May 2020

Heterogeneity, Measurement Error, and Misallocation:

Evidence from African Agriculture

Douglas Gollin
Oxford University

Christopher Udry[*]
Northwestern University

ABSTRACT

Standard measures of productivity display enormous dispersion across farms in Africa. Crop yields and input intensities appear to vary greatly, seemingly in conflict with a model of efficient allocation across farms. In this paper, we present a theoretical framework for distinguishing between measurement error, unobserved heterogeneity, and potential misallocation. Using rich panel data from farms in Tanzania and Uganda, we estimate our model using a flexible specification in which we allow for several kinds of measurement error and heterogeneity. We find that measurement error and heterogeneity together account for a large fraction of the dispersion in measured productivity.

---

1. Introduction

How important is misallocation in explaining the income differences across countries? A recent literature in development and growth economics has focused on misallocation across sectors, firms, and plants.[1] This literature has found evidence that the dispersion of total factor productivity (TFP) across production units seems to be consistently higher in poor countries than in rich ones. Such productivity differences have the potential to account for a large fraction of the cross-country income differences. In an aggregate sense, misallocation across sectors or firms can significantly reduce aggregate TFP.

A challenge in this literature is to distinguish misallocation from other sources of dispersion in productivity, such as technology shocks, measurement error, and adjustment costs of various kinds. Several recent papers have taken up this issue in relation to data from the manufacturing sector; e.g., Bils et al (2017), Haltiwanger et al. (2018), Pellegrino and Zheng (2018), Rotemberg and White (2017), White et al. (2018). These papers all point out that measurement error can lead to problems in identifying the extent and severity of misallocation.

In this paper, we seek to disentangle these different sources of productivity dispersion in an environment where measured cross-firm dispersion is very large, aggregate productivity is low, and market failures undoubtedly contribute to cross-firm frictions in the allocation of resources. Specifically, we take advantage of extraordinarily rich data from farms in two countries in Africa, for which we have detailed panel observations on individual

---

[1] See, for example, Hsieh and Klenow 2009; Restuccia and Rogerson 2008, 2013; McMillan et al. 2014; Porzio 2016; Bento and Restuccia 2017; Hicks et al. 2017; Restuccia 2018.

farms. Many of these farms produce identical and homogeneous outputs on different plots within each growing season. This allows us to observe within-season variation across plots for a given farmer in the input intensity and output of the same crop. We cannot interpret this variation as the result of misallocation, since farmers presumably face no market imperfections in allocating resources across their own plots. As a result, these data allow us to identify and quantify misallocation more precisely. Our strategy allows us to disentangle the productivity dispersion that arises from misallocation from that stemming from measurement error or heterogeneity in technology and inputs (including production shocks).

The agricultural sector provides a valuable window through which to study firm-level misallocation. Most firm surveys have relatively few observations on different plants or establishments operated by the same firm, and this makes it difficult to disentangle firm management from any unobservable characteristics of the plant or factory; in contrast, we observe farmers concurrently operating multiple plots.[2] Another advantage we have, relative to firm surveys, is that our firms are producing highly homogeneous products, with little market power. Consequently, we can compare the output of different firms (farms) without worrying about

---

[2] Recent literature offers several interesting exceptions, with a number of papers looking at productivity differences across divisions of industrial conglomerates or across plants of manufacturing firms; e.g., Maksimovic and Phillips (2002), Matvos and Seru (2014), Ševčík (2015). However, these comparisons are complicated by the difficulty of comparing productivity differences across different outputs, especially in an environment with imperfect competition, where firms have some market power in setting prices. An interesting paper, directly relevant for ours, is Kehrig and Vincent (2019), which looks at firms that operate multiple plants (establishments) producing the same good. Because of lumpiness in investment at the plant level, Kehrig and Vincent point out that efficient internal allocation of capital may lead firms not to equalize revenue measures of productivity (TFPR, to use the Hsieh-Klenow terminology) across plants. The authors revisit the cross-country comparisons of Hsieh and Klenow (2009) and estimate the gains from reallocation after accounting for the role of multi-plant firms.

mark-ups and pricing strategies. Thus, in our context, there is no need to distinguish between revenue-based TFP measures and quantity-based TFP measures. To use the terminology of Hsieh and Klenow (2009), in our setting, TFPR=TFPQ=TFP.

The agricultural sector is also an interesting context for studying misallocation because of evidence that low agricultural productivity can explain – at least in a mechanical sense – a large fraction of the cross-country dispersion of output per worker (Caselli 2005, Restuccia et al. 2008, Restuccia and Rogerson 2017). A cluster of recent papers has suggested that there may be very large dispersion in productivity at the level of farms and farmers, potentially indicative of misallocation.[3] These papers point out that in poor economies, large fractions of the workforce are employed in agriculture. In economies where two-thirds of the people are farmers, such as those of sub-Saharan Africa, it is reasonable to ask whether they are all good at farming – and whether market failures of various kinds may induce too many low-skill farmers to remain in agriculture.

Restuccia and Santaeulalia-Llopis (2017), in particular, have raised the intriguing possibility that much of Africa's productivity deficit might be attributable to misallocation within the agricultural sector. They find suggestive evidence, in data from Malawi, that too much farmland is managed by low-skill farmers. If true, this finding might offer an explanation for sub-Saharan Africa's low productivity in agriculture. By extension, misallocation might thus help explain the region's low levels of income per capita. The finding also suggests a relatively straightforward

---

[3] See, for example, Adamopoulos and Restuccia 2014, Adamopoulos and Restuccia 2015, Adamopoulos et al. 2017, Bento and Restuccia 2017, Restuccia and Santaeulalia-Llopis 2017.

solution – albeit one with great political complexity – namely, the liberalization of land and input markets, so that the best farmers can eventually buy out those farmers who lack the skill to farm productively. Restuccia and Santaeulalia-Llopis calculate that reassigning land from low productivity establishments to high productivity establishments (or equivalently from bad farmers to good farmers) could result in more than a three-fold increase in aggregate agricultural output.

A number of other studies explore the potential importance of misallocation across farms in the developing world; e.g., Adamopoulos et al. (2017), Chen (2016), Emran and Shilpe (2015), Gottlieb and Grobovsek (2018), Shenoy (2017), and Foster and Rosenzweig (2017). These papers build on a broader literature that has looked at productivity differences outside agriculture – typically across firms and plants in advanced countries. Key works include Syverson (2004, 2011), Petrin et al. (2011), Petrin and Sivadasin (2013).

A broader literature in macro and growth economics emphasizes the importance of misallocation for aggregate productivity; e.g., Banerjee and Moll (2010), Bento and Restuccia (2017), Da-Rocha et al. (2017), Garcia-Santana and Pijoan-Mas (2014), Guner et al. (2008), Hopenhayn (2014), Hsieh and Klenow (2009), Kalemli-Ozcan and Sorensen (2012), Midrigan and Xu (2014), Restuccia and Rogerson (2008, 2013). A recurring theme in this literature is that the misallocation of productive resources into low productivity firms can lead to low aggregate productivity. Empirical analysis generally supports the idea that poor countries have many firms with low measured TFP. The reasons for the persistence of these low productivity firms are not always clear, but a sufficient explanation would be frictions or policies that induce distortions to the efficient size distribution of firms.

A challenge to all this literature is the measurement of productivity at the level of individual firms. Typically, the data used for these analyses come from firm surveys that vary in quality and in coverage. To calculate measures of productivity for the individual firm requires a series of strong assumptions about the firm-level production function and about the quality of data. In particular, methods used widely in the macro literature on misallocation have been criticized on methodological grounds; e.g., by Asker et al. (2014), Foster et al. (2016), and Haltiwanger (2016). Our approach addresses some of the concerns raised by these critiques. In particular, our approach recognizes that firm-level productivity may vary for many reasons other than misallocation.

Our paper makes use of panel data from two countries (Tanzania and Uganda) for which we can observe production in great detail. In these data, we can observe the inputs and outputs for specific crops cultivated by individual farmers – not simply households -- on specific plots of land. The data are similar to those used by Restuccia and Santaeulalia-Llopis (2017). The rich detail of the data allows us to disentangle misallocation from three other important sources of variation in measured productivity at the farm level. The first of these is simply the stochastic nature of agricultural production. Farmers face large shocks to production that are not well observed in the data, related to weather, pests, crop diseases, and so on. A second source of variation in productivity is measurement error; in spite of the high quality of the data that we work with, reporting is imperfect and measurement is imprecise.[4] Finally, the third source of variation in productivity is heterogeneity in unobserved land quality.[5] All

[4] See, for example, De Nicola and Giné 2014, Deininger et al. 2012, and Beegle et al. 2012b; although Beegle et al. 2012a offer a more positive view.

[5] The problem of unobserved land quality was recognized by Benjamin 1995 and Udry 1996. More recent surveys often collect quite detailed data on soil quality, but the

will give rise to dispersion in measured total factor productivity (TFP) at the farm level, as well as to dispersion in input intensity. Because of this, any estimates of the potential gains from reallocation need to account carefully for production shocks, mismeasurement and heterogeneity.

In this paper, we propose a theoretical framework that models the processes by which farmers select plots, allocate inputs to individual plots, and subsequently realize output. Our theoretical framework explicitly recognizes the stochastic nature of agricultural production and the sequencing of farm decision-making. We then show how this model can help distinguish empirically between misallocation, mismeasurement, and heterogeneity, given plot-level data.

Drawing on the model, we assess the relative importance of different sources of dispersion in measured productivity. Our results suggest that idiosyncratic shocks, measurement error, and heterogeneity in land quality are important sources of dispersion in measured productivity across farms. We find that when these are taken into account, the potential significance of misallocation drops substantially. Late-season production shocks, measurement error, and heterogeneity in inputs together account for as much as seventy percent of the variance in measured productivity.[6] Since these are not susceptible to reallocation, our estimates for the aggregate

---

dimensionality of soil quality measurement can be overwhelming; see, for example, Tittonell et al. 2008.

[6] By "late-season" shocks, we mean those shocks that affect production *after* the farmer has made most or all of her input choices. We implicitly (and realistically) assume a production process in which significant amounts of labor and other inputs are applied early in the season for land clearing and planting; and then additional inputs are applied during the growing season based on observed growing conditions, market prices, etc. Late-season shocks might correspond to weather, pest, or disease shocks that happen sufficiently late in the growing season that farmers cannot effectively respond to them.

productivity gains that could be attained from a reallocation exercise are correspondingly smaller. Our results suggest that efficient reallocation of land and other agricultural inputs would not dramatically close the income gaps between African countries and the world's rich economies.

An important caveat of our work is that we consider only the effects of static misallocation. Implicitly, this holds constant the existing institutions and technologies. With improved technologies and different institutions, one might expect that the efficient allocation of land and inputs across farms and farmers would look very different. In this sense, our results are not necessarily inconsistent with those of Adamopoulos and Restuccia (2014), who ask how agricultural production would change if all countries had the same size distribution of farms that is observed in the United States. Our data include no observations on farms of this size, making it impossible for us to discipline estimates of such a dramatic change in farm size.

The remainder of this paper proceeds as follows. Section 2 provides some descriptive background and characterizes the dispersion of partial productivity measures (output per unit land and labor per unit land) across farms in our data. In Section 3, we construct a theoretical framework that models the ways in which farmers choose their plots, select crops, apply inputs, and realize output. In Section 4, we use this model to motivate the estimation of agricultural production functions for our two countries. Working with these estimates, we show that the measured dispersion of TFP depends on how we control for heterogeneity and measurement error. This matters in turn for our understanding of the importance of misallocation as a cause of low aggregate productivity. Section 5 discusses these results, and Section 6 concludes.

2. Dispersion in productivity across and within farms

Our paper draws on two nationally representative multi-year panel data sets, for Tanzania and Uganda. These data were collected by government statistical agencies in collaboration with the World Bank's program on Living Standards Measurement Surveys – Integrated Surveys of Agriculture (LSMS-ISA). Both surveys collected data on all plots cultivated by the household. For each plot, the survey identifies the individual or individuals within the household who farm the plot. Detailed information was collected at the plot level on inputs used and output harvested. Depending on the survey, some or all plots were measured by GPS, and data were collected using state-of-the-art survey techniques. The data are freely available online, and all data and documentation are available for open access.[7]

The survey data include detailed descriptors of both the households and the farms. For households, data are available on household composition and the age, education, and health characteristics of each household member; the relationship of each member to the household head; and the allocation of each person's time to household production and market labor, among many other variables. For the farm, data were collected at the plot level on crops cultivated, soil characteristics, toposequence, location, soil quality (including measures of erosion and tree cover), land rights, and a variety of observed shocks, including rainfall.

An important feature of our data – and one that helps us significantly in terms of our identification strategy – is that we have many instances in each country in which we observe the same farmer cultivating the same

---

[7] For information on the LSMS-ISA project and links to the data, see:
http://go.worldbank.org/BCLXW38HY0.

crop on multiple plots within the same season. For instance, we may observe a single farmer growing maize on each of two or three distinct plots in the same growing season.[8]

The plot is the basic unit of farm operations, and we define it as a contiguous area on which a specific crop (or crop mixture) is grown by a particular farmer. Consistent with the definitions in the data, we assume that there is approximate uniformity within a plot in the timing of farm operations and the application of inputs.[9]

Tanzania and Uganda differ to some degree in the types of production systems that we observe. Some crops are common to both countries (e.g., maize), while others (e.g., *matoke*, a kind of cooking banana) are of importance only in a single country (in this case, Uganda). For most purposes, however, the two countries are quite similar in the farming systems and production environment. Key points to note are that these are smallholder farming systems that use few inputs other than human labor and hand tools. Almost none of the farms in our data use irrigation or machinery; commercial fertilizer or other agrochemicals are each used

---

[8] For convenience, we speak of "a farmer" as an individual. But our data sets actually distinguish the person who owns the land from the person who manages the plot and the person who keeps most of the revenue from the plot. Where these differ, we define the farmer as the person who manages the plot. An added level of complexity is that the data often allow for up to two household members to be designated as the manager of the plot. We use the term "farmer" to refer to distinct individuals or pairs of household members. When we speak of a farmer cultivating the same crop on different plots, it could thus be a husband and wife (or father and son, or two brothers, etc.) operating as a pair. (This is the case for about 50% of the plots in both Tanzania and Uganda).

[9] This corresponds precisely to the definition of a plot in the Uganda data. The Tanzanian data uses different terminology but is consistent with this interpretation as well. All of the plots in Uganda and 96 percent of Tanzanian plots are cultivated with a single crop, or with a single intercropped mixture of crops. Tanzanian measured plot sizes are adjusted by the share of the plot area reported as cultivated; Ugandan plot areas are defined as cultivated area.

on less than ten percent of Tanzanian plots and two percent of Ugandan plots. In Uganda, most farmers cultivate crops in two growing seasons per year; in Tanzania, our data primarily reflect cultivation in the main growing season.[10]

Tables 1A and 1B show key descriptive statistics for our two data sets. As Table 1A shows, within some households, there are multiple farmers. Individual plots are quite small, with a median plot size of 0.20 ha in Uganda and 0.40 ha in Tanzania. The majority of farmers cultivate multiple plots within each season.

Table 1B shows yields (output per hectare) for each of the data sets. These are given in value terms because of the prevalence of intercropping (i.e., several crops being cultivated at the same time on a given piece of land). Intercropping makes it difficult (or irrelevant) to measure yield in physical quantities. Instead, we report value per hectare, with the physical quantities of different crops priced using median values reported by all farmers in a community.[11] We define labor input to be all forms of labor (hired and family labor) used on the plot.

---

[10] Tanzania also has a minor growing season, and the data for crops cultivated during this season are somewhat inconsistently handled in the survey. We focus on the main growing season, although the data for this season appear to include some inputs and outputs that should properly be attributed to the minor season. All our results are robust to aggregation to the full year.

[11] Although we make use of price data in valuing joint outputs of specific plots, we note that we are using prices here simply as aggregation weights. We do not need to worry about the endogeneity of prices to plot-level or farm-level input choices, because the marketable surplus of any single farmer is so small.

It is immediately apparent from the yield data that reported yields are wildly skewed. The mean yield is typically around twice the median, and the large standard deviations are indicative of very long right-hand tails of the distributions. This is true even after the data have been winsorized at the 0.01 level. Because there are biophysical constraints on maximum yield, we look skeptically at some of the very high reported values of yield in these data, and we view this as *prima facie* evidence that measurement error is likely to be an important feature of the data.[12]

2.1 Efficient static allocation

As a benchmark, we consider an efficient static allocation of inputs across plots that are homogeneous in quality. For any efficient allocation, by the second welfare theorem, there will be shadow prices common to all farmers such that profits are maximized on each plot, and factor marginal value products are equalized across all plots. This will be true even if farmers differ in ability. For example, suppose the physical production function for crop output Y_{hit} of plot i of farmer h in season t, using land L_{hit} and labor X_{hit} is:

$$Y_{hit} = e^{\omega_h}(L_{hit})^{\alpha_L}(X_{hit})^{\alpha_X}, \qquad (1)$$

with ω_h the total factor productivity of farmer h.[13] The input choices on that plot maximize profits at shadow prices $(p_{Yhit}, p_{Lhit}, p_{Xhit})$. In an efficient allocation these shadow prices are common across all farmers. In familiar Cobb-Douglas fashion, factor demands and output are characterized by:

---

[12] We note that these LSMS-ISA data sets rely on farmer self-reporting of yield, which may be one source of measurement error, as suggested by Gourlay et al. (2017). Enumerator error and data entry mistakes may also be present, even with the most diligent efforts at quality control.

[13] The production function (1) is a gross simplification, of course. Functional form aside, it abstracts from the multi-stage process of farming, treating labor (and land) inputs over each farming season as undifferentiated.

$$x_{hit} - l_{hit} - (\ln(\alpha_X) - \ln(\alpha_L)) = \ln\left(\frac{p_{Lhit}}{p_{Xhit}}\right) = \ln\left(\frac{p_L}{p_X}\right);$$

$$y_{hit} - l_{hit} + \ln(\alpha_L) = \ln\left(\frac{p_{Lhit}}{p_{Yhit}}\right) = \ln\left(\frac{p_L}{P_Y}\right)$$

$$y_{hit} - y_{kit} = \frac{(\omega_h - \omega_k) + \alpha_L \ln\left(\frac{p_{Yhit}}{p_{Lhit}}\right) + \alpha_X \ln\left(\frac{p_{Yhit}}{p_{Xhit}}\right)}{1 - \alpha_L - \alpha_X} \qquad (2)$$

$$= \frac{(\omega_h - \omega_k) + \alpha_L \ln\left(\frac{p_Y}{p_L}\right) + \alpha_X \ln\left(\frac{p_Y}{p_X}\right)}{1 - \alpha_L - \alpha_X}$$

(As is standard, lower case variables denote log values; l_{hit} indicates the log of land area for plot i of farmer h in season t). Efficiency requires that the scale of production should vary across farmers according to their productivity, with that variation limited by decreasing returns to scale. However, factor intensity ratios and the value of output per hectare would be identical across all plots planted with the same crop at the same date. In this riskless world with perfect measurement, any variation across plots in factor ratios or the value of output per hectare would reflect misallocation.[14]

This description does not characterize the world particularly well, and our data from Tanzania and Uganda show marked deviation from this benchmark. There is wide dispersion in factor ratios across plots as well as in realized output per unit land.

Figure 1 shows, in two subgraphs (for Tanzania and Uganda respectively), Epanechnikov kernel estimates of the density of the plot-level deviation of log output per hectare from its sample mean. The different lines on the Figure correspond to dispersions calculated with differing controls. Figure 2 similarly illustrates the plot-level density of the deviation of log labor per hectare in each country.

---

[14]This conclusion is general to any homothetic production function.

Consider first the raw dispersions across plots of output per hectare in Figure 1 and labor per hectare in Figure 2. The variances of log output per hectare are 1.49 for Tanzania and 1.82 for Uganda. The corresponding variance of log labor input per hectare is 1.06 in Tanzania and 1.00 in Uganda. It is noteworthy that the variance of log labor input is quite high; yield dispersion is not coming entirely from shocks affecting final harvest.[15]

The raw data on output and input do not account for variation in observable heterogeneity across plots. Land characteristics such as slope, soil type, and location affect farmers' optimal allocations of inputs and their expected yields. These land characteristics are measured in each of our data sets. Characteristics of the farmer such as gender, education, and experience are also components of productivity that we observe. Moreover, agriculture in each of our settings is almost exclusively rainfed. Rainfall thus affects plot productivity both by affecting the overall level of plot productivity and through unanticipated shocks to output. The data include measurements of rainfall totals at the community level; we condition on measures of rainfall and their interactions with land characteristics as well. If these observed characteristics fully account for the variation in productivity, then in an efficient allocation output per hectare and labor per hectare would not vary across plots, once we control for observables.

To do so, land and labor inputs to production ($J \in \{L, X\}$) are modeled as the observed quantity of that factor (J^o_{hit}), observed as hectares or days of input J on plot i of household h in season t, corrected for a factor-specific set of observables (W_{Jhit}) $J_{hit} = J^o_{hit} e^{W_{Jhit}\beta_J}$. (In essence, this is defining "effective labor units" and "effective land units," based on observables.)

---

[15] As a different measure of dispersion, consider the 90-10 ratio of output per hectare (labour per hectare); i.e., the 90th percentile of output divided by the 10th percentile. These numbers are 16.00 (13.00) for Tanzania, and 24.13(11.50) for Uganda.

Let W_{Yhit} be a set of observable determinants of total factor productivity, including detailed measures of weather shocks realized after factor inputs are chosen, so that the production function now becomes:

$$Y_{hit} = e^{W_{Yhit}\beta_Y + \omega_{Yh}}\left(L_{hit}^o e^{W_{Lhit}\beta_L}\right)^{\alpha_L}\left(X_{hit}^o e^{W_{Xhit}\beta_X}\right)^{\alpha_X}. \qquad (1')$$

Equations (2) continue to hold, with observed factor inputs and observed output adjusted for these observed determinants of productivity:

$$x_{hit}^o + W_{Xhit}\beta_X - l_{hit}^o - W_{Lhit}\beta_L - (\ln(\alpha_X) - \ln(\alpha_L)) = \ln\left(\frac{p_{Lhit}}{p_{Xhit}}\right)$$

$$= \ln\left(\frac{p_L}{p_X}\right);$$

$$y_{hit} + W_{Yhit}\beta_Y - l_{hit} - W_{Lhit}\beta_L + \ln(\alpha_L) = \ln\left(\frac{p_{Lhit}}{p_{Yhit}}\right) = \ln\left(\frac{p_L}{p_Y}\right)$$

$$y_{hit} - y_{kit} + (W_{Yhit} - W_{Ykit})\beta_Y \qquad (2')$$

$$= \frac{(\omega_h - \omega_k) + \alpha_L \ln\left(\frac{p_{Yhit}}{p_{Lhit}}\right) + \alpha_X \ln\left(\frac{p_{Yhit}}{p_{Xhit}}\right)}{1 - \alpha_L - \alpha_X}$$

$$= \frac{(\omega_h - \omega_k) + \alpha_L \ln\left(\frac{p_Y}{p_L}\right) + \alpha_X \ln\left(\frac{p_Y}{p_X}\right)}{1 - \alpha_L - \alpha_X}$$

Tables 2A and 2B report a set of regressions for each country, with output per hectare as the dependent variable in all regressions. Observations are for individual plots in specific years/seasons. In each of these tables, the first column shows selected coefficients from a regression of output per hectare on cultivated area and the large set of observable land characteristics and exogenous shocks that are available in these datasets. The estimated density of the residuals from these regressions is illustrated as the red dashed line in each of the subgraphs of Figure 1. The plot characteristics and shock variables are highly jointly significant in each regression, and the estimated variance of the residuals is significantly smaller than the variance of the raw data in each case. This tells us that the observable plot characteristics are indeed explaining part of the dispersion in yield. Nevertheless, as is apparent from Figure 1, including

these observable plot characteristics does not alter the overall pattern of dispersion in productivity.

The first column of Table 3A and Table 3B reports the same subset of coefficients of the parallel regression of labor input per hectare on the observable exogenous shocks, farmer characteristics, and land characteristics. The estimated density of the residuals from these regressions is the red dashed line in each of the subgraphs of Figure 2. Again, the set of observable characteristics is highly jointly significant in each regression, and the estimated variance of the residuals is significantly smaller that of the raw data. The variation in observable characteristics, including shocks, is an important determinant of the variation in both output and labor input per hectare across plots in each of these samples. But again, these observables do not generate much difference in the pattern of residuals as shown in Figure 2.

Differences in technology across farming systems and crops will presumably affect yield and input intensity, even in an efficient allocation. Similarly, variation over time in the shadow costs of factors of production or the shadow value of output could generate time variation in output or labor per hectare. Therefore, Column (2) in each of the Tables 2A-B and 3A-B reports coefficients from regressions of log output per hectare and log labor per hectare on the same set of plot characteristics with year-season-region-crop fixed effects. Estimates of the density of the residuals from these regressions are the blue dashed lines in each of the subgraphs of Figures 1 and 2.

Qualitatively speaking, these tables provide evidence that observable characteristics of plots and shocks have a statistically meaningful effect on input intensity and yield. However, we note that *quantitatively* speaking,

these observables do not account for a very large fraction of the total dispersion.

One way to see this is to note that the magnitude of the remaining variation is large: the log variance of the residuals is 1.15 in Tanzania and 1.42 in Uganda. In comparison, the variance of log output per hectare for farms in the United States is 0.05 for corn in the Corn Belt and 0.14 for wheat in the Northern Plains (Claassen and Just 2011).[16] The variance of log labor per hectare also remains substantial: it is 0.65 in Tanzania and 0.48 in Uganda.

It is apparent that substantial variation in output per hectare and labor per hectare remains after we account for a rich set of observable characteristics of land, including detailed measures of rainfall variation. The variance remains large even when we add year-season-crop-region fixed effects (Column 2). This remaining variation is sometimes characterized as reflecting the effects of factor and output market distortions that prevent the efficient match of factor inputs to dispersion in total factor productivity (Hsieh and Klenow 2009; Adamopoulos et al., 2017; Restuccia and Santaeulalia-Llopis 2017). For this reason, we will refer to the estimated residuals from the regressions based on equations (2') and reported in columns (2) of Tables 2A-B and 3A-B as our baseline measures of dispersion in productivity across plots.[17]

---

[16] Claassen and Just (2011) report that for more than 500,000 observations in their US data, the 95th percentile corn yield is 190% higher than the 5th percentile, which they view as "quite wide" (p. 148). By contrast, we find 95-5 ratios for Uganda of 9304%, and for Tanzania of 2558%. This reinforces our perception that the dispersion of yield across plots is quite high.

[17] An alternative baseline could be provided by examining the residuals from a similar regression with village-crop-year fixed effects. This would absorb the effects of unobserved village-level shocks which might otherwise be misinterpreted as misallocation, but it would also absorb any real misallocation of resources across

The variation remains substantial as we move from the baseline specification to tighter specifications, adding fixed effects at progressively narrower geographic units. In Tables 2A-B and 3A-B, Column 3 adds village fixed effects. The dispersion falls with successively narrower fixed effects, but it remains non-trivial.

This baseline dispersion could be attributed to misallocation. However, even if factors of production are allocated with full efficiency, output shocks and measurement error generate output dispersion across farms. Consider three components of total factor productivity. The first is the set of observable characteristics of the plot, farmer, or community, W_{Yhit}, which may include both permanent and transitory components. These transitory components may be realized either before or after factor inputs are chosen for the plot. The second is a component that is unobserved in the data but known to the farmer at the time factor inputs are chosen, ω_{Yhit}. Finally, there is a component that is unobserved in the data and unknown to the farmer at the time of input application, ϵ_{Yhit}. This final component could be an actual output shock that is realized late in the season, or it could be pure measurement error in output. (From the production function alone, these cannot be distinguished.)

$$\omega_{hit} = W_{Yhit}\beta_Y + \omega_{Yhit} + \epsilon_{Yhit} \qquad (3)$$

Baseline dispersion in factor ratios could reflect *unobserved* characteristics of land or labor, or measurement error in either input. Land and labor inputs to production ($J \in \{L, X\}$) are modeled as the observed quantity of that factor (J_{hit}^o), observed as hectares or days of input J on plot i of

---

villages. As can be seen in Figures 1 and 2, the estimated dispersion of the residuals from these two specifications is similar.

household h in season t, corrected for a factor-specific set of observables (W_{Jhit}) and subject to classical measurement error ϵ_{Jhit}:

$$J_{hit} = J^o_{hit} e^{W_{Jhit}\beta_J - \epsilon_{Jhit}}. \tag{4}$$

Using (3) and (4) in (2), effective units of labor to land remain equal to the common ratio of factor prices:

$$
\begin{aligned}
x^o_{hit} &+ W_{Xhit}\beta_X - l^o_{hit} - W_{Lhit}\beta_L - (\ln(\alpha_X) - \ln(\alpha_L)) \\
&= \ln\left(\frac{p_{Lhit}}{p_{Xhit}}\right) + \epsilon_{Xhit} - \epsilon_{Lhit} \\
&= \ln\left(\frac{p_L}{p_X}\right) + \epsilon_{Xhit} - \epsilon_{Lhit}.
\end{aligned}
\tag{2''(a)}
$$

but measurement error implies that observed ratios of labor to land vary. Similarly for output per land

$$
\begin{aligned}
y_{hit} &+ W_{Yhit}\beta_Y - l^o_{hit} - W_{Lhit}\beta_L + \ln(\alpha_L) \\
&= \ln\left(\frac{p_{Lhit}}{p_{Yhit}}\right) + \epsilon_{Yhit} - \epsilon_{Lhit} \\
&= \ln\left(\frac{p_L}{p_Y}\right) + \epsilon_{Yhit} - \epsilon_{Lhit}
\end{aligned}
\tag{2''(b)}
$$

Output ratios across plots vary with TFP, and with risk and measurement error:

$$
\begin{aligned}
y_{hit} &- y_{kit} + (W_{Yhit} - W_{Ykit})\beta_Y \\
&= \frac{(\omega_h - \omega_k) + \alpha_L \ln\left(\frac{p_{Yhit}}{p_{Lhit}}\right) + \alpha_X \ln\left(\frac{p_{Yhit}}{p_{Xhit}}\right)}{1 - \alpha_L - \alpha_X} \\
&\quad + \epsilon_{Yhit} - \epsilon_{Ykit} = \\
&= \frac{(\omega_h - \omega_k) + \alpha_L \ln\left(\frac{p_Y}{p_L}\right) + \alpha_X \ln\left(\frac{p_Y}{p_X}\right)}{1 - \alpha_L - \alpha_X} + \epsilon_{Yhit} \\
&\quad - \epsilon_{Ykit}
\end{aligned}
\tag{2''(c)}
$$

In an efficient allocation, the variance of shadow prices is zero; but it is also true that *any* amount of observed dispersion in residuals is consistent with an efficient allocation when there is measurement error or risk.

In order to draw useful conclusions regarding the extent of factor misallocation and its implications for aggregate output loss, it is necessary to disentangle the effects of measurement error and stochastic output from variation due to factor misallocation. To do so, we rely on an assumption that *within a farm,* the allocation of resources across plots is efficient.

We define a farm as the set of plots cultivated under the management of a single farmer in a single season. Any reallocation of factors across plots within a farm requires no market intermediation or other exchange, only rational decision-making by the farmer. While we acknowledge that there may in fact be behavioural limits on the rationality of input decisions by farmers, we abstract away from these sources of efficiency loss for this paper and maintain the Schultzian "poor but efficient" assumption (Schultz 1964). This assumption does not imply that all farmers are equally productive or knowledgeable. One farmer may have superior technical knowledge to another; this difference would be reflected in higher total factor productivity.

If the allocation of factors across plots within a farm (during a single season) is efficient, then the dispersion of factor- and output-factor ratios across plots *within* a farm is generated by (a) imperfect measurement of factor inputs; (b) imperfect measurement of output; or (c) varying realizations of risk.[18]

---

[18] It is of course possible that some farmers are systematically worse than others at allocating efficiently across plots. But it is not obvious that this should have a strong correlation with productivity *levels*. A bad farmer is arguably one who realizes equally

The final two columns of Tables 2A-B and 3A-B show coefficients from regressions of log output per hectare and log labor per hectare with the same set of plot characteristics and within-farm fixed effects. To be precise, Column (4) reports the regressions with crop-season-household fixed effects, and Column (5) is based on crop-season-farmer fixed effects, where we are now looking at variation across plots farmed by the same individual within the households. The residuals from these regressions are again shown in each of the subgraphs of Figures 1 and 2.[19]

In each country, when we consider either the yield regressions of Table 2 or the labor intensity regressions of Table 3, approximately one-fifth of the baseline dispersion reported in Column (2) remains after we focus attention on variation in output per hectare across plots *within a farm.*

Given our assumption of efficient within-farm allocation, we conclude that this residual variation is evidence for significant heterogeneity or measurement error in factors of production or output. Alternatively, it could reflect unobserved shocks to output that do not affect the marginal product of factor inputs or which occur following the application of inputs to different plots within a farm. If the variance of these errors of measurement, or of shocks to output, is at least as large across farmers as it is across plots of a given farmer, then interpreting the residuals of the

poor yields across all plots, based on allocating inputs with the same (improper) intensity across all plots.

[19] There is no evidence of systematic differences in yield and labor intensity on the plots of men and women farmers within the same household in Uganda or Tanzania. Even in Burkina Faso, where there is such evidence, Udry (1996) finds that the magnitude of the dispersion generated by this difference is very small relative to other sources of variation.

equations estimated in columns (2) of Tables 2 and 3 as misallocation would result in an overestimate the importance of misallocation.

To estimate the magnitude of misallocation, we need to know more about the production function and the magnitude of measurement error in factor inputs, which we address in Section 3.

2.2 Plot Size, Yield, and Factor Intensity

We observe a strong and consistent negative relationship between output per hectare and plot size. While this pattern is reminiscent of the long-standing discussion of an inverse *farm* size-yield correlation, we find in the final column of Tables 2 and 3 that this pattern holds across *plots* (planted with the same crop in the same season) within a farm. Across farms, factor market imperfections might explain an inverse relationship between land area and yield, but these market imperfections cannot explain this relationship across plots within a farm. Both Tanzania and Uganda exhibit fairly extreme negative relationships between log yield and log plot size within a farm; the estimated elasticity is -0.73 (s.e.=0.02) for Uganda and -0.50 (s.e.=0.03) for Tanzania.

This pattern of a strong negative relationship between crop yields and plot size within a farm has been observed in multiple data sets from Africa (Carletto et al., 2015; Carletto et al., 2017; Bevis and Barrett, 2017). One source of this estimated inverse relationship might be measurement error in plot size. Kilic et al. (2016) provide a careful account of the role of this kind of measurement error using these same Uganda and Tanzania data sets. They show that while measurement error does contribute to the estimated inverse plot-size relationship, the relationship remains strong after using objective GPS measures of plot area and correcting for selection bias in the subset of plots measured with GPS. A number of other studies

consider other reasons why yield might be negatively related to plot size; e.g., Bevis and Barrett (2017). Most relevant for us, Gourlay et al. (2017) report the results of a methodological experiment, using different data from Uganda, which carefully examined the reporting of output data from farmers. Their findings suggest that self-reported yields are biased upward compared to measurement of crop cuts at harvest – and that this effect is stronger on smaller plots, perhaps fully explaining the observed relationship.

We find, however, that labor per hectare is also strongly declining in plot size within a farm (Column 5 of Table 3). The decline in both yield and labor per hectare with plot size suggests a different interpretation: namely, that smaller plots have higher unmeasured land quality.[20]

These correlations lead us to hypothesize that there is a substantial degree of unmeasured heterogeneity in land quality across the land of a given farmer. This is consistent with the patterns in Figures 1 and 2 documenting important dispersion in yield and factor intensity across the plots of an individual farmer. It may plausibly play a role in the strong inverse relationship observed between cultivation intensity and plot size across the plots of a farmer.

---

[20] Barrett et al. (2010) argue against this hypothesis using data from Madagascar, showing that the introduction of a vector of objective measures of soil quality from soil tests has no effect on the estimated inverse yield – plot size relationship. However, in their data, the measures of land quality are not jointly significant predictors of yield, nor are they jointly significant in the production function estimated. This is a frequent characteristic of observed measures of land quality. In our data, however, the land quality measures are strongly jointly significant, perhaps reflecting the high quality of data collection in the LSMS-ISA data. This encourages us to think that land quality may have some role to play in the relationship between plot size and observed yield.

3. Theoretical framework

Our central argument is that heterogeneity in land quality and growing conditions play an important role in explaining the dispersion of productivity at the level of plots and farms. This heterogeneity is unobservable to the econometrician but may be well recognized by farmers. Some of the unobservables involve intrinsic properties of the soil or land, such as the physical and chemical properties of the soil, or the slope and topography. Other unobservables relate to highly localized shocks – such as hail that strikes one plot on a farm but spares another. Still others may involve complex interactions between shocks and plot characteristics: a heavy early-season rain makes one low-lying plot unworkable at the start of the season because of mud; but the same rainstorm is actually beneficial for another plot that is well-drained.

The importance of this kind of heterogeneity, often at a very fine-grained spatial level, is well documented in agronomic and economic studies.[21] Farmers can and do modify their practices to reflect heterogeneity of this kind; e.g., choosing plot boundaries that reflect spatial differences in soil

---

[21] The importance of heterogeneity in agricultural systems at highly localized spatial scale has been shown previously in numerous contexts; e.g., by Hanna et al. (2014). A data collection experiment conducted in Uganda (Lobell et al. 2018) found that the yields on randomly chosen 8 x8 meter squares within typical maize plots correlated only weakly with the yields on the entire plots (which had a median size of 0.11 ha) because of "substantial intra-plot heterogeneity of yields in these systems." For African crop agriculture, see the work on agronomy by Tittonell et al. (2005), Tittonell et al. (2007), Tittonell et al. (2008), Vanlauwe et al. (2006), and Vanlauwe et al. (2015), along with papers in economics such as Suri (2011) and Tjernstrom et al. (2015). For U.S. crop agriculture, Hurley et al. (2004) document high levels of agronomic heterogeneity within farmers' fields. This indeed is the premise for the emergence of "precision agriculture" technologies, as discussed in Stoorvogel et al. (2015). Recent empirical work on precision agriculture in the United States shows the profitability of fine-tuning inputs to within-plot variation in land quality (Schimmelpfennig 2016). Commercial systems typically fine-tune applications of chemicals at a spatial resolution of less than 1.0 m^2, reflecting meaningful differences in soil properties at that scale.

type. But even in detailed household surveys such as the LSMS-ISA data, the available measures of plot-level land quality do not adequately capture these dimensions of heterogeneity.

To help us understand the significance of this kind of location-specificity, we develop a model of agricultural production on heterogeneous land in which farmers can endogenously choose plot sizes and boundaries. This proves to be a good description of the realities of agriculture in our two countries. In the Uganda data, 62 percent of plots are subdivided from larger parcels of contiguous land controlled by an individual farmer, implying that these plots are literally or approximately adjacent.[22]

3.1 Agricultural production with continuous variation in land quality

The farmer, indexed by h, holds a fixed endowment of land denoted by L_h. This land parcel consists of a continuum of locations that can be indexed by k on the interval $[0, L_h]$.

At a location k, the quality of the land in effective units is denoted by $\zeta(k)$. For simplicity, assume that the function $\zeta(\cdot)$ is continuous and integrable. Land is used for producing an agricultural good. The production process uses a bundle of inputs that, in principle, could be applied on a location-specific basis. We denote the inputs used at a particular location as $\xi(k)$. Output is also affected by a location-specific productivity shock that depends on the state of the world, which we denote

---

[22] A parcel is defined as a contiguous area under uniform land tenure held by a given farmer. Not all parcels are subdivided into plot; some are farmed as single plots. But just around one-third of parcels are subdivided into plots, with an average of 2.8 plots per multi-plot parcel.

by $\gamma(k, s)$. The state of the world s is distributed according to $\Delta(s)$ over support S. This shock is observed by the farmer before she chooses the input bundle. For example, this shock could consist of early-season rain – or perhaps the timing of the onset of the rainy season. A given state of the world may have different productivity implications for different locations on the farmer's land and for different farmers.

Given this notation, we define a simple production technology in which the output obtained by farmer h at location k conditional on the shock s having been realized will be given by:

$$q_h(k, s) = \gamma_h(k, s)\zeta_h(k)\big(\xi_h(k, s)\big)^\theta. \tag{5}$$

If a profit-maximizing farmer were to farm only this single point, facing a farmer-specific shadow price w_h for inputs, the farmer would solve:

$$\max_{\xi_h(k,s)} \left[\gamma_h(k, s)\zeta_h(k)\big(\xi_h(k, s)\big)^\theta - w_h\xi_h(k, s)\right]. \tag{6}$$

As an elementary optimality condition, this would give an optimum of $\xi_h^*(k, s) = \left(\frac{\theta\gamma_h(k,s)\zeta_h(k)}{w_h}\right)^{\frac{1}{1-\theta}}$. The corresponding profit-maximizing output at that location would be:

$$q_h^*(k, s) = \zeta_h(k)\gamma_h(k, s)\left(\frac{\theta\gamma_h(k, s)\zeta_h(k)}{w_h}\right)^{\frac{\theta}{1-\theta}}. \tag{7}$$

3.2   Plot-level production with continuous variation in land quality

Production could, in principle, be fine-tuned in this fashion to match the precise characteristics of each location, with inputs varying continuously

across space. However, farming takes place at the level of a *plot*. A gardener nurtures each plant; a farmer has a goal of routinizing operations across the plot, transforming artisanal attention to each plant into systematic tasks that can be performed across the mass of plants growing together. Therefore, a plot is characterized by synchronicity of a sequence of tasks: land preparation, sowing, thinning, applying inputs, successive rounds of weeding, harvesting, processing, etc. We define a plot as a set of contiguous locations that are managed uniformly.

For a farmer in our model, a plot will be defined as a contiguous interval $[\underline{k}, \overline{k}] \subseteq [0, L_h]$. The farmer faces a fixed cost c to create and farm a plot of land within its overall land holding. Because of this fixed cost, there will be finitely many plots per farm. On a previously defined plot i, the farmer h now applies inputs with the same intensity across every location on the plot. The intensity at each location can be written as ξ_{hi} s . Define the size of the plot to be the distance between its two endpoints; i.e., $L_{hi} = (\overline{k} - \underline{k})$. Define the aggregate inputs used on the plot as $X_{hi} = (\xi_{hi}(s))(\overline{k} - \underline{k})$.[23] Then output at any point on that plot is given by:

$$q_h(k, s) = \gamma_h(k, s)\zeta_h(k)\left(\frac{X_{hi}}{L_{hi}}\right)^{\theta}. \qquad (8)$$

Without loss of generality, assume that $\underline{k} = 0$ and $\overline{k} = L_{hi}$. Total output on the plot is thus:

$$
\begin{aligned}
Y_{hi}(s) &= \int_0^{L_{hi}} \gamma_h(k, s)\zeta_h(k)\left(\frac{X_{hi}}{L_{hi}}\right)^{\theta} dk \\
&= \left(\frac{X_{hi}}{L_{hi}}\right)^{\theta} \int_0^{L_{hi}} \gamma_h(k, s)\zeta_h(k)dk.
\end{aligned}
\qquad (9)
$$

---

[23] We note that as a simple extension of the analysis, we can let the input vector ξ be a Cobb-Douglas composite of two or more other inputs; e.g., labor N and chemicals V, such that $\xi = N^{\alpha}V^{1-\alpha}$. The analysis will go through unchanged.

Defining average land quality as $\zeta_{hi}(s) = \frac{1}{L_{hi}} \int_0^{L_{hi}} \gamma_{hi}(k,s)\zeta_{hi}(k)\,dk$ and substituting:

$$Y_{hi}(s) = L_{hi}\zeta_{hi}(s)\left(\frac{X_{hi}}{L_{hi}}\right)^\theta = \zeta_{hi}(s)L_{hi}^{1-\theta}(X_{hi})^\theta \tag{10}$$

The corresponding profit-maximization problem involves a trade-off between the fixed cost of creating a plot (which incentivizes larger plot sizes) and the fine-tuning of inputs that is possible on a smaller plot. This trade-off is clearly visible when the profit-maximization problem is given in terms of input intensity:

$$\max_{\xi_{hi}(s)}\left[[\overline{k} - \underline{k}]\left(\xi_{hi}(s)\right)^\theta \int_{\underline{k}}^{\overline{k}} \gamma_{hi}(k,s)\zeta_{hi}(k)\,dk \right.$$
$$\left. - w_h[\xi_{hi}(s)][\overline{k} - \underline{k}] - c \right] \tag{11}$$

Note that in equation (11), the bluntness of input use means that the profit-maximizing input bundle ξ_{hi}^* s differs from the "precision agriculture" levels that would be chosen if the farmer were maximizing at each location separately. Output will differ correspondingly. The lone exception is the case in which the fixed cost $c \to 0$, in which case $(\overline{k} - \underline{k}) \to 0$ and $\left[Y_{hi}^* - \int_{\underline{k}}^{\overline{k}} q_h^*(k)dk\right] \to 0$. With $c > 0$, the farmer chooses to divide its land into a finite number of plots. Appendix A1 describes the problem of endogenous plot selection. For the moment, we simply note that under quite general conditions, a farmer will choose a determinate number of plots, with the size and location of these plots reflecting the level and variability of land quality.

3.3   Land quality and plot size

Without imposing some further restrictions on the patterns of land quality, we cannot make any statements about the relationship between land quality and plot size. But we can offer a few relevant observations. First, we show in Appendix A1 that the maximum number of plots that could be cultivated profitably by a farmer depends inversely on the fixed cost and is also positively related to the average land quality across the farm. A farmer with very poor average land quality will *ceteris paribus* have a smaller maximum number of plots than a farmer with the same total land area but better quality land. This does not necessarily give rise to an empirical prediction, because farms will not in general cultivate the maximum possible number of plots. But it does point to an underlying pattern that holds more generally: everything else equal, poor quality plots must be sufficiently large that they will earn positive profits.

Consider the profit maximization for the ith plot cultivated by farmer h. The size of this plot is $\tilde{L}_{hi}$, with its boundaries at L_{hi-1} and L_{hi}, respectively. The average productivity of this plot, conditional on the realization of the shock $\gamma_h(k, s)$, is $\zeta_{hi} = \frac{1}{\tilde{L}_{hi}} \int_{L_{hi-1}}^{L_{hi}} \gamma_h(k, s) \zeta_h(k) \, dk$. We can solve the profit maximization problem and then ask, for a given value of ζ_{hi} what is the smallest plot size that will yield non-negative profits – or in other words, what threshold plot size will be needed to cover the fixed cost c. We can then ask how this plot size threshold changes in relation to ζ_{hi}. The formulation of this is straightforward. Substituting in the optimized value of $X^* \left(= \tilde{L}_{hi} \left(\frac{\theta \zeta_{hi}}{w_h}\right)^{\frac{1}{1-\theta}} \right)$ into the zero profit condition, and setting $\zeta_{hi}(s) \tilde{L}_{hi}^{1-\theta} (X_{hi})^{\theta} - w_h X_{hi} = c$, we get a relationship between the threshold plot size and land quality that will sustain non-negative profits:

$$L_{hi}^{min} = \frac{c w^{\frac{\theta}{1-\theta}}}{(\zeta_{hi}(s))^{\frac{1}{1-\theta}} \left(\theta^{\frac{\theta}{1-\theta}} - \theta^{\frac{1}{1-\theta}} \right)}. \tag{12}$$

Within a farm, the optimal size of a plot depends on both the average quality of the land and the heterogeneity of the land quality. Holding average quality constant, the size of the plot will be decreasing in heterogeneity. Holding heterogeneity constant, the size of the plot will be decreasing in average quality (or put differently, it will increase on poorer land). The underlying logic is that there is a trade-off between the benefits gained by fine-tuning the inputs used on a plot, which tends to drive plot size smaller, and the fixed cost, which tends to drive plot size larger. On high-quality land, the fixed cost is a relatively smaller burden, and so plot size will be smaller, *ceteris paribus*. On low-quality land, the fixed cost poses a heavier burden, and so plot size will tend to be larger. We show in Appendix A1 that for any given fixed cost, a farmer will partition a given parcel of heterogeneous land into two plots if the parcel is sufficiently productive and not if productivity is lower. At the same time, however, the more heterogeneous a plot is in terms of land quality, the more costly it will be to have a large plot; a homogeneous plot can be large. In the extreme case of a farm that is entirely uniform in terms of land quality, there is no reason to subdivide this into plots, regardless of the quality.

A farm of given area and heterogeneity with sufficiently higher average land quality, will be divided into more plots. Table 4 shows direct evidence of this pattern in Uganda, where we can identify the specific parcel of land to which each cultivated plot belongs. Labor intensity and output per acre over the entire parcel are higher on parcels that have been divided into more plots, as shown in column 1. In column 2, we condition on all of the measures of land quality, farmer characteristics and weather shocks used in Tables 2B and 3B; again we find that parcels that have been divided into more plots are cultivated with higher intensity and achieve greater yields. Columns 3 through 6 repeat this exercise with successively finer fixed effects, at the district-season, village-season, household-season and farmer-season levels and in each case we observe the same strong positive

correlation between the number of plots into which a parcel has been divided and both input-intensity and yield at the parcel level. This is consistent with the model that, conditional on parcel area and our rich set of observed characteristics of parcel land quality, there is important unobserved variation in land productivity across *parcels* such that farmers subdivide higher quality parcels into more plots.

3.4   Empirical Framework

This theoretical framework allows us to structure an empirical analysis of the patterns of factor intensities and yields across plots within and across farms described in Section 2, based on the plot-level production function derived in equation (10). The structure of our data correspond quite closely to the theoretical framework, with a few modifications. First, farmers in our model grow a single crop – so that output is homogeneous. In the data, farmers grow many crops and intercropped mixes (e.g., maize, beans, and cassava grown together on the same plot). As we did in our analysis of productivity dispersion in Section 2, we aggregate these by value for simplicity. Second, in our model, farmers have a single parcel, which they divide into multiple plots, each of which is cultivated with uniform input intensity. In the data, farmers may have more than one parcel. Since we observe inputs at the plot level, we can observe differences in input intensity across plots; but we cannot rule out the very real possibility that input use varies within the plot. Third, in our model different plots differ only in the intensity of input use, since output is homogeneous; in the data, different plots are almost always allocated to different crops or crop mixtures, so there is a crop choice dimension to the farmer's decision that is absent from our model. [24]

---

[24] Our reason for abstracting from crop choice is not simply for convenience. In our data, because of the ubiquity of intercropping, it is not sensible to think of farmers making a discrete crop choice. Instead, they are frequently choosing complex intercropped mixtures. In the Uganda data, we observe 428 distinct crops and crop

Accepting these slight departures from our model, we rewrite the production function (10) adding a t index to denote the season and year in which production takes place, because we will be working with panel data in which farmers are observed in multiple seasons. Thus output of plot i of farmer h in season t is $Y_{hit} = \zeta_{hit} L_{hit}^{1-\theta} (X_{hit})^{\theta}$. The productivity term $\zeta_{hit} = \frac{1}{L_{hit}} \int_{0}^{L_{hit}} \gamma_{hit}(k, s(t)) \zeta_{hi}(k) \, dk$ incorporates the plot average of land quality at each point, denoted as $\zeta_{hi}(k)$; and the plot average effect of shocks received in season t at each point $\gamma_{hi}(k, s(t))$. Both land quality and the effects of shocks may vary across plots and across households in ways we do not observe but which our model implies will potentially be correlated with plot size. To account for the possible dependence of average plot productivity on plot size, we parametrize productivity as $\frac{1}{L_{hit}} \int_{0}^{L_{hit}} \gamma_{hi}(k, s(t)) \zeta_{hi}(k) \, dk = e^{\omega_{hit}} (L_{hit})^{\varphi_{Lhit}}$.

Our model implies that $\varphi_{Lhit} < 0$; in other words, larger plots will have lower average productivity. However, we will not impose this assumption in the estimation below. The parameter φ_{Lhit} varies across plots because the rate at which productivity falls with plot size depends upon the variability in land quality over space, which need not be uniform. The production function, therefore, becomes:

---

mixes across farmers' plots; in Tanzania, the intercropping permutations give us an even higher number, with 853 distinct crops and crop mixes. Many of these crop mixtures are derived from variations on a much smaller set of distinct commodities. For instance, a farmer might grow maize by itself on one plot and a maize-cassava mix on another plot, with maize-beans-cassava on a third plot. In Tanzania, 43% of plots have some maize on them. In Uganda, 63 percent of plots are planted with one or more of maize, beans, cassava, and cooking bananas. Even beyond that, a maize-cassava intercrop could be 90 percent maize and 10 percent cassava, or the opposite. Given this complexity, we find it simpler to carry out our analysis in terms of a composite agricultural output, which then corresponds very closely to our model.

$$Y_{hit} = e^{\omega_{hit}}(L_{hit})^{\varphi_{Lhit}}(L_{hit})^{1-\theta}(X_{hit})^{\theta}$$
$$= e^{\omega_{hit}}(L_{hit})^{\alpha_{Lhit}}(X_{hit})^{\alpha_{Xhit}}. \tag{13}$$

The parameter ω_{hit} is total factor productivity, which is at least partially known to the farmer. However, at least some of what is known to the farmer is unobserved to us. Given this structure, factor demands are subject to the classic production function endogeneity concern.[25] In addition, unobserved heterogeneity in factor productivity implies that α_{Lhit} and α_{Xhit} may be heterogeneous across plots. We use (3) and (4) to rewrite (13) in terms of observable inputs and observed shocks and take logs:

$$y_{hit} = \alpha_{Lhit}l^o_{hit} + \alpha_{Xhit}x^o_{hit} + W_{Yhit}\beta_Y$$
$$+ \sum\alpha_{Jhit}(W_{Jhit}\beta_J - \epsilon_{Jhit}) + \omega_{Yhit} + \epsilon_{Yhit} \tag{14}$$

The vector of observable determinants of total factor productivity ($W_{Yhit} = (W_{Ehit}, W_{Hhit})$) includes a rich set of indicators of shocks to productivity; most importantly, we have measures of the amount and timing of local rainfall interacted with characteristics of the plot and indicators of specific shocks (fire, flooding) on particular plots. We denote by W_{Ehit} the subset of those shock indicators that occur before the early harvest season begins, sufficiently early that farmers may be able to adjust factor inputs in response. Similarly, W_{Hhit} is the subset of those shock indicators that occur at harvest season, too late for farmers to adjust factor inputs in response.

We assume that farmers know the productivity of the factors they are using in cultivation, so factor demands will in general be correlated with the realizations of the factor productivities.

---

[25] For a recent discussion see Ackerberg et al. (2015), which in turn builds on Olley and Pakes (1992) and Levinsohn and Petrin (2003).

3.5 Identification of the Production Function

Our strategy for identification of the farm production function rests on the imperfection of factor markets across farms, and builds on a strategy from Benjamin (1992). The central idea of Benjamin's identification strategy was that observed shocks to factor productivity -- and hence factor demand -- on one farm affect the shadow cost of factors to other farmers within the same community, but with much smaller effects across a geographically dispersed sample.

In our case, we use a similar strategy, but we have the added benefit of working with plot-level data rather than farm-level data. Observed shocks to factor productivity -- and hence factor demand -- on one plot affect the shadow cost of factors on other plots of the same farmer, and on the plots of other farmers in the same community. The substantive assumption we require is that, conditional on the observed shocks affecting a given plot i, the observed shocks affecting a different plot j are not correlated with unobserved determinants of factor demand on plot i.

To be more precise, if factor markets are imperfect, then conditional on the realization of W_{Ehit} on plot i, the realizations of $W_{Eh,-i,t}$ on plots $-i \neq i$ of farmer h in season t provide variation in the shadow value of factors of production on plot i of farmer h in season t. Similarly, if there is some within-village exchange of labor or land, but inter-village factor markets are imperfect, realizations of $W_{E,-h,j,t}$ on the plots of farmer $-h \neq h$ within the village of farmer h also provide variation in the shadow value factors of production on all the plots of farmer h. Accordingly, $W_{Eh,-i,t}$ and $W_{E-h,j,t}$, along with more conventional measures of household wealth and demographics comprise a set of potential instruments G_{hit} for factor inputs on plot i.

A typical member of W_{Ehit} is a village- or household-level report of a shock interacted with observed characteristics of plot i. For example, village level peak season rainfall interacted with an indicator that the soil on plot i is clay. The identification assumption is that, conditional on the observed set of shocks on plot i of farmer h in season t, a similar shock on plot j of the same farmer in the same season (for example, the same village level peak season rainfall interacted with an indicator that the soil on plot j is loamy) affects the demand for inputs on plot i only through the shadow value of inputs for the farmer. At the village level, the assumption is again that conditional on observed shocks at the plot level (for example, a self-report of drought at the farmer level interacted with a plot soil quality indicator), average shocks on other farmers' plots in the village affect the plot level demand for inputs only through the shadow value of inputs for the farmer. These are substantive assumptions, violated if there are components of ω_{Yhit} that are correlated with $W_{Eh,-i,t}$ or $W_{E-h,j,t}$ (conditional, of course, on W_{Ehit}). More concretely, we require that conditional on a plot's observed characteristics, its unobserved characteristics are uncorrelated with the observed characteristics of land on other plots in the village. For example, the fact that a large proportion of the land in a village not controlled by a farmer is categorized as "loam" is not systematically correlated with the unobserved characteristics of the land of the farmer that is categorized as "loam". It is clear that these shocks will drive variation in farmer-specific shadow costs of inputs and therefore if these assumptions are correct, this is a useful approach to identifying an agricultural production function.

3.6 Correcting Productivity Estimates for Measurement Error, Risk, and Heterogeneity

With estimates of the deterministic production function parameters $\hat{\beta}_Y, \hat{\beta}_L$ and $\hat{\beta}_X$, and estimates $\hat{\alpha}_L, \hat{\alpha}_X$ of the expected values of the random factor

productivities α_{Lhit} and α_{Xhit}, a first approximation to the distribution of log TFP across plots might simply be the residual:

$$\ln \widehat{TFP}_{hit}^a = y_{hit} - \hat{\alpha}_L l_{hit}^o - \hat{\alpha}_X x_{hit}^o + \hat{\alpha}_K k_{hit}^o - W_{Yhit}\hat{\beta}_Y \\ - \hat{\alpha}_L W_{Lhit}\hat{\beta}_L - \hat{\alpha}_X W_{Xhit}\hat{\beta}_X \tag{15}$$

Equating TFP to the difference between observed output and output predicted by substituting observed factor use (and observed enterprise characteristics) into an estimated production function, as is common in the macro literature, attributes all unexplained variation in output to variation in TFP. This approach overstates the variation in TFP if there is measurement error, and it is further misleading in the presence of shocks to output or unobserved variation in factor productivity. Both of these are surely present in our data.

From (14) and (15), we observe:

$$\ln \widehat{TFP}_{hit}^a = \underbrace{\omega_{Yhit}}_{\text{unobs tfp}} + \underbrace{\sum_{J \in \{L,X\}} (\alpha_{Jhit} - \hat{\alpha}_J)(W_{Jhit}\hat{\beta}_J)}_{\substack{\text{unobs productivity of} \\ \text{observed characteristics}}}$$

$$+ \underbrace{\sum_{J \in \{L,X\}} (\alpha_{Jhit} - \hat{\alpha}_J) \ln(J_{hit}^o)}_{\substack{\text{unobs productivity} \\ \text{of factors}}}$$

$$\tag{16}$$

$$- \underbrace{\sum_{J \in \{L,X\}} (\alpha_{Jhit} - \hat{\alpha}_J)\epsilon_{Jhit}}_{\substack{\text{unobs productivity of} \\ \text{factor measurement error}}} - \underbrace{\sum_{J \in \{L,X\}} \hat{\alpha}_J \epsilon_{Jhit}}_{\substack{\text{factor} \\ \text{measurement error}}}$$

$$+ \underbrace{\epsilon_{Yhit}}_{\substack{\text{post-input shocks and} \\ \text{measurement error in y}}}$$

The final three terms are sources of variation in *measured* productivity (i.e., $\ln \widehat{TFP}_{hit}^a$), but they do not give rise to *actual* productivity variation. The dispersion in measured productivity arising from these three terms cannot be remedied through reallocation. Reallocation cannot "solve"

measurement error, nor can reallocation "solve" late-season idiosyncratic shocks that affect yield.

The variance of the production function residual $\ln\widehat{TFP}^a_{hit}$ is

$$var\big(\ln\widehat{TFP}^a_{hit}\big) =$$

$$var\left(\omega_{Yhit} + \sum_{J\in\{L,X\}}\big(\alpha_{Jhit} - \hat\alpha_J\big)\big(W_{Jhit}\hat\beta_J + \ln(J^o_{hit})\big)\right) \qquad (17)$$
$$+ \hat\alpha_L^2 var(\epsilon_{Lhit}) + \hat\alpha_X^2 var(\epsilon_{xhit}) + var(\epsilon_{Yhit})$$

Only the first term of (17) represents cross-farm variation in productivity that is relevant for any assessment of allocative efficiency; we would like to disentangle it from the final three terms, which are the variation due to measurement error and late season shocks. To do so, we use the observed allocation of factors across the plots of a given farmer in a season as a benchmark. The efficient allocation of factors across the plots cultivated by the same farmer implies patterns of behavior that can identify the variances of these late-season productivity shocks and measurement errors. For example, observed variation in labor inputs across the plots of a single farmer that is not correlated with either output or other inputs is attributable to measurement error in labor.

In what follows, we maintain the assumption that the allocation of factors within a farm is efficient, at least in the narrow sense that market failures do not affect farmers' ability to make allocative decisions across their plots. Uninsured risk, imperfect financial markets, labor market frictions and missing markets for land all assuredly influence a farmer's overall choice set – but not their allocation decisions across plots. Within a farm, shadow factor prices are constant across plots.[26] Within farm variation in observed

---

[26] This could, of course, be a problematic assumption if plots are far apart. In Appendix A5, we report a set of robustness checks that address this assumption.

output and observed inputs depend only on (a) unobserved dimensions of risk or measurement error in output (ϵ_{YHit}); (b) measurement error in factor inputs $(\epsilon_{Lhit}, \epsilon_{Xhit})$; or (c) unobserved heterogeneity in factor $(\alpha_{Lhit}, \alpha_{Xhit})$ or total-factor productivity (ω_{Yhit}).

We show in Appendix A2 that the observed covariances of factor demands and output across plots within a farm (along with a normalization discussed below) provide us with sufficient information to identify the average within-farm variances of plot-level total factor productivity (σ_q^2), factor-specific productivity and their covariance $(\sigma_L^2, \sigma_X^2, \sigma_{LX})$, factor measurement error $(\sigma_{\epsilon L}^2, \sigma_{\epsilon X}^2)$ and output measurement error and post-input risk $(\sigma_{\epsilon y}^2)$, as well as the covariance of plot-level total factor productivity and factor-specific productivity $(\sigma_{QL}, \sigma_{QX})$.

We will not separately identify variation in all three types of unobserved heterogeneity in factor $(\omega_{Lhit}, \omega_{Xhit})$ and total-factor productivity (ω_{Yhit}): a parallel increase in ω_{Lhit} and ω_{Xhit} is equivalent to an increase in ω_{Yhit}. Hence, we normalize $\omega_{Lhit} + \omega_{Xhit} = 0$. Intuitively, a change in ω_{Lhit} relative to ω_{Xhit} is a change in the slope of an isoquant; a change in ω_{Yhit} is a shift in or out of an isoquant. This normalization implies $\sigma_L^2 = \sigma_X^2$, $\sigma_{LX} = -\sigma_L^2$, $\sigma_{QL} = -\sigma_{QX}$.

The assumption of an efficient allocation across plots within a farm, therefore, permits us to calculate the parameters $\hat{\sigma}^2 = \left(\hat{\sigma}_q^2, \hat{\sigma}_L^2, \hat{\sigma}_X^2, \hat{\sigma}_{\epsilon y}^2, \hat{\sigma}_{\epsilon L}^2, \hat{\sigma}_{\epsilon X}^2, \hat{\sigma}_{LX}, \hat{\sigma}_{QL}, \hat{\sigma}_{QX}\right)$ that are consistent with the observed covariance of plot level output and inputs across plots within farms, given an estimate of the production function parameters.

The estimated values of $\hat{\sigma}^2$ reflect the mean, across farmers, of the within-farm variances of measurement error, late-season risk and unobserved productivity. If we maintain the assumption of classical measurement error, then the variance of that measurement error is the same across all

plots as it is, on average, across plots within a farm. Similarly, if late season risk is i.i.d. across plots, then its variance across all plots is the same as it is, on average, across plots within a farm. If there are farmer-specific components to either late season risk or to measurement error, then we can expect (as shown in Appendix A3) that our estimate of the variance of unexplained output attributable to late-season risk and measurement error in output and factor inputs $\left(\hat{\sigma}_{\epsilon y}^2 + \hat{\alpha}_L^2 \hat{\sigma}_{\epsilon L}^2 + \hat{\alpha}_X^2 \hat{\sigma}_{\epsilon X}^2\right)$ is a lower bound estimate of the variance in the final three terms of (15). Subtracting $\hat{\sigma}_{\epsilon y}^2 + \hat{\alpha}_L^2 \hat{\sigma}_{\epsilon L}^2 + \hat{\alpha}_X^2 \hat{\sigma}_{\epsilon X}^2$ from $var\left(\widehat{\ln TFP}_{hit}^a\right)$, therefore, provides an upper bound estimate of $var\left(\omega_{Yhit} + \sum_{J \in \{L,X\}} \left(\alpha_{Jhit} - \hat{\alpha}_J\right)\left(W_{Jhit}\hat{\beta}_J + \ln(J_{hit}^o)\right)\right)$. This gives the variation in productivity that is relevant for any assessment of allocative efficiency.

We therefore proceed by beginning with the naïve production function residual, $\widehat{\ln TFP}_{hit}^a$, and shrinking it towards its mean μ_A to account for the variances of the measurement errors and late-season shocks we can measure within the farm. Our revised estimate of unobserved productivity is

$$\widehat{\ln TFP}_{hit}^b = \mu_A + \left(\widehat{\ln TFP}_{hit}^a - \mu_A\right)$$
$$* \left(\frac{var\left(\widehat{\ln TFP}_{hit}^a\right) - \sigma_{\epsilon Y}^2 - \sum_J \alpha_J^2 \sigma_{\epsilon J}^2}{var\left(\widehat{\ln TFP}_{hit}^a\right)}\right)^{\frac{1}{2}} \qquad (18)$$

If there are aggregate late-season shocks, or farmer- or household-specific components to measurement error, then

$$var\left(\widehat{\ln TFP}_{hit}^b\right)$$
$$> var\left(\omega_{Yhit}\right.$$
$$\left. + \sum_{J \in \{L,X\}} \left(\alpha_{Jhit} - \hat{\alpha}_J\right)\left(W_{Jhit}\hat{\beta}_J + \ln(J_{hit}^o)\right)\right) \qquad (19)$$

and our revised estimate of the dispersion of unobserved productivity remains an overestimate of the true variation.

4. Empirical analysis

4.1 Estimation Procedure

We estimate the agricultural production functions, the implied residuals, $\ln \widehat{TFP}_{hit}^{a}$, the associated variances of unobserved heterogeneity and measurement error and the revised estimates of unobserved productivity $\ln \widehat{TFP}_{hit}^{b}$ both by 2SLS and using Masten and Torgovitsky's (2016) IVCRC estimator.

Heckman and Vytlacil (1998) and Wooldridge (2008) show that 2SLS provides consistent estimates of $E(\alpha_{Lhit})$ and $E(\alpha_{Xhit})$ if the effects of the instruments on factor demands are homogeneous; that is, if the coefficients of the first stage equations are not random. However, if farmers have knowledge of the heterogeneity in productivity across their plots, the effects of the instruments on factor demands are also heterogeneous across plots because the effect of a change in the opportunity cost of an input on input demand will vary depending upon the marginal product of that factor. In this case, 2SLS is inconsistent for the expected values of the production function parameters.

Building on Florens et al (2008), Masten and Torgovitsky (2016) suggest an Instrumental Variables Correlated Random Coefficients (IVCRC) estimator, which allows for heterogeneity in the first-stage regressions of the instruments on the endogenous factor demands. The IVCRC estimator is a control function approach and (therefore) relies on an assumption of single-dimensional heterogeneity in the first stage equations. As in any control function estimator, the idea is that it is possible to invert the first-

stage factor demands to reveal the problematic heterogeneity. In the second-stage production function, factor inputs are orthogonal to the heterogeneity in productivity conditional on this control function. The IVCRC estimator uses a sequence of quantile regressions in the first stage to generate conditional (on the instruments) ranks for each observation; in the second stage, OLS estimation of (14) conditional on the estimated rank equal to r identifies $E(\alpha_{Liht}|Rank = r)$ and $E(\alpha_{Xhit}|Rank = r)$, and then averaging over ranks provides $E(\alpha_{Lhit})$ and $E(\alpha_{Xhit})$.

The IVCRC estimator relies on the assumption of single-dimensional unobserved heterogeneity in factor demands, which is equivalent to the assumption of rank invariance. Rank invariance means that the ordinal ranking of the demand for a factor on any two plots would be the same if both plots had the same realization of the instruments, for any realization of the instruments. However, we show in Appendix A4 that the factor demands implied by (14) have two dimensions of heterogeneity, one generated by unobserved variation in ω_{Lhit} and the other by a real-valued function of all the unobserved productivity variation and measurement error in the model. Therefore, in our setting, rank reversals are possible, and the assumptions required for the consistency of the IVCRC estimator are not strictly met. We show in Appendix A4, however, that the assumption of rank invariance holds *approximately*. Given the estimates generated by either the IVCRC or 2SLS estimators, in either Tanzania or Uganda, rank reversals occur in fewer than 1 percent of pairwise comparisons of observations. Therefore, the assumptions required for the consistency of the IVCRC estimator are approximately satisfied in our data. This finding is a consequence of the much smaller variation in one dimension of the unobserved heterogeneity (that driven by factor-specific productivity) than the second (generated by measurement error, risk and productivity).

4.1   Estimation Results

Tables 5A-B present OLS and quantile regression estimates of the determinants of land and labor inputs into production in Tanzania and Uganda, respectively.[27] These estimates serve as the first stages of the 2SLS and IVCRC production function estimates provided below. The first pair of columns in each table report selected coefficients of the OLS regression of log land area and log labor use on each plot; the second through fourth pairs of columns report the same set of coefficients for the 25th, 50th and 75th quantile regressions. The final pair of columns reports the differences of these coefficients in the 75th and 25th quantile regressions.

The penultimate row of Tables 5A-B shows that the instruments are strong predictors of plot-level land and labor demand. For example, in both Tanzania and Uganda, illness in the household reduces labor use in household plots. In Tanzania, when there is abundant growing season rain, farmers reduce labor on a given plot when a higher fraction of that household's other plots have good soil. In Uganda, when peak season rainfall is higher, farmers reduce labor on a given plot when a higher fraction of that household's other plots have fair or poor soil.

In both Tanzania and Uganda, plots controlled by male farmers are larger and use more labor. In Tanzania, farmers reduce the allocation of land to poor quality plots when they experience drought or flood conditions (the questionnaire does not distinguish). In Uganda, farmers increase the allocation of land to poor quality plots when they experience droughts.

If the production function has random coefficients, then the associated input demand functions will have heterogeneous coefficients as well. The IVCRC estimator was developed to allow for heterogeneity in the first

---

[27] Bootstrapped standard errors, clustered at the household-season level, are reported. The full sets of coefficient estimates are presented in Appendix Tables A1A and A1B.

stage regressions. The final pair of columns reports the difference in coefficients at the 75th and 25th percentiles of the factor demand quantile regressions, and the F-test that these differences are jointly zero for the instrumental variables. We strongly reject that these differences are zero. For example, in Tanzania, at the 25th percentile of the demand for land, the effect of more growing season rain when the household has more plots with good soil is positive, while at the 75th percentile, the effect is negative. Nevertheless, we present 2SLS estimates of the production function and their implications for estimates of the dispersion of unobserved productivities for comparison.

The 2SLS and IVCRC estimates of $E(\alpha_{Lhit})$ and $E(\alpha_{Xhit})$ in the production function for agricultural plots in Tanzania and Uganda are presented in Table 6A.

Columns 1 and 3 present 2SLS estimates of the Cobb-Douglas factor coefficients for Tanzania and Uganda; columns 2 and 4 provide the corresponding IVCRC estimates, with the log total value of crop output as the dependent variable. Crop-year-season-region fixed effects are included, as are a rich set of observable characteristics of land and labor, as well as plot-level observable shocks.[28] Reflecting the simple technology of Tanzanian and Ugandan smallholder agriculture, these coefficients imply a much larger share of income for land than is observed in typical

---

[28] Bootstrapped standard errors are reported. The full sets of coefficients are reported in Appendix Tables A2A and A2B. For Tanzania these include the sale value of land, distance of the plot from home and from the nearest road, three levels of soil quality, four soil types, the gender, health status, literacy and age of the plot manager, indicator variables for household-level drought or floods, crop disease or pests, severe water shortage or other shocks that led to crop loss, and growing season rainfall, interacted with soil type and soil quality dummies. For Uganda, four categories of soil quality, three sources of water, 6 indicators of plot toposequence, the level of erosion, indicators of the gender, literacy and access to agricultural advice of the plot manager, household-level indicators of drought and flood and their interactions with plot level soil quality, village level seasonal and annual rain and their interactions with plot level soil quality.

macroeconomic data, and a much smaller share for labor.[29] The preferred IVCRC estimates imply a larger share for labor and a smaller share for land than the 2SLS estimates.

With estimates $\hat{\alpha}_L = E(\alpha_{Lhit})$, $\hat{\alpha}_X = E(\alpha_{Xhit})$ and $\hat{\beta}_Y, \hat{\beta}_L$ and $\hat{\beta}_X$ in hand, we generate a first approximation, $\ln \widehat{TFP}^a_{hit}$, to the distribution of log TFP across plots. Our measure of TFP is the usual production function residual, as in equation (13). Figures 3A-B provide the empirical distribution of $\ln \widehat{TFP}^a_{hit}$ in Tanzania and Uganda using both the IVCRC and 2SLS estimates. The apparent dispersion is high: in Tanzania (Uganda) the variance based on the IVCRC estimates is 1.22 (1.26); the 90–10 log difference in TFP is 2.70 (2.69), corresponding to a 90–10 ratio of 15 (15) in TFP levels.[30]

Equations (16) and (17), however, clarified that $\ln \widehat{TFP}^a_{hit}$ incorporates the effects of late season agricultural shocks and measurement error in output and factors of production, and thus its dispersion is greater than the dispersion of unobserved total factor productivity.

Maintaining the assumption of within-farm efficient factor allocation, Table 6B presents the estimates of within-farm variation generated by risk, measurement error and heterogeneous productivity in the Tanzania and Uganda samples. The most striking feature of the table is the remarkable importance of late-season risk and measurement error in output in driving the apparent variation in output across plots within a farm. In both Tanzania and Uganda, using both the 2SLS and IVCRC production function estimates, this is the largest component of the unobserved variation in productivity across plots. In Uganda this

---

[29] In the United States a labor share in agriculture is often taken to be about 50%, with land perhaps 15%, and capital about 35% (e.g., Valentinyi and Herrendorf 2008).

[30] Using the 2SLS estimates of the production function, the variance for Tanzania (Uganda) is 1.27 (1.26) and the 90–10 log difference in TFP is 2.72 (2.69).

component is especially dominant. There is important measurement error in land and labor inputs as well. Variation in plot-level productivity observed by the farmer, but unobserved to us $\left(\sigma_q^2\right)$ is also important, though less so in Uganda, where as noted above most plots of a given farmer are spatially contiguous or nearly so. Cross-plot within-farm variation in factor specific productivity also exists, but is relatively less important. The results are similar whether they are based on production function parameters estimated using 2SLS or IVCRC techniques, with the exception that the IVCRC estimates in Tanzania do not show any significant role for land measurement error in explaining apparent productivity variation intensity.

Table 6B provides evidence on the across-plot distributions of productivity shocks and measurement errors within farms. We use these estimates in equation (18) to calculate a revised estimate of plot-level total factor productivity based on the assumption that the overall variances of $\sum \alpha_J \epsilon_{Lhit}$ and ϵ_{Yhit} across all farms are no smaller than these within-farm variances. The variance of $\ln \widehat{TFP}_{hit}^b$ is an upper bound on the dispersion of unobserved total factor productivity across plots more generally.[31]

In both Tanzania and Uganda, accounting for measurement error in factors of production and output and for late-season shocks dramatically reduces the apparent dispersion of TFP across plots. In Tanzania, the variance of the naïve log residual of the production function was 1.22. The variance of the estimate corrected for measurement error and late season risk falls to 0.36. The 90–10 log difference in estimated TFP falls from 2.70

---

[31] It is of course possible that a single farmer may, for a variety of reasons, pursue an optimization strategy that would lead to very different outcomes on different plots and thus to a high within-farm variance. But other farmers will then face similar problems and will realize similarly disparate outcomes. Across all farms, this variation will be amplified by the differing location-specific factors that affect production, so that the aggregate variation is higher than the average within-farm variation. This intuition is formalized in Appendix A3.

to 1.48, corresponding to a drop in the 90-10 ratio of TFP from 14.8 to 4.4. In Uganda, the correction is even more dramatic. The variance of the log production function residual is 1.44, falling to 0.13 when corrected. The 90–10 log difference in estimated TFP falls from 2.76 to 0.85, corresponding to a correction in the 90–10 ratio of TFPs from 15.8 to 2.3.[32]

The effect of correcting the estimates of unobserved total factor productivity for measurement error and late season risk is visually apparent in Figures 3A-B. These figures provide kernel estimates of the densities of $\ln \widehat{TFP}^a_{hit}$ and $\ln \widehat{TFP}^b_{hit}$ for Tanzania and Uganda, for both the IVCRC and 2SLS estimates of the production function. We note that the patterns of dispersion are robust to the choice of estimation method.

We have examined three within-farmer frictions that could restrict the efficient reallocation of factors across plots within farms. First, a farmer cultivating widely dispersed plots could face varying input costs across these plots. When we repeat the estimation procedure restricting attention to close-by plots, however, our results change little, as reported in Appendix A5.[33] The IVCRC estimate of the dispersion of $\ln \widehat{TFP}^b_{hit}$ is actually larger when attention is restricted to nearby plots. Second, land tenure restrictions across physical space could interfere with the smooth reallocation of land across plots of a farmer. We repeat the procedure using only plots contained within a single parcel of each farmer, and again find

---

[32] The corrections are similarly dramatic using 2SLS estimates of the production function. Using these estimates in Tanzania, the variance of unobserved log TFP falls from 1.27 to 0.41, and the $90 - 10$ difference falls from 2.72 to 1.55, corresponding to a fall in the $90 - 10$ ratio of TFP from 15.2 to 4.7. Using the 2SLS estimates in Uganda, the variance of unobserved log TFP falls from 1.38 to 0.53. The 90–10 log difference falls from 2.82 to 1.74, corresponding to a fall in the 90–10 ratio of TFP from 16.7 to 1.7.

[33] Close-by plots in Tanzania are defined as plots within 1 kilometer, or alternatively 100 meters of the home. In Uganda they are plots within a 15 minute walk of home, or alternatively plots within the same parcel of contiguous land.

very similar results. Finally, some of the "farmers" are instead particular pairs of individuals farming jointly; there may be inefficiencies in the allocation of resources in such circumstances. The results, however, are robust to restricting attention to single-person "farmers", except again that the IVCRC estimate of the dispersion of $\widehat{\ln TFP}^b_{hit}$ increases.

4.2   Implications for Characterizing Misallocation

Late-season production shocks and measurement error in factors of production and output together account for about half to two-thirds of the variance in log productivity residuals. Since these are not susceptible to reallocation, the aggregate productivity gains that could be attained from a hypothetical reallocation exercise are correspondingly smaller. The effect of correcting estimates of productivity for the effects of risk and measurement error on estimates of the magnitude of misallocation in economy depend, of course, on the specifics of the reallocation exercise. Any reallocation exercise must impose a great deal of structure on what is ultimately an artificial exercise. For example, results will be sensitive to whether land alone is reallocated to the best farmers, or whether labor is allowed to move along with land. If land is to be reallocated, is it limited to within-village or within-region reallocation? The exact magnitude of the reallocation gains will depend on the extent to which misallocation gives rise to systematic rank reversals, as shown in Hopenhayn (2014). Any of these calculations can be carried out with our estimates in hand, but there is no real discipline on the exercise from either theory or practice. As an alternative approach, in what follows, we have preferred to impose less structure and simply to focus on the dispersion of productivity.

However, a simple calculation serves to illustrate the order of magnitude of this effect. Consider a Cobb Douglas production function without factor-specific productivity heterogeneity or measurement error in factors of production:

$$Y_i = e^{\omega_i + \epsilon_i}(L_i)^{\alpha_L}(X_i)^{\alpha_X} \qquad (20)$$

where ω_i is total factor productivity, known to the producer, and ϵ_i is measurement error in output or output risk that is realized after factors are committed.

Relative to an existing baseline allocation, the gains to efficient reallocation are proportional to $e^{\frac{\sigma_\omega^2}{2(1-\alpha_L-\alpha_X)^2}}$, as shown in Appendix 6, and therefore depend on the dispersion of total factor productivity and the concavity of the production function. An overestimate of the variance of TFP across producers, caused, for example, by misinterpreting measurement error or pure risk as variation in total factor productivity, leads directly to an overestimate of the potential gain from reallocation.

Let $\sigma_A^2 = Var\big(\ln \widehat{TFP}_{hit}^a\big) > Var\big(\ln \widehat{TFP}_{hit}^b\big) = \sigma_B^2$. The overstatement of the potential gains from reallocating resources from an existing baseline to the efficient allocation from assuming that the variance of total factor productivity is σ_A^2 rather than σ_B^2 is

$$\frac{\frac{\overline{Y^e}(\sigma_A^2)}{\frac{1}{N}\Sigma_i Y_i}}{\frac{\overline{Y^e}(\sigma_B^2)}{\frac{1}{N}\Sigma_i Y_i}} = e^{\sigma_A^2 - \sigma_B^2}. \qquad (21)$$

In Tanzania, this ratio is 2.6. In Uganda, the overstatement of the gains to correcting misallocation is 3.7. The extent of misallocation is substantially overstated if the contributions of risk and measurement error to the apparent dispersion of total factor productivity are neglected. This calculation is independent of the particular estimate of the production function: the amount of gain from a hypothesized reallocation depends on concavity, but the relative overstatement generated by overestimating the variance of total factor productivity is independent of the production function parameters. Similarly, the degree of overstatement is independent of many of the particulars of the hypothesized reallocation. For example, if the thought experiment is to leave one factor in its current

(mis)allocation and optimally reallocate the other, the conclusion of (21) remains unchanged.

5. Discussion

The results from Tanzania and Uganda show the importance of accounting carefully for measurement error, shocks, and heterogeneity in technology (including input quality) in measuring productivity at the level of individual production units. These issues have previously been raised in critiques of the literature on misallocation, but the data from African farms provide sufficiently rich detail that we can begin to disentangle the different sources of productivity dispersion. Our analysis suggests that previous estimates of misallocation have probably overestimated the potential productivity losses due to misallocation (or, equivalently, the gains from efficient reallocation). The macro significance of any potential gains from reallocation should be reconsidered accordingly.

Given that reallocation would also entail massive costs – not least, in terms of the social welfare implications of reallocating land away from many poor smallholders in Africa – we believe that these findings are important for their own sake. But in addition, we believe that there are additional implications for the broader literature that has grown up around the topic of misallocation in development and growth. Much of this literature has relied on cross-section data and has assumed that firms are observed without error. The literature has also tended to assume that all firms operate precisely the same technology, with all parameters of the production function known exactly. In our context, these assumptions would lead to flawed conclusions. Even though our data have been carefully collected with highly trained enumerators – and although they

are often characterized as "state of the art" surveys – measurement error is pronounced, and shocks are quantitatively important.

There are limits to our analysis. As noted in the introduction, we cannot rule out the importance of misallocation in a dynamic sense. The current allocation of land and labor across farms may be relatively efficient in a static sense, but improved technologies might be well suited to very different allocations. For instance, mechanization and tractor use might increase efficiency in these countries, but it is possible that the current distribution of land might make it unprofitable to use tractors and might thus slow the diffusion of the new technologies. Thus, one could think about a dynamically optimal allocation, which would raise different issues than those we have addressed here.

We also emphasize that our paper does not argue that the current scale of farming is optimal. Our analysis has nothing to say about the efficient allocation of resources across sectors. For example, we cannot say whether there are too many people working in agriculture. It seems plausible that some of the resources currently deployed in agriculture could be productively reallocated to other sectors; however, we cannot assess this proposition with our data.

This paper also suggests that within the literature on agriculture and development, there is a need to pay close attention to heterogeneity in unobservable characteristics of plots. These may be linked to soil and land quality, which vary in quantitatively significant ways at very fine geographic scale. But there may also be a high degree of spatial variation in shadow prices (reflecting, for example, within-farm transport costs). For instance, the distances from one end of a plot to another may create consequential transport and transaction costs for the application of organic

fertilizers or for the shadow price of output that must be carried to the household or to market. The importance of heterogeneity has been emphasized in recent work on technology adoption (e.g., Suri 2011), and it is surely important for other issues in agricultural development.

In further work, an interesting area to explore is the trade-off between farm scale and the precision of input application. Because input use is (optimally) calibrated to the average quality of a plot, there is a trade-off between increasing the size of the plot (which reduces the fixed cost per unit output) and the loss of profits that comes from applying inputs more crudely. This trade-off may have some power in explaining the tendency of smallholder agriculture in the developing world to rely so heavily on very small plots, finely tuned in terms of crop choice and input use. Previous explanations of small plot size have tended to focus on risk and diversification, but our analysis suggests that there may also be important efficiency arguments.

6. Conclusions

This paper has examined the importance of misallocation across firms as an explanation for low aggregate productivity in developing countries, using data from agriculture in Tanzania and Uganda. A challenge in this kind of analysis is that misallocation is not the only potential source of dispersion in productivity. Some of the other sources of dispersion are not susceptible to improvement through efficient reallocation. In particular, reallocation will not lead to increases in output if dispersion is primarily an artefact of measurement error. Reallocation will also prove futile to the extent that dispersion results from idiosyncratic shocks that occur after inputs have been (efficiently) applied.

Our paper takes advantage of rich data at the plot level to disentangle the different sources of productivity dispersion. We begin by showing that dispersion in productivity is not simply a feature of the cross-farm data; perhaps surprisingly, within-farm dispersion is quite large. This suggests that differences in farmer quality are not sufficient to account for the patterns of dispersion that we observe in the data.

We estimate agricultural production functions for Tanzania and Uganda, with a framework that draws on the sequential nature of production decisions. The estimated production functions can be used to assess the potential gains from reallocation. Our finding is that misallocation does indeed affect aggregate agricultural output in these countries, but commonly used approaches in the literature overstate the dispersion of log TFP by about 100 percent. The gains from a hypothetical reallocation are thus correspondingly overstated by a factor of three or four. Based on our estimates, reallocation can generate non-trivial gains in aggregate output, but not enough to narrow significantly the large cross-country income differences.

Beyond the rather special case of African agriculture, this research points to the need for caution in estimating the impact of misallocation. Not all dispersion in productivity at the firm level reflects misallocation. It is important to for researchers to consider other sources of productivity dispersion, including heterogeneity and measurement error.

Appendix A1: Endogenous plot selection

This appendix describes the process through which a farmer (household) chooses the number and locations of its plots.

Consider first the household's option of producing on a single plot, $[0, L_h]$, making use of the entire land endowment. The profit maximization problem is then given by:

$$\max_{X_h} \left[\left(\frac{X_h}{L_h}\right)^{\theta} \int_0^{L_h} \gamma_h(k,s)\zeta_h(k)\, dk - w_h X_h - c \right]. \qquad (A1.1)$$

As an alternative to the single plot, the household could instead farm multiple plots. We assume that the household divides its landholding into plots at the start of the season, before inputs are chosen and – crucially – before the realization of the productivity shock. In modelling the farm in this way, we seek to capture the notion that inputs can be adjusted through most of the growing season, so that the total input vector responds to the shocks. But plot boundaries cannot normally be adjusted once planting has taken place – and indeed, plot boundaries are often set even before planting, with a series of decisions that commit the household to planting certain crops at certain moments. For instance, the timing and techniques of land preparation will be linked to decisions about plot boundaries and potentially also crop choice.

Consider first the problem of a household that is choosing a single boundary that will define two plots. Denote the threshold location between the two plots as L_{h1}, so that the two plots are $[0, L_{h1}]$ and $[L_{h1}, L_h]$. In this case, an interior solution for the size of the two plots must hold; expected total profits could not be increased by moving this location either to the left or the right on the number line.

The profit maximization problem can be written as:

$$
\max_{L_{h1}} \int_{s \in S} \left[\max_{X_{h1}, X_{h2}} \left[\left(\frac{X_{h1}}{L_{h1}} \right)^{\theta} \int_0^{L_{h1}} \gamma_h(k, s) \zeta_h(k) \, dk \right. \right.
$$

$$
+ \left(\frac{X_{h1}}{L_h - L_{h1}} \right)^{\theta} \int_{L_{h1}}^{L_h} \gamma_h(k, s) \zeta_h(k) \, dk - w_h X_{h1} \qquad \text{(A1.2)}
$$

$$
\left. \left. - w_h X_{h2} - 2c \right] \right] d\Delta(s).
$$

In effect, the household chooses the plot boundary L_{h1} to maximize expected profits, knowing what input bundle it would choose for each plot for every realization of the productivity shock $\gamma_h(k, s)$. The problem is well-defined.

Now consider a household that farms I plots, $I > 2$. We use the notation that L_{hi} will denote the right-hand boundary of the ith plot; i.e., the boundary between plot i and plot $i + 1$. For notational convenience, we set $L_{h0} = 0$ and $L_{hI} = L_h$. Then $\{L_{hi}\}_{i=0}^I$ is the sequence of plot boundaries. The first plot is given by the interval $[0, L_{h1}]$, and the ith plot covers the interval $[L_{hi-1}, L_{hi}]$, continuing to the Ith plot, which covers $[L_{hI-1}, L_h]$.

We assume for convenience in what follows that all the plots are of sufficient quality that they will be actively farmed, allowing for an interior solution. The logic of the analysis would extend, however, to a situation in which the household chooses not to cultivate some portion of its land.

For notational convenience, let the size of the ith plot be denoted as $\tilde{L}_{hi} \equiv (L_{hi-1} - L_{hi})$. As before, the average productivity of plot i, conditional on the realization of the shock $\gamma_h(k,s)$, can be written as $\zeta_{hi} = \frac{1}{\tilde{L}_{hi}} \int_{L_{hi-1}}^{L_{hi}} \gamma_i(k,s)\zeta_i(k)\,dk$.

Then the household's problem of choosing the boundaries of I plots can be written as:

$$E\hat{\pi}(I) = \max_{\{L_{hi}\}_{i=1}^{I}} \int_{s \in S} s \left[\max_{\{X_{hi}\}_{i=1}^{I}} \left[\zeta_{hi}\tilde{L}_{ij}X_{hi}^{\theta} - \sum_{i=1}^{I} w_h X_{hi}\tilde{L}_{hi} \right. \right.$$
$$\left. \left. - cI \right] \right] d\Delta(s). \tag{A1.3}$$

How many plots might the household farm? We can identify a finite maximum number of plots for any household. Because the problem in equation (7) is well-defined for any number of plots I, we use this to define an upper bound for I. Recall that for a single location k, the household can maximize profits conditional on the shock s, by choosing a point-specific input bundle. This gives output $q_h^*(k,s) = \zeta_h(k)\gamma_h(k,s)\left(\frac{\theta\gamma_h(k,s)\zeta_h(k)}{w_h}\right)^{\frac{\theta}{1-\theta}}$ with corresponding profits of $\pi_h^*(k,s) = q_h^*(k,s) - w_h\xi_h^*(k,s)$. Across the entire land holding of the household, this gives rise to an expression for the maximum profits that can be earned, conditional on the shock s, with $c = 0$: $\pi_h^*(s) = \int_0^{L_h} \pi_h^*(k,s)\,dk$. This expression can be understood as the "precision agriculture profits" in which every location on the household's land holdings is farmed with optimal point-specific inputs. Integrating over possible realizations of the shock s, then $\pi_h^* = \int_{s \in S} \pi_h^*(k,s)\,d\Delta(s)$ is the expected maximum profits. Given this, $I^* = \left(\frac{\pi_h^*}{c} + 1\right)$ is an upper bound for the number of plots that can be profitably cultivated.

With this upper bound defined, the household's choice of its optimal number of plots reduces to a discrete optimization, with $\hat{I} = \text{argmax}_j \{E\hat{\pi}(j)\}_{j=1}^{I^*}$.

We now consider the relationship between plot quality and plot size within a farm. A simple illustration is provided by the special case of a farmer who has access to multiple physical parcels, each of unit size. Parcel i has average productivity $\zeta_{hi} = \int_0^1 \gamma_i(k,s)\zeta_i(k)\,dk$. If that parcel can be partitioned into two plots (A and B) of any size such that $\zeta_{hi}^A \neq \zeta_{hi}^B$, then there exists a scalar $z^* \geq 0$ such that $\forall z \geq z^*$, if we replace $\zeta_i(k)$ with $\zeta_{zi}(k) = z\zeta_i(k)$, it is optimal to split the parcel into more than one plot. Therefore, if a parcel is divided into multiple plots, then a more productive parcel is also divided. And a sufficiently less productive parcel will not be.

Define $\pi_{1i} = \zeta_{hi}\left(\frac{\zeta_{hi}\theta}{w_h}\right)^{\frac{\theta}{1-\theta}} - w_h\left(\frac{\zeta_{hi}\theta}{w_h}\right)^{\frac{1}{1-\theta}}$ as the profit from farming the parcel as a unit. Let L_{hi}^A and $L_{hi}^B = 1 - L_{hi}^A$ be the areas of the two plots that optimally divide parcel i (the solution to (A1.2)). So $\pi_{1i}^A = L_{hi}^A\zeta_{hi}^A\left(\frac{\zeta_{hi}^A\theta}{w_h}\right)^{\frac{\theta}{1-\theta}} - w_h L_{hi}^A\left(\frac{\zeta_{hi}^A\theta}{w_h}\right)^{\frac{1}{1-\theta}}$ and $\pi_{1i}^B = L_{hi}^B\zeta_{hi}^B\left(\frac{\zeta_{hi}^B\theta}{w_h}\right)^{\frac{\theta}{1-\theta}} - w_h L_{hi}^B\left(\frac{\zeta_{hi}^B\theta}{w_h}\right)^{\frac{1}{1-\theta}}$. Define $\zeta_{zi} = \int_0^1 \gamma_i(k,s)\zeta_{zi}(k)\,dk = z\zeta_{hi}$ as the average productivity of the z-transformed parcel, and π_{zi}, π_{zi}^A, and π_{zi}^B as the profits from farming the full parcel, and optimally subdivided if the productivity process is $\zeta_{zi}(k)$. Finally, define $\tilde{\pi}_{zi}^A = L_{hi}^A\zeta_{zi}^A\left(\frac{\zeta_{zi}^A\theta}{w_h}\right)^{\frac{\theta}{1-\theta}} - w_h L_{hi}^A\left(\frac{\zeta_{zi}^A\theta}{w_h}\right)^{\frac{1}{1-\theta}}$, and similarly $\tilde{\pi}_{zi}^B$ as the maximized profits generated on plots A and B of the z-transformed parcel, where plots A and B are defined by the optimal partition of the parcel given its original productivity.

$\zeta_{hi}^A \neq \zeta_{hi}^B$ implies that $\frac{X_{1i}^A}{L_{1i}^A} \neq \frac{X_{1i}^B}{L_{1i}^B}$ so for $\theta < 1$,

$$\pi_{1i} = L_{hi}^A\pi_{1i} + L_{hi}^B\pi_{1i} < \pi_{1i}^A + \pi_{1i}^B.$$

Suppose $\pi_{1i}^A + \pi_{1i}^B - c > \pi_{1i}$. Then for all $z \geq 1$,

$$\pi_{zi}^A + \pi_{zi}^B - c \geq \tilde{\pi}_{zi}^A + \tilde{\pi}_{zi}^B - c = z^{\frac{1}{1-\theta}}\left(\pi_{1i}^A + \pi_{1i}^B\right) - c$$
$$> z^{\frac{1}{1-\theta}}\pi_{1i} = \pi_{zi}. \tag{A1.4}$$

Therefore, if a parcel is divided into more than one plot, then any more productive parcel is also divided. Conversely, for a sufficiently low value of z, $\pi_{zi}^A + \pi_{zi}^B < c$ and it is not feasible to divide the parcel.

Appendix A2: Estimating the within-farm variances of measurement error, late-season risk, and unobserved productivity

We consider a plot i farmed by household (farmer) h in season t. We define log TFP for the plot, inclusive of the plot-specific factor productivities, as

$$z_{hit} \equiv \frac{1}{1 - \sum_J \alpha_{Jhit}} \left\{ W_{Ehit}\beta_E + \omega_{Yhit} + \alpha_{Lhit} \ln\left(\frac{\alpha_{Lhit}}{p_{Lht}}\right) + \alpha_{Xhit} \ln\left(\frac{\alpha_{Xhit}}{p_{Xht}}\right) \right\} \tag{A2.1}$$

We write log output and (actual, not observed) factor demand on the plot as

$$y_{hit} = W_{Hhit}\beta_H + \epsilon_{Yhit} + z_{hit}$$
$$l_{hit} = \ln(\alpha_{Lhit}) - \ln(p_{lht}) + z_{hit} \tag{A2.2}$$
$$x_{hit} = \ln(\alpha_{Xhit}) - \ln(p_{Xht}) + z_{hit}$$

The IVCRC procedure provides us with an estimate of the means of the distribution of the factor productivity coefficients, $\hat{\alpha}_L$ and $\hat{\alpha}_X$. We will work in terms of observable inputs, and output, adjusted for the estimated effects of observed characteristics

$$y_{hit} - W_{Hhit}\hat{\beta}_H = \epsilon_{Yhit} + z_{hit}$$
$$l^o_{hit} + W_{Lhit}\hat{\beta}_L = \hat{\alpha}_L + \omega_{Lhit} - \ln(p_{Lht}) + \epsilon_{Lhit} + z_{hit} \tag{A2.3}$$
$$x^o_{hit} + W_{Xhit}\hat{\beta}_X = \hat{\alpha}_X + \omega_{Xhit} - \ln(p_{Xht}) + \epsilon_{Xhit} + z_{hit}$$

ω_{Lhit} and ω_{Xhit} are plot level productivities of land and labor and and z_{hit} is plot level total productivity.

$$\tilde{y}_{hit} \equiv y_{hit} - \bar{y}_{h.t} - (W_{Hhit} - \bar{W}_{Hh.t})\hat{\beta}_H$$
$$= \epsilon_{Yhit} - \bar{\epsilon}_{Yh.t} + z_{hit} - \bar{z}_{h.t}$$
$$\tilde{l}_{hit} \equiv l^o_{hit} - \bar{l}^o_{h.t} + (W_{Lhit} - \bar{W}_{Lh.t})\hat{\beta}_L \tag{A2.4}$$
$$= \omega_{Lhit} - \bar{\omega}_{Lh.t} + \epsilon_{Lhit} - \bar{\epsilon}_{Lh.t} + z_{hit} - \bar{z}_{h.t}$$

$$\tilde{x}_{hit} \equiv x_{hit}^o - \bar{x}_{h.t}^o + (W_{Xhit} - \overline{W}_{Xh.t})\hat{\beta}_X$$
$$= \omega_{Xhit} - \bar{\omega}_{Xh.t} + \epsilon_{Xhit} - \bar{\epsilon}_{Xh.t} + z_{hit} - \bar{z}_{h.t}$$

The LHS of these are observable. Their covariance (and a normalization discussed below) provides us with sufficient information to identify the within-farm variances of plot-level total factor productivity (σ_z^2), factor-specific productivity and their covariance ($\sigma_L^2, \sigma_X^2, \sigma_{LX}$), factor measurement error ($\sigma_{\epsilon L}^2, \sigma_{\epsilon X}^2$) and output measurement error and post-input risk $(\sigma_{\epsilon y}^2)$, as well as the covariance of plot-level total factor productivity and factor-specific productivity (σ_{zL}, σ_{zX}):

$$var(\tilde{y}_{hit}) = \sigma_z^2 + \sigma_{\epsilon Yhit}^2$$
$$var(\tilde{l}_{hit}) = \sigma_L^2 + \sigma_{\epsilon L}^2 + \sigma_z^2 + 2\sigma_{zL}$$
$$var(\tilde{x}_{hit}) = \sigma_X^2 + \sigma_{\epsilon X}^2 + \sigma_Q^2 + 2\sigma_{zX}$$
$$cov(\tilde{y}_{hit}, \tilde{l}_{hit}) = \sigma_{zL} + \sigma_z^2$$
$$cov(\tilde{y}_{hit}, \tilde{x}_{hit}) = \sigma_{zX} + \sigma_z^2$$
$$cov(\tilde{l}_{hit}, \tilde{x}_{hit}) = \sigma_{LX} + \sigma_{zL} + \sigma_{zX} + \sigma_z^2$$

(A2.5)

We will not separately identify variation in all three types of unobserved heterogeneity in factor ($\omega_{Lhit}, \omega_{Xhit}$) or total-factor productivity (z_{hit}): a parallel increase in ω_{Lhit} and ω_{Xhit} is equivalent to an increase in z_{hit}. Hence, we normalize $\omega_{Lhit} + \omega_{Xhit} = 0$. Intuitively, a change in ω_{Lhit} relative to ω_{Xhit} is a change in the slope of an isoquant; a change in z_{hit} is a shift in or out of an isoquant. The normalization of factor specific productivities distinguishes these from TFP shocks; this normalization adds the restrictions

$$\sigma_L^2 = \sigma_X^2$$
$$\sigma_{LX} = -\sigma_L^2$$
$$\sigma_{zL} = -\sigma_{zX}$$

(A2.6)

From equations (A2.5)-(A2.6) we calculate the parameters $(\hat{\sigma}_z^2, \hat{\sigma}_L^2, \hat{\sigma}_X^2, \hat{\sigma}_{\epsilon y}^2, \hat{\sigma}_{\epsilon L}^2, \hat{\sigma}_{\epsilon X}^2, \hat{\sigma}_{LX}, \hat{\sigma}_{zL}, \hat{\sigma}_{zX})$ that are consistent with the observed covariance of plot level output and inputs across plots within farms, given an estimate of the production function parameters and the assumption of efficient allocation across plots within a farm.

Appendix A3: Measurement error / shock variances across all plots and average across farmers of within-farmer variances

We estimate the mean, across farmers, of the within-farm, cross-plot variance of measurement errors in factor inputs and of measurement error and late-season shocks to output. How does this compare with the overall variance, across all plots, of these measurement errors/random shocks?

Denote by y_{fi} the realization of any of these errors/shocks $\epsilon_{Yhit}, \epsilon_{Lhit}, \epsilon_{Xhit}$.[34] Let N be the total number of plots, N^f the number of farmers and N_f^i be the number of plots of farmer f. The average across farmers of the cross-plot within-farmer variance of y is $\sigma_F^2 \equiv$ $\frac{1}{N^f}\sum_{f=1}^{N^f}\frac{1}{N_f^i}\sum_{i=1}^{N_f^i}(y_{fi}-y_f)^2$. The variance of y across plots in the sample is $\sigma^2 \equiv \frac{1}{N}\sum_{f=1}^{N^f}\sum_{i=1}^{N_f^i}(y_{fi}-y)^2$. If there are no farmer effects in measurement error or the late season shock to output, then $y_f = y \; \forall f$ and $\sigma_F^2 = \sigma^2$.

However, if there is variation across farmers in the mean level of measurement error or the late season shock, then the average across farmers of the within-farmer variance may differ from the variance across all plots. The largest number of plots cultivated by a single farmer is k. We partition the sample of farmers into sets $\{M_1, M_2, \ldots, M_k\}$ such that each farmer $f \in M_k$ has k plots. With some abuse of notation we denote the cardinality of each set M_k as M_k. Then we have

$$\sigma_{Fk}^2 = \frac{1}{M_k}\sum_{f \in M_k}\frac{1}{k}\sum_{i=1}^{k}(y_{fi}-y_f)^2$$

[34] We drop the t subscript for this appendix; the calculations should be understood as occurring within any season.

$$\sigma_k^2 = \frac{1}{kM_k} \sum_{f \in M_k} \sum_{i=1}^{k} (y_{fi} - y_k)^2$$

With these sets defined, the overall variance of y can be defined as

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^{k} \sum_{f \in M_k} \sum_{i=1}^{k} y_{fi}^2 - \left(\frac{1}{N} \sum_{k=1}^{k} kM_k \overline{y_k} \right)^2$$

$$\geq \frac{1}{N} \sum_{k=1}^{k} \sum_{f \in M_k} \sum_{i=1}^{k} y_{fi}^2 - \frac{1}{N} \sum_{k=1}^{k} M_k k \; \overline{y_k}^{\;2} \qquad \text{(A3.1)}$$

$$= \frac{1}{N} \sum_{k=1}^{k} [M_k k \sigma_k^2].$$

where the inequality follows from convexity (and is a strict equality if $y_f = y \; \forall f$. The average across farmers of the variance of y is

$$\sigma_F^2 = \frac{1}{N_F} \sum_{k=1}^{k} M_k \sigma_{Fk}^2.$$

So

$$\sigma^2 - \sigma_F^2 \geq \sum_{k=1}^{k} \left(\frac{k}{N} - \frac{1}{N_F} \right) M_k \sigma_{Fk}^2. \qquad \text{(A3.2)}$$

If each farmer has the same number of plots, then the weak inequality in (2) is an equality, and $\sum_{k=1}^{k} \left(\frac{k}{N} - \frac{1}{N_F} \right) M_k \sigma_{Fk}^2 = 0$ and the average across farmers of the within-farmer variance of plot yield is the same as the overall variance of plot yields.

Note that $\left(\frac{k}{N} - \frac{1}{N_F} \right)$ is increasing in k. If the average number of plots per farmer is less than or equal to 2, then $\left(\frac{k}{N} - \frac{1}{N_F} \right) \geq 0$ for all k and $\sigma^2 - \sigma_F^2 \geq 0$. The average number of plots per farmer in Tanzania is 1.95. Therefore, the average across farmers of the within-farmer variance of y is less than the overall variance of y in Tanzania.

In Uganda the average number of plots per farmer is 2.7. If the average variance of y across plots of farmers who have only 2 plots is much larger than the average variance of y across plots of farmers who have many more plots, than it is possible that the RHS of (A2) is negative. Given the observed number of plots (N), number of farmers (N_F) and numbers of farmers cultivating k plots (M_k), then we can calculate that if $\sigma_{F2}^2 \leq 3.82 * \sigma_{Fk}^2$ for $k > 2$, then $\sum_{k=1}^{k} \left(\frac{k}{N} - \frac{1}{N_F} \right) M_k \sigma_{Fk}^2 > 0$. That is, as long as the average variance across plots of y of farmers cultivating 2 plots is no more than about 4 times as large as the average variance across plots of y of farmers cultivating more than 2 plots, then the overall variance of y across plots is larger than the average across farmers of within-farmer cross-plot variance of y.

It should noted that the variance across farmers of the mean farmer level shock y_f, which is relevant if one were to conduct the analysis at the farm level, can be either larger or smaller than σ^2, depending upon the covariance of y_{fi} within f. For example, if measurement error in plot size resulted largely from mistakes regarding the boundaries between plots within a parcel, so that land used on one plot was mistakenly attributed to another, then the negative covariance of l_{fi} within farmers would imply that the variance of land measurement area across farms might be less than its variance across plots.

Appendix A4: Consistency of the Instrumental Variables Correlated Random Coefficients Estimator

Assumption I2 of Masten and Torgovitsky (2016) requires that the unobserved heterogeneity in any of the endogenous variables by single dimensional. This is equivalent to the assumption of rank invariance, which "means that the ordinal ranking of any two agents in terms of …[land or labor demand] would be the same if both agents received the same realization of [the instrument set] Z, for any realization of Z." (Masten and Torgovitsky, p. 10)

Our instruments Z_i are determinants of the shadow prices of labor and land.

$$ln(p_{Xht}) = a_{Xi} + Z_i\beta_{pX} + \gamma_{pX}W_i$$

$$ln(p_{Lht}) = a_{Li} + Z_i\beta_{pL} + \gamma_{pL}W_i$$

Substitution in factor demands (A2.3) yields first stage factor demand functions of the form of our demand functions is:

$$l_i^o = f_{l1}(\epsilon_{Yi}, \epsilon_{Li}, \epsilon_{Xi}, \omega_{Yi}, \omega_{Li}; W_i) + f_{l2}(Z_i) + \omega_{Li}\frac{[Z_i(\beta_{pX} - \beta_{pL})]}{1 - \alpha_X - \alpha_L}$$

$$x_i^o = f_{x1}(\epsilon_{Yi}, \epsilon_{Li}, \epsilon_{Xi}, \omega_{Yi}, \omega_{Li}; W_i) + f_{x2}(Z_i) + \omega_{Li}\frac{[Z_i(\beta_{pX} - \beta_{pL})]}{1 - \alpha_X - \alpha_L} \qquad \text{(A4.1)}$$

where $\epsilon_{Yi}, \epsilon_{Li}, \epsilon_{Xi}, \omega_{Yi}, \omega_{Li}$ are unobserved productivities, shocks and measurement errors, W_i are observed characteristics that affect both factor demand and enter the production function, $\beta_{pX}, \beta_{pL}, \alpha_X, \alpha_L$ are estimated parameters, and Z_i is our vector of instruments.[35]

---

[35] For this appendix, we replace the plot-individual-year subscripts with single plot subscripts (e.g., l_i^o instead of l_{hit}^o) because the distinctions between farmers and time are not relevant to the discussion.

$f_{i1}(.)$, for example, is a scalar random variable. So is ω_L. So the unobserved heterogeneity in the demand for land is two dimensional, violating I2. The same is true for the heterogeneity in the demand for labor.

Define

$$l_i^o(Z) \equiv f_{l1}(\epsilon_{Yi}, \epsilon_{Li}, \epsilon_{Xi}, \omega_{Yi}, \omega_{Li}; W_i) + f_{l2}(Z) + \omega_{Li}\frac{[Z(\beta_{pX} - \beta_{pL})]}{1 - \alpha_X - \alpha_L}$$

which defines land demand given $(\epsilon_{Yi}, \epsilon_{Li}, \epsilon_{Xi}, \omega_{Yi}, \omega_{Li}; W_i)$ for any value of the instrument vector Z. The estimation procedure requires rank invariance, which is the requirement that

$$\left[l_i^o(Z') - l_j^o(Z')\right] \times \left[l_i^o(Z'') - l_j^o(Z'')\right] > 0$$

for any values of the instrument vector $\{Z', Z''\}$, and any values of $(\epsilon_{Yi}, \epsilon_{Li}, \epsilon_{Xi}, \omega_{Yi}, \omega_{Li}; W_i)$ and $(\epsilon_{Yj}, \epsilon_{Lj}, \epsilon_{Xj}, \omega_{Yj}, \omega_{Lj}; W_j)$.

Rank invariance does not hold for (A4.1) in general. However, given our estimates, we show that violations are rare. We proceed in two steps. First, we randomly assign values of the instrumental variables (drawn from the sample distribution) to randomly chosen pairs of plots (drawn from the sample distribution and estimated distribution of unobserved heterogeneity) to quantify the frequency of rank reversals. For Tanzania, rank reversals occur in less than .0001 of cases for labor and less than .0003 cases for land. For Uganda, they occur in less than 0.007 cases for either land or labor.

Note as well that if a rank reversal ever occurs for a given pair of plots, it always occurs at the extreme values of $Z_i(\beta_{pX} - \beta_{pL})$. So we couple this analysis with a calculation of the frequency of violations of rank invariance when pairs of plots are assigned values of the instruments that correspond to extreme values of $(\beta_{pX} - \beta_{pL})$. . Even in this case, rank reversals are rare. For Tanzania, they occur in .001 of cases; for Uganda, the reversal rate is 0.050 for land and 0.056 for labor.

Appendix A5: Robustness Checks on Restricted Samples

We relax the assumption that the allocation of factors across plots within a farm is efficient. We replace this assumption with the assumption that

- (a) the allocation of factors is efficient across "nearby" plots within a farm; or
- (b) the allocation of factors is efficient across plots within a given contiguous parcel of a farm; or
- (c) the allocation of factors is efficient across plots within a farm cultivated by a single individual (rather than jointly cultivated by a specific pair of individuals).

Appendix table A3 reports the consequences of these relaxations on our estimates of the variances and covariances of measurement error, late season risk and unobserved heterogeneity. Results are reported for samples restricted to plots in Tanzania within 1 km of the respondent's home, within 100m of the respondent's home, or cultivated by a single individual. The baseline estimates are in column 1 for comparison. Results are reported for Uganda restricting attention to plots within a 15 minute walk of the respondent's home, to plots within contiguous parcel cultivated by a specific farmer, or to plots that are cultivated by a single individual. The baseline estimates for the full sample are provided in the initial columns for comparison. The bottom panel of Appendix table A3 reports the consequences of the same set of relaxations for our estimates of the variances of TFP A and TFP B. Appendix figure A1 shows the implications for productivity dispersion of assumption (a) in Tanzania. Appendix figure A2 shows the implications for productivity dispersion of assumptions (a) (top) and (b) (bottom) in Uganda. Appendix figure A3 shows the implication of assumption (c) in both Tanzania and Uganda.

In all cases, comparison with Figure 3 shows that the changes to the estimates are minor, with the exception that the IVCRC estimate of the

variance of TFP B is lower in the base estimates than in the alternatives.

Appendix A6: Gains from Reallocation

An efficient allocation of factors across plots requires $L_i^e = s_i^e \bar{L}$ and $X_i^e = s_i^e \bar{X}$, where

$$s_i^e = \frac{\exp\left(\frac{1}{1-\alpha_L-\alpha_X}\omega_i\right)}{\bar{\omega}},$$

$\bar{L}$ and $\bar{X}$ are aggregate endowments of land and labor and $\bar{\omega} = \sum_i e^{\left(\frac{1}{1-\alpha_L-\alpha_X}\omega_i\right)}$. Measured output of producer i in an efficient allocation is

$$Y_i^e = \left(\frac{1}{\bar{\omega}}\right)^{\alpha_L+\alpha_X} e^{\frac{\omega_i}{1-\alpha_L-\alpha_X}+\epsilon_i}(\bar{L})^{\alpha_L}(\bar{X})^{\alpha_X}. \tag{A6.1}$$

If ω_i and ϵ_i are normally distributed and independent of each other, then expected output is

$$\begin{aligned}
E(Y_i^e) &= \left(\frac{1}{\bar{\omega}}\right)^{\alpha_L+\alpha_X}(\bar{L})^{\alpha_L}(\bar{X})^{\alpha_X}E(e^{\epsilon_i})E\left(e^{\frac{\omega_i}{1-\alpha_L-\alpha_X}}\right) \\
&= \left(\frac{1}{\bar{\omega}}\right)^{\alpha_L+\alpha_X}(\bar{L})^{\alpha_L}(\bar{X})^{\alpha_X}E(e^{\epsilon_i})e^{\frac{E(\omega_i)}{1-\alpha_L-\alpha_X}}\left(e^{\frac{\sigma_\omega^2}{2(1-\alpha_L-\alpha_X)^2}}\right) \qquad \text{(A6.2)} \\
&\equiv \overline{Y^e}(\sigma_\omega^2),
\end{aligned}$$

where σ_ω^2 is the variance of total factor productivity. The notation $\overline{Y^e}(\sigma_\omega^2)$ emphasizes the dependence of the average output in the efficient allocation on the variance of total factor productivity.

References

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* 104(6): 1667-1697.

Adamopoulos, Tasso and Diego Restuccia. 2015. "Land Reform and Productivity: A Quantitative Analysis with Micro Data." Working Paper 540, University of Toronto, Department of Economics.

Adamopoulos, Tasso, Loren Brandt, Jessica Leight, and Diego Restuccia. 2017. "Misallocation, Selection and Productivity: A Quantitative Analysis with Panel Data from China." Working Paper no. 23039, NBER, Cambridge, MA.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic Inputs and Resource (Mis) Allocation." *Journal of Political Economy* 122(5): 1013-1063.

Banerjee, Abhijit V. and Benjamin Moll. 2010. "Why Does Misallocation Persist?" *American Economic Journal: Macroeconomics* 2(1): 189-206.

Barrett, Christopher B., Marc F. Bellemare, and Janet Y. Hou. 2010. "Reconsidering Conventional Explanations of the Inverse Productivity–Size Relationship." *World Development* 38(1): 88-97.

Beegle, Kathleen, Calogero Carletto, and Kristen Himelein. 2012a. "Reliability of Recall in Agricultural Data." *Journal of Development Economics* 98(1): 34-41.

Beegle, Kathleen, Joachim De Weerdt, Jed Friedman, and John Gibson. 2012b. "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania." *Journal of Development Economics* 98(1): 3-18.

Benjamin, Dwayne. 1992. "Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models." *Econometrica* 60(2): 287–322.

Benjamin, Dwayne. 1995. "Can Unobserved Land Quality Explain the Inverse Productivity Relationship?" *Journal of Development Economics*, 46(1): 51-84.

Bento, Pedro, and Diego Restuccia. 2017. "Misallocation, Establishment Size, and Productivity." *American Economic Journal: Macroeconomics* 9(3): 267-303.

Bevis, Leah and Christopher B. Barrett. 2017. "Close to the Edge: Do Behavioral Explanations Account for the Inverse Productivity Relationship?" Manuscript, Department of Applied Economics and Management, Cornell University.

Bils, Mark, Peter J. Klenow, and Cian Ruane. 2020. "Misallocation or Mismeasurement?" Working Paper no. 26711. NBER, Cambridge, MA

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

Carletto, Calogero, Sydney Gourlay, Siobhan Murray, and Alberto Zezza. 2017. "Cheaper, Faster, and More than Good Enough: Is GPS the New Gold Standard in Land Area Measurement?" *Survey Research Methods* 11(3): 235-65.

Caselli, Francesco. 2005. "Accounting for Cross-Country Income Differences." In Philippe Aghion & Steven Durlauf (eds.), *Handbook of Economic Growth*, Edition 1, Volume 1, Chapter 9, pp. 679-741. Amsterdam: North-Holland.

Chen, Chaoran. 2016. "Untitled Land, Occupational Choice, and Agricultural Productivity." Working Paper 573. University of Toronto, Department of Economics.

Claassen, Roger and Richard E. Just. 2011. "Heterogeneity and Distributional Form of Farm-Level Yields." *American Journal of Agricultural Economics* 93 (1): 144-160.

Da-Rocha, José-María, Marina Mendes Tavares, and Diego Restuccia. 2017. "Policy Distortions and Aggregate Productivity with Endogenous Establishment-Level Productivity." Working Paper no. 23339. NBER, Cambridge, MA.

De Nicola, Francesca, and Xavier Giné. 2014. "How Accurate Are Recall Data? Evidence from Coastal India." *Journal of Development Economics* 106: 52-65.

Deininger, Klaus, Calogero Carletto, Sara Savastano, and James Muwonge. 2012. "Can Diaries Help in Improving Agricultural Production Statistics? Evidence from Uganda." *Journal of Development Economics* 98(1): 42-50.

Emran, M. Shahe and Forhad J. Shilpi. 2015. "Do Land Market Restrictions Hinder Structural Change in a Rural Economy? Evidence From Sri Lanka." Policy Research Working Paper Series 7525. Washington, DC: The World Bank.

Esfahani, Sara. 2018. "Agricultural Misallocation: Mismeasurement, Misspecification, or Market Frictions?" Manuscript, Department of Economics and International Business, Sam Houston State University.

Foster, Andrew D., and Mark R. Rosenzweig. 2017. "Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size." Working Paper no. 23909. NBER, Cambridge, MA.

Foster, Lucia, Cheryl Grim, John Haltiwanger and Zoltan Wolf. 2016. "Firm-Level Dispersion in Productivity: Is the Devil in the Details?" *American Economic Review* 106(5):95-98.

García-Santana, Manuel and Josep Pijoan-Mas. 2014. "The Reservation Laws in India and the Misallocation of Production Factors." *Journal of Monetary Economics* 66(C): 193-209.

Gottlieb, Charles and Jan Grobovsek. 2018. "Communal Land and Agricultural Productivity." Forthcoming, *Journal of Development Economics.*

Gourlay, Sydney, Talip Kilic, and David Lobell. 2017. "Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for Inverse Scale-Productivity Relationship in Uganda." Manuscript, Living

Standards Measurement Study (LSMS) Survey Unit, Development Data Group, World Bank, Rome, Italy.

Guner, Nezih, Gustavo Ventura, and Xu Yi. 2008. "Macroeconomic Implications of Size-Dependent Policies." *Review of Economic Dynamics* 11(4): 721-744.

Haltiwanger, John. 2016. "Firm Dynamics and Productivity: TFPQ, TFPR, and Demand-Side Factors." *Economía,* 17(1): 3-26.

Haltiwanger, John, Robert Kulick, and Chad Syverson. 2018. "Misallocation Measures: The Distortion that Ate the Residual." Working Paper no. 24199. NBER, Cambridge, MA.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning through Noticing: Theory and Evidence from a Field Experiment." *Quarterly Journal of Economics* 129(3): 1311-1353.

Heckman, James J., Jeffrey Smith, and Nancy Clements. 1997. "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies* 64(4): 487-535.

Heckman, James, and Edward Vytlacil. 1998. "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling when the Return Is Correlated with Schooling." *Journal of Human Resources* 33(4): 974-987.

Hicks, Joan Hamory, Marieke Kleemans, Nicholas Y. Li, and Edward Miguel. 2017. "Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata." Working Paper no. 23253. NBER, Cambridge, MA.

Hopenhayn, Hugo A. 2014. "Firms, Misallocation, and Aggregate Productivity: A Review." *Annual Review of Economics* 6(1): 735-770.

Hsieh, Chang-Tai and Peter Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124(4): 1403-1448.

Hurley, Terrance M., Gary L. Malzer, and Bernard Kilian. 2004. "Estimating Site-Specific Nitrogen Crop Response Functions." *Agronomy Journal* 96(5): 1331-1343.

Kalemli-Ozcan, Sebnem and Bent E. Sorensen. 2012. "Misallocation, Property Rights, and Access to Finance: Evidence from Within and across Africa." Working Paper no. 18030, NBER, Cambridge, MA.

Kehrig, Matthias, and Nicolas Vincent. 2019. "Good Dispersion, Bad Dispersion." Working Paper no. 25923, NBER, Cambridge, MA.

Kilic, Talip, Alberto Zezza, Calogero Carletto, and Sara Savastano. 2016. "Missing (ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements." *World Development* 92(April): 143-157.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies* 70(2): 317-341.

Lobell, David B., George Azzari, Marshall Burke, Sydney Gourlay, Zhenong 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." World Bank Policy Research Working Paper Series 8374. Washington, DC: The World Bank.

Masten, Matthew A., and Alexander Torgovitsky. 2016. "Identification of Instrumental Variable Correlated Random Coefficients Models." *Review of Economics and Statistics* 98(5): 1001-1005.

Maksimovic, Vojislav, and Gordon Phillips. 2001. "The Market for Corporate Assets: Who Engages in Mergers and Asset Sales and Are There Efficiency Gains?" *Journal of Finance* 56(6): 2019-2065.

Matvos, Gregor, and Amit Seru. 2014. "Resource Allocation within Firms and Financial Market Dislocation: Evidence from Diversified Conglomerates." *Review of Financial Studies* 27(4): 1143-1189.

McMillan, Margaret, Dani Rodrik, and Íñigo Verduzco-Gallo. 2014. "Globalization, Structural Change, and Productivity Growth, with an Update on Africa." *World Development* 63(C): 11-32.

Midrigan, 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. 1992. "The Dynamics of Productivity in the Telecommunications Equipment Industry." Working Paper no. 3977, NBER, Cambridge, MA.

Pellegrino, Bruno and Geoffery Zheng. 2018. "What Is the Extent of Misallocation?" Manuscript, Anderson School of Management, University of California at Los Angeles.

Petrin, Amil, and Jagadeesh Sivadasan. 2013. "Estimating Lost Output from Allocative Inefficiency, with an Application to Chile and Firing Costs." *Review of Economics and Statistics* 95(1): 286-301.

Petrin, Amil, T. Kirk White, and Jerome P. Reiter. 2011. The Impact of Plant-Level Resource Reallocations and Technical Progress on US Macroeconomic Growth." *Review of Economic Dynamics* 14(1): 3-26.

Porzio, Tommaso. 2016. "Cross-Country Differences in the Optimal Allocation of Talent and Technology." Manuscript, Department of Economics, University of California at San Diego.

Restuccia, Diego. 2018. "Misallocation and Aggregate Productivity across Time and Space." University of Toronto Department of Economics Working Paper No. 596.

Restuccia, Diego and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments." *Review of Economic Dynamics* 11(4): 707–20.

Restuccia, Diego and Richard Rogerson. 2013. "Misallocation and Productivity." *Review of Economic Dynamics* 16(1): 1-10.

Restuccia, Diego and Richard Rogerson. 2017. "The Causes and Costs of Misallocation." *Journal of Economic Perspectives* 31(3): 151-174.

Restuccia, Diego and Raul Santaeulalia-Llopis. 2017. "Land Misallocation and Productivity." Working Paper no. 23128, NBER, Cambridge, MA.

Restuccia, Diego, Dennis Tao Yang, and Xiadong Zhu. 2008. "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis." *Journal of Monetary Economics* 55(2): 234-50.

Rotemberg, Martin, and T. Kirk White. 2017. "Measuring Cross-Country Differences in Misallocation." Manuscript, Department of Economics, New York University.

Schimmelpfennig, David. 2016. *Farm Profits and Adoption of Precision Agriculture.* No. 249773. United States Department of Agriculture, Economic Research Service.

Schultz, T.W. 1964. *Transforming Traditional Agriculture.* New Haven: Yale University Press.

Ševčík, Pavel. 2015. "Financial Frictions, Internal Capital Markets, and the Organization of Production." *Review of Economic Dynamics* 18(3): 505-522.

Shenoy, Ajay. 2017. "Market Failures and Misallocation." *Journal of Development Economics* 128: 65-80.

Stoorvogel, Jetse J., Lammert Kooistra, and Johan Bouma. 2015. "Managing Soil Variability at Different Spatial Scales as a Basis for Precision Agriculture." In *Soil-Specific Farming: Precision Agriculture*, eds. R. Lal and BA Stewart: 37-72.

Suri, Tavneet. 2011. "Selection and Comparative Advantage in Technology Adoption." *Econometrica* 79(1): 159-209.

Syverson, Chad. 2004. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics* 86(2): 534-50.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49(2): 326-365.

Tittonell, Pablo D., Keith D. Shepherd, Bernard Vanlauwe, and Ken E. Giller. 2008. "Unravelling the Effects of Soil and Crop Management on Maize Productivity in Smallholder Agricultural Systems of Western Kenya — An Application of Classification and Regression Tree Analysis." *Agriculture, Ecosystems & Environment* 123(1-3): 137-150.

Tittonell, Pablo, Bernard Vanlauwe, P. A. Leffelaar, Keith D. Shepherd, and Ken E. Giller. 2005. "Exploring Diversity in Soil Fertility Management of Smallholder Farms in Western Kenya: II. Within-Farm Variability in

Resource Allocation, Nutrient Flows and Soil Fertility Status." *Agriculture, Ecosystems and Environment* 110(3): 166-184.

Tittonell, Pablo, Bernard Vanlauwe, N. De Ridder, and Ken E. Giller. 2007. "Heterogeneity of Crop Productivity and Resource Use Efficiency within Smallholder Kenyan Farms: Soil Fertility Gradients or Management Intensity Gradients?" *Agricultural Systems* 94(2): 376-390.

Tjernstrom, Emilia, Michael Carter, and Travis Lybbert. 2015. "The Dirt on Dirt: Soil Characteristics and Variable Fertilizer Returns in Kenyan Maize Systems." Working Paper, USAID Feed the Future Innovation Lab for Assets and Market Access.

Udry, Christopher. 1996. "Gender, Agricultural Production, and the Theory of the Household." *Journal of Political Economy* 104(5): 1010-1046.

Valentinyi, Akos, and Berthold Herrendorf. 2008. "Measuring Factor Income Shares at the Sectoral Level." *Review of Economic Dynamics* 11 (4): 820–835.

Vanlauwe, Bernard, P. Tittonell, and J. Mukalama. 2006. "Within-Farm Soil Fertility Gradients Affect Response of Maize to Fertiliser Application in Western Kenya." *Nutrient Cycling in Agroecosystems* 76(2): 171-182.

Vanlauwe, B., K. Descheemaeker, K. E. Giller, J. Huising, R. Merckx, G. Nziguheba, J. Wendt, and S. Zingore. 2015. "Integrated Soil Fertility Management in Sub-Saharan Africa: Unravelling Local Adaptation." *Soil* 1(1): 491-508.

White, T. Kirk, Jerome P. Reiter, and Amil Petrin. 2018. "Imputation in US Manufacturing Data and Its Implications for Productivity Dispersion." *Review of Economics and Statistics* 100(3): 502-09.

Wooldridge, Jeffrey M. 2008. "Instrumental Variables Estimation of the Average Treatment Effect in the Correlated Random Coefficient Model." In Tom Fomby, R. Carter Hill, Daniel L. Millimet, Jeffrey A. Smith, and Edward J. Vytlacil (eds.) *Modelling and Evaluating Treatment Effects in Econometrics* (Advances in Econometrics, Volume 21). Bingley, United Kingdom: Emerald Group Publishing Limited. pp.93 – 116.

Figure 1: Dispersion of Log Output per Hectare Across Plots

Tanzania

Uganda

Raw data   Observed Plot Chars.
Region–Y–S–C FE   Village–Y–S–C FE
HH–Y–S–C FE   Farmer Y–S–C FE
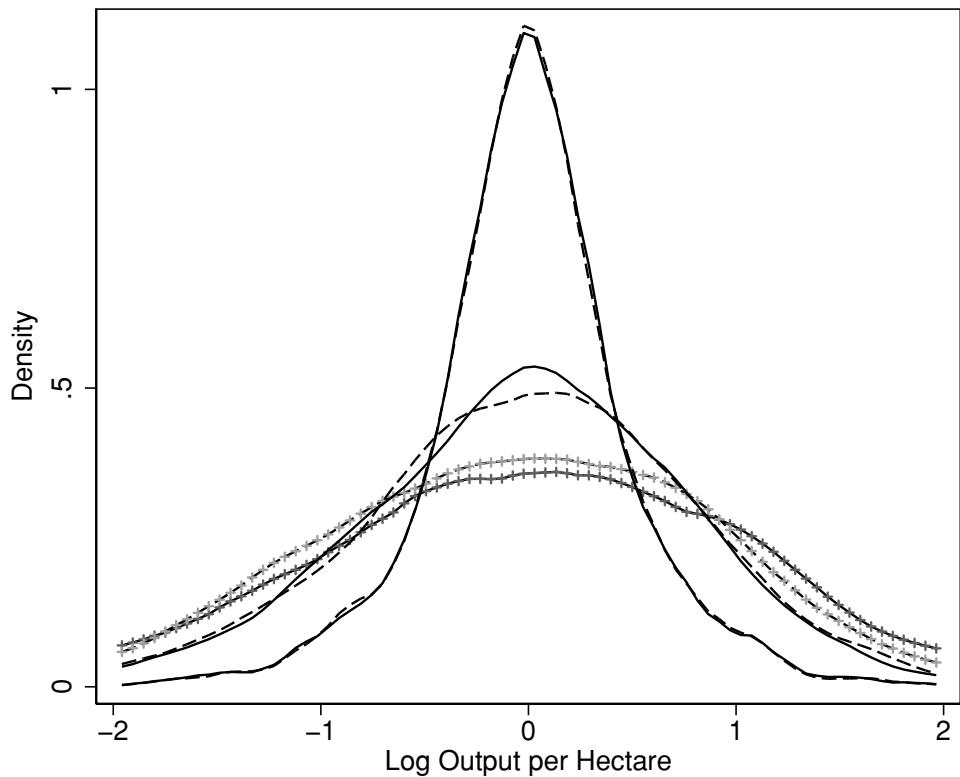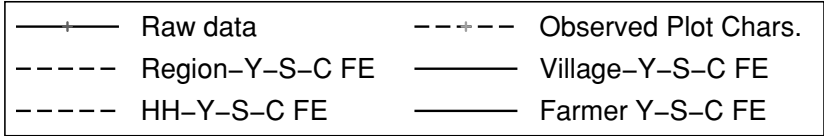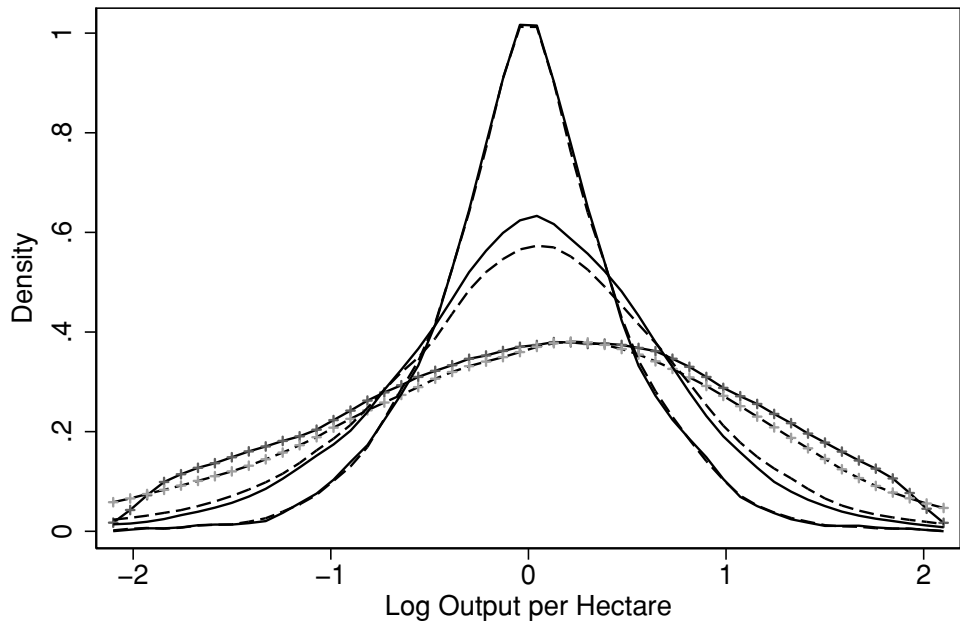
Figure 2: Dispersion of Log Labour per Hectare Across Plots

Tanzania

Uganda

Raw data — Observed Plot Chars.
Region–Y–S–C FE — Village–Y–S–C FE
Household Y–S–C FE — Farmer Y–S–C FE
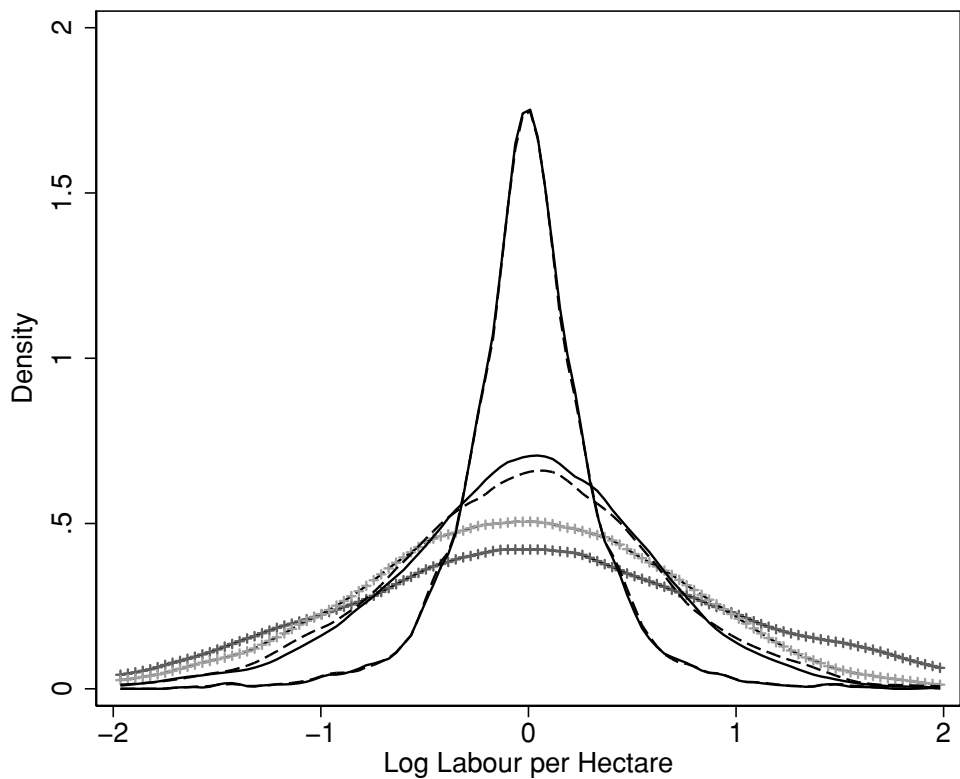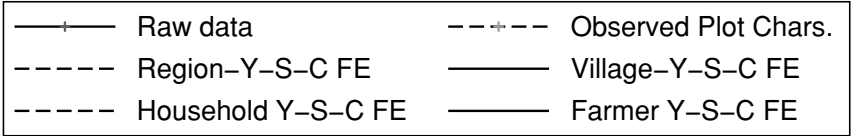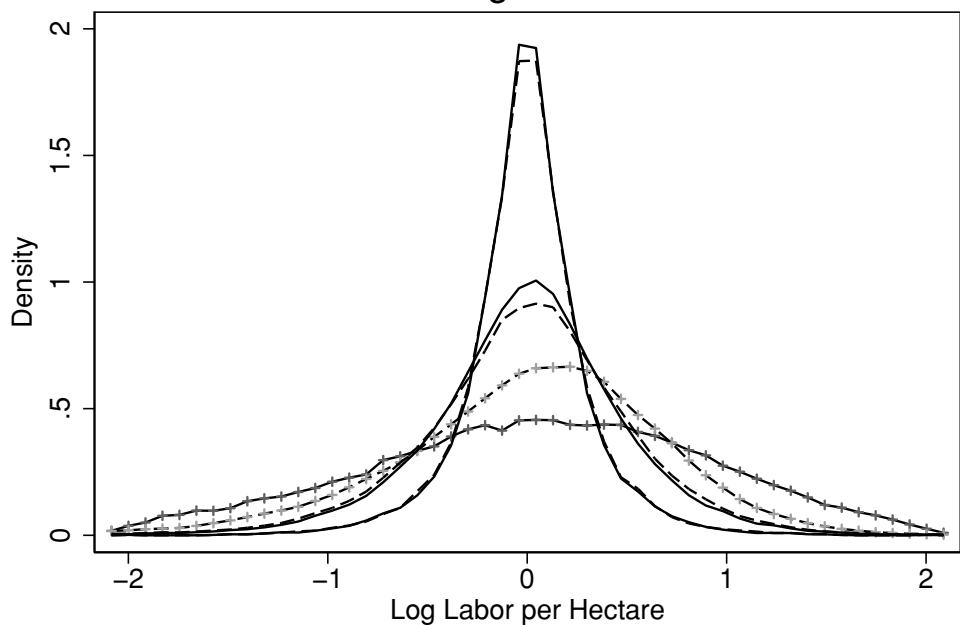
Figure 3: Log Productivity Dispersion

A. Tanzania

B. Uganda

lnTFPA – IVCRC production function residuals

lnTFPB – IVCRC Corrected for risk and measurement error

lnTFPA – 2sls production function residuals

lnTFPB – 2sls Corrected for risk and measurement error
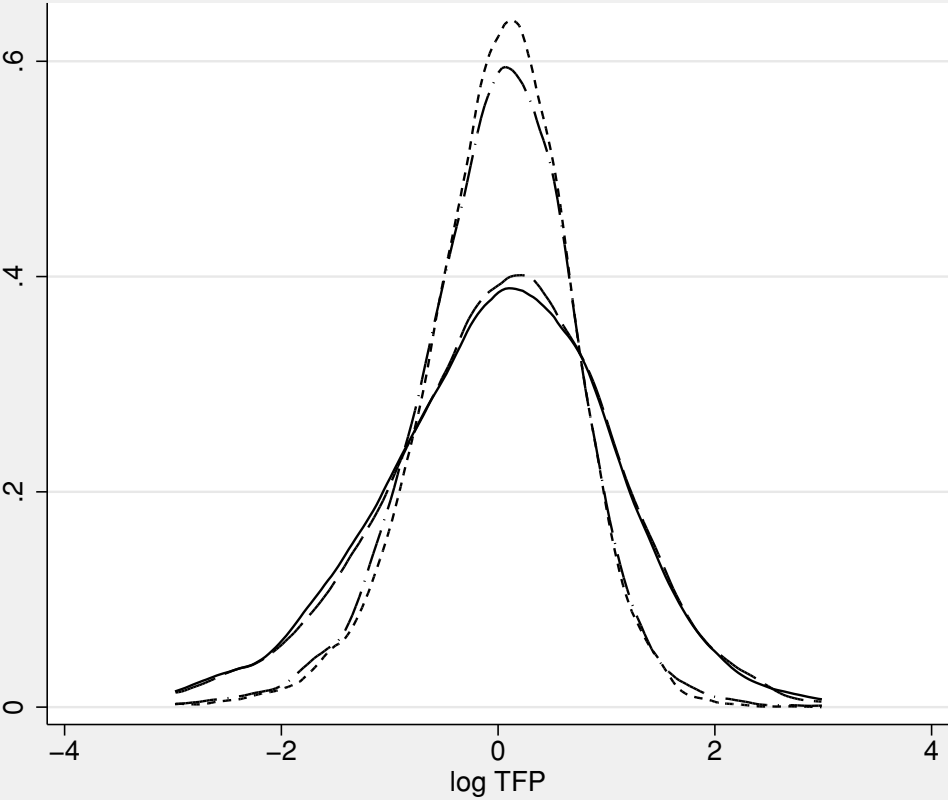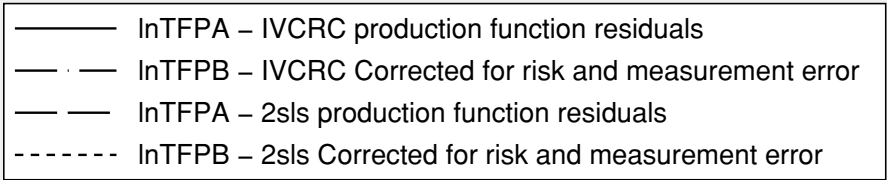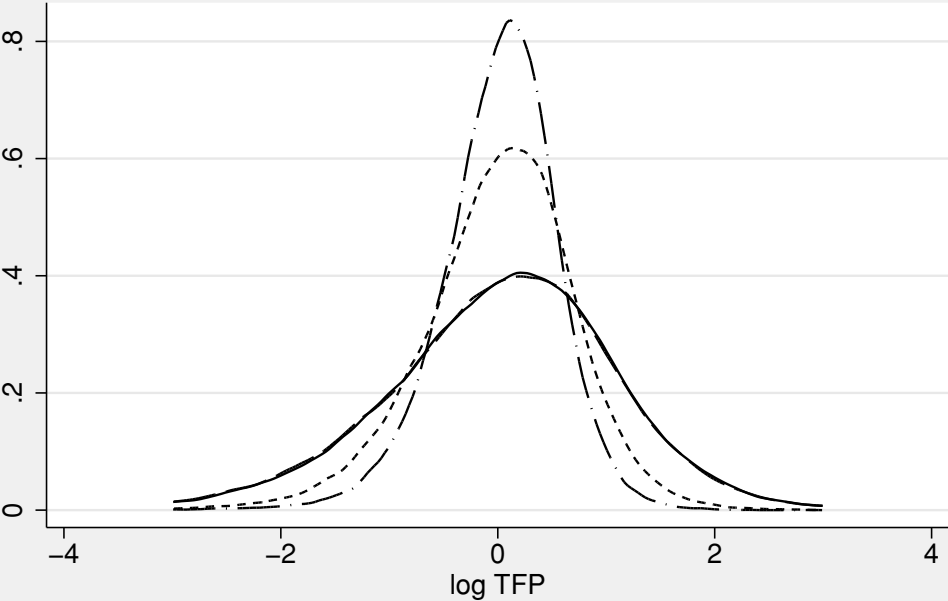
Table 1A: Agriculture in Tanzania and Uganda - Samples

		Tanzania	Uganda
Sample Size			
	Households	5832	4997
	Farmers	7302	10364
	Plot-Seasons	21000	52585
	Seasons	4	8
	Regions	25	6
	Districts	125	81
	Villages	633	1147
Number of Clusters			
	Farmer - Seasons	9821	21793
	Farmer - Crop - Seasons	17089	48347
Size of Clusters (Median)			
	Farmer - Seasons	2	3
	Farmer - Crop - Seasons	1	1

Table 1B: Agriculture in Tanzania and Uganda - Yields

		Tanzania	Uganda
Median Plot Size (ha)		0.40	0.20
Yield ($/ha)			
	N	15635	52334
	mean	903	3061
	median	458	182
	std deviation	5133	345562
Yield on Maize Plots ($/ha)			
	N	7370	10232
	mean	812	519
	median	438	210
	std deviation	6092	3244
Yield on Groundnut/Beans Plots ($/ha)			
	N	988	10760
	mean	754	6892
	median	439	218
	std deviation	1035	656396
Yield on Cassava (TZ) or Banana (UG) Plots ($/ha)			
	N	1172	6083
	mean	689	565
	median	370	314
	std deviation	974	1317
Labour (Days/ha)			
	N	18416	52334
	mean	225	258
	median	128	141
	std deviation	697	1217

*Note: All yields winsorized at the 0.01 level

Table 2A: Log Output per Hectare in Tanzania

	(1) No Fixed Effects	(2) Year-Crop-Region	(3) Year-Crop-Village	(4) Year-Crop-Household	(5) Year-Crop-Farmer
ln(ha)	-0.49	-0.52	-0.48	-0.51	-0.5
	(0.009)	(0.011)	(0.012)	(0.030)	(0.031)
Female Plot	-0.16	-0.12	-0.15	-0.0014	
	(0.026)	(0.028)	(0.031)	(0.200)	
Plot used free of charge	-0.044	-0.021	-0.033	0.073	0.099
	(0.029)	(0.032)	(0.034)	(0.095)	(0.098)
Shared - rent	0.015	-0.23	-0.02	-0.67	-0.62
	(0.140)	(0.160)	(0.160)	(0.700)	(0.710)
Shared - owned	0.14	0.08	0.032	-0.13	-0.098
	(0.038)	(0.048)	(0.049)	(0.140)	(0.140)
Average quality	-0.16	-0.17	-0.15	-0.054	-0.074
	(0.018)	(0.020)	(0.021)	(0.070)	(0.072)
Poor quality	-0.28	-0.36	-0.35	-0.21	-0.21
	(0.038)	(0.040)	(0.046)	(0.120)	(0.120)
Loam	0.17	0.071	0.14	-0.051	-0.063
	(0.024)	(0.027)	(0.029)	(0.081)	(0.084)
Clay	0.2	0.09	0.16	0.012	-0.013
	(0.030)	(0.033)	(0.036)	(0.095)	(0.099)
Distance to market	0.0023	0.0011	0.0011	0.0045	0.0046
	(0.001)	(0.001)	(0.001)	(0.006)	(0.006)
Irrigated	0.46	0.17	0.23	0.69	0.49
	(0.061)	(0.083)	(0.085)	(0.230)	(0.240)
Erosion evident	-0.018	-0.015	-0.044	0.026	0.022

	(0.026)	(0.028)	(0.031)	(0.070)	(0.072)
Sale value	-0.047	0.021	-0.053	0.64	0.62
	(0.076)	(0.087)	(0.077)	(1.850)	(2.040)
Current morbidity (z)	-0.043	-0.044	-0.035	0.22	
	(0.010)	(0.010)	(0.011)	(0.140)	
Morbidity missing	4.2	4.3	3.46	-22.1	
	(0.950)	(1.020)	(1.140)	(14.000)	
R2	0.29	0.52	0.53	0.71	0.71
Log Variance of Residuals	1.15	0.76	0.73	0.24	0.24
F-Stat for Plot Characteristics	122.8	70.8	56.1	10.5	10.3
Corresponding p-Value	0	0	0	0	0

Notes: Standard errors in parentheses. R2 from OLS dummy variable specification. Variance of dependent variable is 1.49 Regressions include 34 plot characteristics - e.g. soil type toposequence boundary markers location, rainfall and interactions with rain and 14 household and farmer characteristics - e.g. housing education age.

Table 2B: Log Output per Hectare in Uganda

	(1) No Fixed Effects	(2) Year-Crop-District	(3) Year-Crop-Village	(4) Year-Crop-Household	(5) Year-Crop-Farmer
ln(ha)	-0.56	-0.69	-0.7	-0.73	-0.73
	(0.006)	(0.008)	(0.008)	(0.019)	(0.020)
Male Plot	0.14	0.11	0.095	0.33	
	(0.024)	(0.026)	(0.028)	(0.170)	
Leasehold Plot	0.0001	0.055	0.011	-0.094	-0.21
	(0.032)	(0.036)	(0.039)	(0.110)	(0.120)
Customary Plot	-0.024	0.047	0.012	-0.14	-0.17
	(0.045)	(0.050)	(0.052)	(0.150)	(0.150)
Mailo Plot	-0.06	0.021	0.0091	-0.037	-0.0074
	(0.030)	(0.036)	(0.042)	(0.160)	(0.180)
Plot via Occupancy	-0.015	0.0017	-0.025	-0.041	-0.033
	(0.030)	(0.035)	(0.037)	(0.110)	(0.120)
Customary	-0.3	0.0079	0.021	-0.041	0.041
	(0.016)	(0.026)	(0.028)	(0.092)	(0.099)
No document	0.0045	-0.019	-0.028	-0.0058	-0.022
	(0.024)	(0.027)	(0.029)	(0.093)	(0.097)
Fair Soil	-0.033	-0.022	-0.013	0.0012	0.11
	(0.040)	(0.044)	(0.043)	(0.120)	(0.130)
Poor Soil	-0.23	-0.2	-0.16	-0.26	-0.52
	(0.097)	(0.100)	(0.100)	(0.310)	(0.360)
Sandy clay loam	0.01	-0.01	-0.0018	-0.014	-0.0089
	(0.014)	(0.016)	(0.016)	(0.048)	(0.051)
Black clay	-0.0041	-0.00071	0.019	0.07	0.084

	(0.018)	(0.020)	(0.022)	(0.064)	(0.067)
Fair Soil*Total Rain	-0.15	-0.15	-0.17	0.064	0.031
	(0.056)	(0.067)	(0.070)	(0.200)	(0.220)
Poor Soil*Total Rain	0.024	0.12	0.048	-0.63	-0.65
	(0.160)	(0.170)	(0.180)	(0.490)	(0.520)
Fair Soil*Flood duration	0.016	0.043	0.033	0.19	0.25
	(0.029)	(0.032)	(0.033)	(0.110)	(0.150)
R2	0.33	0.69	0.73	0.79	0.79
Log Variance of Residuals	1.42	0.66	0.52	0.26	0.25
F-Stat for Plot Characteristics	192.5	171.0	158.1	37.5	36.5
Corresponding p-Value	0	0	0	0	0

Notes: Standard errors in parentheses. R2 from OLS dummy variable specification.
Variance of dependent variable is 1.82. Regressions include: 37 plot characteristics - e.g.
dummies for soil type toposequence boundary markers location; 11 interactions of plot
characteristics with weather shocks; 12 household and farmer characteristics e.g.
education age dummies for morbidity housing; 38 household level self-reported shocks
e.g. drought floods job loss livestock disease; 3 community level remote sensing rainfall
variables annual and seasonal rainfall totals.

Table 3A: Log Labor per Hectare in Tanzania

	(1)	(2)	(3)	(4)	(5)
	No Fixed Effects	Year-Crop-Region	Year-Crop-Village	Year-Crop-Household	Year-Crop-Farmer
ln(ha)	-0.63	-0.61	-0.61	-0.61	-0.6
	(0.006)	(0.008)	(0.008)	(0.018)	(0.019)
Female Plot	-0.12	-0.12	-0.12	0.12	
	(0.018)	(0.019)	(0.021)	(0.110)	
Plot used free of charge	0.066	-0.016	0.028	0.041	0.012
	(0.020)	(0.022)	(0.023)	(0.055)	(0.056)
Shared - rent	-0.059	-0.018	-0.16	-0.96	-0.97
	(0.099)	(0.110)	(0.110)	(0.470)	(0.460)
Shared - owned	-0.048	-0.022	-0.022	-0.018	-0.047
	(0.025)	(0.032)	(0.033)	(0.083)	(0.084)
Average quality	-0.0053	0.0079	0.0071	0.043	0.038
	(0.013)	(0.014)	(0.015)	(0.041)	(0.042)
Poor quality	-0.022	-0.0095	-0.029	0.079	0.1
	(0.025)	(0.027)	(0.031)	(0.070)	(0.072)
Loam	0.011	0.025	0.037	0.027	0.03
	(0.016)	(0.018)	(0.020)	(0.047)	(0.048)
Clay	0.053	0.059	0.074	-0.015	-0.0019
	(0.021)	(0.023)	(0.025)	(0.057)	(0.058)
Distance to market	-0.00015	-0.00011	-0.00033	0.02	0.021
	(0.000)	(0.001)	(0.001)	(0.004)	(0.004)
Irrigated	-0.075	0.018	-0.12	0.096	0.067
	(0.042)	(0.059)	(0.059)	(0.140)	(0.150)
Erosion evident	-0.019	-0.066	-0.056	-0.19	-0.18

	(0.018)	(0.020)	(0.021)	(0.042)	(0.043)
Sale value	-0.14	-0.19	-0.14	0.95	1.11
	(0.057)	(0.062)	(0.059)	(1.140)	(1.230)
Current morbidity (z)	-0.0052	-0.0083	-0.011	0.14	
	(0.007)	(0.007)	(0.008)	(0.078)	
Morbidity missing	0.56	0.86	1.21	-13.3	
	(0.650)	(0.700)	(0.780)	-7.76	
R2	0.34	0.46	0.46	0.61	0.61
Log Variance of Residuals	0.65	0.45	0.44	0.11	0.11
F-Stat for Plot Characteristics	333.8	202.7	181.4	37.6	37.6
Corresponding p-Value	0	0	0	0	0

Notes: Standard errors in parentheses. R2 from OLS dummy variable specification. Variance of dependent variable is 1.06. Regressions include 34 plot characteristics - e.g. soil type toposequence boundary markers location, rainfall and interactions with rain and 14 household and farmer characteristics - e.g. housing education age.

Table 3B: Log Labour per Hectare in Uganda

	(1)	(2)	(3)	(4)	(5)
	No Fixed Effects	Year-Crop-District	Year-Crop-Village	Year-Crop-Household	Year-Crop-Farmer
ln(ha)	-0.72	-0.77	-0.78	-0.79	-0.79
	(0.004)	(0.005)	(0.006)	(0.012)	(0.012)
Male Plot	-0.014	0.012	-0.015	-0.077	
	(0.013)	(0.016)	(0.018)	(0.096)	
Leasehold Plot	-0.088	-0.093	-0.11	0.065	0.15
	(0.019)	(0.024)	(0.027)	(0.066)	(0.075)
Customary Plot	-0.2	-0.17	-0.15	-0.13	-0.15
	(0.026)	(0.032)	(0.035)	(0.087)	(0.090)
Mailo Plot	-0.1	-0.12	-0.08	0.21	0.14
	(0.017)	(0.023)	(0.028)	(0.092)	(0.100)
Plot via Occupancy	-0.14	-0.16	-0.11	0.011	0.024
	(0.018)	(0.023)	(0.026)	(0.069)	(0.074)
Customary	-0.12	-0.11	-0.12	0.15	0.18
	(0.009)	(0.017)	(0.019)	(0.055)	(0.059)
No document	-0.066	-0.041	-0.01	0.0025	0.0015
	(0.014)	(0.018)	(0.020)	(0.056)	(0.059)
Fair Soil	0.059	0.04	0.038	-0.19	-0.2
	(0.023)	(0.028)	(0.029)	(0.071)	(0.076)
Poor Soil	-0.024	-0.028	-0.017	0.11	0.15
	(0.057)	(0.067)	(0.068)	(0.180)	(0.220)
Sandy clay loam	0.00055	-0.0088	-0.00014	0.023	0.034
	(0.008)	(0.010)	(0.011)	(0.028)	(0.029)
Black clay	-0.004	-0.013	-0.0014	0.016	0.028

	(0.010)	(0.013)	(0.015)	(0.036)	(0.038)
Fair Soil*Total Rain	-0.028	-0.037	-0.046	0.04	0.13
	(0.033)	(0.044)	(0.048)	(0.120)	(0.130)
Poor Soil*Total Rain	-0.036	-0.11	-0.16	-0.43	-0.36
	(0.090)	(0.110)	(0.120)	(0.290)	(0.300)
Fair Soil*Flood duration	0.019	0.048	0.033	0.075	0.063
	(0.016)	(0.019)	(0.021)	(0.059)	(0.076)
Poor Soil*Flood duration	0.066	0.1	0.025		
	(0.110)	(0.140)	(0.150)		
R2	0.41	0.51	0.53	0.60	0.60
Log Variance of Residuals	0.48	0.28	0.25	0.09	0.09
F-Stat for Plot Characteristics	719.8	487.3	402.9	110.8	107.7
Corresponding p-Value	0	0	0	0	0

Notes: Standard errors in parentheses. R2 from OLS dummy variable specification. Variance of dependent variable is 1.00. Regressions include: 37 plot characteristics - e.g. dummies for soil type toposequence boundary markers location; 11 interactions of plot characteristics with weather shocks; 12 household and farmer characteristics e.g. education age dummies for morbidity housing; 38 household level self-reported shocks e.g. drought floods job loss livestock disease; 3 community level remote sensing rainfall variables annual and seasonal rainfall totals.

Table 4: Intensity of Cultivation and Parcel Subdivision in Uganda

A. Log Output per Hectare

	(1)	(2)	(3)	(4)	(5)	(6)
	No Plot Characteristics	No Fixed Effects	Year-District	Year-Village	Year-Household	Year-Farmer
Log Parcel Area	-0.60	-0.75	-0.75	-0.74	-0.81	-0.82
	(0.0068)	(0.010)	(0.010)	(0.011)	(0.017)	(0.018)
Number of Plots in Parcel	0.30	0.36	0.36	0.36	0.42	0.42
	(0.0066)	(0.0075)	(0.0078)	(0.0080)	(0.014)	(0.015)
R2	0.24	0.34	0.45	0.49	0.65	0.65
Log Variance of Residuals	1.34	1.20	1.01	0.93	0.57	0.56
F-Stat for Plot Characteristics		23.3	13.8	12.7	8.77	7.97
Corresponding p-Value		0	0	0	0	0

B. Log Labor per Hectare

	(1)	(2)	(3)	(4)	(5)	(6)
	No Plot Characteristics	No Fixed Effects	Year-District	Year-Village	Year-Household	Year-Farmer
Log Parcel Area	-0.87	-0.82	-0.83	-0.83	-0.85	-0.85
	(0.0044)	(0.0061)	(0.0063)	(0.0065)	(0.0082)	(0.0086)
Number of Plots in Parcel	0.43	0.41	0.39	0.39	0.46	0.46
	(0.0043)	(0.0046)	(0.0047)	(0.0048)	(0.0067)	(0.0071)
R2	0.51	0.57	0.61	0.62	0.69	0.69
Log Variance of Residuals	0.59	0.45	0.37	0.34	0.14	0.14
F-Stat for Plot Characteristics		16.2	10.9	9.93	11.1	9.87
Corresponding p-Value		0	0	0	0	0

Standard errors in parentheses. R2 defined as R2 from OLS dummy variable specification.
Regressions include:
37 plot characteristics - e.g. dummies for soil type toposequence boundary markers location
12 interactions of plot characteristics with weather shocks

12 household and farmer characteristics e.g. education age dummies for morbidity housing
38 household level self-reported shocks e.g. drought floods job loss livestock disease
3 community level remote sensing rainfall variables annual and seasonal rainfall totals

Table 5A: OLS and Quantile Regression Determinants of Log Land and Log Labor Inputs in Tanzania

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile range	
	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor
Male Manager	0.35	0.23	0.37	0.28	0.32	0.22	0.33	0.22	-0.034	-0.069
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Land Value	0.21	0.1	0.19	0.11	0.2	0.1	0.24	0.1	0.044	-0.0056
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Drought/Flood *Good Soil	0.096	0.024	0.14	0.011	0.045	-0.03	0.045	0.0023	-0.091	-0.0085
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Drought/Flood *Avg Soil	0.073	0.035	0.12	-0.0043	-0.0038	0.055	0.025	0.036	-0.1	0.04
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.07)
Drought/Flood *Poor Soil	-0.027	-0.035	-0.014	-0.083	-0.031	-0.082	0.0056	-0.016	0.0083	0.067
	(0.08)	(0.07)	(0.08)	(0.09)	(0.09)	(0.10)	(0.07)	(0.08)	(0.10)	(0.11)
Illness/accident of HH member	0.035	-0.039	0.042	-0.075	0.056	-0.0095	0.0073	0.018	-0.035	0.093
	(0.03)	(0.03)	(0.05)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.06)	(0.04)
GS Rain *Good Soil in HH	-0.038	-0.0035	0.042	0.037	-0.045	-0.0067	-0.059	-0.03	-0.1	-0.067
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)
GS Rain *Avg Soil in HH	-0.0058	0.017	0.078	0.032	-0.027	0.011	-0.025	0.015	-0.1	-0.017
	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)
GS Rain *Poor Soil in HH	-0.061	-0.05	-0.093	-0.081	-0.11	-0.065	-0.019	0.028	0.075	0.11
	(0.05)	(0.05)	(0.06)	(0.07)	(0.04)	(0.06)	(0.04)	(0.04)	(0.09)	(0.08)
	0.0053	0.038	0.02	0.046	0.011	0.051	0.0067	0.032	-0.014	-0.014

Drought/Flood *Good Soil in HH	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
Drought/Flood *Avg Soil in HH	-0.0089	0.02	-0.029	0.04	0.019	0.026	0.0011	0.011	0.03	-0.029
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Drought/Flood *Poor Soil in HH	0.11	0.1	0.11	0.097	0.08	0.17	0.072	0.025	-0.042	-0.072
	(0.05)	(0.04)	(0.04)	(0.05)	(0.08)	(0.05)	(0.03)	(0.04)	(0.07)	(0.05)
Adverse shock to household plots†	-0.089	-0.032	-0.11	-0.046	-0.081	-0.032	-0.061	-0.022	0.052	0.023
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Pseudo R2	0.13	0.09	0.07	0.05	0.06	0.05	0.08	0.05		
F statistic for joint significance of instruments.	17.1	11.5	13.7	7.32	24.4	11.3	19.4	11.2		
p-value	0	0	0	0	0	0	0	0		
F statistic for h0: Coefficients equal for 25th and 75th percentiles									30.1	5.46
p-value									0	0

Note: Bootstrapped (500 samples) standard errors, clustered at the household level, in parantheses. The full sets of coefficients are reported in Appendix Table A1a. These include the sale value of land, distance of the plot from home and from the nearest road, three levels of soil quality, four soil types, the gender, health status, literacy and age of the plot manager, indicator variables for household-level drought or floods, crop disease or pests, severe water shortage or other shocks that led to crop loss, and growing season rainfall, interacted with soil type and soil quality dummies.

†Variable serves as instrument in Table 6

Table 5B: OLS and Quantile Regression Determinants of Log Land and Log Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile range	
	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor
Male Plot	0.14	0.062	0.13	0.044	0.14	0.067	0.14	0.083	0.015	0.039
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Fair Soil	-0.092	-0.078	-0.1	-0.08	-0.14	-0.098	-0.1	-0.12	-0.0045	-0.036
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Poor Soil	-0.2	-0.11	-0.19	-0.19	-0.31	-0.18	-0.13	-0.022	0.064	0.17
	(0.07)	(0.06)	(0.08)	(0.05)	(0.06)	(0.07)	(0.06)	(0.05)	(0.14)	(0.07)
Fair Soil *Drought duration	0.026	0.024	0.032	0.021	0.032	0.024	0.025	0.035	-0.0071	0.013
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Poor Soil *Drought duration	0.035	0.027	0.035	0.027	0.036	0.021	0.04	0.0044	0.0049	-0.023
	(0.010)	(0.010)	(0.020)	(0.010)	(0.010)	(0.010)	(0.010)	(0.00)	(0.020)	(0.010)
Illness Incidence in HH	0.0027	-0.051	-0.0019	-0.067	0.013	-0.066	-0.0078	-0.064	-0.0059	0.0027
	(0.010)	(0.010)	(0.020)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.020)	(0.010)
HH Fair Soil *Peak Rain	8.2E-05	-7.4E-05	0.0001	-7.3E-05	4.5E-05	-0.00007	4.5E-05	-3.2E-05	-5.8E-05	4.2E-05
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HH Poor Soil *Peak Rain	0.00021	-5.3E-05	0.00022	-3.7E-05	0.00025	-0.00011	0.00022	-3.6E-05	5.7E-06	7.5E-08
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HH Fair Soil *Drought duration	-0.0053	-0.0069	-0.0067	-0.0076	-0.0061	-0.0078	-0.0074	-0.0083	-0.00062	-0.00073
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
HH Fair Soil *Total Rain	-0.048	0.017	-0.053	0.018	-0.045	0.012	-0.041	-0.00038	0.012	-0.018

	(0.008)	(0.007)	(0.008)	(0.009)	(0.010)	(0.007)	(0.010)	(0.006)	(0.011)	(0.009)
HH Poor Soil *Total Rain	-0.083	0.023	-0.07	0.018	-0.097	0.037	-0.091	-0.0038	-0.02	-0.022
	(0.025)	(0.020)	(0.031)	(0.031)	(0.019)	(0.029)	(0.022)	(0.018)	(0.027)	(0.036)
Shocks on Household Plots†	0.023	-0.00034	0.017	-0.0011	0.069	0.00065	0.1	-0.00049	0.087	0.00065
	(0.001)	(0.001)	(0.001)	(0.001)	(0.010)	(0.001)	(0.010)	(0.001)	(0.010)	(0.00)
Pseudo R2	0.08	0.02	0.038	0.013	0.048	0.013	0.052	0.015		
F statistic for joint significance of instruments.	37.1	14.7	30.4	20.9	34.6	14.6	34.9	11.6		
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
F statistic for h0: Coefficients equal for 25th and 75th percentiles									71.2	36
p-value									0.00	0.00

Note: Bootstrapped (500 samples) standard errors, clustered at the household level, in parantheses. The full sets of coefficients are reported in Appendix Table A5B. These include four categories of soil quality, three sources of water, 6 indicators of plot toposequence, the level of erosion, indicators of the gender, literacy and access to agricultural advice of the plot manager, household-level indicators of drought and flood and their interactions with plot level soil quality, village level seasonal and annual rain and their interactions with plot level soil quality.

†Variable serves as instrument in Table 6

Table 6: Production Function and Variance Components

	Tanzania		Uganda	
	2SLS	IVCRC	2SLS	IVCRC
A. Cobb-Douglas Factor Coefficients				
Land	0.73	0.61	0.68	0.53
	(0.17)	(0.01)	(0.05)	(0.00)
Labor	0.28	0.26	0.23	0.40
	(0.23)	(0.02)	(0.12)	(0.01)
B. Implied Plot Level Variances of Productivity, Risk and Measurement Errors				
Plot TFP	0.35	0.38	0.18	0.18
	(0.01)	(0.02)	(0.00)	(0.01)
Land/Labor Productivity	0.07	0.10	0.07	0.07
	(0.01)	(0.02)	(0.01)	(0.01)
Late Season Risk and Output Measurement Error	0.65	0.70	0.67	0.67
	(0.02)	(0.02)	(0.01)	(0.01)
Land Measurement Error	0.08	0.03	0.13	0.13
	(0.02)	(0.02)	(0.01)	(0.01)
Labor Measurement Error	0.27	0.32	0.18	0.18
	(0.02)	(0.02)	(0.01)	(0.01)
Cov of TFP and Land/labor Productivity	0.06	0.09	0.05	0.05
	(0.01)	(0.02)	(0.00)	(0.01)
N	14535	14535	43187	43187

Note: Bootstrapped (500 samples) standard errors, clustered at the household level, in parantheses. Full production function results shown in Appendix Tables A2a and A2b.

Fig. A1: Tanzania Distance Effects
Log Productivity Dispersion

1 Kilometer

log TFP

500 Meters

log TFP

lnTFPA – IVCRC prod func residuals

lnTFPB – IVCRC Corrected for risk and measurement error

lnTFPA – 2sls prod func residuals

lnTFPB – 2sls Corrected for risk and measurement error
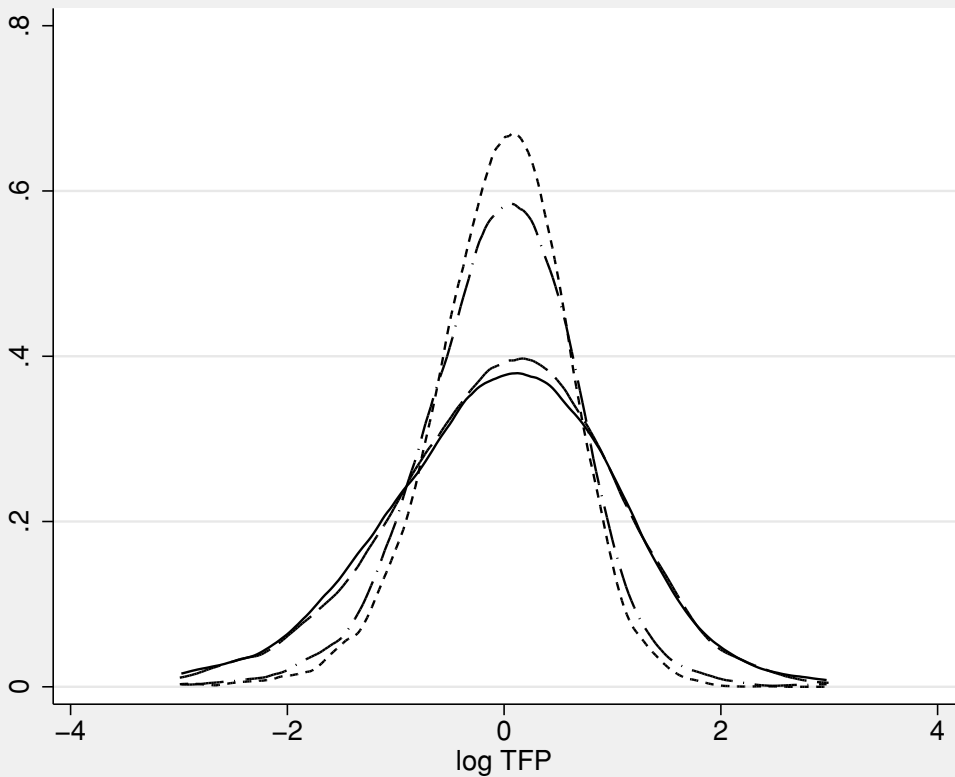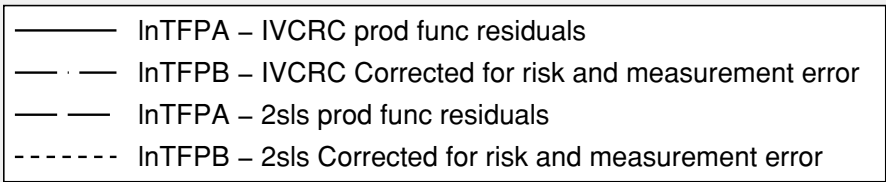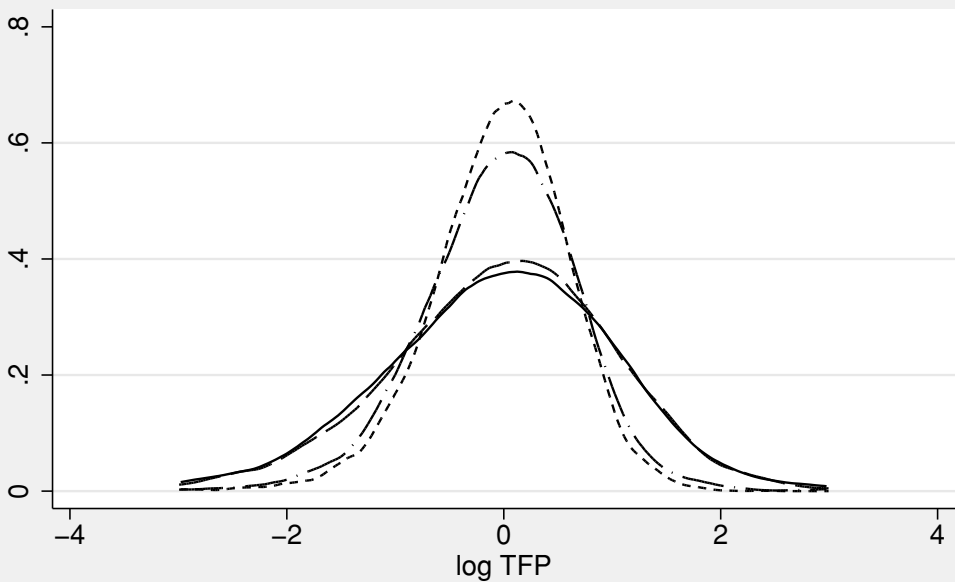
Fig. A2: Uganda Distance and Parcel Effects
Log Productivity Dispersion

Within 15 mins

———	lnTFPA – IVCRC production function residuals
—·—·—	lnTFPB – IVCRC Corrected for risk and measurement error
— — —	lnTFPA – 2sls production function residuals
- - - - -	lnTFPB – 2sls Corrected for risk and measurement error

Within Parcel

———	lnTFPA – IVCRC production function residuals
—·—·—	lnTFPB – IVCRC Corrected for risk and measurement error
— — —	lnTFPA – 2sls production function residuals
- - - - -	lnTFPB – 2sls Corrected for risk and measurement error

Fig A3: Single Managers Only
Log Productivity Dispersion

Tanzania
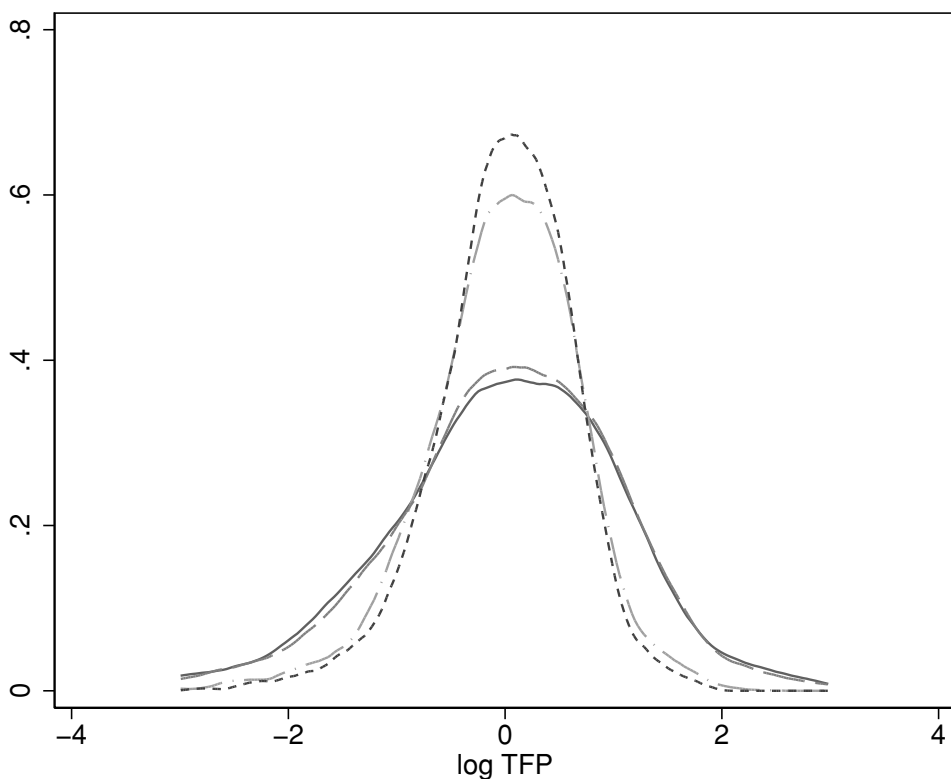
Uganda
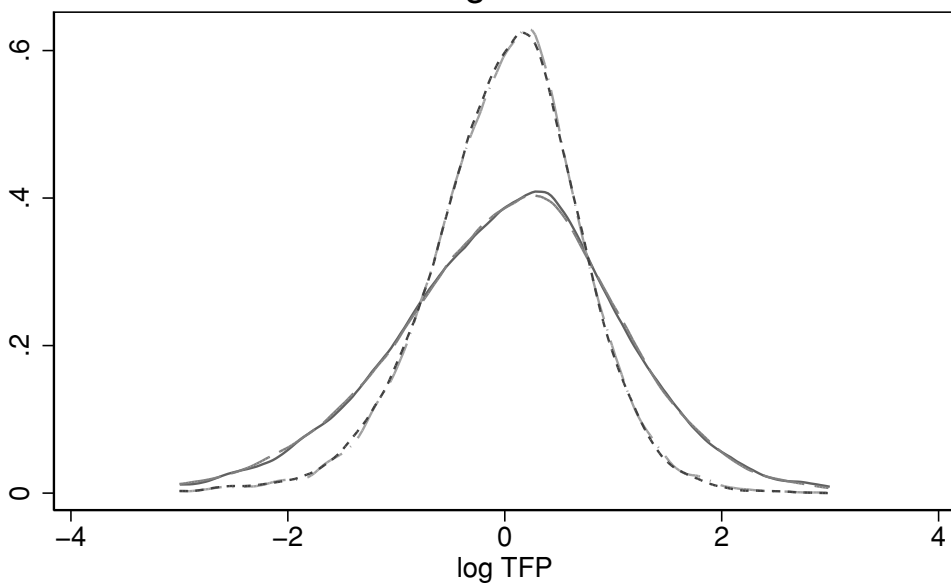
lnTFPA – IVCRC production function residuals

lnTFPB – IVCRC Corrected for risk and measurement error

lnTFPA – 2sls production function residuals

lnTFPB – 2sls Corrected for risk and measurement error

Appendix Table A1A: OLS and Quantile Regression Determinants of Log Land and Log Labor Inputs in Tanzania

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile range	
	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor
Male Manager	0.35	0.23	0.37	0.28	0.32	0.22	0.33	0.22	-0.034	-0.069
	(0.019)	(0.018)	(0.023)	(0.025)	(0.023)	(0.021)	(0.023)	(0.020)	(0.028)	(0.031)
Land Value	0.21	0.1	0.19	0.11	0.2	0.1	0.24	0.1	0.044	-0.0056
	0.009)	(0.008)	0.010)	(0.010)	(0.010)	0.009)	0.010)	0.009)	(0.014)	(0.011)
Drought/ Flood *Good Soil	0.096	0.024	0.14	0.011	0.045	-0.03	0.045	0.0023	-0.091	-0.0085
	(0.034)	0.031)	(0.040)	(0.040)	(0.041)	(0.036)	(0.044)	(0.036)	(0.044)	(0.048)
Drought/ Flood *Avg Soil	0.073	0.035	0.12	-0.0043	-0.0038	0.055	0.025	0.036	-0.1	0.04
	(0.033)	(0.031)	(0.039)	(0.043)	(0.042)	(0.036)	(0.039)	(0.036)	(0.040)	(0.067)
Drought/ Flood *Poor Soil	-0.027	-0.035	-0.014	-0.083	-0.031	-0.082	-0.0056	-0.016	0.0083	0.067
	(0.077)	(0.072)	(0.078)	(0.085)	(0.089)	(0.098)	(0.065)	(0.080)	(0.095)	(0.110)
Illness/ accident of HH mem	0.035	-0.039	0.042	-0.075	0.056	-0.0095	0.0073	0.018	-0.035	0.093

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	(0.031)	(0.029)	(0.046)	(0.040)	(0.031)	(0.034)	(0.036)	(0.036)	(0.056)	(0.043)
GS Rain *Good Soil in HH†	-4E-05	-4E-06	4E-05	4E-05	-5E-05	-7E-06	-6E-05	-3E-05	-0.0001	-7E-05
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GS Rain *Avg Soil in HH†	-6E-06	2E-05	8E-05	3E-05	-3E-05	1E-05	-3E-05	2E-05	-0.0001	-2E-05
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GS Rain *Poor Soil in HH†	-6E-05	-5E-05	-9E-05	-8E-05	-0.0001	-7E-05	-2E-05	3E-05	8E-05	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Drought/ Flood *Good Soil in HH†	0.0053	0.038	0.02	0.046	0.011	0.051	0.0067	0.032	-0.014	-0.014
	(0.017)	(0.016)	(0.017)	(0.025)	(0.018)	(0.014)	(0.016)	(0.019)	(0.022)	(0.025)
Drought/ Flood *Avg Soil in HH†	-0.009	0.02	-0.029	0.04	0.019	0.026	0.0011	0.011	0.03	-0.029
	(0.017)	(0.016)	(0.015)	(0.025)	(0.020)	(0.019)	(0.019)	(0.016)	(0.025)	(0.033)
	0.11	0.1	0.11	0.097	0.08	0.17	0.072	0.025	-0.042	-0.072
Drought/ Flood *Poor Soil in HH†	(0.045)	(0.042)	(0.038)	(0.047)	(0.075)	(0.051)	(0.028)	(0.040)	(0.072)	(0.045)
Adverse shock to Hh plots†	-0.089	-0.032	-0.11	-0.046	-0.081	-0.032	-0.061	-0.022	0.052	0.023
	(0.011)	(0.010)	(0.012)	(0.013)	(0.011)	(0.010)	(0.012)	(0.011)	(0.016)	(0.013)
Land Value Missing	1.87	0.87	1.72	0.93	1.82	0.86	2.2	0.93	0.48	-0.0055
	(0.099)	(0.092)	(0.110)	(0.120)	(0.110)	(0.100)	(0.120)	(0.092)	(0.130)	(0.140)

Distance Home	0.0017	8E-05	0.0021	0.0001	0.0017	1E-04	0.0019	-0.0003	-0.0002	-0.0004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Distance to Road	0.021	0.018	0.024	0.021	0.023	0.02	0.024	0.018	0.0007	-0.0026
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)
Good Soil	-0.093	-0.12	-0.14	-0.19	-0.067	-0.18	-0.17	-0.18	-0.029	0.0096
	(0.100)	(0.097)	(0.120)	(0.120)	(0.091)	(0.120)	(0.110)	(0.094)	(0.160)	(0.180)
Average Soil	-0.19	-0.22	-0.28	-0.29	-0.23	-0.32	-0.24	-0.25	0.036	0.038
	(0.110)	(0.098)	(0.120)	(0.120)	(0.093)	(0.120)	(0.110)	(0.095)	(0.160)	(0.160)
Sandy Soil	-0.38	-0.4	-0.43	-0.36	-0.38	-0.54	-0.5	-0.51	-0.069	-0.15
	(0.150)	(0.140)	(0.110)	(0.130)	(0.220)	(0.230)	(0.110)	(0.100)	(0.220)	(0.200)
Loamy Soil	-0.3	-0.44	-0.34	-0.36	-0.33	-0.48	-0.26	-0.45	0.077	-0.091
	(0.140)	(0.130)	(0.089)	(0.120)	(0.210)	(0.220)	(0.110)	(0.094)	(0.230)	(0.200)
Clay Soil	-0.27	-0.3	-0.26	-0.24	-0.35	-0.32	-0.23	-0.31	0.026	-0.068
	(0.150)	(0.140)	(0.100)	(0.130)	(0.210)	(0.220)	(0.120)	(0.100)	(0.240)	(0.200)
Single Manager	-0.037	-0.13	-0.02	-0.13	-0.044	-0.15	-0.045	-0.15	-0.025	-0.029
	(0.018)	(0.017)	(0.023)	(0.022)	(0.022)	(0.020)	(0.022)	(0.019)	(0.026)	(0.023)

Poor Health†	-0.017	-0.0056	-0.022	0.0017	-0.016	-0.0011	-0.019	-0.0047	0.0031	-0.0064
	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.010)	(0.009)
Missing Health†	-0.038	0.06	-0.049	0.079	-0.086	0.077	-0.031	0.027	0.018	-0.052
	(0.030)	(0.028)	(0.038)	(0.032)	(0.034)	(0.033)	(0.038)	(0.033)	(0.054)	(0.055)
Illiterate†	0.057	0.1	0.088	0.13	0.067	0.085	0.056	0.083	-0.032	-0.042
	(0.026)	(0.024)	(0.036)	(0.034)	(0.030)	(0.028)	(0.030)	(0.027)	(0.051)	(0.036)
Lit Missing†	-0.22	0.022	0.028	0.093	-0.33	-0.23	-0.11	-0.13	-0.14	-0.22
	(0.220)	(0.200)	(0.160)	(0.089)	(0.095)	(0.190)	(0.300)	(0.110)	(0.450)	(0.320)
Manager Age†	0.0041	0.0077	0.0037	0.0077	0.0033	0.0084	0.0053	0.0083	0.0015	0.0005
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age Missing†	-4.150	-7.490	-4.010	-7.390	-3.340	-8.100	-5.380	-8.120	-1.370	-0.730
	(0.560)	(0.520)	(0.690)	(0.670)	(0.660)	(0.610)	(0.700)	(0.550)	(0.800)	(0.930)
Crop Shock	0.08	0.14	0.14	0.17	0.058	0.17	0.053	0.096	-0.083	-0.079
	(0.017)	(0.015)	(0.021)	(0.021)	(0.019)	(0.018)	(0.020)	(0.017)	(0.019)	(0.021)
Crop Disease *Good Soil	-0.027	0.077	-0.033	0.067	0.034	0.096	-0.0063	0.065	0.026	-0.0014
	(0.036)	(0.033)	(0.052)	(0.043)	(0.043)	(0.038)	(0.043)	(0.038)	(0.055)	(0.048)
Crop Disease *Average Soil	0.039	0.12	0.11	0.091	0.061	0.077	0.021	0.12	-0.086	0.032

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	(0.035)	(0.032)	(0.048)	(0.047)	(0.042)	(0.036)	(0.040)	(0.036)	(0.051)	(0.045)
Crop Disease *Poor Soil	-0.16	-0.054	-0.2	-0.036	-0.12	-0.03	-0.14	-0.095	0.054	-0.059
	(0.078)	(0.072)	(0.077)	(0.100)	(0.067)	(0.087)	(0.061)	(0.074)	(0.110)	(0.130)
(soil_quality==1) *wetQ	-0.0002	-0.0002	-0.0002	-0.0003	-0.0002	-3E-05	0.0002	2E-05	0.0004	0.0003
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(soil_quality==2) *wetQ	-1E-04	-1E-04	-0.0002	-0.0001	-2E-05	0.0001	0.0002	4E-05	0.0004	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(soil_quality==3) *wetQ	-0.0003	-0.0004	-0.0005	-0.0007	-0.0004	-0.0003	-0.0001	-0.0003	0.0003	0.0004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GS Rainfall *Loamy Soil	2E-06	0.0003	0.0001	0.0003	-6E-06	7E-05	-0.0004	6E-05	-0.0005	-0.0002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GS Rainfall *Clay Soil	-4E-05	0.0002	3E-05	0.0002	-4E-07	-4E-05	-0.0005	-7E-05	-0.0005	-0.0003
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GS Rainfall *Other Soil	-0.0006	-0.0006	-0.0006	-0.0005	-0.0008	-0.0009	-0.001	-0.0008	-0.0004	-0.0003
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Water Shortage	0.12	0.049	0.099	0.078	0.14	0.045	0.14	0.065	0.043	-0.013
	(0.018)	(0.017)	(0.026)	(0.024)	(0.022)	(0.019)	(0.022)	(0.019)	(0.028)	(0.034)

Livestock death or stolen†	0.065 (0.020)	0.055 (0.018)	0.039 (0.027)	0.028 (0.023)	0.082 (0.024)	0.038 (0.021)	0.075 (0.022)	0.088 (0.023)	0.037 (0.039)	0.06 (0.037)
Death of HH member†	0.1 (0.026)	0.099 (0.024)	0.09 (0.024)	0.12 (0.039)	0.075 (0.031)	0.12 (0.027)	0.061 (0.032)	0.091 (0.022)	-0.029 (0.034)	-0.029 (0.048)
Property Crime in HH†	-0.002 (0.032)	-0.055 (0.030)	0.008 (0.034)	-0.095 (0.048)	-0.019 (0.040)	-0.038 (0.034)	-0.013 (0.029)	-0.047 (0.027)	-0.021 (0.036)	0.048 (0.051)
HH Good Soil *Shock on Plots in HH†	0.0015 (0.017)	-0.044 (0.015)	-0.031 (0.023)	-0.074 (0.025)	-0.022 (0.021)	-0.045 (0.014)	0.0072 (0.017)	-0.024 (0.020)	0.038 (0.035)	0.05 (0.020)
HH Avg Soil *Shock on Plots in HH†	-0.0052 (0.017)	-0.0072 (0.016)	0.0003 (0.019)	-0.0021 (0.023)	-0.011 (0.015)	0.011 (0.017)	-0.01 (0.019)	-0.0062 (0.015)	-0.011 (0.025)	-0.0041 (0.023)
HH Poor Soil *Shock on Plots in HH†	-0.086 (0.046)	-0.045 (0.043)	-0.034 (0.043)	-0.017 (0.050)	-0.082 (0.053)	-0.05 (0.041)	-0.12 (0.030)	-0.07 (0.048)	-0.09 (0.090)	-0.052 (0.060)
HH Loamy Soil **GS rain†	-7E-05 (0.000)	-0.0001 (0.000)	-0.0001 (0.000)	-0.0001 (0.000)	-6E-05 (0.000)	-0.0001 (0.000)	-1E-04 (0.000)	-0.0001 (0.000)	2E-05 (0.000)	-1E-05 (0.000)
HH Clay Soil * GS rain†	-6E-05 (0.000)	-7E-05 (0.000)	-0.0001 (0.000)	-0.0001 (0.000)	-6E-05 (0.000)	-8E-05 (0.000)	-6E-05 (0.000)	-8E-05 (0.000)	4E-05 (0.000)	2E-05 (0.000)

HH Other Soil * GS rain†	-0.0001	-9E-05	-0.0002	-8E-05	-0.0001	-7E-05	-0.0002	-0.0001	2E-05	-3E-05
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-2E-08	-9E-09	-0.58	-0.55	0.028	0.062	0.62	0.61	1.19	1.16
	(0.008)	(0.007)	(0.011)	(0.011)	(0.010)	(0.009)	(0.010)	(0.009)	(0.013)	(0.011)
F statistic for joint significance of insturments.	17.1	11.5	13.7	7.32	24.4	11.3	19.4	11.2		
p-value	0	0	0	0	0	0	0	0		
F statistic for h0: Coefficients equal for 25th and 75th percentiles									30.1	5.46
p-value									0	0

Note: N=14535. Bootstrapped (500 samples) standard errors, clustered at the household level, in parentheses.

†Variable serves as instrument in Table 6

Appendix Table A1B: OLS and Quantile Regression Determinants of Log Land and Log Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile range	
	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor	Log Land	Log Labor
Male Plot	0.14	0.062	0.13	0.044	0.14	0.067	0.14	0.083	0.015	0.039
	(0.0086)	(0.0071)	(0.0097)	(0.0093)	(0.0093)	(0.0075)	(0.0093)	(0.0075)	(0.014)	(0.011)
Fair Soil	-0.092	-0.078	-0.100	-0.080	-0.14	-0.098	-0.10	-0.12	-0.0045	-0.036
	(0.024)	(0.021)	(0.030)	(0.029)	(0.027)	(0.018)	(0.024)	(0.023)	(0.025)	(0.024)
Poor Soil	-0.20	-0.11	-0.19	-0.19	-0.31	-0.18	-0.13	-0.022	0.064	0.17
	(0.067)	(0.062)	(0.082)	(0.046)	(0.056)	(0.066)	(0.060)	(0.047)	(0.14)	(0.065)
Fair Soil*Drought duration	0.026	0.024	0.032	0.021	0.032	0.024	0.025	0.035	-0.0071	0.013
	(0.0052)	(0.0046)	(0.0056)	(0.0060)	(0.0057)	(0.0048)	(0.0052)	(0.0050)	(0.0081)	(0.0059)
Poor Soil*Drought duration	0.035	0.027	0.035	0.027	0.036	0.021	0.040	0.0044	0.0049	-0.023
	(0.012)	(0.0092)	(0.015)	(0.010)	(0.012)	(0.0097)	(0.010)	(0.0040)	(0.018)	(0.013)
Illness Incidence in HH†	0.0027	-0.051	-0.0019	-0.067	0.013	-0.066	-0.0078	-0.064	-0.0059	0.0027
	(0.013)	(0.0098)	(0.015)	(0.013)	(0.014)	(0.011)	(0.012)	(0.010)	(0.016)	(0.013)
HH Fair Soil*Peak Rain†	0.000082	-0.000074	0.00010	-0.000073	0.000045	-0.000070	0.000045	-0.000032	-0.000058	0.000042

	(0.000023)	(0.000019)	(0.000023)	(0.000024)	(0.000026)	(0.000020)	(0.000027)	(0.000018)	(0.000035)	(0.000027)
HH Poor Soil*Peak Rain†	0.00021	-0.000053	0.00022	-0.000037	0.00025	-0.00011	0.00022	-0.000036	0.0000057	0.000000075
	(0.000073)	(0.000055)	(0.000086)	(0.000087)	(0.000058)	(0.000081)	(0.000067)	(0.000055)	(0.00010)	(0.00011)
HH Fair Soil*Drought duration†	-0.0053	-0.0069	-0.0067	-0.0076	-0.0061	-0.0078	-0.0074	-0.0083	-0.00062	-0.00073
	(0.0018)	(0.0015)	(0.0018)	(0.0017)	(0.0019)	(0.0014)	(0.0015)	(0.0017)	(0.0026)	(0.0027)
HH Fair Soil*Total Rain†	-0.048	0.017	-0.053	0.018	-0.045	0.012	-0.041	-0.00038	0.012	-0.018
	(0.0083)	(0.0069)	(0.0082)	(0.0086)	(0.0097)	(0.0071)	(0.0098)	(0.0059)	(0.011)	(0.0092)
HH Poor Soil*Total Rain†	-0.083	0.023	-0.070	0.018	-0.097	0.037	-0.091	-0.0038	-0.020	-0.022
	(0.025)	(0.020)	(0.031)	(0.031)	(0.019)	(0.029)	(0.022)	(0.018)	(0.027)	(0.036)
Shocks on Household Plots†	0.023	-0.00034	0.017	-0.0011	0.069	0.00065	0.10	-0.00049	0.087	0.00065
	(0.0032)	(0.0013)	(0.0018)	(0.0035)	(0.0072)	(0.0012)	(0.0080)	(0.0022)	(0.0090)	(0.0029)
Household Non- Ag Shock†	0.084	-0.0063	0.083	-0.0067	0.059	-0.020	0.093	-0.024	0.010	-0.018
	(0.026)	(0.021)	(0.019)	(0.026)	(0.030)	(0.021)	(0.032)	(0.022)	(0.031)	(0.030)
Value of household assets†	1.6e-09	1.5e-10	1.7e-09	1.0e-10	1.6e-09	2.8e-10	1.5e-09	2.9e-10	-2.4e-10	1.9e-10
	(1.4e-10)	(1.3e-10)	(1.5e-10)	(1.5e-10)	(1.5e-10)	(1.2e-10)	(1.1e-10)	(1.5e-10)	(2.7e-10)	(1.7e-10)
Cement Construction†	0.068	0.023	0.059	0.019	0.052	0.0052	0.054	0.027	-0.0051	0.0077
	(0.012)	(0.0097)	(0.013)	(0.013)	(0.013)	(0.010)	(0.013)	(0.010)	(0.017)	(0.017)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
No of household members†	0.025 (0.0020)	0.010 (0.0017)	0.023 (0.0024)	0.014 (0.0022)	0.022 (0.0022)	0.011 (0.0018)	0.022 (0.0022)	0.010 (0.0018)	-0.00035 (0.0027)	-0.0041 (0.0026)
No of adults in hh†	0.017 (0.0029)	0.0081 (0.0025)	0.018 (0.0036)	0.0054 (0.0031)	0.018 (0.0031)	0.0078 (0.0026)	0.016 (0.0033)	0.0075 (0.0026)	-0.0022 (0.0043)	0.0022 (0.0041)
Every member of the household has at least one pair of shoes†	-0.0074 (0.0092)	-0.070 (0.0078)	-0.017 (0.010)	-0.094 (0.010)	0.0014 (0.0099)	-0.058 (0.0080)	0.018 (0.0098)	-0.034 (0.0081)	0.035 (0.012)	0.060 (0.013)
DOES THIS HOUSE HAVE ELECTRICITY? †	0.078 (0.020)	0.053 (0.016)	0.12 (0.022)	0.044 (0.019)	0.054 (0.021)	0.021 (0.015)	0.026 (0.023)	0.038 (0.017)	-0.094 (0.033)	-0.0062 (0.027)
Literacy of plot managers†	0.062 (0.012)	0.042 (0.011)	0.054 (0.015)	0.049 (0.013)	0.069 (0.013)	0.039 (0.011)	0.063 (0.014)	0.025 (0.011)	0.0081 (0.017)	-0.024 (0.019)
Schooling level of plot managers†	-0.0016 (0.013)	-0.00099 (0.011)	-0.021 (0.014)	0.0025 (0.015)	0.0016 (0.014)	-0.0050 (0.011)	-0.0061 (0.012)	-0.012 (0.012)	0.015 (0.019)	-0.014 (0.012)
Plot manager is recent resident†	0.015 (0.019)	0.074 (0.016)	0.022 (0.020)	0.076 (0.023)	0.038 (0.018)	0.077 (0.017)	0.045 (0.018)	0.069 (0.016)	0.023 (0.024)	-0.0064 (0.022)
One or both manager serve on committee†	0.022 (0.014)	0.013 (0.011)	0.0092 (0.016)	-0.0048 (0.014)	0.0047 (0.015)	0.029 (0.012)	-0.024 (0.015)	0.025 (0.012)	-0.034 (0.019)	0.029 (0.015)

Household Adult Equivalence Scale†	0.000025	0.000091	0.000012	0.000099	-0.000063	0.000065	-0.000014	0.000045	-0.000026	-0.000054
	(0.000048)	(0.000038)	(0.000044)	(0.000065)	(0.000058)	(0.000058)	(0.000081)	(0.000024)	(0.000074)	(0.000045)
V Fair Soil*Drought duration†	0.0031	0.00047	0.0024	0.00067	0.0030	0.00053	0.0027	0.00030	0.00030	-0.00037
	(0.00037)	(0.00030)	(0.00041)	(0.00040)	(0.00039)	(0.00033)	(0.00036)	(0.00034)	(0.00051)	(0.00031)
HH Poor Soil*Drought duration†	-0.0029	-0.0035	-0.0024	-0.0039	-0.0057	-0.0042	-0.0070	0.0033	-0.0046	0.0072
	(0.0043)	(0.0037)	(0.0053)	(0.0041)	(0.0038)	(0.0043)	(0.0049)	(0.0035)	(0.0068)	(0.0056)
V Poor Soil*Drought duration†	0.0066	-0.0012	0.0050	-0.00063	0.0019	-0.0012	0.0031	-0.0024	-0.0019	-0.0018
	(0.0013)	(0.00100)	(0.0012)	(0.0013)	(0.0011)	(0.00098)	(0.0012)	(0.0011)	(0.0018)	(0.0020)
HH Missing Soil*Drought duration†	-0.028	0.019	-0.036	0.0044	-0.023	0.013	-0.010	0.027	0.026	0.023
	(0.0099)	(0.0061)	(0.0094)	(0.0063)	(0.0081)	(0.0058)	(0.0078)	(0.011)	(0.018)	(0.011)
V Missing Soil*Drought duration†	-0.0020	0.00061	-0.0023	0.0044	-0.0016	-0.00083	-0.0033	-0.00075	-0.0010	-0.0052
	(0.0022)	(0.0017)	(0.0019)	(0.0019)	(0.0020)	(0.0015)	(0.0013)	(0.0015)	(0.0034)	(0.0022)
HH Fair Soil*Flood duration†	0.016	0.0088	0.025	0.0083	0.016	0.011	0.0067	0.0032	-0.019	-0.0052
	(0.0054)	(0.0056)	(0.012)	(0.0091)	(0.0056)	(0.0031)	(0.0060)	(0.0068)	(0.012)	(0.012)
V Fair Soil*Flood duration†	-0.0037	-0.0017	-0.0035	-0.0025	-0.0052	-0.0017	-0.0038	-0.0010	-0.00034	0.0015
	(0.0013)	(0.0011)	(0.0022)	(0.00041)	(0.0022)	(0.0012)	(0.0018)	(0.0013)	(0.0021)	(0.0020)

HH Poor Soil*Flood duration†	0.056	0.022	0.028	0.054	0.069	0.029	0.019	0.00091	-0.0090	-0.053
	(0.043)	(0.027)	(0.17)	(0.023)	(0.23)	(0.020)	(0.044)	(0.031)	(0.12)	(0.029)
V Poor Soil*Flood duration†	0.0029	0.018	0.0031	0.026	0.011	0.022	0.016	0.013	0.013	-0.013
	(0.012)	(0.0094)	(0.0075)	(0.022)	(0.0054)	(0.0099)	(0.013)	(0.013)	(0.015)	(0.019)
HH Missing Soil*Flood duration†	0.14	0.079	0.15	0.096	0.12	0.11	0.16	0.074	0.015	-0.022
	(0.037)	(0.028)	(0.083)	(0.071)	(0.28)	(0.042)	(0.074)	(0.093)	(0.079)	(0.075)
V Missing Soil*Flood duration†	-0.077	-0.018	-0.035	-0.016	-0.068	-0.019	-0.058	-0.0079	-0.023	0.0085
	(0.020)	(0.015)	(0.0098)	(0.017)	(0.018)	(0.016)	(0.0078)	(0.011)	(0.030)	(0.026)
V Fair Soil*Total Rain†	-0.0047	-0.00077	-0.0054	-0.00017	-0.0065	-0.00019	-0.0055	-0.00051	-0.000097	-0.00034
	(0.0010)	(0.00089)	(0.0013)	(0.0011)	(0.0011)	(0.00084)	(0.0010)	(0.00087)	(0.0018)	(0.0012)
V Poor Soil*Total Rain†	-0.036	-0.0030	-0.030	0.0017	-0.030	-0.0076	-0.031	-0.0063	-0.00094	-0.0080
	(0.0059)	(0.0044)	(0.0069)	(0.0060)	(0.0061)	(0.0053)	(0.0068)	(0.0046)	(0.0093)	(0.0083)
HH Missing Soil*Total Rain†	-0.00011	-0.0000067	-0.00012	0.000066	-0.000075	-0.000021	-0.00019	-0.000085	-0.000067	-0.00015
	(0.000056)	(0.000040)	(0.000068)	(0.000063)	(0.000081)	(0.000036)	(0.00006)	(0.00004)	(0.00009)	(0.000078)
V Missing Soil*Total Rain†	0.000025	0.0000059	0.000034	-0.000014	0.000036	0.000010	0.000028	0.000012	-0.000006	0.000026
	(0.000008)	(0.000006)	(0.000010)	(0.000008)	(0.000009)	(0.000007)	(0.00001)	(0.00001)	(0.00001)	(0.000007)
V Fair Soil*Peak Rain†	0.000004	0.0000040	0.000008	0.0000031	0.000010	0.000002	0.000002	0.000000	-0.000006	-0.0000030

	(0.000003)	(0.000002)	(0.000004)	(0.0000030)	(0.000003)	(0.000002)	(0.000003)	(0.00000)	(0.000004)	(0.000003)
V Poor Soil*Peak Rain†	0.000044	-0.000027	0.000032	-0.000042	0.000044	-0.000021	0.000047	-0.000012	0.000016	0.000031
	(0.000016)	(0.000012)	(0.000019)	(0.000017)	(0.000017)	(0.000015)	(0.000020)	(0.000013)	(0.000026)	(0.000022)
HH Missing Soil*Peak Rain†	0.00038	-0.00015	0.00045	-0.00028	0.00027	-0.00012	0.00041	0.000087	-0.000042	0.00036
	(0.00016)	(0.00012)	(0.00021)	(0.00016)	(0.00023)	(0.00010)	(0.00016)	(0.00012)	(0.00019)	(0.00025)
V Missing Soil*Peak Rain†	-0.000040	0.0000024	-0.000060	0.000018	-0.000070	0.0000082	-0.000039	0.0000042	0.000022	-0.000014
	(0.000017)	(0.000013)	(0.000020)	(0.000016)	(0.000017)	(0.000013)	(0.000018)	(0.000013)	(0.000017)	(0.000018)
Shocks on Village Plots†	0.0025	0.00067	0.0026	0.00091	0.0025	0.00075	0.0039	0.00084	0.0013	-0.000066
	(0.00044)	(0.00029)	(0.00055)	(0.00045)	(0.00029)	(0.00027)	(0.00054)	(0.00029)	(0.00071)	(0.00045)
Missing Soil*Drought duration	0.028	-0.017	0.013	-0.032	0.012	0.000006	0.026	0.0022	0.014	0.034
	(0.027)	(0.020)	(0.023)	(0.018)	(0.021)	(0.015)	(0.038)	(0.027)	(0.044)	(0.031)
Fair Soil*Flood duration	-0.021	-0.024	-0.057	-0.024	-0.0082	-0.023	-0.0023	-0.0095	0.055	0.015
	(0.019)	(0.018)	(0.026)	(0.011)	(0.029)	(0.015)	(0.012)	(0.021)	(0.034)	(0.034)
Poor Soil*Flood duration	-0.12	0.10	0.047	0.091	-0.14	-0.0021	-0.18	0.016	-0.23	-0.075
	(0.084)	(0.078)	(0.94)	(0.22)	(0.47)	(0.025)	(0.092)	(0.061)	(0.25)	(0.19)
Missing Soil*Flood duration	-0.46	-0.22	-0.63	-0.21	-0.31	-0.39	-0.59	-0.45	0.042	-0.25
	(0.17)	(0.12)	(0.48)	(0.42)	(0.53)	(0.25)	(0.48)	(0.50)	(0.52)	(0.29)

Fair Soil*Total Rain	0.087	-0.0070	0.15	-0.013	0.12	0.037	0.10	0.083	-0.045	0.095
	(0.029)	(0.027)	(0.036)	(0.035)	(0.032)	(0.026)	(0.034)	(0.028)	(0.043)	(0.037)
Poor Soil*Total Rain	0.087	-0.027	0.12	0.031	0.15	0.060	0.060	-0.015	-0.058	-0.046
	(0.079)	(0.068)	(0.089)	(0.062)	(0.071)	(0.068)	(0.085)	(0.073)	(0.13)	(0.084)
Missing Soil*Total Rain	-0.0011	0.00012	-0.0014	-0.00016	-0.00064	0.000064	-0.00069	0.00047	0.00067	0.00063
	(0.00026)	(0.00020)	(0.00022)	(0.00024)	(0.00021)	(0.00018)	(0.00029)	(0.00021)	(0.00033)	(0.00042)
Fair Soil*Peak Rain	-0.00011	0.00013	-0.00025	0.00016	-0.000075	0.000071	-0.000086	-0.000083	0.00016	-0.00024
	(0.00006)	(0.00006)	(0.00008)	(0.00008)	(0.00007)	(0.00006)	(0.00008)	(0.00006)	(0.00011)	(0.000073)
Poor Soil*Peak Rain	-0.00020	0.000014	-0.00039	-0.000000070	-0.000059	0.000016	-0.00019	-0.00013	0.00020	-0.00013
	(0.00017)	(0.00013)	(0.00018)	(0.00016)	(0.00017)	(0.00013)	(0.00021)	(0.00017)	(0.00025)	(0.00025)
Missing Soil*Peak Rain	-0.00029	0.00011	-0.000047	0.00018	-0.00020	0.00021	-0.00016	0.00017	-0.00011	-0.000014
	(0.00022)	(0.00018)	(0.00024)	(0.00023)	(0.00025)	(0.00018)	(0.00025)	(0.00021)	(0.00034)	(0.00043)
Missing Soil	1.48	0.24	1.54	0.56	0.82	0.16	1.52	-0.55	-0.024	-1.11
	(0.41)	(0.26)	(0.33)	(0.22)	(1.34)	(0.26)	(0.41)	(0.96)	(0.69)	(0.56)
Rainfed	-0.15	-0.062	-0.14	-0.14	-0.14	-0.057	-0.23	-0.030	-0.083	0.11
	(0.038)	(0.027)	(0.050)	(0.033)	(0.040)	(0.020)	(0.049)	(0.020)	(0.063)	(0.036)

Wetland	-0.20	-0.047	-0.18	-0.13	-0.18	-0.083	-0.26	-0.035	-0.079	0.097
	(0.049)	(0.038)	(0.068)	(0.047)	(0.048)	(0.038)	(0.065)	(0.036)	(0.083)	(0.062)
Missing water	-0.55	-0.51	-0.75	-0.45	-0.80	-0.48	-0.89	-0.20	-0.14	0.25
	(0.22)	(0.18)	(0.055)	(0.11)	(1.46)	(0.072)	(0.052)	(0.95)	(0.41)	(0.27)
Flat Land	-0.025	-0.036	-0.020	-0.043	-0.0083	-0.012	0.027	-0.019	0.048	0.025
	(0.018)	(0.013)	(0.021)	(0.019)	(0.018)	(0.014)	(0.019)	(0.014)	(0.022)	(0.026)
Gentle Slope	-0.066	-0.0087	-0.091	-0.020	-0.078	0.0035	-0.026	-0.010	0.065	0.010
	(0.017)	(0.013)	(0.020)	(0.018)	(0.017)	(0.014)	(0.019)	(0.014)	(0.023)	(0.023)
Steep Slope	-0.072	-0.029	-0.100	-0.055	-0.051	-0.0031	-0.038	0.015	0.062	0.070
	(0.027)	(0.022)	(0.034)	(0.029)	(0.028)	(0.021)	(0.032)	(0.023)	(0.041)	(0.034)
Valley	-0.024	-0.0051	-0.064	-0.023	-0.058	0.00070	-0.032	0.0067	0.032	0.029
	(0.035)	(0.025)	(0.035)	(0.034)	(0.041)	(0.029)	(0.049)	(0.026)	(0.054)	(0.047)
Other Slope	-0.33	0.29	-0.56	0.14	-0.15	0.53	-0.27	0.43	0.29	0.29
	(0.24)	(0.23)	(0.51)	(0.95)	(0.15)	(0.051)	(0.60)	(0.17)	(0.45)	(0.42)
Missing Toposequence	0.31	0.035	0.54	-0.12	0.50	0.12	0.16	0.082	-0.38	0.20
	(0.22)	(0.16)	(0.25)	(0.20)	(0.22)	(0.17)	(0.047)	(0.051)	(0.43)	(0.33)
No Erosion	0.013	-0.021	0.018	-0.031	0.026	-0.023	-0.022	-0.021	-0.041	0.0096

	(0.012)	(0.0093)	(0.013)	(0.012)	(0.012)	(0.0095)	(0.014)	(0.010)	(0.013)	(0.010)
Missing Erosion	-0.12	-0.0047	-0.0054	0.00078	-0.0038	0.037	-0.25	0.078	-0.24	0.077
	(0.11)	(0.088)	(0.13)	(0.086)	(0.098)	(0.17)	(0.23)	(0.034)	(0.23)	(0.13)
hh received advice on ag production (AGSEC9)	0.13	0.043	0.14	0.051	0.13	0.041	0.090	0.027	-0.048	-0.024
	(0.0098)	(0.0085)	(0.012)	(0.011)	(0.011)	(0.0088)	(0.010)	(0.0093)	(0.014)	(0.011)
Constant	-7.3e-09	1.0e-09	-0.52	-0.38	0.012	0.059	0.54	0.45	1.06	0.83
	(0.0041)	(0.0034)	(0.0051)	(0.0047)	(0.0048)	(0.0037)	(0.0051)	(0.0038)	(0.0052)	(0.0044)
F statistic for joint significance of insturments.	37.1	14.7	30.4	20.9	34.6	14.6	34.9	11.6		
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
F statistic for h0: Coefficients equal for 25th and 75th percentiles									71.2	36
p-value									0.00	0.00

Note: N= 43187. Bootstrapped (500 samples) standard errors, clustered at the household level, in parantheses.

†Variable serves as instrument in Table 6

Appendix Table A2A: Tanzania Production Function Estimates

	2SLS	IVCRC
Land	0.73	0.61
	(0.19)	(0.05)
Labor	0.28	0.26
	(0.26)	(0.05)
Land Value	0.04	0.07
	(0.03)	(0.01)
Land Value Missing	0.41	0.91
	(0.26)	(0.33)
Distance Home	0.00	0.00
	(0.00)	(0.00)
Distance to Road	-0.01	-0.01
	(0.00)	(0.00)
Good Soil	0.35	1.56
	(0.12)	(3.41)
Average Soil	0.20	1.39
	(0.12)	(3.43)
Sandy Soil	-0.06	-2.37
	(0.21)	(4.55)
Loamy Soil	0.02	-2.27
	(0.21)	(4.53)
Clay Soil	0.06	-2.28
	(0.21)	(4.54)
Single Manager	0.02	0.04
	(0.04)	(0.02)
Poor Health	-0.03	-0.03
	(0.01)	(0.00)
Missing Health	-0.12	-0.17
	(0.05)	(0.09)
Illiterate	-0.06	-0.09
	(0.04)	(0.08)
Lit Missing	0.07	-6.23
	(0.33)	(6.34)
Male Manager	0.02	0.07
	(0.04)	(0.02)
Manager Age	0.00	0.00
	(0.00)	(0.00)
Age Missing	3.70	4.26
	(1.45)	(6.75)
Crop Shock	-0.02	-0.02
	(0.04)	(0.01)
Drought*Good Soil	-0.07	-0.06
	(0.04)	(0.04)
Drought*Average Soil	-0.07	-0.07
	(0.04)	(0.07)
Drought*Poor Soil	0.01	0.34
	(0.09)	(3.39)
Crop Disease*Good Soil	0.02	0.01
	(0.04)	(0.06)

Crop Disease*Average Soil	-0.05	-0.06
	(0.04)	(0.03)
Crop Disease*Poor Soil	0.17	-1.54
	(0.09)	(2.65)
GS Rainfall*Good Soil	0.00	0.00
	(0.00)	(0.00)
GS Rainfall*Average Soil	0.00	0.00
	(0.00)	(0.00)
GS Rainfall*Poor Soil	0.00	0.00
	(0.00)	(0.00)
GS Rainfall*Loamy Soil	0.00	0.00
	(0.00)	(0.00)
GS Rainfall*Clay Soil	0.00	0.00
	(0.00)	(0.00)
GS Rainfall*Other Soil	0.00	-0.01
	(0.00)	(0.01)
Water Shortage	0.00	0.01
	(0.03)	(0.03)
Constant	0.00	0.01
	(0.00)	(0.24)
N	14535	14535

Appendix Table A2B: Uganda Production Function Estimates

	2SLS	IVCRC
Land	0.69	0.53
	(0.06)	(0.01)
Labor	0.23	0.43
	(0.15)	(0.01)
Fair Soil*Drought duration	-0.03	-0.03
	(0.01)	(0.00)
Poor Soil*Drought duration	-0.03	-0.03
	(0.02)	(0.01)
Missing Soil*Drought duration	0.00	-0.17
	(0.03)	(0.06)
Fair Soil*Flood duration	0.02	0.07
	(0.02)	(0.02)
Poor Soil*Flood duration	0.22	0.43
	(0.17)	(0.57)
Missing Soil*Flood duration	-0.23	-2.16
	(0.31)	(0.68)
Fair Soil*Total Rain (x10)	0.00	0.00
	(0.00)	(0.00)
Poor Soil*Total Rain (x10)	0.03	0.02
	(0.01)	(0.01)
Missing Soil*Total Rain (x10)	0.04	0.03
	(0.03)	(0.08)
Fair Soil*Peak Rain (x10)	0.00	0.01
	(0.01)	(0.00)
Poor Soil*Peak Rain (x10)	-0.07	-0.05
	(0.02)	(0.01)
Missing Soil*Peak Rain (x10)	0.06	0.01
	(0.02)	(0.01)
Fair Soil	-0.09	-0.07
	(0.04)	(0.01)
Poor Soil	-0.23	-0.17
	(0.10)	(0.04)
Missing Soil	-1.02	0.52
	(0.50)	(1.38)
Rainfed	0.19	0.18
	(0.06)	(0.02)
Wetland	0.35	0.27
	(0.08)	(0.04)
Missing water	0.51	1.22
	(0.23)	(0.77)
Flat Land	0.03	0.02
	(0.03)	(0.00)
Gentle Slope	0.08	0.07
	(0.03)	(0.01)
Steep Slope	0.10	0.08
	(0.04)	(0.01)
Valley	-0.01	0.02
	(0.05)	(0.02)

Other Slope	0.28	1.43
	(0.28)	(1.33)
Missing Toposequence	-0.20	-1.62
	(0.25)	(1.01)
No Erosion	0.07	0.07
	(0.02)	(0.00)
Missing Erosion	0.53	0.34
	(0.17)	(0.28)
Male Plot	0.10	0.11
	(0.01)	(0.00)
Ag advice	0.09	0.11
	(0.02)	(0.00)
Constant	0.00	-0.02
	(0.00)	(0.01)
N	43187	43187

Appendix Table A3: Variance Components and Productivity Dispersion for Restricted Samples

	Base		1 km		100m		Single Manager	
Tanzania	2SLS	IVCRC	2SLS	IVCRC	2SLS	IVCRC	2SLS	IVCRC
A. Variance Components								
Plot TFP	0.35	0.38	0.36	0.39	0.36	0.39	0.35	0.39
Land/Labor Productivity	0.07	0.10	0.07	0.10	0.08	0.10	0.06	0.10
Late Season Risk and Output Measurement Error	0.65	0.70	0.67	0.71	0.67	0.71	0.68	0.73
Land Measurement Error	0.08	0.03	0.08	0.03	0.08	0.03	0.13	0.07
Labor Measurement Error	0.27	0.32	0.27	0.34	0.27	0.33	0.31	0.31
Cov of TFP and Land/labor Productivity	0.06	0.09	0.06	0.09	0.06	0.09	0.07	0.10
B. Productivity Dispersion								
var(TFP A)	1.19	1.29	1.13	1.29	1.13	1.29	1.17	1.29
var(TFP B)	0.47	0.55	0.40	0.54	0.39	0.54	0.39	0.51
Implied Share of Measurement Error/Unobserved Heterogeneity in var(TFP A)	0.61	0.57	0.65	0.58	0.65	0.58	0.67	0.60

	Base		15 Minutes		Within Parcel		Single Manager	
Uganda	2SLS	IVCRC	2SLS	IVCRC	2SLS	IVCRC	2SLS	IVCRC
A. Variance Components								
Plot TFP	0.18	0.18	0.18	0.18	0.18	0.18	0.17	0.17
Land/Labor Productivity	0.07	0.07	0.08	0.08	0.08	0.08	0.06	0.06
Late Season Risk and Output Measurement Error	0.67	0.67	0.69	0.69	0.71	0.71	0.63	0.63
Land Measurement Error	0.13	0.13	0.16	0.16	0.18	0.18	0.14	0.14
Labor Measurement Error	0.18	0.18	0.20	0.20	0.18	0.18	0.20	0.20

Cov of TFP and Land/labor Productivity	0.05	0.05	0.06	0.06	0.05	0.05	0.06	0.06
B. Productivity Dispersion								
var(TFP A)	1.26	1.25	1.34	1.31	1.38	1.35	1.21	1.22
var(TFP B)	0.52	0.29	0.56	0.54	0.58	0.56	0.51	0.52
Implied Share of Measurement Error/Unobserved Heterogeneity in var(TFP A)	0.59	0.77	0.58	0.59	0.58	0.59	0.58	0.57