CSAE Working Paper WPS/2019-1

December 2018

Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture

Douglas Gollin
Oxford University

Christopher Udry[1]
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 – as much as ninety percent -- of the dispersion in measured productivity. In contrast to some previous estimates, we suggest that the potential for efficiency gains through reallocation of land across farms and farmers may be relatively modest.

[1] Authors' addresses: Oxford Department of International Development, Queen Elizabeth House, 3 Mansfield Road, Oxford OX1 3TB, United Kingdom; Department of Economics, Northwestern University, 2211 Campus Drive, 3rd Floor, Evanston, IL 60208. E-mail addresses: *douglas.gollin@qeh.ox.ac.uk* and *christopher.udry@northwestern.edu*. We are grateful to Philipp Wollburg for excellent research assistance. We have received helpful comments from seminar participants at Oxford CSAE, UAB Barcelona, Princeton, Heidelberg, UCLA, Yale, CEMFI-Madrid, Cambridge I-NET, University of St Andrew's, University of Bristol, Graduate Institute in Geneva, University of Houston, the Society for Economic Dynamics in Edinburgh, Hebrew University, the University of Manchester, Tufts University, PUC, and the Edinburgh Workshop on Agricultural Productivity, Rural-Urban Migration, and Structural Transformation. We received many helpful comments from seminar participants and colleagues, but we particularly acknowledge Stefano Caria, Cheryl Doss, Andrew Foster, Sydney Gourlay, Talip Kilic, R. Vijay Krishna, Rocco Macchiavello, Dilip Mukherjee, Gabriella Santangelo, and Chris Woodruff.

Centre for the Study of African Economies
Department of Economics · University of Oxford · Manor Road Building · Oxford OX1 3UQ
T: +44 (0)1865 271084 · F: +44 (0)1865 281447 · E: csae.enquiries@economics.ox.ac.uk · W: www.csae.ox.ac.uk

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 outputs on different plots within each growing season. This allows us to observe within-season variation 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. 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 mark-ups and pricing strategies.

Understanding the extent of misallocation in agriculture is also of interest 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 at this micro level.[2] These papers point out that in poor economies, large fractions of the workforce are employed in agriculture, in contrast to rich countries, where few people earn a living from farming. In economies where two-thirds of the people are farmers, 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. Indeed, it might by extension help explain the region's low levels of income per capita. The finding also suggests a relatively straightforward 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 the social planner would reassign land from low productivity establishments to high productivity establishments (or equivalently from bad farmers to good farmers), resulting in more than a three-fold increase in aggregate agricultural output.

The misallocation hypothesis for African agriculture is particularly plausible because of abundant evidence that the continent's agricultural markets work poorly – for land, rural labor, intermediate goods, and

---

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

output. Much land lacks formal title, and rural labor markets are often poorly integrated. Empirical tests consistently reject the hypothesis that African agricultural markets are complete.[3]

At the same time, market failure need not lead to misallocation. Development economists have repeatedly and convincingly documented the existence and effectiveness of rural customs and institutions that can stand in for complete markets, with at least limited effectiveness. Informal credit markets appear to substitute imperfectly for both formal credit markets and formal insurance markets.[4] This literature has argued that informal institutions can often succeed in avoiding gross inefficiencies – perhaps as the result of some evolutionary pressures that shape these institutions over time. From this perspective, the persistence of very costly land misallocation across farmers would pose a puzzle.

Our paper addresses the measurement of misallocation using 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), although we exploit the panel dimension of these data sets rather than the cross-section. For each of our countries, we can observe many of the same individual farmers in at least three periods.

---

[3] See, for example, Dillon and Barrett (2017) for a set of African countries; Karlan et al. (2014) for Ghana, Udry (1996a) for Burkina Faso, and Udry (1996b) for Kenya. Similar findings are common for other parts of the developing world as well; see LaFave and Thomas (2016) for Indonesia; Kaur (2016) and Jayachandran (2006) for India.

[4] Early papers in this literature included Townsend (1994) and Udry (1994); this theme was also central to numerous papers by Jean-Philippe Platteau, synthesized in Platteau (2000).

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 a large number of 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.[5] Finally, the third source of variation in productivity is heterogeneity in unobserved land quality.[6] All 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 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

[5] 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.

[6] The problem of unobserved land quality was recognized by Benjamin 1995 and Udry 1996a. More recent surveys often collect quite detailed data on soil quality, but the dimensionality of soil quality measurement can be overwhelming; see, for example, Tittonell et al. 2008.

idiosyncratic shocks, measurement error, and heterogeneity in land quality are important sources of dispersion in 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 ninety percent of the variance in measured productivity.[7] Since these are not susceptible to reallocation, our estimates for the aggregate 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.

Although our work focuses on agriculture, many of the same issues clearly matter for the broader literature on the importance of misallocation across firms in the developing world. Much of the macro literature on misallocation has abstracted entirely from measurement error and heterogeneity. Our results show that estimates of the gains from reallocation are highly sensitive to assumptions about these other sources of dispersion in measured productivity. Put simply, if most of the cross-firm dispersion in productivity arises from manager characteristics, the gains from reallocation will be large. But if most of this dispersion arises from shocks and mismeasurement, or from heterogeneity in input quality, then reallocating assets to different firms and managers will have little impact.

---

[7] 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.

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. For instance, with different market structures and institutions, farmers in our two countries might find it worthwhile to mechanize and to use tractors for land preparation and other farming activities. Given a shift from human power to mechanical power, the efficient operational size of a farm might change quite dramatically, and labor might be replaced by capital, with farm size increasing as it has in Europe and North America. Our analysis does not consider this hypothetical case. Neither do we ask whether technology adoption would take place more rapidly if farms were consolidated. 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 reviews related literature. We show how our paper connects to a number of strands in both the micro and macro literature. Section 3 presents some descriptive features of the data. In Section 4, we consider the dispersion of some partial productivity measures (output per unit land and labor per unit land) across farms. Dispersion is very high across all farms. What is perhaps more surprising is that this dispersion remains large as we zoom in from the national level to increasingly disaggregated geographic levels. We show that dispersion in productivity at the district and village levels is almost as large as that at the national level. Even within farms, the dispersion on different plots cultivated by the same farmer is substantial. We interpret this as implying that farmer characteristics do not account for the bulk of the dispersion in productivity that we observe in the national data. Moreover, we find that there are important patterns of productivity variation across plots within farms. In Section 5, we draw on these patterns of productivity variation within farms to construct a theoretical framework that models the ways in which farmers choose their plots, select the crops (or crop combinations) that they cultivate on each plot, apply inputs, and realize output. In Section 6, we use this model to motivate the estimation of agricultural production functions for our two countries. These production function estimates, together with the assumption that the allocation of inputs across the different plots of an individual farmer at a given point in time is efficient, identify the extent of measurement error and heterogeneity in technology and input quality (including production shocks) across plots. We use the estimated production functions and variances of measurement error and unobserved heterogeneity to show that the measured dispersion of TFP depends on how we control for heterogeneity and measurement error. This matters in turn for our estimates of the impact of reallocating factors of production to the most productive farmers. Section 7 discusses these results, and Section 8 concludes.

2. Background and literature review

Across sub-Saharan Africa, over 60 percent of the population lives in rural areas, and agriculture remains the dominant source of employment in most countries of the region (World Bank, World Development Indicators). Measured agricultural productivity levels are extremely low. Value added per worker in African agriculture appears to be less than half the level attained in other sectors, even after adjusting for differences in input quantity and quality (Gollin et al., 2014). In a proximate sense, these two facts imply an unpleasant agricultural arithmetic for African income levels: if many people earn their living from agriculture, and if agricultural incomes are low, then aggregate incomes will be correspondingly low.

The disparities in average labor productivity across sectors do not necessarily imply misallocation, however. Average productivity is not the same as marginal productivity, so sectoral differences in productivity could arise efficiently from sectoral differences in capital intensity, to give one example. Average labor productivity could also differ across sectors due to unobserved differences in worker skills. For instance, higher-skill individuals might tend to leave agriculture, so that average productivity would differ across the two sectors -- but there might be no difference for a worker of a particular skill level (Hicks et al. 2017).

Nevertheless, there are many reasons to consider seriously the possibility that misallocation could be an important factor in explaining sectoral differences in productivity. For a start, the sheer number of people working in agriculture suggests the possibility of misallocation. In rich countries, only one or two percent of the workforce are engaged in farming; in sub-Saharan Africa, nearly two-thirds of the workforce consists of farmers. Presumably not all the people working in African agriculture are particularly gifted as farmers. Some will surely be better than others. But for a variety of reasons, many people born in rural areas find it difficult to leave, and rural institutions in much of sub-Saharan Africa are designed to share farmland and other resources among those who remain.

This view of misallocation has motivated a series of recent papers that have explored the possibility that there are too many small farms in the developing world, with too many of these farms operated by poorly skilled farmers. This view is at the heart of work by Adamopoulos and Restuccia (2014, 2015) and Restuccia and Santaeulalia-Llopis (2017), among others. These papers explore the hypothesis that distortions in farm size may account for a large fraction of cross-country differences in agricultural productivity. Similar issues are explored in Adamopoulos et al. (2015), Chen (2016), Emran and Shilpe (2015), Gottlieb and Grobovsek (2018), Shenoy (2017), and Foster and Rosenzweig (2017).

This literature builds on a broader literature in growth economics that has emphasized the importance of misallocation across firms and plants as a potential source of cross-country productivity differences or macro fluctuations in rich countries; e.g., Syverson (2004, 2011), Petrin et al. (2011), Petrin et al. (2013). An important branch of this literature has viewed misallocation as a plausible -- and indeed likely – explanation for low aggregate productivity in widely varying contexts, including developing countries; 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 in poor countries that induce distortions to the efficient size distribution of firms.

A challenge in 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 may 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 idiosyncratic shocks (such as weather shocks), unobserved variation in input quality, and measurement error may give rise to apparent dispersion in productivity but would not necessarily indicate misallocation.

Our paper also connects with a long strand of micro development literature that has examined some of the same questions around allocative efficiency that have been taken up in the recent macro misallocation literature. The literature on efficiency within and across farms in developing countries dates back to a large literature on the rationality of farms in developing countries, starting perhaps with Chayanov's work on peasant economies (republished in English 1966) and taken up again in the work of Schultz (1964). A large literature from the past half century has tried to understand efficiency in the context of agricultural household models. This literature has explored the ways in which agricultural households facing incomplete markets may make choices that are efficient subject to a variety of constraints. Our theoretical framework and analytic approach are consistent with this literature.

3. Data and Settings

Our paper draws on two nationally representative data sets, for Tanzania and Uganda, collected by government statistical agencies in collaboration with the World Bank's program on Living Standards Measurement

Surveys – Integrated Surveys of Agriculture (LSMS-ISA). The first of these is the Uganda National Panel Survey (UNPS), which has followed about 3,200 households that were interviewed in 2009-2010, 2010-2011, 2011-2012, and 2013-14. The second is the Tanzania National Panel Survey (TZNPS), which has followed about 3,300 households that were interviewed in 2008-2009, 2010-2011, and 2012-2013. 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 for each plot 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.[8]

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 crop on multiple plots within the same year. For instance, we may observe

---

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

a single farmer growing maize on each of two or three distinct plots in the same growing season.[9] This is not particularly surprising; in many African production environments, farmers may farm non-contiguous plots because of complex patterns of inheritance and land rights, as well as for purposes of risk management. Even when the plots are contiguous, farmers may plant different plots with the same crop but at different dates or with different varieties, due to the micro characteristics of the plots or as an effort to diversify against shocks that might occur at different points in the season.

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; very few farmers use chemical fertilizers, pesticide, or herbicide. In Uganda, most farmers cultivate crops in two growing seasons per year; in Tanzania, there is only a single growing season.

[9] For convenience, we speak of "a farmer" as an individual. But our data sets actually provide quite rich data that distinguishes 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. We focus here on 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).

3.1 Descriptive statistics

Tables 1a and 1b show key descriptive statistics for our two data sets. As Table 1a shows, within some households, there are multiple farmers; for instance, within the Uganda data, the 2,592 farm households correspond to 4,989 distinct farmers (where "farmer" is defined, as per footnote 9, either as an individual or a pair of individuals). We observe these farmers over six seasons, and we end up with nearly 40,000 plot-season observations. For Tanzania, we have almost 17,000 plot-season observations. 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. Thus, for Uganda, the median number of plots per farmer-season is 4; for Tanzania, the median farmer cultivates 2 plots per season. Not all of these plots are cultivated with the same crops; the median number of plots that a given farmer cultivates with a given crop in a given season is one.

Our samples are geographically quite dispersed and are representative at the national level. The Uganda data covers over 600 villages across 81 districts; the Tanzania data come from 184 villages across 140 districts in Tanzania.

Table 1b shows yields (output per hectare) for each of the data sets. These are given in value terms because of the prevalence of multiple cropping (i.e., several crops being cultivated at the same time on a given piece of land). Multiple cropping 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.

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.[10]

The data on input intensity are somewhat less skewed. We define labor input to be all forms of labor (hired and family labor) that is reported to have been used on the plot. Median days of labor per plot are not very different across the two countries, with 32 days per plot in Uganda and 47 in Tanzania.

4. Heterogeneity, allocative efficiency, and variation in the intensity of cultivation

In this section, we document the dispersion in measured productivity across farms and plots, and we explore patterns that are evident in the data. It is useful to begin with a simple benchmark model of efficient static allocation.

[10] 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).

4.1 Efficient static allocation

If the allocation of resources were efficient, then by the second welfare theorem there are shadow prices common to all farmers such that profits are maximized on each plot, and factor marginal value products are equalized across all plots. Perfectly-measured factor ratios would be identical across all plots planted with the same crop at the same time. Note that this condition would hold even if farmers were pursuing complex diversification strategies. If in addition, there were no risk (i.e., no stochasticity) or measurement error in output, then perfectly measured output-factor ratios would also be identical across all plots.[11]

Needless to say, 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. This dispersion is large and ubiquitous, and it remains even after controlling for a variety of observable plot characteristics and observable shocks.

4.2 Dispersion of yield and factor intensity

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. This is a measure of input intensity, which is a useful alternative to the measure of realized yield. One might

[11] In this context, perfect measurement would account for differential quality of inputs and outputs.

imagine, for instance, that yield is a noisier measure, given that output realizations necessarily embody all the shocks that have occurred during the growing season. By contrast, much of the labor applied to each plot is realized before harvest and hence should not reflect all of the shocks that might alter yield.

Consider first the solid black lines in Figure 1 and Figure 2. These are the raw dispersions across plots. The corresponding variances of log output per hectare are 1.74 for Tanzania and 1.98 for Uganda. The variance of log labor input per hectare is 1.09 in Tanzania to 0.92 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.[12]

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 plot 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 shocks; we condition on measures of these shocks, and their interactions with land characteristics as well. If these observed characteristics fully account for the variation in

[12] As a different measure of dispersion, note that in the data, we can 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 17.99 (13.84) for Tanzania, and 27.32(10.00) for Uganda.

productivity, then in an efficient allocation output per hectare and labor per hectare would not vary across plots, once we control for observables.

Tables 2(a) and 2(b) 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 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 3(a) and Table 3(b) reports the same subset of coefficients of the parallel regression of labor input per hectare on the observable land characteristics, farmer characteristics, and exogenous shocks. The estimated density of the residuals from these regressions is the red/orange 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.

In the analysis thus far, the within-country data pool observations across farming systems and over multiple growing seasons. 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 2(a)-(b) and 3(a)-(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 0.81 in Tanzania and 1.28 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).[13] The variance of

[13] 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 4691%, and for Tanzania of 2376%. This reinforces our perception that the dispersion of yield across plots is quite high.

log labor per hectare also remains substantial: it is 0.44 in Tanzania and 0.43 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 reported in columns (2) of Tables 2(a)-(b) and 3(a)-(b) as our baseline measures of dispersion in productivity across plots.[14]

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

This baseline dispersion might be a consequence of *unobserved* characteristics of land; it might also reflect unobserved dimensions of risk, or measurement error in output or factor inputs. These could drive

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

dispersion, even if factors of production are allocated with full efficiency. In order to draw useful conclusions regarding the extent of factor misallocation and its implications for aggregate output loss, it is vital to disentangle these sources of variation. To do so, we rely on an assumption that *within a farm*, the allocation of resources across plots is efficient.

A farm is defined 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. 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 of the first.

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.[15]

---

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

The final two columns of Tables 2(a)-(b) and 3(a)-(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.[16]

In each country, when we consider the yield regressions of Table 2, approximately one-quarter of the baseline dispersion from the specification reported in Column (2) remains after we focus attention on variation in output per hectare across plots *within a farm*. When we look instead at the labor intensity regressions of Table 3, about one-quarter of the baseline dispersion remains after we restrict attention to variation within farms.

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 measurements 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

[16] The fourth, penultimate column reports the results of regressions with household-crop-season fixed effects. 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 (1996a) 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 improve our estimate of 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 5.

4.3 Plot Size, Yield, and Factor Intensity

As we seek to disentangle the different sources of dispersion in yield and input intensity, we next turn to a clue that emerges from a simple reduced form analysis of the data. 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 -0.75 (s.e.=0.02) for Uganda and -0.61 (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 the same Uganda and Tanzania data sets that form part of our analysis. 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.

Bevis and Barrett (2017) hypothesize that there is an "edge effect" on land productivity, in which plants near the boundary of a plot receive more attention in cultivation from the farmer, and perhaps have access to better nutrients and water then plants in the center of a plot. They provide evidence from the shape of plots that this edge effect explains part of the negative plot size – yield relationship in Uganda. Finally, Gourlay et al. (2017) report the results of a methodological experiment, also in Uganda, which carefully examined the problem of misreporting output data from farmers. They argue that farmers misreport crop harvests and that this measurement error is not random. Self-reported yields are biased upward compared to measurement of crop cuts at harvest. This effect is stronger on smaller plots. Gourlay et al. (2017) find that taking into account this measurement error fully explains the inverse plot size – yield relationship in a sample of farms in eastern Uganda.

We find, however, that labor per hectare is also strongly declining in plot size within a farm (Column 5 of Table 3). In Uganda and Tanzania, the elasticity of labor intensity with respect to plot size is almost identical to that of yield with respect plot size. The consistency of the estimates of the correlations of plot size with labor intensity and with yield in Uganda and Tanzania suggests a potentially different interpretation: namely, that smaller plots have higher unmeasured land quality.[17]

---

[17] 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, 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

These correlations lead us to hypothesize that there is an important 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 play a role in the strong inverse relationship observed between cultivation intensity and plot size across the plots of a farmer.

5. 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 part of a farm but spares another part. 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.

---

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.

The importance of this kind of heterogeneity, often at a very fine-grained spatial level, is well documented in agronomic and economic studies.[18] Farmers can and do modify their practices to reflect heterogeneity of this kind; e.g., choosing plot boundaries that reflect spatial differences in soil 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 locations.

5.1 Agricultural production with continuous variation in land quality

For the purposes of the model, let us assume without loss of generality that a household consists of a single farmer.[19] The farm household, indexed

[18] 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). 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.

[19] In what follows, we will use the terms "household" and "farmer" interchangeably. As noted above, the data allow us to distinguish multiple plot managers within a given household, and we will ultimately treat individuals and pairs of individuals as "farmers."

by h, holds a fixed endowment of land denoted by L_h. This land consists of a continuum of locations that can be indexed by k on the interval $[0, \mathrm{L}_h]$.

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 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 household h at location k conditional on the shock s having been realized will be given by:

$$\mathrm{q}_h(k, s) = \gamma_h(k, s)\zeta_h(k)\big(\xi_h(k, s)\big)^{\theta}. \tag{1}$$

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

$$\max_{\xi_h(k,s)} \left[\gamma_h(k, s)\zeta_h(k)\big(\xi_h(k, s)\big)^{\theta} - w_h\xi_h(k, s)\right]. \tag{2}$$

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:

$$\mathrm{q}_h^*(k,s) = \zeta_h(k)\gamma_h(k,s)\left(\frac{\theta\gamma_h(k,s)\zeta_h(k)}{w_h}\right)^{\frac{\theta}{1-\theta}}. \tag{3}$$

5.2 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, production takes place at the level of a *plot*. We define a plot (consistent with the definition used in most surveys) as a set of contiguous locations that are treated with an identical input bundle. In the rural contexts that we seek to model, a plot is a unit of land that is prepared in the same way and at the same time. The same crop (or crop mix) is planted across the plot, and the same inputs – including labor – are used across the plot.[20]

For a household in our model, a plot will be defined as a contiguous set of locations $[\underline{k}, \overline{k}] \subseteq [0, \mathrm{L}_h]$. The household 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 household h now applies inputs with the same intensity across

[20] We note that in some farm surveys, nomenclature varies. Our usage is consistent with the Uganda LSMS-ISA data, in which contiguous parcels of land are divided into plots. The individual plots may be planted with different crops and may be farmed with different inputs and management techniques.

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 $X_{hi} = \xi_{hi}(s)(\overline{k} - \underline{k})$.[21] Then output at any point on that plot is given by:

$$q_h(k,s) = \gamma_h(k,s)\zeta_h(k)\left(\frac{X_{hi}}{L_{hi}}\right)^{\theta}. \tag{4}$$

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

$$\begin{aligned} Y_{hi}(s) &= \int_0^{L_{hi}} \gamma_h(k,s)\zeta_h(k)\left(\frac{X_{hi}}{L_{hi}}\right)^{\theta} dk \\ &= \left(\frac{X_{hi}}{L_{hi}}\right)^{\theta} \int_0^{L_{hi}} \gamma_h(k,s)\zeta_h(k)dk. \end{aligned} \tag{5}$$

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, we get:

$$Y_{hi}(s) = L_{hi}\zeta_{hi}(s)\left(\frac{X_{hi}}{L_{hi}}\right)^{\theta} = \zeta_{hi}(s)L_{hi}^{1-\theta}(X_{hi})^{\theta} \tag{6}$$

[21] 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.

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:

$$\max_{\xi_{hi}(s)}\left[[\bar{k}-\underline{k}]\left(\xi_{hi}(s)\right)^{\theta}\int_{\underline{k}}^{\bar{k}}\gamma_{hi}(k,s)\zeta_{hi}(k)\,dk - w_h[\xi_{hi}(s)][\bar{k}-\underline{k}]-c\right] \tag{7}$$

Note that in equation (7), the bluntness of input use means that the profit-maximizing input bundle $\xi^*_{hi}\ s$ differs from the "precision agriculture" levels that would be chosen if the household 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 $\left(\bar{k}-\underline{k}\right) \to 0$ and $\left[\mathrm{Y}^*_{hi} - \int_{\underline{k}}^{\bar{k}} \mathrm{q}^*_{h}(k)dk\right] \to 0$. With $c > 0$, the household 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.

5.3 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 household depends inversely on the fixed cost and is also positively related to the average land quality across the farm. A farm household with very poor average land quality will *ceteris paribus* have a smaller maximum number of plots than a farm household 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 should hold more generally: everything else equal, poor quality plots must be sufficiently large that they will earn positive profits.

Consider the profit maximization from above for the ith plot cultivated by household 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, 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:

$$L_{hi}^{min} = \frac{c w^{\frac{\theta}{1-\theta}}}{\left(\zeta_{hi}(s)\right)^{\frac{1}{1-\theta}}\left(\theta^{\frac{\theta}{1-\theta}} - \theta^{\frac{1}{1-\theta}}\right)}. \tag{8}$$

This relationship is illustrated in Figure 4, which shows the negative relationship between the average quality of the plot and the minimum profitable plot size. This threshold plot size will also vary with the fixed cost. The higher the fixed cost of cultivating a plot, the larger will a plot needs to be, for any given quality, to make it profitably cultivable.

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.

5.4 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 4, based on the plot-level production function derived in Section 5.2.

We rewrite the production function derived in (6) 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 average of land quality at each point, denoted as $\zeta_{hi}(k)$; and the average effect of shocks received in season t at each point $\gamma_{hi}\big(k, s(t)\big)$. 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:

$$\begin{aligned} Y_{hit} &= e^{\omega_{hit}} (L_{hit})^{\varphi_{Lhit}} (L_{hit})^{1-\theta} (X_{hit})^{\theta} \\ &= e^{\omega_{hit}} (L_{hit})^{\alpha_{Lhit}} (X_{hit})^{\alpha_{Xhit}}. \end{aligned} \tag{9}$$

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.[22] In

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

addition, unobserved heterogeneity in factor productivity gives us α_{Lhit} and α_{Xhit} that may be heterogeneous across plots.

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. Therefore, Equation (9) is an example of a model with correlated random coefficients.

Total factor productivity has three components. The first is a 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.

$$\omega_{hit} = W_{Yhit}\beta_Y + \omega_{Yhit} + \epsilon_{Yhit} \tag{10}$$

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}) and subject to classical measurement error ϵ_{Jhit}:

$$J_{hit} = J^o_{hit} e^{W_{Jhit}\beta_J - \epsilon_{Jhit}}. \tag{11}$$

As a notational convenience, we use lower-case y to represent log output; similarly, x and l represent respectively log labor and land. The production function we estimate in logs, therefore is:

$$y_{hit} = \alpha_{Lhit} l^o_{hit} + \alpha_{Xhit} x^o_{hit} + W_{Yhit}\beta_Y + \sum \alpha_{Jhit}\left(W_{Jhit}\beta_J - \epsilon_{Lhit}\right) + \omega_{Yhit} + \epsilon_{Yhit} \quad (12)$$

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.

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 measures of household wealth and demographics comprise the set of instruments G_{hit} for plot i.

The heterogeneity across plots in factor productivity (α_{Lhit} and α_{Xhit}) implies that the effects of the instruments on factor demands are also heterogeneous across plots. Therefore, we estimate the parameters β_Y, β_J and the expected values of α_{Lhit} and α_{Xhit} using the Instrumental Variables Correlated Random Coefficients (IVCRC) estimator of Masten and Torgovitsky (2016), which allows for heterogeneity in the first-stage regressions of the instruments on the endogenous factor demands.

5.5 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:

$$\begin{aligned} \ln \widehat{TFP}^a_{hit} = y_{hit} - \hat{\alpha}_L l^o_{hit} - \hat{\alpha}_X x^o_{hit} + \hat{\alpha}_K k^o_{hit} - W_{Yhit}\hat{\beta}_Y \\ - \hat{\alpha}_L W_{Lhit}\hat{\beta}_L - \hat{\alpha}_X W_{Xhit}\hat{\beta}_X \end{aligned} \tag{13}$$

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 (12) and (13), we observe:

$$\begin{aligned}\ln\widehat{TFP}^{a}_{hit} = &\underbrace{\omega_{Yhit}}_{\text{unobs tfp}} + \underbrace{\sum_{J\in\{L,X\}}\left(\alpha_{Jhit}-\hat{\alpha}_J\right)\left(W_{Jhit}\hat{\beta}_J\right)}_{\substack{\text{unobs productivity of}\\ \text{observed characteristics}}} \\ &+ \underbrace{\sum_{J\in\{L,X\}}\left(\alpha_{Jhit}-\hat{\alpha}_J\right)\ln(J^{o}_{hit})}_{\substack{\text{unobs productivity}\\ \text{of factors}}} \\ &- \underbrace{\sum_{J\in\{L,X\}}(\alpha_{Jhit}-\hat{\alpha}_J)\epsilon_{Jhit}}_{\substack{\text{unobs productivity of}\\ \text{factor measurement error}}} - \underbrace{\sum_{J\in\{L,X\}}\hat{\alpha}_J\epsilon_{Jhit}}_{\substack{\text{factor}\\ \text{measurement error}}} \\ &+ \underbrace{\epsilon_{Yhit}}_{\substack{\text{post-input shocks and}\\ \text{measurement error in y}}}\end{aligned} \tag{14}$$

The final three terms are sources of variation in *measured* productivity (i.e., $\ln\widehat{TFP}^{a}_{hit}$), but they do not give rise to *actual* productivity variation. This implies, for instance, that 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

$$\begin{aligned} var\left(\ln\widehat{TFP}^{a}_{hit}\right) = \\ var\left(\omega_{Yhit} + \sum_{J\in\{L,X\}}\left(\alpha_{Jhit}-\hat{\alpha}_J\right)\left(W_{Jhit}\hat{\beta}_J + \ln(J^{o}_{hit})\right)\right) \\ + \hat{\alpha}_L^2 var(\epsilon_{Lhit}) + \hat{\alpha}_X^2 var(\epsilon_{xhit}) + var(\epsilon_{Yhit})\end{aligned} \tag{15}$$

Only the first term of (15) 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 these 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. There should be no market failures that affect farmers' allocative decisions across his/her plots.[23] Uninsured risk, imperfect financial markets, labor market frictions and missing markets for land all assuredly influence a farmer's input choices. However, by the second welfare theorem, there exist farm-season-specific shadow prices (p_{Lht}, p_{Xht}), with $p_{Yht} = 1$ the numeraire, such that the input choices maximize plot profits on all the plots of a farm. Therefore, log factor inputs satisfy:

$$\begin{aligned} l^o_{hit} + W_{Lhit}\beta_L = l_{hit} = \ln\left(\frac{\alpha_{Lhit}}{p_{Lht}}\right) + y_{hit} - W_{Hhit}\beta_H - \epsilon_{Yhit}; \\ x^o_{hit} + W_{Xhit}\beta_X = x_{hit} = \ln\left(\frac{\alpha_{Xhit}}{p_{Xht}}\right) + y_{hit} - W_{Hhit}\beta_H - \epsilon_{Yhit}, \end{aligned} \tag{16}$$

in familiar Cobb-Douglas fashion. Similarly, log output is:

$$\begin{aligned} y_{hit} = W_{Hhit}\beta_H + \epsilon_{Yhit} \\ + \frac{1}{1-\sum_J \alpha_{Jhit}}\left\{W_{Ehit}\beta_E + \omega_{Yhit} \right. \\ \left. + \alpha_{Lhit}\ln\left(\frac{\alpha_{Lhit}}{p_{Lht}}\right) + \alpha_{Xhit}\ln\left(\frac{\alpha_{Xhit}}{p_{Xht}}\right)\right\} \end{aligned} \tag{17}$$

Observed factor inputs and output, unsurprisingly, depend on the costs of the factors, on the realization of factor productivity and total productivity, and the realizations of factor and output measurement error and risk.

Factor shadow costs p_{Lht} and p_{Xht} vary in unspecified and unobserved ways across farmers as a consequence of factor and other market imperfections. Within a farm, however, factor prices are constant across

[23] Clearly it is possible for market failures to affect allocations within the *household* (e.g., between husband and wife, as in Udry (1996a). But our analysis will focus on individual farmers, rather than on households.

plots. Therefore, differencing plot output and input demands from within-farm means eliminates the shadow prices. Within farm variation in observed 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 $\left(\sigma_{\epsilon y}^2\right)$, as well as the covariance of plot-level total factor productivity and factor-specific productivity $\left(\sigma_{QL}, \sigma_{QX}\right)$.

We will not separately identify variation in all three types of unobserved heterogeneity in factor $(\omega_{Lhit}, \omega_{Xhit})$ or 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 iid 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 A2) 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 $var\left(\ln\widehat{TFP}_{hit}^a\right)$, 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, $\ln\widehat{TFP}_{hit}^a$, 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

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

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

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

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

6. Empirical analysis

6.1 Estimation

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.

The 2SLS estimates are consistent for the expected value of the production function coefficients if the effects of the instruments on land and labor demand are homogeneous (Heckman and Vytlacil, 1998; Wooldridge 2008). However, we do not expect this homogeneity, because the effect of a change in the opportunity cost of an input on input demand should vary depending upon the marginal product of that factor. The IVCRC estimator uses a control function estimated with first stage quantile regressions, with the conditioning approximated with kernel weights. The expected values of the production function factor coefficients are estimated from an average of weighted linear regressions, with the weights determined by the first stage quantile regression (Masten and Torgovitsky 2016).

Tables 4(a)–(b) present OLS and quantile regression estimates of the determinants of land and labor inputs into production in Tanzania and Uganda, respectively.[24] 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 labour 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 4(a)–(b) shows that the instruments are strong predictors of plot-level land and labor demand. For example, when growing conditions are good, demand for both land and labor on a specific plot is lower when the other plots in the household have loamy soil (in Tanzania), and when household illness raises the shadow price of labor (in Uganda).

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 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 Uganda, at the 25th percentile of the demand for land, the effect of adverse shocks on other plots is less than one-quarter as strong

[24] Bootstrapped standard errors, clustered at the household-season level, are reported. The full sets of coefficient estimates are presented in Appendix Tables A4a and A4b.

as it is for plots at the 75th percentile. 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 5 (a).

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 by year-season by region fixed effects are included, as are a rich set of observable characteristics of land and labor, as well as plot-level observable shocks.[25] 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 macroeconomic data, and a much smaller share for labor.[26] The preferred IVCRC estimates imply a larger share for labour and a smaller share for land than the 2SLS, and returns to scale that are much closer to unity.

---

[25] Bootstrapped standard errors clustered at the household-season level are reported. The full sets of coefficients are reported in Appendix Tables A5a and A5b. For Tanzania, 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 drought or floods, crop disease or pests, severe water shortage or other shocks that led to crop loss, and the seasonal maximum enhanced vegetation index for the village, interacted with soil type and soil quality dummies. For Uganda, six categories of soil type, four categories of soil quality, three sources of water, 6 indicators of plot toposequence, 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 season rain and enhanced vegetation index and their interactions with plot level soil quality.

[26] In the United States a labor share is often taken to be about 50%, with land perhaps 15%, and capital about 35%.

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 3(a)–(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 (1.44); the 90–10 log difference in TFP is 2.70 (2.76), corresponding to a 90–10 ratio of 15 (16) in TFP levels.[27]

Equations (14) and (15), however, clarified that $\ln TFP^a$ 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 5(b) 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 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, (σ_q^2), is also important. Cross-plot within-farm variation in factor specific productivity also exists, but is relatively less important. The results are similar whether they are

---

[27] Using the 2SLS estimates of the production function, the variance for Tanzania (Uganda) is 1.27 (1.38) and the 90–10 log difference in TFP is 2.72 (2.82).

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 5(b) 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 $\ln\widehat{TFP}^b_{hit}$ is an upper bound on the dispersion of unobserved total factor productivity across plots more generally.[28]

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 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.[29]

[28] 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.

[29] 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

The effect of correcting the estimates of unobserved total factor productivity for measurement error and late season risk is visually apparent in Figures 3(a)–(b). These figures provide kernel estimates of the densities of $\ln\widehat{TFP}_{hit}^{a}$ and $\ln\widehat{TFP}_{hit}^{b}$ 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.

6.2 Implications for Characterizing Misallocation

Late-season production shocks and measurement error in factors of production and output together account for about two-thirds to ninety percent 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. However, a simple calculation serves to illustrate the order of magnitude of this effect.[30]

---

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.

[30] 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? If labor is reallocated along with land, how should one account for the shifting allocations of population – or even for the potential exit from agriculture of some of the labor force? Many other potential questions arise, and 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.

Consider a Cobb Douglas production function without factor-specific productivity heterogeneity or measurement error in factors of production:

$$Y_i = e^{\omega_i+\epsilon_i}(L_i)^{\alpha_L}(X_i)^{\alpha_X} \tag{18}$$

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. 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

$$Y_i^e = \left(\frac{1}{\bar{\omega}}\right)^{\alpha_L+\alpha_X} e^{\frac{\omega_i}{1-\alpha_L-\alpha_X}+\epsilon_i}(\bar{L})^{\alpha_L}(\bar{X})^{\alpha_X}. \tag{19}$$

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

$$\begin{aligned} E(Y_i^e) &= \left(\frac{1}{\bar{\omega}}\right)^{\alpha_L+\alpha_X}(\bar{L})^{\alpha_L}(\bar{X})^{\alpha_X}E(e^{\epsilon_i})E\left(e^{\frac{\omega_i}{1-\alpha_L-\alpha_X}}\right) \\ &= \left(\frac{1}{\bar{\omega}}\right)^{\alpha_L+\alpha_X}(\bar{L})^{\alpha_L}(\bar{X})^{\alpha_X}E(e^{\epsilon_i})e^{\frac{E(\omega_i)}{1-\alpha_L-\alpha_X}}\left(e^{\frac{\sigma_\omega^2}{2(1-\alpha_L-\alpha_X)^2}}\right) \\ &\equiv \overline{Y^e}(\sigma_\omega^2), \end{aligned} \tag{20}$$

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. Relative to an existing baseline allocation, the gains to efficient reallocation are proportional to $e^{\frac{\sigma_\omega^2}{2(1-\alpha_L-\alpha_X)^2}}$, and therefore depend on the dispersion of total factor productivity and the concavity of the production function. An overestimate of the variance of total factor productivity 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 $\sigma_A^2 = Var(\ln\widehat{TFP}_{hit}^a) > Var(\ln\widehat{TFP}_{hit}^b) = \sigma_B^2$. 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

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

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.

7. 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). We do find that the gains from reallocation are non-trivial, but they are certainly not of such a magnitude as to account in a macro sense for the aggregate differences in agricultural productivity – or income per capita – between rich and poor countries.

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.

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.

8. 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 as much as an order of magnitude. 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.

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. NBER Working Paper 23039. Cambridge, Mass.: National Bureau of Economic Research.

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. 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? Mimeo: Cornell University. Paper presented at the 2017 Centre for the Study of African Economies (CSAE) Conference.

Bils, Mark, Peter J. Klenow, and Cian Ruane. 2017. Misallocation or mismeasurement? Mimeo: Department of Economics, Stanford University.

Carletto, C., Gourlay, S. & Winters, P., 2015. From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis. *Journal of African Economies*, Volume 24(5): 593-628.

Carletto, C., Gourlay, S., Murray, S. & Zezza, A., 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.

Chayanov, A. V. 1966. *The Theory of Peasant Economy*. New York: Irwin. (Originally published in Russian in 1925.)

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.

Collard-Wexler, Allan and Jan De Loecker. 2016. Production function estimation with measurement error in inputs. NBER Working Paper No. w22437. Cambridge, Mass.: National Bureau of Economic Research.

Da-Rocha, José-María, Marina Mendes Tavares, and Diego Restuccia. 2017. Policy distortions and aggregate productivity with endogenous establishment-level productivity. NBER Working Paper No. w23339. Cambridge, Mass.: National Bureau of Economic Research.

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.

Dillon, Brian and Christopher B. Barrett. 2017. Agricultural factor markets in Sub-Saharan Africa: An updated view with formal tests for market failure. *Food Policy* 67(February): 64-77.

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.

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. NBER Working Paper No. w23909. Cambridge, Mass.: National Bureau of Economic Research.

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.

Gollin, Douglas, David Lagakos, and Michael Waugh. 2014. The agricultural productivity gap. *Quarterly Journal of Economics* 129(2): 939-993.

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. Mimeo: 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): pp. 3-26.

Haltiwanger, John, Robert Kulick, and Chad Syverson. 2018. Misallocation measures: The distortion that ate the residual. NBER Working Paper w24199. Cambridge, Mass.: National Bureau of Economic Research.

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, 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. NBER Working Paper No. w23253. Cambridge, Mass.: National Bureau of Economic Research.

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.

Jayachandran, Seema. 2006. Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of Political Economy* 114(3): 538-575.

Kalemli-Ozcan, Sebnem and Bent E. Sorensen. 2012. Misallocation, property rights, and access to finance: Evidence from within and across Africa. NBER Working Paper 18030. Cambridge, Mass.: National Bureau of Economic Research.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry. 2014. Agricultural decisions after relaxing credit and risk constraints. *Quarterly Journal of Economics* 129 (2): 597-652.

Kaur, Supreet. 2016. Nominal wage rigidity in village labor markets. NBER Working Paper No. 20770. Cambridge, Mass.: National Bureau of Economic Research.

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.

LaFave, Daniel, and Duncan Thomas. 2016. Farms, families, and markets: New evidence on completeness of markets in agricultural settings. *Econometrica* 84(5): 1917-1960.

Levinsohn, James, and Amil Petrin. 2003. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70(2): 317-341.

Masten, Matthew A., and Alexander Torgovitsky. 2016. Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics* 98(5): 1001-1005.

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. NBER Working Paper No. w3977. Cambridge, Mass.: National Bureau of Economic Research.

Pellegrino, Bruno and Geoffery Zheng. 2018. What is the extent of misallocation? Mimeo: 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.

Platteau, Jean-Philippe. 2000. *Institutions, Social Norms and Economic Development*. London: Routledge.

Porzio, Tommaso. 2016. Cross-country differences in the optimal allocation of talent and technology. Mimeo: 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. NBER Working Paper 23128. Cambridge, Mass.: National Bureau of Economic Research.

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. Mimeo: 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.

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, P.D., K.D. Shepherd, B. Vanlauwe, and K.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. Mimeo: USAID Feed the Future Innovation Lab for Assets and Market Access Working Paper.

Townsend, Robert M. 1994. Risk and insurance in village India. *Econometrica* 62(3): 539-591.

Udry, Christopher. 1994. Risk and insurance in a rural credit market: An empirical investigation in northern Nigeria. *The Review of Economic Studies*, 61(3): 495-526.

Udry, Christopher. 1996a. Gender, agricultural production, and the theory of the household. *Journal of Political Economy* 104(5): 1010-1046.

Udry, Christopher. 1996b. Efficiency in market structure: testing for profit maximization in African agriculture. Mimeo: Northwestern University.

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

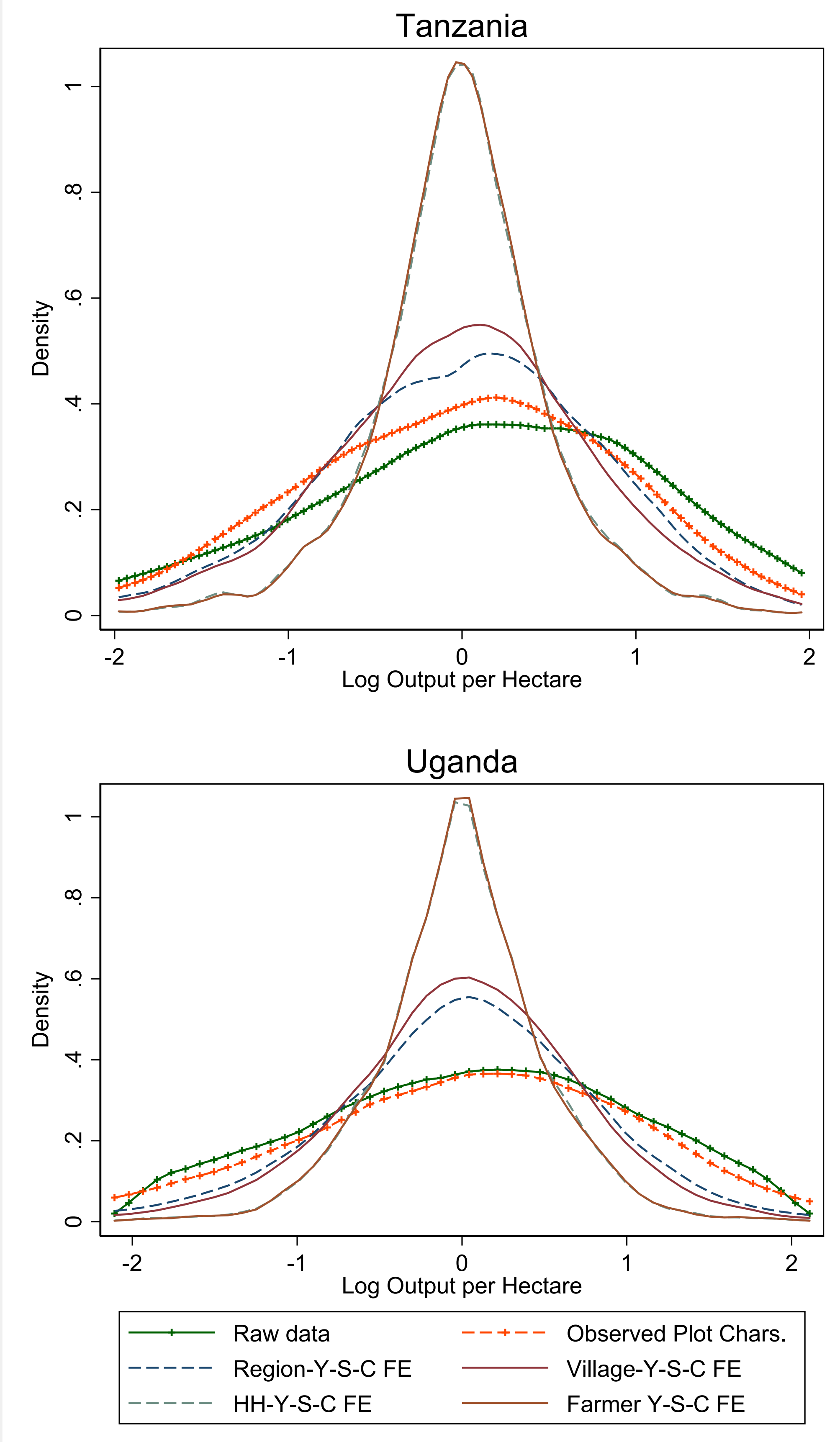

Figure 2: Dispersion of Log Labour per Hectare Across Plots

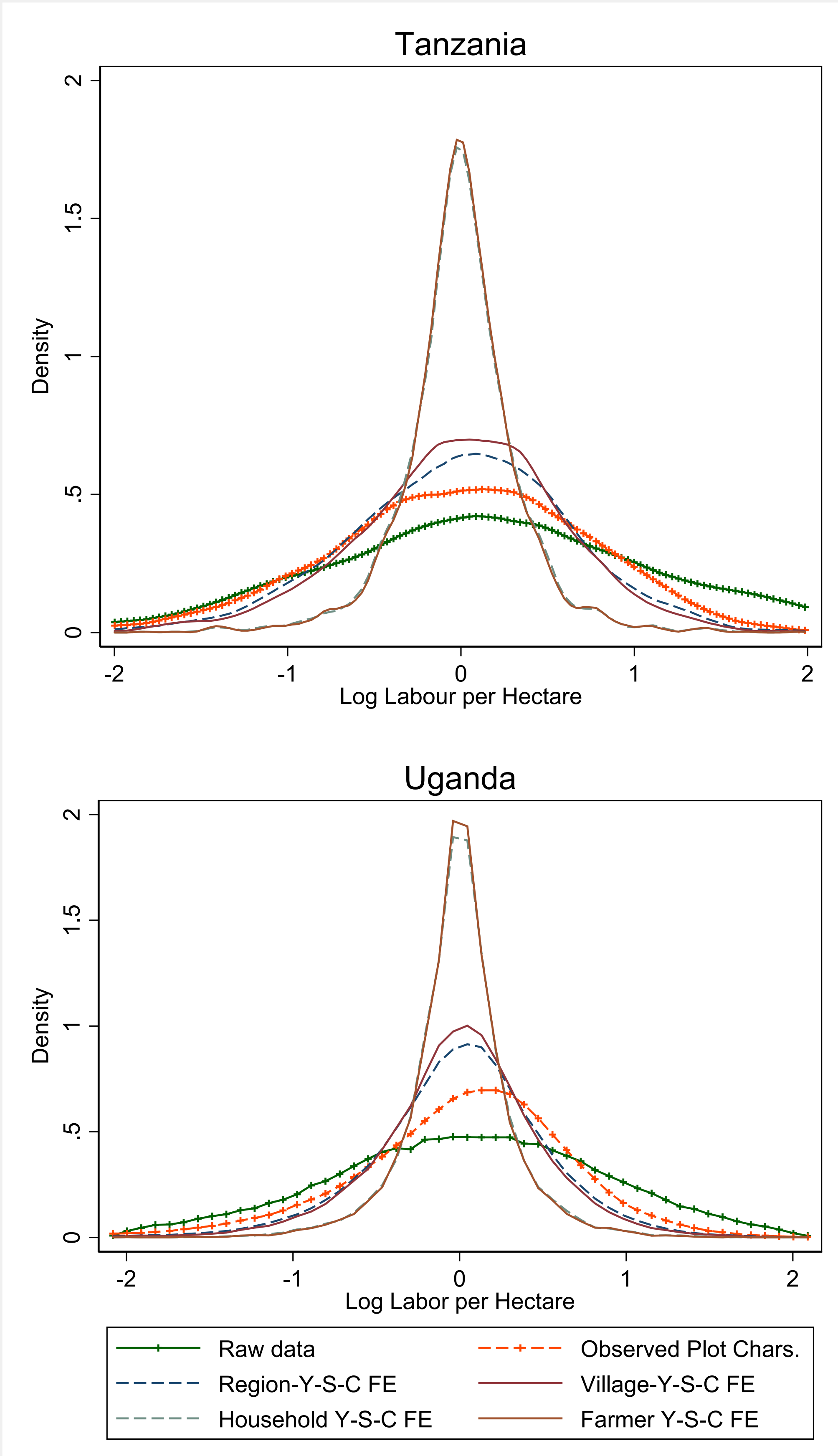

Figure 3: Log Productivity Dispersion

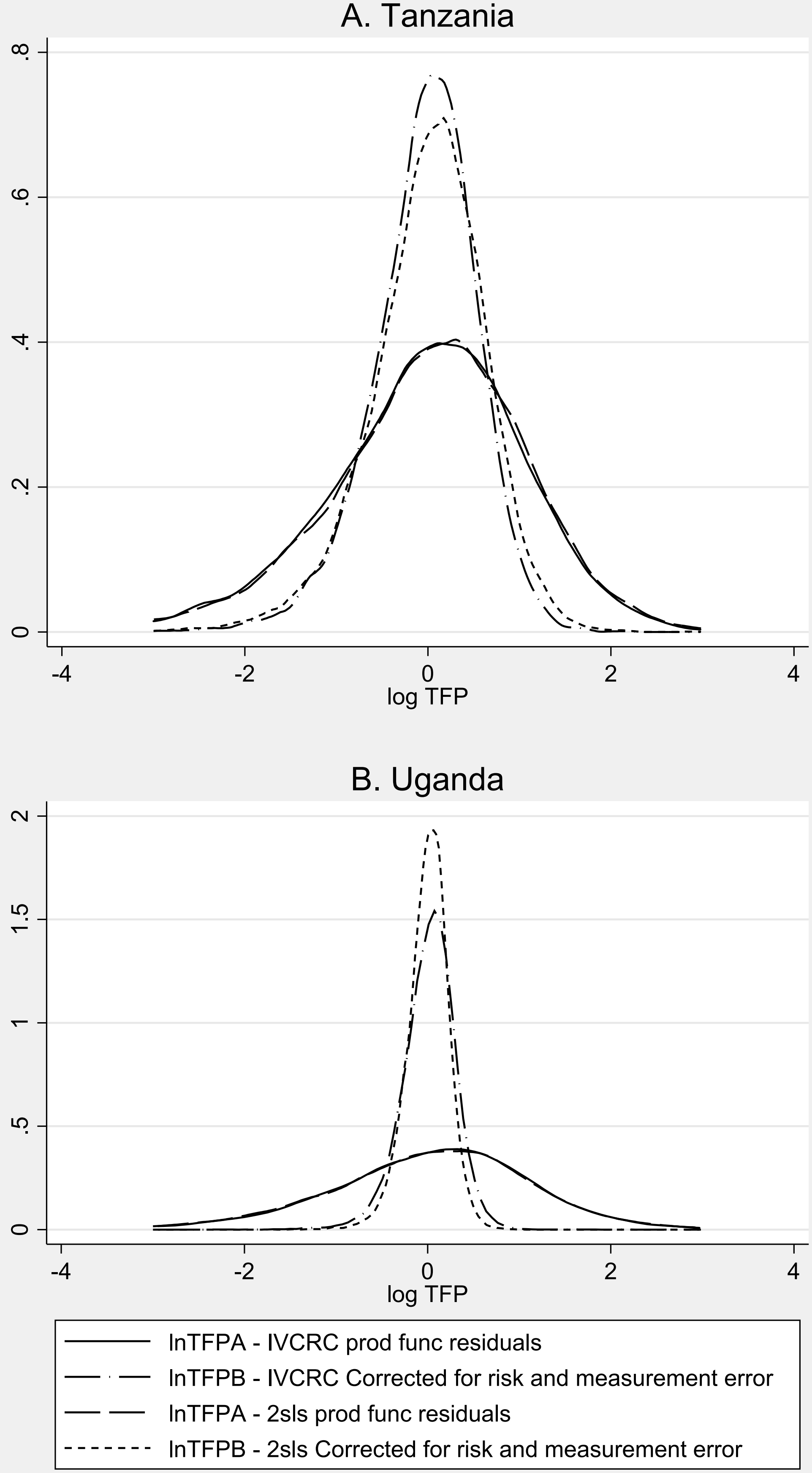

Figure 4: Minimum profitable plot size in relation to average plot quality

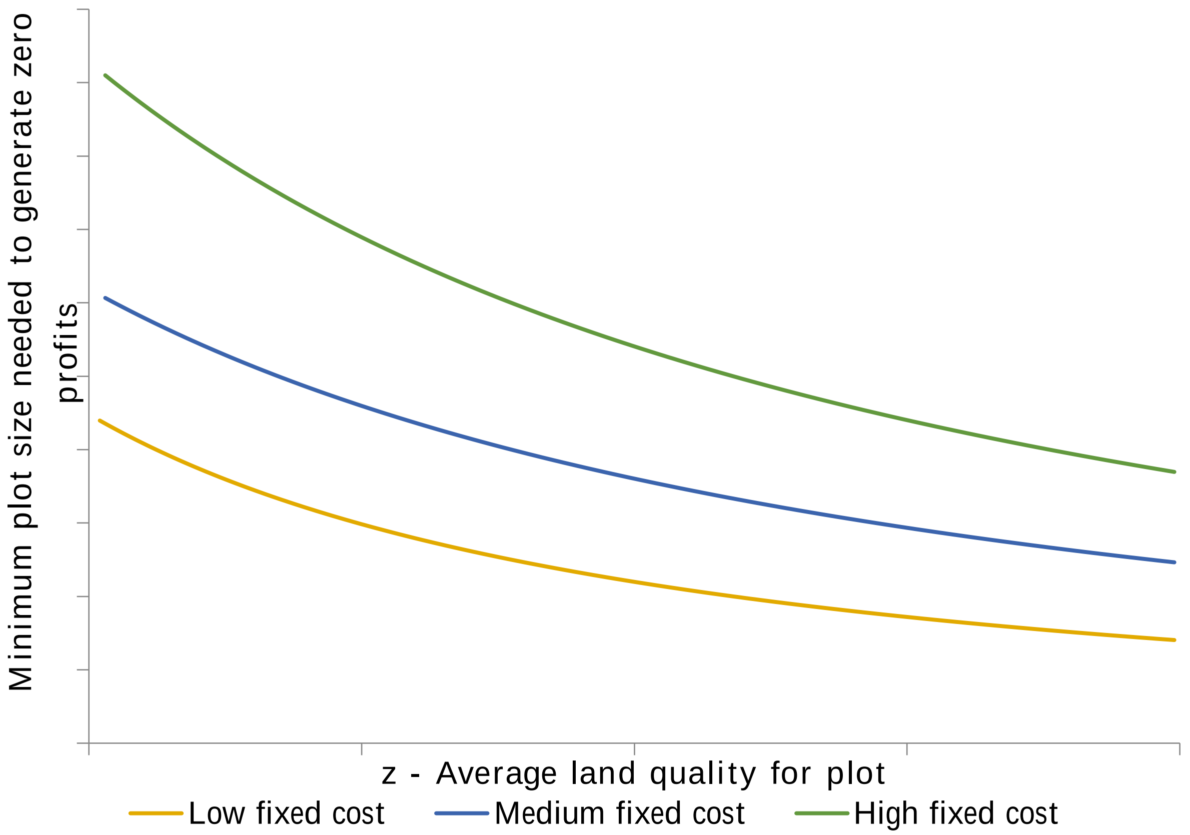

Table 1a: Agriculture in Tanzania and Uganda - Samples

	Tanzania	Uganda
Sample Size		
Households	5,791	2,592
Farmers	7,089	4,989
Plot-Seasons	16,998	39,290
Seasons	3	6
Regions	26	6
Districts	140	81
Villages	184	622
Number of Clusters		
Farmer - Seasons	7,732	13,810
Farmer - Crop - Seasons	13,589	34,366
Size of Clusters (Median)		
Farmer - Seasons	2	4
Farmer - Crop - Seasons	1	1

Table 1b: Agriculture in Tanzania and Uganda - Yields

	Tanzania	Uganda
Median Plot Size (ha)	0.40	0.20
Yield ($/ha)		
N	14,777	39,065
mean	844	3,992
median	442	181
std deviation	3,256	399,964
Yield on Maize Plots ($/ha)		
N	5,795	7,768
mean	814	579
median	464	206
std deviation	3,985	3,715
Yield on Groundnut/Beans Plots ($/ha)		
N	784	8,051
mean	723	9,103
median	465	221
std deviation	867	758,833
Yield on Cassava (Tz) or Banana (UG) Plots ($/ha)		
N	1,386	4,487
mean	606	570
median	347	287
std deviation	897	1,490
Labour (Days/ha)		
N	14,582	39,065
mean	182	299
median	106	165
std deviation	693	1,400

Note: All yields winsorized at the 0.01 level

Table 2a: Log Output per Hectare in Tanzania

	(1)	(2)	(3)	(4)	(5)
Fixed effects:	None	Year-Season-Crop-Region	Year-Season-Crop-Village	Year-Season-Crop-Household	Year-Season-Crop-Farmer
ln(ha)	-0.58	-0.59	-0.57	-0.61	-0.61
	(0.0092)	(0.011)	(0.013)	(0.031)	(0.032)
Female Plot	-0.070	-0.090	-0.097	-0.15	
	(0.028)	(0.030)	(0.035)	(0.23)	
Plot used free of charge	-0.17	-0.090	-0.12	-0.056	-0.031
	(0.030)	(0.033)	(0.037)	(0.098)	(0.10)
Shared - rent	-1.13	-1.30	-0.94	0	0
	(0.37)	(0.36)	(0.40)	(.)	(.)
Shared - owned	-0.23	-0.16	-0.20	0.71	0.68
	(0.11)	(0.12)	(0.14)	(0.31)	(0.32)
Average quality	-0.17	-0.13	-0.14	-0.060	-0.086
	(0.020)	(0.021)	(0.024)	(0.071)	(0.073)
Poor quality	-0.27	-0.26	-0.28	-0.28	-0.30
	(0.041)	(0.043)	(0.052)	(0.12)	(0.12)
Loam	0.17	0.069	0.11	-0.063	-0.097
	(0.025)	(0.028)	(0.032)	(0.081)	(0.083)
Clay	0.18	0.071	0.12	-0.070	-0.099
	(0.032)	(0.035)	(0.040)	(0.097)	(0.10)
Distance to market	0.0023	0.00100	0.00047	0.0044	0.00043
	(0.00070)	(0.00084)	(0.00087)	(0.0079)	(0.0082)
Irrigated	0.48	0.10	0.24	0.050	-0.18
	(0.068)	(0.093)	(0.099)	(0.26)	(0.28)
Erosion evident	-0.019	-0.046	-0.057	-0.071	-0.072
	(0.029)	(0.031)	(0.035)	(0.072)	(0.074)
Sale value	-0.033	-0.14	-0.21	-0.14	0.73
	(0.078)	(0.089)	(0.080)	(1.75)	(1.91)
Current morbidity (z)	-0.053	-0.046	-0.045	0.23	0
	(0.010)	(0.011)	(0.013)	(0.12)	(.)
Morbidity missing	5.21	4.54	4.47	-22.4	0
	(1.01)	(1.09)	(1.26)	(11.9)	(.)
Log variance of residuals	1.23	0.81	0.79	0.24	0.24
F-stat for plot characteristics	177.8	104.7	84.1	17.1	17.1
Corresponding P-value	0	0	0	0	0

Notes: Standard errors in parentheses. Variance of dependent variable is 1.09. Regressions include: 28 plot characteristics - e.g. dummies for soil type, toposequence, presence of boundary markers, location, etc.; most also interacted with weather shocks. Regressions also include 14 household and farmer characteristics; e.g., housing, education, age, etc.

Table 2b: Log Output per Hectare in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)
Fixed effects:	None	Year-Season-Crop-Region	Year-Season-Crop-District	Year-Season-Crop-Village	Year-Season-Crop-Household	Year-Season-Crop-Farmer
ln(ha)	-0.58	-0.66	-0.71	-0.72	-0.75	-0.75
	(0.0071)	(0.0071)	(0.0082)	(0.0087)	(0.021)	(0.022)
Male Plot	0.17	0.15	0.10	0.11	0.21	0
	(0.023)	(0.024)	(0.025)	(0.026)	(0.16)	(.)
Leasehold Plot	-0.020	-0.050	0.12	0.061	-0.13	-0.30
	(0.038)	(0.036)	(0.042)	(0.045)	(0.12)	(0.14)
Customary Plot	-0.017	0.029	0.051	-0.016	-0.18	-0.22
	(0.057)	(0.054)	(0.060)	(0.063)	(0.20)	(0.23)
Mailo Plot	-0.099	-0.052	0.0086	0.012	0.0035	0.015
	(0.034)	(0.032)	(0.040)	(0.047)	(0.16)	(0.20)
Plot via Occupancy	-0.023	-0.0036	0.011	-0.0055	-0.033	-0.063
	(0.035)	(0.033)	(0.039)	(0.042)	(0.13)	(0.14)
Customary	-0.23	-0.20	0.0093	0.031	-0.0097	0.042
	(0.021)	(0.020)	(0.032)	(0.034)	(0.10)	(0.12)
No document	0.0083	0.032	-0.040	-0.057	-0.0070	-0.045
	(0.027)	(0.026)	(0.031)	(0.032)	(0.11)	(0.12)
Fair Soil	-0.12	-0.10	-0.075	-0.058	-0.16	-0.13
	(0.036)	(0.034)	(0.038)	(0.040)	(0.12)	(0.13)
Poor Soil	-0.014	-0.013	0.0069	0.020	-0.035	-0.081
	(0.090)	(0.085)	(0.091)	(0.096)	(0.25)	(0.26)
Sandy clay loam	-0.00032	-0.010	-0.0010	0.0069	-0.016	-0.016
	(0.016)	(0.016)	(0.017)	(0.018)	(0.052)	(0.056)
Black clay	-0.037	-0.074	0.012	0.023	0.12	0.14
	(0.020)	(0.019)	(0.022)	(0.023)	(0.067)	(0.073)
Flood*Avg Soil	0.100	0.19	0.25	0.10	2.09	2.46
	(0.14)	(0.13)	(0.14)	(0.15)	(1.11)	(1.48)
Flood*Poor Soil	-0.30	-0.30	0.0030	-0.10	0.55	0.50
	(0.44)	(0.41)	(0.46)	(0.46)	(1.03)	(1.04)
Drought*Avg Soil	-0.00056	0.00039	0.00067	0.0021	0.0043	-0.014
	(0.0071)	(0.0066)	(0.0071)	(0.0076)	(0.020)	(0.021)
Drought*Poor Soil	-0.010	-0.0090	-0.0085	0.0052	-0.030	-0.027
	(0.014)	(0.013)	(0.014)	(0.015)	(0.041)	(0.044)
Log variance of residuals	1.52	1.28	0.71	0.56	0.27	0.27
F-stat for plot characteristics	141.4	169.6	142.8	133.4	28.2	28.0
Corresponding P-value	0	0	0	0	0	0

Notes: Standard errors in parentheses. Variance of dependent variable is 1.98. Regressions include: 54 plot characteristics - e.g. dummies for soil type, toposequence, presence of bound+A25ary markers, location, etc.; most also interacted with weather shocks. Regressions also include 57 household and farmer characteristics; e.g., education, age, weather shocks, dummies for morbidity and housing, etc.

Table 3a: Log Labor per Hectare in Tanzania

	(1)	(2)	(3)	(4)	(5)
Fixed effects:	None	Year-Season-Crop-Region	Year-Season-Crop-Village	Year-Season-Crop-Household	Year-Season-Crop-Farmer
ln(ha)	-0.63	-0.62	-0.63	-0.62	-0.61
	(0.0067)	(0.0082)	(0.0090)	(0.020)	(0.020)
Female Plot	-0.13	-0.11	-0.11	0.050	0
	(0.020)	(0.021)	(0.025)	(0.13)	(.)
Plot used free of charge	0.13	0.0047	0.074	0.044	0.023
	(0.021)	(0.024)	(0.026)	(0.059)	(0.060)
Shared - rent	0.29	0.35	0.24	0	0
	(0.24)	(0.23)	(0.26)	(.)	(.)
Shared - owned	-0.077	-0.20	-0.13	0.36	0.31
	(0.084)	(0.092)	(0.11)	(0.21)	(0.22)
Average quality	-0.000048	0.0072	0.011	0.036	0.032
	(0.014)	(0.015)	(0.017)	(0.044)	(0.044)
Poor quality	-0.044	-0.024	-0.050	0.040	0.063
	(0.028)	(0.031)	(0.036)	(0.077)	(0.080)
Loam	0.0083	0.036	0.050	0.067	0.062
	(0.018)	(0.020)	(0.022)	(0.051)	(0.052)
Clay	0.033	0.051	0.068	-0.017	-0.0059
	(0.022)	(0.025)	(0.028)	(0.062)	(0.063)
Distance to market	-0.00086	-0.00044	-0.00044	0.023	0.024
	(0.00049)	(0.00059)	(0.00061)	(0.0049)	(0.0050)
Irrigated	-0.15	-0.063	-0.22	-0.029	-0.099
	(0.046)	(0.064)	(0.068)	(0.16)	(0.17)
Erosion evident	-0.0084	-0.060	-0.046	-0.21	-0.20
	(0.020)	(0.022)	(0.025)	(0.046)	(0.047)
Sale value	-0.27	-0.34	-0.28	-0.24	-0.25
	(0.056)	(0.061)	(0.058)	(1.15)	(1.24)
Current morbidity (z)	-0.013	-0.013	-0.016	0.14	0
	(0.0071)	(0.0079)	(0.0090)	(0.085)	(.)
Morbidity missing	1.30	1.32	1.65	-13.8	0
	(0.71)	(0.78)	(0.89)	(8.39)	(.)
Log variance of residuals	0.62	0.44	0.41	0.11	0.11
F-stat for plot characteristics	379.4	222.3	191.9	42.6	42.7
Corresponding P-value	0	0	0	0	0

Notes: Standard errors in parentheses. Variance of dependent variable is 1.09. Regressions include: 28 plot characteristics - e.g. dummies for soil type, toposequence, presence of boundary markers, location, etc.; most also interacted with weather shocks. Regressions also include 14 household and farmer characteristics; e.g., housing, education, age, etc.

Table 3b: Log Labour per Hectare in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)
Fixed effects:	None	Year-Season-Crop-Region	Year-Season-Crop-District	Year-Season-Crop-Village	Year-Season-Crop-Household	Year-Season-Crop-Farmer
ln(ha)	-0.72	-0.73	-0.76	-0.77	-0.78	-0.78
	(0.0041)	(0.0042)	(0.0053)	(0.0058)	(0.012)	(0.013)
Male Plot	-0.016	0.0073	-0.0035	-0.017	-0.062	0
	(0.012)	(0.013)	(0.015)	(0.016)	(0.090)	(.)
Leasehold Plot	-0.026	-0.046	-0.018	-0.048	0.061	0.16
	(0.021)	(0.021)	(0.026)	(0.029)	(0.073)	(0.085)
Customary Plot	-0.10	-0.11	-0.11	-0.100	-0.17	-0.25
	(0.031)	(0.031)	(0.037)	(0.040)	(0.12)	(0.13)
Mailo Plot	-0.12	-0.11	-0.14	-0.089	0.19	0.13
	(0.019)	(0.018)	(0.025)	(0.030)	(0.095)	(0.11)
Plot via Occupancy	-0.24	-0.22	-0.24	-0.22	-0.064	-0.059
	(0.019)	(0.019)	(0.025)	(0.027)	(0.076)	(0.084)
Customary	-0.048	-0.062	-0.065	-0.070	0.14	0.21
	(0.012)	(0.011)	(0.020)	(0.022)	(0.061)	(0.069)
No document	-0.14	-0.16	-0.11	-0.083	-0.076	-0.068
	(0.016)	(0.015)	(0.020)	(0.022)	(0.066)	(0.070)
Fair Soil	-0.031	-0.043	-0.059	-0.029	-0.077	-0.029
	(0.020)	(0.019)	(0.023)	(0.025)	(0.063)	(0.068)
Poor Soil	-0.092	-0.11	-0.15	-0.11	-0.34	-0.28
	(0.051)	(0.050)	(0.058)	(0.064)	(0.15)	(0.15)
Sandy clay loam	-0.020	-0.022	-0.015	-0.017	0.013	0.031
	(0.0090)	(0.0087)	(0.010)	(0.011)	(0.029)	(0.031)
Black clay	-0.041	-0.032	-0.023	-0.015	-0.0033	0.033
	(0.011)	(0.011)	(0.013)	(0.015)	(0.037)	(0.039)
Flood*Avg Soil	0.013	0.051	0.020	0.085	0.19	-0.58
	(0.075)	(0.073)	(0.086)	(0.091)	(0.35)	(0.47)
Flood*Poor Soil	0.0079	-0.013	0.10	0.075	0.28	0.27
	(0.24)	(0.23)	(0.28)	(0.29)	(0.62)	(0.61)
Drought*Avg Soil	0.0072	0.0057	0.0043	0.0037	0.016	0.0095
	(0.0039)	(0.0038)	(0.0045)	(0.0049)	(0.012)	(0.013)
Drought*Poor Soil	0.017	0.015	0.015	-0.00041	0.025	0.025
	(0.0078)	(0.0076)	(0.0092)	(0.010)	(0.025)	(0.026)
Log variance of residuals	0.47	0.43	0.29	0.25	0.10	0.090
F-stat for plot characteristics	594.4	576.4	399.6	337.3	81.3	81.2
Corresponding P-value	0	0	0	0	0	0

Notes: Standard errors in parentheses. Variance of dependent variable is 0.92. Regressions include: 54 plot characteristics - e.g. dummies for soil type, toposequence, presence of boundary markers, location, etc.; most also interacted with weather shocks. Regressions also include 57 household and farmer characteristics; e.g., education, age, weather shocks, dummies for morbidity and housing, etc.

Table 4a: OLS and Quantile Regression Determinants of Land and Labor Inputs in Tanzania

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
Male Plot	0.36	0.21	0.31	0.25	0.32	0.19	0.37	0.21	0.059	-0.037
	(0.022)	(0.019)	(0.027)	(0.030)	(0.024)	(0.023)	(0.026)	(0.020)	(0.029)	(0.031)
EVI*Good Soil in HH*	-0.10	-0.025	-0.060	-0.029	-0.11	-0.021	-0.13	-0.027	-0.068	0.0025
	(0.030)	(0.027)	(0.039)	(0.043)	(0.034)	(0.031)	(0.035)	(0.029)	(0.046)	(0.038)
EVI*Avg Soil in HH*	-0.056	0.027	0.011	0.049	-0.073	0.042	-0.11	0.012	-0.12	-0.037
	(0.029)	(0.026)	(0.036)	(0.042)	(0.031)	(0.030)	(0.037)	(0.027)	(0.035)	(0.043)
EVI*Poor Soil in HH*	-0.048	0.018	-0.0033	0.021	-0.054	0.056	-0.049	0.050	-0.046	0.029
	(0.048)	(0.043)	(0.037)	(0.074)	(0.064)	(0.056)	(0.040)	(0.042)	(0.052)	(0.067)
EVI*Loam in HH*	-0.057	-0.11	-0.096	-0.11	-0.056	-0.12	-0.087	-0.11	0.0085	0.0020
	(0.028)	(0.025)	(0.034)	(0.040)	(0.031)	(0.029)	(0.033)	(0.027)	(0.044)	(0.034)
EVI*Clay in HH*	-0.048	-0.061	-0.099	-0.049	-0.050	-0.046	-0.067	-0.081	0.032	-0.031
	(0.036)	(0.033)	(0.044)	(0.052)	(0.041)	(0.036)	(0.042)	(0.033)	(0.047)	(0.048)
EVI*Other in HH*	-0.13	-0.16	-0.15	-0.10	-0.16	-0.15	-0.093	-0.15	0.055	-0.049
	(0.062)	(0.056)	(0.081)	(0.077)	(0.053)	(0.035)	(0.068)	(0.083)	(0.092)	(0.083)
Livestock death or stolen*	0.098	0.014	0.082	-0.00037	0.087	0.017	0.12	0.048	0.037	0.049
	(0.022)	(0.020)	(0.030)	(0.026)	(0.027)	(0.024)	(0.029)	(0.023)	(0.029)	(0.035)
Illness/accident of hh member*	0.070	-0.023	0.049	-0.044	0.11	-0.035	0.081	0.0024	0.031	0.046
	(0.035)	(0.031)	(0.059)	(0.046)	(0.040)	(0.042)	(0.046)	(0.025)	(0.053)	(0.053)
Death of hh member*	0.15	0.071	0.12	0.085	0.14	0.099	0.13	0.073	0.012	-0.012
	(0.028)	(0.025)	(0.033)	(0.031)	(0.030)	(0.030)	(0.035)	(0.025)	(0.049)	(0.040)
Victim of property crime or attack*	0.036	-0.073	0.063	-0.13	0.017	-0.077	0.052	-0.047	-0.012	0.079
	(0.035)	(0.032)	(0.039)	(0.039)	(0.041)	(0.037)	(0.051)	(0.035)	(0.064)	(0.048)
Adverse shock to other household plots*	-0.075	-0.021	-0.078	-0.024	-0.065	-0.020	-0.069	-0.022	0.0086	0.0026
	(0.012)	(0.011)	(0.012)	(0.015)	(0.013)	(0.012)	(0.014)	(0.011)	(0.017)	(0.016)
F statistic; see note #.	26.8	12.2	15.6	7.85	25.8	10.8	26.8	15.2		
Corresponding p value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
F statistic; see note §.									9.28	17.4
Corresponding p value									0.00	0.00

Notes:
Standard errors in parentheses.
*Variables used as instruments in Table 5.
#F statistic for joint significance of variables used as production function instruments in Table 5.
§F stat for h0 – coefficients of instruments equal for 25th and 75th percentile.
All regressions also include 28 plot and household characteristics.

Table 4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
Male Plot	0.14	0.1	0.13	0.092	0.14	0.11	0.14	0.11	0.014	0.021
	(0.010)	(0.0080)	(0.012)	(0.010)	(0.011)	(0.0082)	(0.011)	(0.0083)	(0.013)	(0.015)
EVI*Avg Soil in household*	-0.0002	-0.0004	-0.0003	-0.0004	-0.0001	-0.0004	-0.0005	-0.0005	-0.00022	-0.000074
	(0.00018)	(0.00016)	(0.00023)	(0.00019)	(0.00019)	(0.00014)	(0.00022)	(0.00015)	(0.00033)	(0.00031)
EVI*Poor Soil in household*	0.00041	0.00039	0.00062	-0.0005	0.00049	-0.0005	-9E-05	0.0003	-0.00071	0.00075
	(0.00059)	(0.00052)	(0.00068)	(0.00048)	(0.00075)	(0.00046)	(0.00072)	(0.00047)	(0.00090)	(0.00082)
EVI*Missing Soil in household*	-0.0066	-0.0004	-0.0051	-0.0005	-0.0043	0.00092	-0.0049	-0.0017	0.00024	-0.0012
	(0.0019)	(0.0015)	(0.0024)	(0.0016)	(0.0016)	(0.0012)	(0.0017)	(0.0012)	(0.0033)	(0.0021)
Drought*Avg Soil in household*	-0.0063	-0.0071	-0.01	-0.0075	-0.0057	-0.0068	-0.0051	-0.0075	0.0052	-0.000077
	(0.0020)	(0.0016)	(0.0025)	(0.0018)	(0.0016)	(0.0015)	(0.0018)	(0.0014)	(0.0030)	(0.0017)
Drought*Poor Soil in household*	-0.0029	0.00075	-0.0045	0.000049	-0.0063	0.0032	-0.0021	0.0053	0.0024	0.0053
	(0.0044)	(0.0037)	(0.0050)	(0.0025)	(0.0047)	(0.0037)	(0.0049)	(0.0035)	(0.0060)	(0.0067)
Drought*Missing Soil in household*	-0.031	0.023	-0.041	0.018	-0.024	0.017	-0.0017	0.029	0.039	0.011
	(0.0089)	(0.0059)	(0.013)	(0.0070)	(0.0084)	(0.0072)	(0.0075)	(0.0044)	(0.016)	(0.011)
Illness Incidence in household*	-0.0072	-0.073	-0.0059	-0.084	0.0014	-0.088	-0.013	-0.096	-0.0071	-0.011
	(0.014)	(0.011)	(0.016)	(0.015)	(0.013)	(0.012)	(0.015)	(0.012)	(0.017)	(0.015)
Household shock to non-agricultural income*	0.088	-0.0023	0.080	-0.0011	0.072	-0.0041	0.090	-0.0097	0.0100	-0.0086
	(0.029)	(0.023)	(0.023)	(0.031)	(0.033)	(0.022)	(0.026)	(0.021)	(0.038)	(0.034)
Number of household members*	0.026	0.017	0.024	0.020	0.023	0.016	0.026	0.017	0.0015	-0.0030
	(0.0024)	(0.0019)	(0.0030)	(0.0027)	(0.0024)	(0.0019)	(0.0026)	(0.0020)	(0.0033)	(0.0031)
Number of adults in household*	0.011	0.0082	0.011	0.0041	0.016	0.0075	0.0093	0.0078	-0.0019	0.0037
	(0.0033)	(0.0027)	(0.0042)	(0.0037)	(0.0034)	(0.0027)	(0.0037)	(0.0029)	(0.0044)	(0.0047)
Shocks on other plots*	0.021	-0.0005	0.018	-0.0035	0.048	-0.0006	0.081	-3E-05	0.063	0.0035
	(0.0029)	(0.0013)	(0.0024)	(0.0032)	(0.0099)	(0.00046)	(0.0072)	(0.0012)	(0.016)	(0.0027)
F statistic; see note #.	23	16.7	55.7	27.2	28.5	67.5	30.4	59.4		
Corresponding p value	0	0	0	0	0	0	0	0		
F statistic; see note §.									112	7.15
Corresponding p value									0	1.20E-19

Notes:
Standard errors in parentheses.
*Variables used as instruments in Table 5.
#F statistic for joint significance of variables used as production function instruments in Table 5.
§F stat for h0 – coefficients of instruments equal for 25th and 75th percentile.
All regressions also include 34 plot and household characteristics.
Not shown are 28 additional variables used as instruments in 5b. These are: 6 community average soil characteristics interacted with EVI and household reported drought, 22 household and community average soil characteristics interacted with household reported flood, and community early season rain, household asset value, dummies for electricity, roofing, literacy, schooling, generations resident in community, community leadership position.

Table 5: Production Function and Variance Components

	Tanzania		Uganda	
	2SLS	IVCRC	2SLS	IVCRC
A. Cobb-Douglas Factor Coefficients				
Land	0.76	0.57	0.70	0.53
	(0.15)	(0.01)	(0.05)	(0.00)
Labor	0.21	0.37	0.14	0.38
	(0.23)	(0.03)	(0.10)	(0.00)
B. Implied Plot Level Variances of Productivity, Risk and Measurement Errors				
Plot TFP	0.36	0.61	0.26	0.26
	(0.01)	(0.05)	(0.01)	(0.01)
Land Productivity (= Labor Productivity)	0.04	0.29	0.10	0.10
	(0.01)	(0.05)	(0.01)	(0.01)
Late Season Risk and Output Measurement Error	0.68	0.93	1.20	1.20
	(0.02)	(0.17)	(0.02)	(0.01)
Land Measurement Error	0.28	-0.21	0.25	0.25
	(0.03)	(0.32)	(0.01)	(0.01)
Labor Measurement Error	0.37	0.37	0.21	0.21
	(0.02)	(0.02)	(0.01)	(0.01)
Covariance of TFP and Land Prod. (= -Covariance with Labour Prod.)	0.08	0.34	0.06	0.06
	(0.01)	(0.06)	(0.00)	(0.00)

Notes: Full production function results presented in Appendix tables A5a and A5b. Bootstrapped standard errors (50 bootstrap iterations) in parentheses.

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, $[0, \mathrm{L}_h]$, making use of the entire land endowment. The profit maximization problem is then given by:

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

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 $[0, \mathrm{L}_{h1}]$ and $[\mathrm{L}_{h1}, \mathrm{L}_h]$. 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:

$$\max_{\mathrm{L}_{h1}} \int_{s\in S} \left[\max_{X_{h1},X_{h2}} \left[\left(\frac{X_{h1}}{L_{h1}}\right)^{\theta} \int_{0}^{\mathrm{L}_{h1}} \gamma_h(k,s)\zeta_h(k)\,dk + \left(\frac{X_{h1}}{L_h - L_{h1}}\right)^{\theta} \int_{\mathrm{L}_{h1}}^{\mathrm{L}_h} \gamma_h(k,s)\zeta_h(k)\,dk - w_h X_{h1} - w_h X_{h2} - 2c \right]\right] d\Delta(s). \tag{A1.2}$$

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 $\mathrm{L}_{h0} = 0$ and $\mathrm{L}_{hI} = \mathrm{L}_h$. Then $\{\mathrm{L}_{hi}\}_{i=0}^{I}$ is the sequence of plot boundaries. The first plot is given by the interval $[0, \mathrm{L}_{h1}]$, and the ith plot covers the interval $[\mathrm{L}_{hi-1}, \mathrm{L}_{hi}]$, continuing to the Ith plot, which covers $[\mathrm{L}_{hI-1}, \mathrm{L}_h]$.

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:

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

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 $\hat{I} = \mathrm{argmax}_j \{E\hat{\pi}(j)\}_{j=1}^{I^*}$.

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 $z \geq 1$,

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

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

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

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 define the plot-specific factor productivities $\omega_{Lhit} = \ln(\alpha_{Lhit}) - \hat{\alpha}_L$ and $\omega_{Xhit} = \ln(\alpha_{Xhit}) - \hat{\alpha}_X$. We will work in terms of observable inputs, and output, adjusted for the estimated effects of observed characteristics

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

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

$$\begin{aligned} \tilde{y}_{hit} &\equiv y_{hit} - \bar{y}_{h.t} - (W_{Hhit} - \overline{W}_{Hh.t})\hat{\beta}_H \\ &= \epsilon_{Yhit} - \bar{\epsilon}_{Yh.t} + z_{hit} - \bar{z}_{h.t} \end{aligned} \tag{A2.4}$$

$$
\begin{aligned}
\tilde{l}_{hit} &\equiv l^{o}_{hit} - \bar{l}^{o}_{h.t} + (W_{Lhit} - \bar{W}_{Lh.t})\hat{\beta}_L \\
&= \omega_{Lhit} - \bar{\omega}_{Lh.t} + \epsilon_{Lhit} - \bar{\epsilon}_{Lh.t} + q_{hit} - \bar{q}_{h.t} \\
\tilde{x}_{hit} &\equiv x^{o}_{hit} - \bar{x}^{o}_{h.t} + (W_{Xhit} - \bar{W}_{Xh.t})\hat{\beta}_X \\
&= \omega_{Xhit} - \bar{\omega}_{Xh.t} + \epsilon_{Xhit} - \bar{\epsilon}_{Xh.t} + q_{hit} - \bar{q}_{h.t}
\end{aligned}
$$

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_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 $\left(\sigma_{\epsilon y}^2\right)$, as well as the covariance of plot-level total factor productivity and factor-specific productivity $\left(\sigma_{QL}, \sigma_{QX}\right)$:

$$
\begin{aligned}
var(\tilde{y}_{hit}) &= \sigma_Q^2 + \sigma_{\epsilon Yhit}^2 \\
var\left(\tilde{l}_{hit}\right) &= \sigma_L^2 + \sigma_{\epsilon L}^2 + \sigma_Q^2 + 2\sigma_{QL} \\
var(\tilde{x}_{hit}) &= \sigma_X^2 + \sigma_{\epsilon X}^2 + \sigma_Q^2 + 2\sigma_{QX} \\
cov\left(\tilde{y}_{hit}, \tilde{l}_{hit}\right) &= \sigma_{QL} + \sigma_Q^2 \\
cov(\tilde{y}_{hit}, \tilde{x}_{hit}) &= \sigma_{QX} + \sigma_Q^2 \\
cov\left(\tilde{l}_{hit}, \tilde{x}_{hit}\right) &= \sigma_{LX} + \sigma_{QL} + \sigma_{QX} + \sigma_Q^2
\end{aligned}
\qquad \text{(A2.5)}
$$

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

$$
\begin{aligned}
\sigma_L^2 &= \sigma_X^2 \\
\sigma_{LX} &= -\sigma_L^2 \\
\sigma_{QL} &= -\sigma_{QX}
\end{aligned}
\qquad \text{(A2.6)}
$$

From equations (A2.5)-(A2.6) we calculate the parameters ($\hat{\sigma}_q^2, \hat{\sigma}_L^2, \hat{\sigma}_X^2, \hat{\sigma}_{\epsilon y}^2, \hat{\sigma}_{\epsilon L}^2, \hat{\sigma}_{\epsilon X}^2, \hat{\sigma}_{LX}, \hat{\sigma}_{QL}, \hat{\sigma}_{QX}$) 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 the overall variance, across all plots, of these measurement errors/random shocks?

Denote by y_{fi} the realization of any of these errors/shocks $(\epsilon_{Yhit}, \epsilon_{Lhit}, \epsilon_{Xhit})$.[31] 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 $\sigma_F^2 \equiv$ $\frac{1}{N^f}\sum_{f=1}^{Nf}\frac{1}{N_f^i}\sum_{i=1}^{N_f^i}(y_{fi}-\bar{y}_f)^2$. The variance of y across plots in the sample is $\sigma^2 \equiv \frac{1}{N}\sum_{f=1}^{Nf}\sum_{i=1}^{N_f^i}(y_{fi}-\bar{y})^2$. If there are no farmer effects in measurement error or the late season shock to output, then $\bar{y}_f = \bar{y}\ \forall f$ 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 $\bar{k}$. We partition the sample of farmers into sets $\{M_1, M_2, \dots, M_{\bar{k}}\}$ 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

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

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

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

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

$$\begin{aligned}
\sigma^2 &= \frac{1}{N} \sum_{k=1}^{\bar{k}} \sum_{f \in M_k} \sum_{i=1}^{k} y_{fi}^2 - \left(\frac{1}{N} \sum_{k=1}^{\bar{k}} k M_k \overline{y_k} \right)^2 \\
&\geq \frac{1}{N} \sum_{k=1}^{\bar{k}} \sum_{f \in M_k} \sum_{i=1}^{k} y_{fi}^2 - \frac{1}{N} \sum_{k=1}^{\bar{k}} M_k k (\overline{y_k})^2 \\
&= \frac{1}{N} \sum_{k=1}^{\bar{k}} [M_k k \sigma_k^2].
\end{aligned} \tag{A3.1}$$

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

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

So

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

If each farmer has the same number of plots, then the weak inequality in (2) is an equality, and $\sum_{k=1}^{\bar{k}} \left(\frac{k}{N} - \frac{1}{N_F} \right) M_k \sigma_{Fk}^2 = 0$ 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 $\sum_{k=1}^{\bar{k}}\left(\frac{k}{N} - \frac{1}{N_F}\right)M_k\sigma_{Fk}^2 > 0$. 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.

Appendix Table A4a: OLS and Quantile Regression Determinants of Land and Labor Inputs in Tanzania

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
Male Plot	0.36	0.21	0.31	0.25	0.32	0.19	0.37	0.21	0.059	-0.037
	(.022)	(.019)	(.027)	(.03)	(.024)	(.023)	(.026)	(.02)	(.03)	(.021)
EVI*Good Soil in HH	-0.1	-0.025	-0.06	-0.029	-0.11	-0.021	-0.13	-0.027	-0.068	0.0025
	(.03)	(.027)	(.039)	(.043)	(.034)	(.031)	(.035)	(.029)	(.039)	(.039)
EVI*Avg Soil in HH	-0.056	0.027	0.011	0.049	-0.073	0.042	-0.11	0.012	-0.12	-0.037
	(.029)	(.026)	(.036)	(.042)	(.031)	(.03)	(.037)	(.027)	(.05)	(.054)
EVI*Poor Soil in HH	-0.048	0.018	-0.0033	0.021	-0.054	0.056	-0.049	0.05	-0.046	0.029
	(.048)	(.043)	(.037)	(.074)	(.064)	(.056)	(.04)	(.042)	(.071)	(.063)
EVI*Loam in HH	-0.057	-0.11	-0.096	-0.11	-0.056	-0.12	-0.087	-0.11	0.0085	0.002
	(.028)	(.025)	(.034)	(.04)	(.031)	(.029)	(.033)	(.027)	(.037)	(.039)
EVI*Clay in HH	-0.048	-0.061	-0.099	-0.049	-0.05	-0.046	-0.067	-0.081	0.032	-0.031
	(.036)	(.033)	(.044)	(.052)	(.041)	(.036)	(.042)	(.033)	(.059)	(.042)
EVI*Other in HH	-0.13	-0.16	-0.15	-0.1	-0.16	-0.15	-0.093	-0.15	0.055	-0.049
	(.062)	(.056)	(.081)	(.077)	(.053)	(.035)	(.068)	(.083)	(.099)	(.11)
Livestock death or stolen	0.098	0.014	0.082	-0.00037	0.087	0.017	0.12	0.048	0.037	0.049
	(.022)	(.02)	(.03)	(.026)	(.027)	(.024)	(.029)	(.023)	(.04)	(.029)
illness/accident of hh member	0.07	-0.023	0.049	-0.044	0.11	-0.035	0.081	0.0024	0.031	0.046
	(.035)	(.031)	(.059)	(.046)	(.04)	(.042)	(.046)	(.025)	(.057)	(.049)
death of hh member	0.15	0.071	0.12	0.085	0.14	0.099	0.13	0.073	0.012	-0.012
	(.028)	(.025)	(.033)	(.031)	(.03)	(.03)	(.035)	(.025)	(.044)	(.046)

Appendix Table A4a: OLS and Quantile Regression Determinants of Land and Labor Inputs in Tanzania

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
Property Crime	0.036	-0.073	0.063	-0.13	0.017	-0.077	0.052	-0.047	-0.012	0.079
	(.035)	(.032)	(.039)	(.039)	(.041)	(.037)	(.051)	(.035)	(.06)	(.062)
Zcrops_lost	-0.075	-0.021	-0.078	-0.024	-0.065	-0.02	-0.069	-0.022	0.0086	0.0026
	(.012)	(.011)	(.012)	(.015)	(.013)	(.012)	(.014)	(.011)	(.021)	(.02)
log Plot Value	2.30E-01	9.80E-02	2.10E-01	1.20E-01	2.40E-01	1.10E-01	2.70E-01	9.50E-02	5.80E-02	-2.40E-02
	(.0078)	(.007)	(.01)	(.0097)	(.0083)	(.0082)	(.0094)	(.0073)	(.0097)	(.012)
lnsale_valuem	-0.055	-0.004	-0.016	-0.042	-0.02	-0.028	-0.11	-0.02	-0.094	0.022
	(.067)	(.06)	(.068)	(.073)	(.1)	(.051)	(.057)	(.034)	(.095)	(.098)
(first) dist_home	0.0022	0.000051	0.0024	0.00073	0.0024	-0.00025	0.0027	-0.0009	0.00034	-0.0016
	(.0004)	(.00036)	(.00024)	(.0002)	(.00076)	(.00014)	(.00042)	(.00016)	(.00057)	(.0003)
(first) dist_road	0.024	0.018	0.028	0.02	0.028	0.022	0.028	0.019	0.000002	-0.00049
	(.0018)	(.0016)	(.0029)	(.0016)	(.0017)	(.0014)	(.0022)	(.00093)	(.0042)	(.0028)
soil_quality==1	-0.17	-0.42	-0.13	-0.58	-0.2	-0.22	-0.2	-0.51	-0.073	0.069
	(.28)	(.25)	(.24)	(.34)	(.37)	(.3)	(.33)	(.31)	(.5)	(.44)
soil_quality==2	0.43	-0.096	0.9	-0.088	0.53	0.13	0.22	-0.33	-0.68	-0.24
	(.28)	(.25)	(.24)	(.34)	(.37)	(.3)	(.34)	(.31)	(.47)	(.52)
soil_type==1	-2.22	-1.36	-1.79	-1.34	-2.41	-2.38	-2.05	-0.91	-0.26	0.43
	(.59)	(.53)	(.57)	(.74)	(.37)	(.37)	(.6)	(.45)	(1.41)	(.73)
soil_type==2	-1.98	-1.2	-1.89	-1.41	-2.21	-1.93	-1.73	-0.62	0.15	0.8
	(.58)	(.52)	(.55)	(.73)	(.34)	(.34)	(.58)	(.43)	(1.42)	(.7)

Appendix Table A4a: OLS and Quantile Regression Determinants of Land and Labor Inputs in Tanzania

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
soil_type==3	-1.59	-1.23	-1.59	-1.5	-1.8	-1.99	-1.31	-0.63	0.28	0.88
	(.6)	(.53)	(.58)	(.74)	(.38)	(.37)	(.59)	(.44)	(1.46)	(.64)
(first) single_manager	-0.045	-0.12	-0.059	-0.13	-0.045	-0.11	-0.051	-0.12	0.0084	0.016
	(.021)	(.019)	(.027)	(.025)	(.023)	(.021)	(.025)	(.02)	(.041)	(.028)
health	-0.015	-0.0074	-0.027	-0.00099	-0.018	0.0028	-0.015	-0.0063	0.012	-0.0053
	(.0056)	(.005)	(.007)	(.0077)	(.0067)	(.0063)	(.0072)	(.005)	(.0081)	(.009)
healthm	-0.1	0.059	-0.17	0.11	-0.11	0.09	-0.095	0.015	0.072	-0.099
	(.033)	(.03)	(.047)	(.032)	(.037)	(.038)	(.043)	(.031)	(.048)	(.04)
literacy==1	0.027	0.071	0.024	0.12	0.013	0.053	-0.017	0.044	-0.041	-0.077
	(.026)	(.024)	(.037)	(.032)	(.032)	(.029)	(.035)	(.024)	(.052)	(.038)
literacy==9	-0.15	-0.012	-0.036	0.042	-0.29	-0.16	-0.25	-0.13	-0.22	-0.18
	(.21)	(.19)	(.14)	(.077)	(.24)	(.08)	(.36)	(.13)	(.37)	(.19)
Average age of plot managers	0.0061	0.0064	0.0043	0.0062	0.0051	0.0064	0.0071	0.0074	0.0029	0.0012
	(.00062)	(.00056)	(.00083)	(.00079)	(.00072)	(.00067)	(.0008)	(.00057)	(.00085)	(.0011)
av_agem	-5.95	-6.31	-4.36	-5.87	-4.89	-6.18	-6.96	-7.27	-2.6	-1.4
	(.62)	(.55)	(.8)	(.76)	(.7)	(.64)	(.76)	(.55)	(.95)	(1.05)
Crops Lost	0.12	0.091	0.12	0.11	0.1	0.1	0.11	0.088	-0.0065	-0.022
	(.019)	(.017)	(.02)	(.024)	(.022)	(.02)	(.023)	(.018)	(.036)	(.028)
Adverse plot shock	0.26	0.14	0.29	0.2	0.25	0.15	0.21	0.07	-0.08	-0.13
	(.022)	(.02)	(.027)	(.026)	(.025)	(.022)	(.027)	(.019)	(.038)	(.031)
(soil_quality==1)*evimax	0.14	1.48	0.16	1.66	0.51	2.2	0.23	1.47	0.064	-0.18

Appendix Table A4a: OLS and Quantile Regression Determinants of Land and Labor Inputs in Tanzania

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
	(.31)	(.28)	(.37)	(.32)	(.37)	(.34)	(.38)	(.32)	(.57)	(.38)
(soil_quality==2)*evimax	-1.28	0.77	-2.17	0.62	-1.14	1.47	-0.78	1.04	1.39	0.42
	(.3)	(.27)	(.37)	(.36)	(.36)	(.31)	(.38)	(.32)	(.51)	(.49)
(soil_quality==3)*evimax	-0.4	0.5	-0.37	0.27	-0.064	1.6	-0.37	0.28	-0.0051	0.011
	(.56)	(.5)	(.43)	(.71)	(.74)	(.59)	(.63)	(.61)	(.91)	(.77)
(soil_type==2)*evimax	-0.48	-0.14	0.25	0.34	-0.44	-0.69	-0.64	-0.43	-0.89	-0.78
	(.32)	(.28)	(.37)	(.35)	(.38)	(.33)	(.4)	(.32)	(.62)	(.38)
(soil_type==3)*evimax	-1.28	0.026	-0.3	0.64	-1.31	-0.47	-1.59	-0.33	-1.29	-0.97
	(.4)	(.36)	(.51)	(.46)	(.48)	(.42)	(.47)	(.37)	(.71)	(.59)
(soil_type==4)*evimax	-4.38	-2.45	-3.42	-2.5	-4.88	-4.54	-4.31	-1.68	-0.9	0.82
	(1.16)	(1.04)	(1.16)	(1.41)	(.86)	(.75)	(1.13)	(.96)	(2.65)	(1.43)
Drought/Floods	0.044	0.12	0.047	0.14	0.015	0.15	0.014	0.085	-0.033	-0.053
	(.02)	(.018)	(.025)	(.024)	(.023)	(.021)	(.027)	(.019)	(.038)	(.022)
Crop disease or pest	-0.0036	0.063	-0.015	0.032	0.0048	0.039	-0.022	0.071	-0.0069	0.039
	(.021)	(.019)	(.028)	(.027)	(.024)	(.023)	(.026)	(.019)	(.038)	(.026)
severe water shortage	0.14	0.011	0.17	0.021	0.15	0.0067	0.13	0.026	-0.042	0.0056
	(.021)	(.019)	(.028)	(.026)	(.023)	(.02)	(.025)	(.021)	(.033)	(.033)
Adverse plot shock in HH	-0.02	-0.023	-0.024	-0.029	-0.017	-0.015	-0.011	-0.027	0.013	0.0019
	(.014)	(.012)	(.017)	(.016)	(.012)	(.013)	(.018)	(.01)	(.022)	(.018)
Constant	-4E-08	-8E-08	-0.62	-0.54	-0.0037	0.063	0.63	0.63	1.25	1.16
	-0.0085	-0.0076	-0.012	-0.012	-0.01	-0.0096	-0.012	-0.0089	-0.017	-0.0089

Appendix Table A4a: OLS and Quantile Regression Determinants of Land and Labor Inputs in Tanzania

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
F statistic; see note #.	26.8	12.2	15.6	7.85	25.8	10.8	26.8	15.2		
Corresponding p value	8.8E-61	5.2E-25	2.1E-33	8.5E-15	2.3E-58	8.1E-22	9.5E-61	2.9E-32		
F statistic; see note §.									23.2	11.1
Corresponding p value									3E-80	5.6E-34

Standard errors in parentheses
#F statistic for joint significance of variables used as production function instruments in Table 5b.
§F stat for h0 – coefficients of instruments equal for 25th and 75th percentile.

Appendix Table A4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
Male Plot	0.14	0.1	0.13	0.092	0.14	0.11	0.14	0.11	0.014	0.021
	(.01)	(.008)	(.012)	(.01)	(.011)	(.0082)	(.011)	(.0083)	(.013)	(.013)
EVI*Avg Soil in HH	-0.00018	-0.00037	-0.00031	-0.00038	-0.00013	-0.00039	-0.00053	-0.00046	-0.00022	-0.000074
	(.00018)	(.00016)	(.00023)	(.00019)	(.00019)	(.00014)	(.00022)	(.00015)	(.00027)	(.00022)
EVI*Poor Soil in HH	0.00041	0.00039	0.00062	-0.00045	0.00049	-0.00052	-0.000092	0.0003	-0.00071	0.00075
	(.00059)	(.00052)	(.00068)	(.00048)	(.00075)	(.00046)	(.00072)	(.00047)	(.00068)	(.00067)
EVI*Missing Soil in HH	-0.0066	-0.00038	-0.0051	-0.00048	-0.0043	0.00092	-0.0049	-0.0017	0.00024	-0.0012
	(.0019)	(.0015)	(.0024)	(.0016)	(.0016)	(.0012)	(.0017)	(.0012)	(.0039)	(.0016)
Drought*Avg Soil in HH	-0.0063	-0.0071	-0.01	-0.0075	-0.0057	-0.0068	-0.0051	-0.0075	0.0052	-0.000077
	(.002)	(.0016)	(.0025)	(.0018)	(.0016)	(.0015)	(.0018)	(.0014)	(.0023)	(.0025)
Drought*Poor Soil in HH	-0.0029	0.00075	-0.0045	0.000049	-0.0063	0.0032	-0.0021	0.0053	0.0024	0.0053
	(.0044)	(.0037)	(.005)	(.0025)	(.0047)	(.0037)	(.0049)	(.0035)	(.0058)	(.0057)
Drought*Missing Soil in HH	-0.031	0.023	-0.041	0.018	-0.024	0.017	-0.0017	0.029	0.039	0.011
	(.0089)	(.0059)	(.013)	(.007)	(.0084)	(.0072)	(.0075)	(.0044)	(.015)	(.0086)
Illness Incidence in household	-0.0072	-0.073	-0.0059	-0.084	0.0014	-0.088	-0.013	-0.096	-0.0071	-0.011
	(.014)	(.011)	(.016)	(.015)	(.013)	(.012)	(.015)	(.012)	(.022)	(.02)
Household non-agric shock	0.088	-0.0023	0.08	-0.0011	0.072	-0.0041	0.09	-0.0097	0.01	-0.0086
	(.029)	(.023)	(.023)	(.031)	(.033)	(.022)	(.026)	(.021)	(.034)	(.033)
No of household members	0.026	0.017	0.024	0.02	0.023	0.016	0.026	0.017	0.0015	-0.003
	(.0024)	(.0019)	(.003)	(.0027)	(.0024)	(.0019)	(.0026)	(.002)	(.003)	(.0029)

Appendix Table A4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
No of adults in household	0.011	0.0082	0.011	0.0041	0.016	0.0075	0.0093	0.0078	-0.0019	0.0037
	(.0033)	(.0027)	(.0042)	(.0037)	(.0034)	(.0027)	(.0037)	(.0029)	(.005)	(.0039)
Housing Quality	0.079	0.025	0.059	0.016	0.063	0.0058	0.063	0.025	0.0042	0.0088
	(.012)	(.01)	(.015)	(.013)	(.013)	(.0099)	(.013)	(.01)	(.021)	(.013)
Every household member has at least o	-0.015	-0.076	-0.022	-0.093	-0.0034	-0.052	0.019	-0.03	0.041	0.062
	(.011)	(.0087)	(.014)	(.011)	(.012)	(.0088)	(.012)	(.0091)	(.02)	(.012)
Household electricity access dummy	0.087	0.052	0.13	0.035	0.093	0.03	0.015	0.026	-0.11	-0.0093
	(.022)	(.017)	(.021)	(.018)	(.025)	(.017)	(.019)	(.016)	(.035)	(.024)
Literacy of plot managers	0.037	0.016	0.035	0.026	0.046	0.024	0.025	0.017	-0.01	-0.0095
	(.014)	(.012)	(.018)	(.02)	(.015)	(.013)	(.02)	(.014)	(.016)	(.023)
Schooling level of plot managers	0.00018	0.0079	-0.016	0.017	-0.0042	0.0063	-0.0058	-0.01	0.01	-0.028
	(.014)	(.012)	(.018)	(.017)	(.015)	(.014)	(.016)	(.013)	(.019)	(.016)
Plot manager is recent resident	0.061	0.099	0.061	0.097	0.09	0.091	0.086	0.1	0.025	0.006
	(.022)	(.018)	(.04)	(.021)	(.021)	(.018)	(.026)	(.019)	(.034)	(.025)
One or both manager serve on committ	0.019	0.002	0.0097	-0.008	-0.0055	0.021	-0.018	0.013	-0.028	0.021
	(.014)	(.012)	(.017)	(.014)	(.014)	(.012)	(.014)	(.013)	(.018)	(.022)
Household Adult Equivalence Scale (*	0.028	0.092	0.0013	0.12	0.0075	0.09	-0.046	0.018	-0.048	-0.098
	(.05)	(.04)	(.048)	(.044)	(.076)	(.045)	(.019)	(.034)	(.066)	(.05)
Average Shocks on Other Plots in Household:										
Low Early Rain * Avg Soil	0.086	0.07	0.083	0.065	0.098	0.075	0.079	0.076	-0.0043	0.011

Appendix Table A4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
	(.035)	(.025)	(.036)	(.02)	(.079)	(.0092)	(.048)	(.018)	(.05)	(.043)
Low Early Rain * Poor Soil	0.41	-0.036	0.23	0.068	0.37	-0.11	0.66	-0.082	0.43	-0.15
	(.14)	(.073)	(.08)	(.16)	(.36)	(.2)	(.16)	(.049)	(.29)	(.11)
Low Early Rain * Missing Soil	-0.21	0.029	-0.2	-0.013	-0.32	-0.059	-0.094	0.034	0.1	0.047
	(.093)	(.074)	(.18)	(.11)	(2.73)	(.038)	(.3)	(.14)	(.15)	(.11)
Flood*Avg Soil	0.0015	-0.0022	0.016	-0.0093	-0.00018	-0.0013	-0.01	-0.0046	-0.027	0.0047
	(.0098)	(.0097)	(.013)	(.013)	(.018)	(.0058)	(.0085)	(.0044)	(.016)	(.021)
Flood*Poor Soil	-0.091	0.041	-0.039	0.022	-0.081	0.056	-0.19	0.04	-0.15	0.018
	(.06)	(.034)	(.078)	(.092)	(.13)	(.048)	(.14)	(.079)	(.19)	(.074)
Flood*Missing Soil	0.27	0.12	0.33	0.13	0.22	0.14	0.21	0.16	-0.12	0.029
	(.069)	(.048)	(.098)	(.11)	(2.71)	(.035)	(.21)	(.12)	(.12)	(.12)
High Early Rain * Avg Soil	-0.000015	0.000009	-3.4E-06	0.000018	-0.000038	1.5E-06	0.000012	0.000015	0.000016	-0.0000029
	(.000023)	(.00002)	(.000028)	(.000026)	(.000026)	(.000019)	(.000029)	(.00002)	(.000034)	(.000025)
High Early Rain * Poor Soil	-0.000099	-0.000078	-0.00005	0.000028	-0.00013	4.4E-06	-0.000073	-0.00012	-0.000023	-0.00015
	(.000077)	(.000068)	(.000087)	(.000054)	(.0001)	(.000064)	(.000096)	(.000056)	(.0001)	(.000094)
High Early Rain * Other Soil	-0.00052	-0.00053	-0.00089	-0.00047	-0.00061	-0.00052	-0.00025	-0.00061	0.00064	-0.00014
	(.00036)	(.00011)	(.000031)	(.087)	(.00012)	(.00002)	(.000038)	(.000094)	(.0039)	(.0025)
High Early Rain * Missing Soil	0.0011	-0.0001	0.00094	-0.000034	0.00083	-0.00028	0.0006	0.0001	-0.00034	0.00013
	(.00028)	(.00023)	(.0003)	(.00021)	(.00026)	(.00018)	(.00026)	(.00019)	(.00056)	(.00025)
Adverse shock	-0.0011	0.00054	-0.00083	0.00031	-0.00078	0.00029	-0.0012	0.00052	-0.0004	0.00021
	(.00038)	(.00028)	(.00041)	(.0003)	(.00034)	(.00023)	(.00032)	(.00025)	(.00056)	(.00043)

Appendix Table A4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
Average Shocks on Other Plots in Village:										
Drought*Avg Soil	0.0022	0.0011	0.002	0.0011	0.0021	0.00097	0.0017	0.00078	-0.00033	-0.00031
	(.00038)	(.00029)	(.00048)	(.00038)	(.00039)	(.00029)	(.00038)	(.00029)	(.00048)	(.00039)
Drought*Poor Soil	0.0031	-0.0035	0.0023	-0.0029	-0.00064	-0.0038	-0.000009	-0.0048	-0.0023	-0.0019
	(.0012)	(.00091)	(.0022)	(.0012)	(.001)	(.00092)	(.0013)	(.00095)	(.0019)	(.0016)
Drought*Missing Soil	0.0058	-0.0024	0.006	-0.00077	0.0056	-0.0034	-0.00045	-0.0033	-0.0064	-0.0025
	(.0024)	(.0019)	(.0027)	(.0023)	(.0024)	(.0015)	(.0014)	(.0019)	(.0031)	(.0019)
Flood*Avg Soil	-0.038	-0.0041	-0.051	-0.011	-0.036	-0.0055	-0.021	0.01	0.029	0.021
	(.0093)	(.0071)	(.012)	(.0096)	(.012)	(.0078)	(.011)	(.0083)	(.015)	(.012)
Flood*Poor Soil	-0.072	-0.0087	-0.036	-0.018	-0.021	0.016	-0.12	0.034	-0.087	0.052
	(.032)	(.037)	(.037)	(.058)	(.024)	(.047)	(.064)	(.024)	(.056)	(.039)
Flood*Missing Soil	0.082	0.029	0.14	0.037	0.062	0.01	0.038	0.041	-0.099	0.0039
	(.035)	(.027)	(.046)	(.054)	(.055)	(.025)	(.023)	(.028)	(.046)	(.038)
Low Early Rain * Avg Soil	0.0039	0.00029	0.0073	0.0012	0.003	0.00021	0.0012	-0.0034	-0.0062	-0.0046
	(.0025)	(.002)	(.0034)	(.0031)	(.0035)	(.0025)	(.0033)	(.0021)	(.0046)	(.004)
Low Early Rain * Poor Soil	0.017	0.021	0.0065	0.039	0.014	0.021	0.044	0.0025	0.038	-0.036
	(.018)	(.017)	(.024)	(.029)	(.011)	(.016)	(.025)	(.015)	(.023)	(.019)
Low Early Rain * Missing Soil	-0.018	-0.041	-0.045	-0.029	-0.01	-0.037	-0.0022	-0.057	0.043	-0.028
	(.019)	(.014)	(.036)	(.046)	(.035)	(.012)	(.016)	(.012)	(.022)	(.022)
High Early Rain * Avg Soil	3.5E-06	2.4E-06	6.8E-06	6.1E-06	0.0000041	2.4E-06	-6.8E-06	-2.4E-06	-0.000014	-0.0000085

Appendix Table A4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
	(.000003)	(.000003)	(.000004)	(.000004)	(.000004)	(.000002)	(.000004)	(.000002)	(.000004)	(.000004)
High Early Rain * Poor Soil	0.000039	-0.000003	0.000035	-7.7E-06	0.000014	-0.00002	0.000022	-0.000016	-0.000013	-0.0000085
	(.000019)	(.000015)	(.000025)	(.00002)	(.000019)	(.000015)	(.000022)	(.000016)	(.000024)	(.00002)
High Early Rain * Other Soil	0.00034	0.00034	0.00031	0.00038	0.00032	0.00033	0.00034	0.00024	0.000023	-0.00013
	(.0001)	(.000075)	(.000034)	(.00049)	(.000068)	(.000016)	(.00002)	(.000093)	(.00019)	(.00011)
High Early Rain * Missing Soil	0.00012	-0.000067	0.000091	-0.00007	0.000062	-0.000013	0.00018	-0.000043	0.000087	0.000027
	(.000055)	(.000042)	(.00006)	(.000048)	(.000049)	(.000033)	(.000048)	(.000037)	(.00007)	(.000063)
EVI*Avg Soil in Village	-0.000041	0.000022	-0.000057	5.7E-06	-0.000044	6.6E-06	3.9E-06	0.000022	0.000061	0.000016
	(.000021)	(.000018)	(.00003)	(.000023)	(.000026)	(.000016)	(.000025)	(.000017)	(.000028)	(.000025)
EVI*Poor Soil in Village	-0.00063	-0.000086	-0.00064	-0.000059	-0.00033	0.00002	-0.00038	0.000083	0.00026	0.00014
	(.00014)	(.00011)	(.00022)	(.00014)	(.00013)	(.00012)	(.00017)	(.00012)	(.00019)	(.00015)
EVI*Missing Soil in Village	0.021	-0.00052	0.018	-0.0035	0.048	-0.00058	0.081	-0.000032	0.063	0.0035
	(.0029)	(.0013)	(.0024)	(.0032)	(.0099)	(.00046)	(.0072)	(.0012)	(.01)	(.0032)
Adverse Shock	0.0012	0.00055	0.0014	0.0012	0.0015	0.00077	0.0023	0.00024	0.00087	-0.001
	(.00048)	(.00031)	(.00061)	(.00053)	(.00044)	(.00031)	(.00046)	(.00018)	(.00065)	(.0005)
Own Plot Shocks:										
Drought*Avg Soil	0.026	0.027	0.038	0.025	0.024	0.022	0.018	0.036	-0.02	0.011
	(.0058)	(.0049)	(.0064)	(.0062)	(.0057)	(.0049)	(.0059)	(.0051)	(.0077)	(.0081)
Drought*Poor Soil	0.041	0.036	0.042	0.031	0.042	0.024	0.021	0.017	-0.021	-0.014
	(.013)	(.0098)	(.012)	(.0094)	(.014)	(.0097)	(.014)	(.012)	(.016)	(.017)

Appendix Table A4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
Drought*Missing Soil	0.019	-0.011	0.006	-0.02	-0.0015	0.016	0.007	0.01	0.001	0.03
	(.026)	(.02)	(.04)	(.023)	(.028)	(.015)	(.026)	(.012)	(.04)	(.022)
Flood*Avg Soil	-0.036	-0.048	-0.065	-0.04	-0.044	-0.04	-0.02	-0.048	0.044	-0.0086
	(.02)	(.021)	(.021)	(.03)	(.018)	(.025)	(.013)	(.026)	(.029)	(.041)
Flood*Poor Soil	-0.14	0.099	0.074	0.092	-0.0077	0.0088	-0.16	-0.015	-0.24	-0.11
	(.099)	(.074)	(1.65)	(.15)	(.37)	(.065)	(.13)	(.13)	(.34)	(.17)
Flood*Missing Soil	-0.63	-0.3	-1.09	-0.24	-0.32	-0.38	-0.65	-0.56	0.44	-0.32
	(.21)	(.13)	(.5)	(.22)	(.76)	(.12)	(.9)	(.71)	(.52)	(.25)
High Early Rain * Avg Soil	-0.00018	8.2E-06	-0.00017	-0.000038	-0.000086	-0.000038	-0.00024	-0.000066	-0.000065	-0.000028
	(.000075)	(.000066)	(.000094)	(.0001)	(.000088)	(.000065)	(.000098)	(.000068)	(.000075)	(.000092)
High Early Rain * Poor Soil	-0.0002	0.00022	0.000023	-0.000097	0.00002	0.000065	-0.00028	0.00023	-0.0003	0.00033
	(.00021)	(.00018)	(.00029)	(.00021)	(.00019)	(.00018)	(.00025)	(.00019)	(.00032)	(.00023)
High Early Rain * Other Soil	-0.0013	0.00029	-0.0018	-0.00073	-0.0023	0.0011	-0.00058	0.0014	0.0012	0.0021
	(.0019)	(.00076)	(.26)	(1.03)	(.029)	(.012)	(.00026)	(.19)	(.031)	(.014)
High Early Rain * Missing Soil	-0.0028	0.0018	-0.0029	0.0012	-0.0015	0.0016	-0.0011	0.0016	0.0018	0.00039
	(.00086)	(.00063)	(.0012)	(.00057)	(.0004)	(.0008)	(.00068)	(.00063)	(.0015)	(.0012)
EVI*Avg Soil	0.00063	0.00066	0.00025	0.00076	0.00056	0.00096	0.0015	0.00097	0.0013	0.00021
	(.00054)	(.00048)	(.00065)	(.00069)	(.00062)	(.00047)	(.00071)	(.00049)	(.00065)	(.00068)
EVI*Poor Soil	0.0018	-0.0021	-0.0015	-0.0002	0.0029	-0.00082	0.0025	-0.0019	0.004	-0.0017
	(.0015)	(.0012)	(.0022)	(.0017)	(.0014)	(.0012)	(.002)	(.0013)	(.0022)	(.0016)
EVI*Missing Soil	0.0065	-0.011	0.0073	-0.0081	0.002	-0.0086	0.006	-0.0065	-0.0013	0.0016

Appendix Table A4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
	(.0052)	(.0043)	(.0073)	(.0047)	(.0032)	(.0053)	(.0031)	(.0038)	(.0079)	(.0087)
soil_type==2	-0.0076	-0.016	-0.016	-0.0061	0.013	-0.0096	0.0096	-0.0061	0.025	-0.000035
	(.012)	(.0093)	(.014)	(.012)	(.013)	(.0098)	(.013)	(.0098)	(.015)	(.019)
soil_type==3	-0.0018	-0.031	0.012	-0.029	0.011	-0.031	0.011	-0.023	-0.0011	0.0063
	(.014)	(.012)	(.017)	(.016)	(.016)	(.012)	(.016)	(.013)	(.015)	(.014)
soil_type==4	-0.052	0.018	-0.051	0.0013	-0.085	-0.015	-0.053	0.016	-0.0018	0.015
	(.025)	(.02)	(.032)	(.026)	(.032)	(.023)	(.032)	(.025)	(.037)	(.035)
soil_type==6	0.093	0.12	0.056	0.11	0.028	0.13	0.12	0.17	0.06	0.061
	(.023)	(.02)	(.039)	(.025)	(.024)	(.02)	(.025)	(.02)	(.034)	(.031)
soil_type==99999	0.13	0.045	0.17	-0.087	0.023	0.078	0.045	0.14	-0.13	0.23
	(.083)	(.086)	(.078)	(.045)	(.048)	(.039)	(.057)	(.13)	(.12)	(.16)
soil_quality==2	-0.015	-0.073	0.015	-0.055	-0.033	-0.059	-0.019	-0.072	-0.035	-0.016
	(.024)	(.02)	(.029)	(.026)	(.026)	(.021)	(.029)	(.022)	(.025)	(.024)
soil_quality==3	-0.21	-0.16	-0.18	-0.15	-0.32	-0.12	-0.072	-0.13	0.11	0.023
	(.067)	(.051)	(.07)	(.072)	(.082)	(.06)	(.1)	(.048)	(.07)	(.062)
soil_quality==5	0.97	0.08	0.94	0.39	0.2	0.14	0.81	-0.26	-0.13	-0.65
	(.4)	(.082)	(141.2)	(130.1)	(6.45)	(1.18)	(.11)	(106.2)	(5.93)	(7.25)
soil_quality==99999	0.59	0.33	0.59	0.35	0.5	0.054	0.55	0.033	-0.037	-0.32
	(.41)	(.21)	(.38)	(.18)	(.26)	(.3)	(.33)	(.2)	(.73)	(.37)
water_source==2	-0.14	-0.0089	-0.13	-0.063	-0.12	-0.024	-0.21	0.019	-0.082	0.082
	(.041)	(.029)	(.068)	(.035)	(.036)	(.027)	(.072)	(.021)	(.086)	(.038)

Appendix Table A4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
water_source==3	-0.22	-0.0035	-0.15	-0.094	-0.21	-0.055	-0.38	0.032	-0.22	0.13
	(.054)	(.042)	(.082)	(.055)	(.056)	(.037)	(.09)	(.044)	(.12)	(.057)
water_source==99999	-0.48	-0.48	-0.77	-0.35	-0.74	-0.47	-0.91	-0.32	-0.15	0.024
	(.23)	(.19)	(.11)	(.16)	(.055)	(.049)	(.18)	(.41)	(.38)	(.23)
slope==2	0.0091	-0.037	0.023	-0.031	0.039	-0.0081	0.07	-0.012	0.047	0.018
	(.02)	(.015)	(.031)	(.02)	(.021)	(.016)	(.022)	(.016)	(.026)	(.016)
slope==3	-0.04	0.012	-0.059	0.013	-0.045	0.018	0.0026	0.018	0.062	0.0044
	(.019)	(.014)	(.031)	(.019)	(.02)	(.015)	(.021)	(.015)	(.025)	(.021)
slope==4	-0.069	0.016	-0.13	-0.032	-0.091	0.0092	-0.025	0.071	0.11	0.1
	(.033)	(.025)	(.043)	(.033)	(.035)	(.029)	(.045)	(.028)	(.065)	(.042)
slope==5	0.075	0.055	0.015	0.037	0.031	0.055	0.13	0.076	0.12	0.039
	(.039)	(.028)	(.039)	(.033)	(.05)	(.026)	(.064)	(.032)	(.055)	(.032)
slope==6	-0.29	0.53	-0.64	0.79	-0.55	0.69	-0.3	0.49	0.33	-0.3
	(.29)	(.19)	(73.2)	(.33)	(76.6)	(.029)	(.052)	(.036)	(.6)	(.44)
slope==99999	0.4	0.027	0.58	-0.016	0.55	0.23	0.47	-0.014	-0.11	0.0017
	(.24)	(.16)	(.31)	(.24)	(.15)	(.24)	(.071)	(.41)	(.36)	(.27)
erosion==2	-0.02	-0.028	-0.024	-0.039	-0.0097	-0.028	-0.06	-0.025	-0.037	0.014
	(.013)	(.01)	(.017)	(.013)	(.014)	(.011)	(.015)	(.011)	(.021)	(.011)
erosion==99999	-0.13	-0.011	-0.0032	0.019	-0.059	0.0035	-0.19	-0.064	-0.19	-0.083
	(.12)	(.09)	(.24)	(.1)	(.069)	(.19)	(.21)	(.13)	(.25)	(.14)

Appendix Table A4b: OLS and Quantile Regression Determinants of Land and Labor Inputs in Uganda

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS		25th percentile		50th percentile		75th percentile		Interquartile Range	
	Land	Labor	Land	Labor	Land	Labor	Land	Labor	Land	Labor
Extension services	0.16	0.052	0.16	0.064	0.15	0.051	0.11	0.036	-0.051	-0.028
	(.011)	(.0095)	(.014)	(.012)	(.013)	(.0094)	(.012)	(.01)	(.015)	(.015)
Constant	0.00	0.00	-0.53	-0.37	0.019	0.059	0.55	0.43	1.08	0.8
	(.0048)	(.0038)	(.023)	(.021)	(.023)	(.0042)	(.006)	(.0078)	(.0091)	(.0063)
F statistic; see note #.	23	16.7	55.7	27.2	28.5	67.5	30.4	59.4		
Corresponding p value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
F statistic; see note §.									45.9	14.5
Corresponding p value									0.00	0.00

Standard errors in parentheses
#F statistic for joint significance of variables used as production function instruments in Table 5b.
§F stat for h0 – coefficients of instruments equal for 25th and 75th percentile.

Appendix Table A5a: Tanzania Production Function Estimates

	Tanzania	
	2SLS	IVCRC
Land	0.76	0.57
	(0.15)	(0.01)
Labor	0.21	0.37
	(0.23)	(0.03)
Land Value	0.00	0.03
	(0.02)	(0.01)
Land Value Missing	0.14	0.09
	(0.08)	(0.10)
Distance Home	0.00	0.00
	(0.00)	(0.00)
Distance to Road	-0.02	-0.02
	(0.00)	(0.00)
Good Soil	0.14	-4.55
	(0.32)	(0.60)
Average Soil	-0.17	-4.86
	(0.33)	(0.55)
Sandy Soil	0.47	5.88
	(0.88)	(3.28)
Loamy Soil	0.84	6.46
	(0.91)	(3.29)
Clay Soil	0.51	6.60
	(0.96)	(3.25)
Single Manager	-0.05	-0.01
	(0.03)	(0.01)
Poor Health	-0.03	-0.03
	(0.01)	(0.00)
Missing Health	-0.11	-0.15
	(0.05)	(0.05)
Illiterate Manager	-0.06	-0.08
	(0.04)	(0.02)
Literacy Missing	0.07	0.13
	(0.30)	(1.40)
Male Manager	-0.06	0.00
	(0.03)	(0.01)

Appendix Table A5a: Tanzania Production Function Estimates

	Tanzania	
	2SLS	IVCRC
Manager Age	0.00	0.00
	(0.00)	(0.00)
Age Missing	3.66	4.01
	(1.02)	(0.52)
Crops Lost	0.07	0.07
	(0.03)	(0.01)
Bad Shock	-0.27	-0.27
	(0.03)	(0.01)
EVI*Good Soil	0.14	0.01
	(0.54)	(0.46)
EVI*Average Soil	0.51	0.37
	(0.58)	(0.31)
EVI*Poor Soil	0.00	-8.67
	(0.63)	(1.34)
EVI*Loamy Soil	-0.58	-0.94
	(0.32)	(0.32)
EVI*Clay Soil	0.03	-1.35
	(0.59)	(0.31)
EVI*Other Soil	1.11	12.37
	(1.65)	(5.68)
Drought/flood Severity	-0.07	-0.05
	(0.03)	(0.01)
Crop Disease Severity	0.02	0.03
	(0.02)	(0.01)
Water Shortage	-0.07	-0.06
	(0.03)	(0.02)
Constant	0.00	0.03
	(0.00)	(0.04)

Notes: Bootstrapped standard errors (50 bootstrap iterations) in parentheses.

Appendix Table A5b: Uganda Production Function Estimates

	Uganda	
	2SLS	IVCRC
Land	0.70	0.53
	(0.05)	(0.00)
Labour	0.14	0.38
	(0.10)	(0.00)
Drought*Avg Soil	-0.03	-0.03
	(0.01)	(0.00)
Drought*Poor Soil	-0.04	-0.05
	(0.02)	(0.01)
Drought*Missing Soil	0.00	-0.01
	(0.03)	(0.05)
Flood*Avg Soil	0.01	-0.05
	(0.03)	(0.03)
Flood*Poor Soil	0.26	0.46
	(0.31)	(0.63)
Flood*Missing Soil	-0.22	1.16
	(0.17)	(0.41)
High Early Rain * Avg Soil	0.00	0.00
	(0.00)	(0.00)
High Early Rain * Poor Soil	0.00	0.00
	(0.00)	(0.00)
High Early Rain * Other Soil	0.00	-0.04
	(0.00)	(0.03)
High Early Rain * Missing Soil	0.00	0.00
	(0.00)	(0.00)
EVI*Avg Soil	0.00	0.00
	(0.00)	(0.00)
EVI*Poor Soil	0.00	0.00
	(0.00)	(0.00)
EVI*Missing Soil	0.00	0.00
	(0.00)	(0.00)
soil_type==2	0.00	0.03
	(0.01)	(0.01)
soil_type==3	0.02	0.02
	(0.02)	(0.00)
soil_type==4	0.02	0.01
	(0.02)	(0.00)
soil_type==6	0.01	-0.03

Appendix Table A5b: Uganda Production Function Estimates

	Uganda	
	2SLS	IVCRC
	(0.04)	(0.02)
soil_type==99999	-0.02	-0.01
	(0.04)	(0.00)
soil_quality==2	-0.24	-0.34
	(0.11)	(0.14)
soil_quality==3	-0.13	-0.11
	(0.03)	(0.00)
soil_quality==5	0.07	0.04
	(0.09)	(0.03)
soil_quality==99999	-1.09	0.51
	(0.61)	(0.74)
water_source==2	-0.54	3.78
	(0.38)	(0.81)
water_source==3	0.18	0.12
	(0.06)	(0.01)
water_source==99999	0.39	0.29
	(0.08)	(0.01)
slope==2	0.52	1.32
	(0.32)	(0.53)
slope==3	0.01	0.01
	(0.03)	(0.00)
slope==4	0.08	0.07
	(0.02)	(0.00)
slope==5	0.11	0.04
	(0.04)	(0.02)
slope==6	-0.05	-0.04
	(0.05)	(0.02)
slope==99999	0.34	4.10
	(0.29)	(2.31)
erosion==2	-0.19	-5.95
	(0.27)	(0.73)
erosion==99999	0.08	0.07
	(0.02)	(0.00)

Appendix Table A5b: Uganda Production Function Estimates

	Uganda	
	2SLS	IVCRC
Male plot	0.56	0.94
	(0.18)	(0.26)
Extension	0.09	0.09
	(0.01)	(0.00)
Constant	0.09	0.12
	(0.02)	(0.00)

Notes: Bootstrapped standard errors (50 bootstrap iterations) in parentheses.