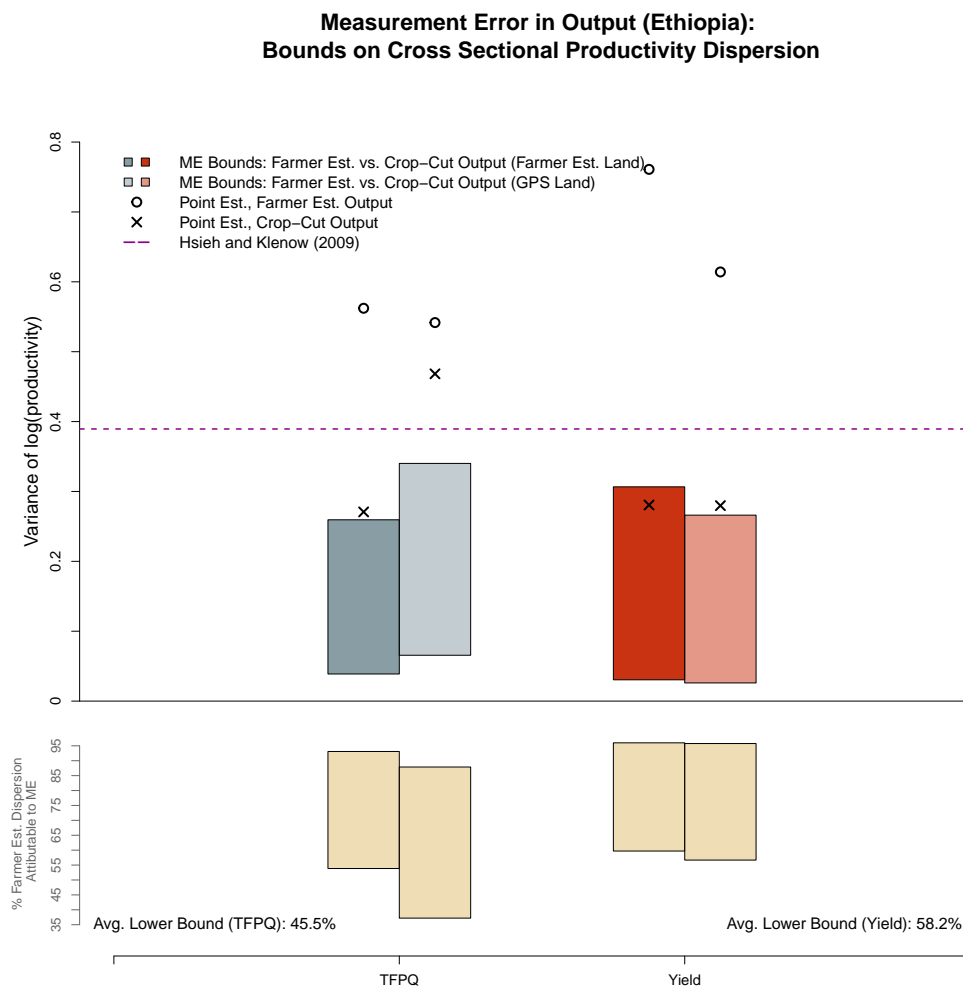
Figure A5: Bounds on cross-sectional productivity dispersion due to measurement error in output. Errors in measures of crop output in the Ethiopian LSMS data significantly magnify measured dispersion in physical productivity, and accounting for measurement error, dispersion appears lower than the benchmark value for non-agricultural firms. Plotted points indicate the variance of the distributions of productivity residuals computed using farmer-estimated (◦'s) and crop-cut based (×'s) measures of crop output individually. Colored polygons bound the variance range between those associated with the distributions of the average (upper bound) and projected (lower bound) composites. For each measure of productivity, points and polygons on the left were calculated using output measures based on farmer-estimated plot size, and those on the right were calculated using output based on GPS-measured plot size. The tan polygons in lower panel indicate the percent of dispersion in productivity computed using farmer-estimated output that is attributable to measurement error implied by the estimates in the top panel. For example, the lower bound of each tan polygon is the difference between the ◦'s and the upper value of the colored bounded area, expressed as percentage of the ◦ values. The dashed magenta line represents the variance of the distribution of TFPR observed among Indian manufacturing firms in Hsieh and Klenow (2009).

Figure A6: Bounds on dispersion due to measurement error in yield. Errors in measures of crop output in our non-LSMS data from Kenya and Uganda significantly magnify measured dispersion in maize yields, similar to the effect observed in Figure 7. Plotted points indicate the variances of the distributions of yields computed using farmer-estimated (○'s) and satellite-based (×'s) measures of crop output individually. Colored polygons bound the variance range between those associated with the distributions of the average (upper bound) and projected (lower bound) composites. The tan polygons in lower panel indicate the percent of dispersion in productivity computed using farmer-estimated yield that is attributable to measurement error implied by the estimates in the top panel. For example, the lower bound of each tan polygon is the difference between the ○'s and the upper value of the colored bounded area, expressed as percentage of the ○ values. The dashed magenta line represents the variance of the distribution of TFPR observed among Indian manufacturing firms in Hsieh and Klenow (2009).

Measurement Error in Land:
Bounds on Cross Sectional Productivity Dispersion
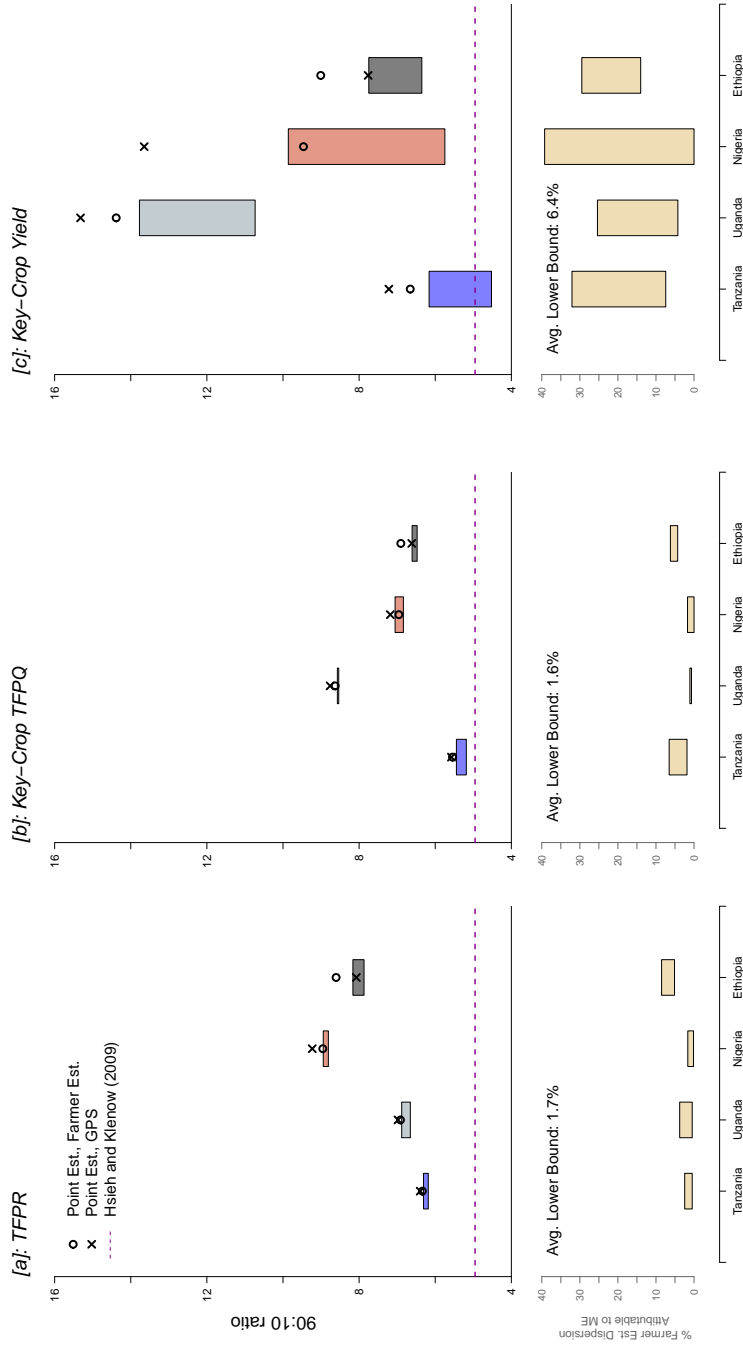
Figure A7: Effect of measurement error in land on dispersion estimates. Measurement errors in plot size have a small effect on measured dispersion in TFPR and TFPQ, and a larger (but still small), effect on measured dispersion in yields. In each panel ([a]-[c]), plotted points indicate the 90:10 ratios of the distributions of productivity residuals computed using farmer-estimated (○'s) and GPS-measured (×'s) plot area individually. Colored polygons bound the range of 90:10 ratios between those associated with the distributions of the average (upper bound) and projected (lower bound) composites. The tan polygons in lower panels indicate the percent of dispersion in productivity computed using farmer-estimated land that is attributable to measurement error implied by the estimates in the top panel. For example, the lower bound of each tan polygon is the difference between the ○'s and the upper value of the colored bounded area, expressed as percentage of the ○ values. As a benchmark, the 90:10 ratio of the distribution of TFPR observed among Indian manufacturing firms in Hsieh and Klenow (2009) is plotted as a dashed magenta line.

59

Figure A8: Effect of measurement error in land on dispersion estimates. Measurement errors in plot size have a small effect on measured dispersion in TFPR and TFPQ, and a larger (but still small), effect on measured dispersion in yields. In each panel ([a]-[c]), plotted points indicate the variances of the distributions of productivity residuals computed using farmer-estimated (○'s) and GPS-measured (×'s) plot area individually. Colored polygons bound the variance range between those associated with the distributions of the average (upper bound) and projected (lower bound) composites. The tan polygons in lower panels indicate the percent of dispersion in productivity computed using farmer-estimated land that is attributable to measurement error implied by the estimates in the top panel. For example, the lower bound of each tan polygon is the difference between the ○'s and the upper value of the colored bounded area, expressed as percentage of the ○ values. As a benchmark, the variance of the distribution of TFPR observed among Indian manufacturing firms in Hsieh and Klenow (2009) is plotted as a dashed magenta line.

A.6 Detailed IV Regression Results

Table A8: IV Persistence, Tanzania (Farmer-Estimated vs. GPS-Measured Land)

	Naive			Reduced-Form			1st-Stage			2nd-Stage		
	TFPR_t^{FE}	TFPQ_t^{FE}	Yield_t^{FE}	TFPR_t^{FE}	TFPQ_t^{FE}	Yield_t^{FE}	TFPR_{t-2}^{FE}	TFPQ_{t-2}^{FE}	Yield_{t-2}^{FE}	TFPR_t^{FE}	TFPQ_t^{FE}	Yield_t^{FE}
TFPR_{t-2}^{FE}	0.143*** (0.014)									0.217*** (0.021)		
TFPQ_{t-2}^{FE}		0.244*** (0.015)									0.310*** (0.024)	
Yield_{t-2}^{FE}			0.151*** (0.016)									1.085 (0.682)
TFPR_{t-2}^{GPS}				0.212*** (0.020)			0.985*** (0.007)					
TFPQ_{t-2}^{GPS}					0.284*** (0.022)			0.924*** (0.008)				
Yield_{t-2}^{GPS}						0.044** (0.021)			0.020 (0.017)			
Observations:	4062	3748	3748	2507	2203	2203	3134	3598	3597	2507	2201	2201
R²:	0.026	0.07	0.023	0.041	0.072	0.002	0.853	0.778	0	0.043	0.074	0.028
Adjusted R²:	0.026	0.07	0.023	0.041	0.072	0.002	0.853	0.778	0	0.043	0.073	0.028
Fixed Effects:	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	None	None	None	Crop-Year	Crop-Year	Crop-Year
Cluster Var:	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village
F statistic:	109.96	281.05	88.58	107.36	171.89	4.57	18154.09	12573.97	1.42	107.61	171.66	2.53
F.statistic (df1, df2):	(1,4060)	(1,3746)	(1,3746)	(1,2505)	(1,2201)	(1,2201)	(1,3132)	(1,3596)	(1,3595)	(1,2505)	(1,2201)	(1,2201)
F.statistic (p-value):	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.033	< 0.001	< 0.001	0.233	< 0.001	< 0.001	0.112

Note:

*p<0.1; **p<0.05; ***p<0.01
Productivity estimated with CropMix and Village-Year fixed effects, and all available geo-variables.
Standard errors clustered at the Village level.

Table A9: IV Persistence, Uganda (Farmer-Estimated vs. GPS-Measured Land)

	Naive			Reduced-Form			1st-Stage			2nd-Stage		
	TFPR$_t^{FE}$	TFPQ$_t^{FE}$	Yield$_t^{FE}$	TFPR$_t^{FE}$	TFPQ$_t^{FE}$	Yield$_t^{FE}$	TFPR$_{t-1}^{FE}$	TFPQ$_{t-1}^{FE}$	Yield$_{t-1}^{FE}$	TFPR$_t^{FE}$	TFPQ$_t^{FE}$	Yield$_t^{FE}$
TFPR$_{t-1}^{FE}$	0.168***									0.186***		
	(0.012)									(0.014)		
TFPQ$_{t-1}^{FE}$		0.198***									0.213***	
		(0.011)									(0.012)	
Yield$_{t-1}^{FE}$			0.085***									7.734
			(0.013)									(27.400)
TFPR$_{t-1}^{GPS}$				0.182***			0.973***					
				(0.013)			(0.004)					
TFPQ$_{t-1}^{GPS}$					0.210***			0.991***				
					(0.012)			(0.002)				
Yield$_{t-1}^{GPS}$						0.028**			0.021**			
						(0.013)			(0.010)			
Observations:	5881	6078	6076	4940	5264	5262	5719	9173	9151	4931	5249	5247
R²:	0.032	0.047	0.007	0.037	0.052	0.001	0.924	0.975	0	0.035	0.052	0.007
Adjusted R²:	0.032	0.047	0.007	0.037	0.052	0.001	0.924	0.975	0	0.035	0.052	0.006
Fixed Effects:	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	None	None	None	Crop-Year	Crop-Year	Crop-Year
Cluster Var:	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village
F statistic:	192.71	299.7	45.23	189.55	290.72	4.87	69725.87	359444.59	4.22	187.4	292.36	0.08
F.statistic (df1, df2):	(1,5879)	(1,6076)	(1,6074)	(1,4938)	(1,5262)	(1,5260)	(1,5717)	(1,9171)	(1,9149)	(1,4938)	(1,5262)	(1,5260)
F.statistic (p-value):	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.027	< 0.001	< 0.001	0.04	< 0.001	< 0.001	0.778

Note:

*p<0.1; **p<0.05; ***p<0.01
Productivity estimated with CropMix and Village-Year fixed effects, and all available geo-variables.
Standard errors clustered at the Village level.

Table A10: IV Persistence, Nigeria (Farmer-Estimated vs. GPS-Measured Land)

	Naive			Reduced-Form			1st-Stage			2nd-Stage		
	TFPR_t^{FE}	TFPQ_t^{FE}	Yield_t^{FE}	TFPR_t^{FE}	TFPQ_t^{FE}	Yield_t^{FE}	TFPR_{t-3}^{FE}	TFPQ_{t-3}^{FE}	Yield_{t-3}^{FE}	TFPR_t^{FE}	TFPQ_t^{FE}	Yield_t^{FE}
TFPR_{t-3}^{FE}	0.092*** (0.019)									0.102*** (0.020)		
TFPQ_{t-3}^{FE}		0.090*** (0.026)									0.096*** (0.028)	
Yield_{t-3}^{FE}			0.085*** (0.026)									2.727 (5.677)
TFPR_{t-3}^{GPS}				0.099*** (0.020)			0.970*** (0.006)					
TFPQ_{t-3}^{GPS}					0.092*** (0.027)			0.956*** (0.005)				
Yield_{t-3}^{GPS}						0.033 (0.023)			0.057*** (0.017)			
Observations:	2058	1386	1386	1962	1333	1333	2371	2623	2605	1962	1331	1331
R²:	0.011	0.009	0.007	0.013	0.009	0.002	0.929	0.929	0.004	0.012	0.009	0.008
Adjusted R²:	0.011	0.008	0.007	0.012	0.008	0.001	0.929	0.929	0.004	0.012	0.008	0.007
Fixed Effects:	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	None	None	None	Crop-Year	Crop-Year	Crop-Year
Cluster Var:	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village
F statistic:	22.94	11.95	10.35	25.53	11.65	2.01	30986.87	34248.64	11.2	25.52	11.47	0.23
F.statistic (df1, df2):	(1,2056)	(1,1384)	(1,1384)	(1,1960)	(1,1331)	(1,1331)	(1,2369)	(1,2621)	(1,2603)	(1,1960)	(1,1331)	(1,1331)
F.statistic (p-value):	< 0.001	0.001	0.001	< 0.001	0.001	0.157	< 0.001	< 0.001	0.001	< 0.001	0.001	0.631

Note:

* p<0.1; ** p<0.05; *** p<0.01
Productivity estimated with CropMix and Village-Year fixed effects, and all available geo-variables.
Standard errors clustered at the Village level.

Table A11: IV Persistence, Ethiopia (Farmer-Estimated vs. GPS-Measured Land)

	Naive			Reduced-Form			1st-Stage			2nd-Stage		
	TFPR$_t^{FE}$	TFPQ$_t^{FE}$	Yield$_t^{FE}$	TFPR$_t^{FE}$	TFPQ$_t^{FE}$	Yield$_t^{FE}$	TFPR$_{t-2}^{FE}$	TFPQ$_{t-2}^{FE}$	Yield$_{t-2}^{FE}$	TFPR$_t^{FE}$	TFPQ$_t^{FE}$	Yield$_t^{FE}$
TFPR$_{t-2}^{FE}$	0.098*** (0.012)									0.102*** (0.013)		
TFPQ$_{t-2}^{FE}$		0.155*** (0.013)									0.152*** (0.015)	
Yield$_{t-2}^{FE}$			0.040*** (0.014)									0.141 (2.342)
TFPR$_{t-2}^{GPS}$				0.099*** (0.013)			0.982*** (0.004)					
TFPQ$_{t-2}^{GPS}$					0.143*** (0.014)			0.953*** (0.005)				
Yield$_{t-2}^{GPS}$						0.002 (0.015)			0.013 (0.012)			
Observations:	4409	5386	5317	4317	5144	5076	4885	7388	7267	4317	5094	5029
R²:	0.014	0.025	0.002	0.014	0.02	0	0.921	0.831	0	0.015	0.025	0.002
Adjusted R²:	0.014	0.025	0.001	0.014	0.02	0	0.921	0.831	0	0.015	0.025	0.002
Fixed Effects:	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	None	None	None	Crop-Year	Crop-Year	Crop-Year
Cluster Var:	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village
F statistic:	63.75	139.69	8.18	62.9	103.33	0.02	57028.29	36210.74	1.19	62.93	105.15	0
F.statistic (df1, df2):	(1,4407)	(1,5384)	(1,5315)	(1,4315)	(1,5142)	(1,5074)	(1,4883)	(1,7386)	(1,7265)	(1,4315)	(1,5142)	(1,5074)
F.statistic (p-value):	< 0.001	< 0.001	0.004	< 0.001	< 0.001	0.887	< 0.001	< 0.001	0.275	< 0.001	< 0.001	0.952

Note:

*p<0.1; **p<0.05; ***p<0.01
Productivity estimated with CropMix and Village-Year fixed effects, and all available geo-variables.
Standard errors clustered at the Village level.

Table A12: IV Persistence, Tanzania (2× Lagged Productivity)

	Naive			Reduced-Form			1st-Stage			2nd-Stage		
	TFPR_t	TFPQ_t	Yield_t	TFPR_t	TFPQ_t	Yield_t	TFPR_{t-2}	TFPQ_{t-2}	Yield_{t-2}	TFPR_t	TFPQ_t	Yield_t
TFPR_{t-2}	0.143*** (0.014)									1.128*** (0.244)		
TFPQ_{t-2}		0.243*** (0.015)									0.704*** (0.103)	
Yield_{t-2}			0.206*** (0.015)									0.687*** (0.119)
TFPR_{t-4}				0.124*** (0.019)			0.115*** (0.018)					
TFPQ_{t-4}					0.166*** (0.021)			0.244*** (0.019)				
Yield_{t-4}						0.143*** (0.021)			0.202*** (0.020)			
Observations:	4062	3748	3748	1669	1461	1461	1775	1652	1652	1501	1170	1170
R^2:	0.026	0.069	0.05	0.026	0.039	0.03	0.024	0.087	0.061	0.02	0.084	0.064
Adjusted R^2:	0.026	0.069	0.05	0.025	0.039	0.029	0.023	0.086	0.06	0.02	0.083	0.063
Fixed Effects:	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	None	None	None	Crop-Year	Crop-Year	Crop-Year
Cluster Var:	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village
F statistic:	110.31	276.8	197.48	44.14	59.98	44.92	43.03	156.25	106.37	21.39	46.29	33.07
F.statistic (df1, df2):	(1,4060)	(1,3746)	(1,3746)	(1,1667)	(1,1459)	(1,1459)	(1,1773)	(1,1650)	(1,1650)	(1,1667)	(1,1459)	(1,1459)
F.statistic (p-value):	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Note:

*p<0.1; **p<0.05; ***p<0.01
Productivity estimated with CropMix and Village-Year fixed effects, and all available geo-variables.
Standard errors clustered at the Village level.

Table A13: IV Persistence, Uganda (2× Lagged Productivity)

	Naive			Reduced-Form			1st-Stage			2nd-Stage		
	$TFPR_t$	$TFPQ_t$	$Yield_t$	$TFPR_t$	$TFPQ_t$	$Yield_t$	$TFPR_{t-1}$	$TFPQ_{t-1}$	$Yield_{t-1}$	$TFPR_t$	$TFPQ_t$	$Yield_t$
$TFPR_{t-1}$	0.168*** (0.012)									0.897*** (0.136)		
$TFPQ_{t-1}$		0.198*** (0.011)									0.924*** (0.123)	
$Yield_{t-1}$			0.145*** (0.012)									0.691*** (0.132)
$TFPR_{t-2}$				0.153*** (0.017)			0.163*** (0.016)					
$TFPQ_{t-2}$					0.171*** (0.015)			0.231*** (0.017)				
$Yield_{t-2}$						0.107*** (0.016)			0.187*** (0.018)			
Observations:	5881	6078	6078	2695	2676	2676	2659	2722	2722	2379	1795	1795
R^2:	0.032	0.047	0.025	0.03	0.044	0.016	0.036	0.063	0.037	0.03	0.046	0.018
Adjusted R^2:	0.032	0.047	0.024	0.03	0.044	0.016	0.036	0.062	0.037	0.03	0.046	0.017
Fixed Effects:	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	None	None	None	Crop-Year	Crop-Year	Crop-Year
Cluster Var:	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village
F statistic:	192.72	299.71	153.07	83.01	123.88	43.62	100.01	181.72	104.32	43.78	56.09	27.25
F.statistic (df1, df2):	(1,5879)	(1,6076)	(1,6076)	(1,2693)	(1,2674)	(1,2674)	(1,2657)	(1,2720)	(1,2720)	(1,2657)	(1,2674)	(1,2674)
F.statistic (p-value):	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Note:

*p<0.1; **p<0.05; ***p<0.01
Productivity estimated with CropMix and Village-Year fixed effects, and all available geo-variables.
Standard errors clustered at the Village level.

Table A14: IV Persistence, Ethiopia (2× Lagged Productivity)

	Naive			Reduced-Form			1st-Stage			2nd-Stage		
	TFPR_t	TFPQ_t	Yield_t	TFPR_t	TFPQ_t	Yield_t	TFPR_{t-2}	TFPQ_{t-2}	Yield_{t-2}	TFPR_t	TFPQ_t	Yield_t
TFPR_{t-2}	0.098***									1.184***		
	(0.012)									(0.329)		
TFPQ_{t-2}		0.155***									0.871**	
		(0.013)									(0.384)	
Yield_{t-2}			0.079***									0.674
			(0.013)									(0.960)
TFPR_{t-4}				0.069***			0.069***					
				(0.015)			(0.014)					
TFPQ_{t-4}					0.057***			0.075***				
					(0.021)			(0.019)				
Yield_{t-4}						0.018			0.020			
						(0.020)			(0.019)			
Observations:	4409	5386	5386	1931	1541	1541	2061	1779	1779	1833	1339	1339
R^2:	0.014	0.025	0.007	0.011	0.005	0	0.012	0.008	0.001	0.017	0.051	0.014
Adjusted R^2:	0.014	0.025	0.007	0.01	0.004	0	0.011	0.008	0	0.016	0.05	0.013
Fixed Effects:	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	Crop-Year	None	None	None	Crop-Year	Crop-Year	Crop-Year
Cluster Var:	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village	Village
F statistic:	63.75	139.74	36.27	20.93	7.34	0.76	24.08	14.94	1.1	12.96	5.16	0.49
F.statistic (df1, df2):	(1,4407)	(1,5384)	(1,5384)	(1,1929)	(1,1539)	(1,1539)	(1,2059)	(1,1777)	(1,1777)	(1,1929)	(1,1539)	(1,1539)
F.statistic (p-value):	< 0.001	< 0.001	< 0.001	< 0.001	0.007	0.383	< 0.001	< 0.001	0.294	< 0.001	0.023	0.482

Note:

*p<0.1; **p<0.05; ***p<0.01
Productivity estimated with CropMix and Village-Year fixed effects, and all available geo-variables.
Standard errors clustered at the Village level.

Table A15: IV Persistence, Ethiopia (Farmer-Estimated vs. Crop-Cut Output)

	Naive		Reduced-Form		1st-Stage		2nd-Stage	
	TFPQ^{FE}_t	Yield^{FE}_t	TFPQ^{FE}_t	Yield^{FE}_t	TFPQ^{FE}_{t-2}	Yield^{FE}_{t-2}	TFPQ^{FE}_t	Yield^{FE}_t
TFPQ^{FE}_{t-2}	0.155***						0.382***	
	(0.013)						(0.068)	
Yield^{FE}_{t-2}		−0.021						−1.903
		(0.014)						(6.550)
$\text{TFPQ}^{Cropcut}_{t-2}$			0.176***		0.469***			
			(0.030)		(0.025)			
$\text{Yield}^{Cropcut}_{t-2}$				−0.027		0.001		
				(0.037)		(0.032)		
Observations:	5386	5304	2347	1866	3521	2680	2347	1851
R^2:	0.025	0	0.014	0	0.088	0	0.022	0
Adjusted R^2:	0.025	0	0.014	0	0.088	0	0.021	0
Fixed Effects:	Crop-Year	Crop-Year	Crop-Year	Crop-Year	None	None	Crop-Year	Crop-Year
Cluster Var:	Village	Village	Village	Village	Village	Village	Village	Village
F statistic:	139.69	2.33	33.3	0.54	339	0	31.48	0.08
F.statistic (df1, df2):	(1,5384)	(1,5302)	(1,2345)	(1,1864)	(1,3519)	(1,2678)	(1,2345)	(1,1864)
F.statistic (p-value):	< 0.001	0.127	< 0.001	0.463	< 0.001	0.972	< 0.001	0.771

Note:

*p<0.1; **p<0.05; ***p<0.01
Productivity estimated with CropMix and Village-Year fixed effects, and all available geo-variables.
Standard errors clustered at the Village level.