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

Douglas Gollin

*Oxford University*

Christopher Udry

*Northwestern University*

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

## I. Introduction

How important is misallocation in explaining the income differences across countries? A recent literature in development and growth economics

We are grateful to Philip Wollburg and Eduardo Campillo Betancourt for excellent research assistance. We received helpful comments from many 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. Data are provided as supplementary material online.

Electronically published November 30, 2020
[*Journal of Political Economy*, 2021, vol. 129, no. 1]

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. In an aggregate sense, misallocation across sectors or firms can significantly reduce TFP. Such productivity differences have the potential to account for a large fraction of the cross-country income differences.

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: for example, Rotemberg and White (2017), Haltiwanger, Kulick, and Syverson (2018), Pellegrino and Zheng (2018), White, Reiter, and Petrin (2018), and Bils, Klenow, and Ruane (2020). 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 farms. Many of these farms produce identical and homogeneous outputs on different plots within each growing season. This allows us to observe within-season variation across plots for a given farmer in the input intensity and output of the same crop. We cannot interpret this variation as the result of misallocation, since farmers presumably face no market imperfections in allocating resources across their own plots. As a result, these data allow us to identify and quantify misallocation more precisely. Our strategy allows us to disentangle the productivity dispersion that arises from misallocation from that stemming from measurement error or heterogeneity in technology and inputs (including production shocks).

The agricultural sector provides a valuable window through which to study firm-level misallocation. Most firm surveys have relatively few observations on different plants or establishments operated by the same firm, and this makes it difficult to disentangle firm management from any unobservable characteristics of the plant or factory; in contrast, we observe farmers concurrently operating multiple plots.[2] Another advantage we have,

[1] See, e.g., Restuccia and Rogerson (2008, 2013), Hsieh and Klenow (2009), McMillan, Rodrik, and Verduzco-Gallo (2014), Porzio (2016), Bento and Restuccia (2017), Hicks et al. (2017), and Restuccia (2018).

[2] Recent literature offers several interesting exceptions, with a number of papers looking at productivity differences across divisions of industrial conglomerates or across plants of manufacturing firms: e.g., Maksimovic and Phillips (2001), Matvos and Seru (2014), and

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

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

Restuccia and Santaeulalia-Llopis (2017), in particular, have raised the intriguing possibility that much of Africa's productivity deficit might be attributable to misallocation within the agricultural sector. They find suggestive evidence, in data from Malawi, that too much farmland is managed by low-skill farmers. If true, this finding might offer an explanation for sub-Saharan Africa's low productivity in agriculture. By extension, misallocation might thus help explain the region's low levels of income per capita. The finding also suggests a relatively straightforward solution—albeit one with great political complexity—namely, the liberalization of land and input markets, so that the best farmers can eventually buy out those farmers who lack the skill to farm productively. Restuccia and Santaeulalia-Llopis calculate that reassigning land from low-productivity establishments to high-productivity establishments (or equivalently from bad farmers to good farmers) could result in more than a threefold increase in aggregate agricultural output.

Ševčík (2015). However, these comparisons are complicated by the difficulty of comparing productivity differences across different outputs, especially in an environment with imperfect competition, where firms have some market power in setting prices. An interesting paper, directly relevant for ours, is Kehrig and Vincent (2019), which looks at firms that operate multiple plants (establishments) producing the same good. Because of lumpiness in investment at the plant level, Kehrig and Vincent point out that efficient internal allocation of capital may lead firms not to equalize revenue measures of productivity (TFPR, to use the Hsieh-Klenow terminology) across plants. The authors revisit the cross-country comparisons of Hsieh and Klenow (2009) and estimate the gains from reallocation after accounting for the role of multiplant firms.

[3] See, e.g., Adamopoulos and Restuccia (2014, 2015), Adamopoulos et al. (2017), Bento and Restuccia (2017), and Restuccia and Santaeulalia-Llopis (2017).

A number of other studies explore the potential importance of misallocation across farms in the developing world: for example, Emran and Shilpi (2015), Chen (2016), Adamopoulos et al. (2017), Foster and Rosenzweig (2017), Shenoy (2017), and Gottlieb and Grobovšek (2018). A paper close to ours in approach—in that it seeks to distinguish misallocation from measurement error—is Esfahani (2018). These papers build on a broader literature that has looked at productivity differences outside agriculture—typically across firms and plants in advanced countries. Key works include Syverson (2004, 2011), Petrin, White, and Reiter (2011), and Petrin and Sivadasan (2013).

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

A challenge to all this literature is the measurement of productivity at the level of individual firms. Typically, the data used for these analyses come from firm surveys that vary in quality and 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 by, for example, Asker, Collard-Wexler, and De Loecker (2014), Foster et al. (2016), and Haltiwanger (2016). Our approach addresses some of the concerns raised by these critiques. In particular, our approach recognizes that firm-level productivity may vary for many reasons other than misallocation.

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

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

Drawing on the model, we assess the relative importance of different sources of dispersion in measured productivity. Our results suggest that idiosyncratic shocks, measurement error, and heterogeneity in land quality are important sources of dispersion in measured productivity across farms. We find that when these are taken into account, the potential significance of misallocation drops substantially. Late-season production shocks, measurement error, and heterogeneity in inputs together account for as much as 70% of the variance in measured productivity.[6] 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.

An important caveat of our work is that we consider only the effects of static misallocation. Implicitly, this holds constant the existing institutions and technologies. With improved technologies and different institutions, one might expect that the efficient allocation of land and inputs across farms and farmers would look very different. In this sense, our results are not necessarily inconsistent with those of Adamopoulos and

[4] See, e.g., Beegle et al. (2012), Deininger et al. (2012), and De Nicola and Giné (2014), although Beegle, Carletto, and Himelein (2012) offer a more positive view.

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

[6] By "late-season" shocks, we mean those shocks that affect production after the farmer has made most or all of her input choices. We implicitly (and realistically) assume a production process in which significant amounts of labor and other inputs are applied early in the season for land clearing and planting and then additional inputs are applied during the growing season based on observed growing conditions, market prices, and so on. 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.

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

## II. Dispersion in Productivity across and within Farms

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

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

[7] For information on the LSMS-ISA project and links to the data, see https://www.worldbank.org/en/programs/lsms/initiatives/lsms-ISA.

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

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

Tanzania and Uganda differ to some degree in the types of production systems that we observe. Some crops are common to both countries (e.g., maize), while others (e.g., *matoke*, a kind of cooking banana) are of importance in only a single country (in this case, Uganda). For most purposes, however, the two countries are quite similar in the farming systems and production environment. Key points to note are that these are smallholder farming systems that use few inputs other than human labor and hand tools. Almost none of the farms in our data use irrigation or machinery; commercial fertilizer and other agrochemicals are each used on less than 10% of Tanzanian plots and 2% of Ugandan plots. In Uganda, most farmers cultivate crops in two growing seasons per year; in Tanzania, our data primarily reflect cultivation in the main growing season.[10]

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

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

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

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

TABLE 1
Agriculture in Tanzania and Uganda

| | Tanzania | Uganda |
|---|---|---|
| | A. Samples | |
| Sample size: | | |
| Households | 5,832 | 4,997 |
| Farmers | 7,302 | 10,364 |
| Plot-seasons | 21,000 | 52,585 |
| Seasons | 4 | 8 |
| Regions | 25 | 6 |
| Districts | 125 | 81 |
| Villages | 633 | 1,147 |
| Farmer-seasons | 9,821 | 21,793 |
| Farmer-crop-seasons | 17,089 | 48,343 |
| Size of clusters (median): | | |
| Farmer-seasons | 2 | 3 |
| Farmer-crop-seasons | 1 | 1 |
| | B. Yields | |
| Median plot size (ha) | .40 | .20 |
| Yield ($/ha): | | |
| Observations | 15,635 | 52,334 |
| Mean | 903 | 3,061 |
| Median | 458 | 182 |
| Standard deviation | 5,133 | 345,562 |
| Yield on maize plots ($/ha): | | |
| Observations | 7,370 | 10,232 |
| Mean | 812 | 519 |
| Median | 438 | 210 |
| Standard deviation | 6092 | 3244 |
| Yield on groundnut/beans plots ($/ha): | | |
| Observations | 988 | 10,760 |
| Mean | 754 | 6,892 |
| Median | 439 | 218 |
| Standard deviation | 1,035 | 656,396 |
| Yield on cassava (Tanzania) or banana (Uganda) plots ($/ha): | | |
| Observations | 1,172 | 6,083 |
| Mean | 689 | 565 |
| Median | 370 | 314 |
| Standard deviation | 974 | 1,317 |
| Labor (days/ha): | | |
| Observations | 18,416 | 52,334 |
| Mean | 225 | 258 |
| Median | 128 | 141 |
| Standard deviation | 697 | 1,217 |

Note.—All yields are winsorized at the 0.01 level.

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

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

### A. *Efficient Static Allocation*

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

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

with $\omega_h$ the TFP of farmer $h$.[13] The input choices on that plot maximize profits at shadow prices $(p_{Y_{hit}}, p_{L_{hit}}, p_{X_{hit}})$. In an efficient allocation, these shadow prices are common across all farmers. In familiar Cobb-Douglas fashion, factor demands and output are characterized by

$$\begin{aligned} x_{hit} - l_{hit} - (\ln(\alpha_X) - \ln(\alpha_L)) &= \ln\left(\frac{p_{L_{hit}}}{p_{X_{hit}}}\right) = \ln\left(\frac{p_L}{p_X}\right), \\ y_{hit} - l_{hit} + \ln(\alpha_L) &= \ln\left(\frac{p_{L_{hit}}}{p_{Y_{hit}}}\right) = \ln\left(\frac{p_L}{p_Y}\right), \end{aligned} \tag{2}$$

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

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

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

$$y_{hit} - y_{kit} = \frac{(\omega_h - \omega_k) + \alpha_L \ln(p_{Y_{hit}}/p_{L_{hit}}) + \alpha_X \ln(p_{Y_{hit}}/p_{X_{hit}})}{1 - \alpha_L - \alpha_X}$$
$$= \frac{(\omega_h - \omega_k) + \alpha_L \ln(p_Y/p_L) + \alpha_X \ln(p_Y/p_X)}{1 - \alpha_L - \alpha_X}.$$

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

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

Figure 1 shows, in two panels (for Tanzania and Uganda), 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.

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

The raw data on output and input do not account for variation in observable heterogeneity across plots. Land characteristics such as slope, soil type, and location affect farmers' optimal allocations of inputs and their expected yields. These land characteristics are measured in each of our data sets. Characteristics of the farmer, such as gender, education, and experience, are also components of productivity that we observe. Moreover, agriculture in each of our settings is almost exclusively rain fed. Rainfall thus affects plot productivity both by affecting the overall level of plot productivity and through unanticipated shocks to output. The data include measurements of rainfall totals at the community level;

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

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

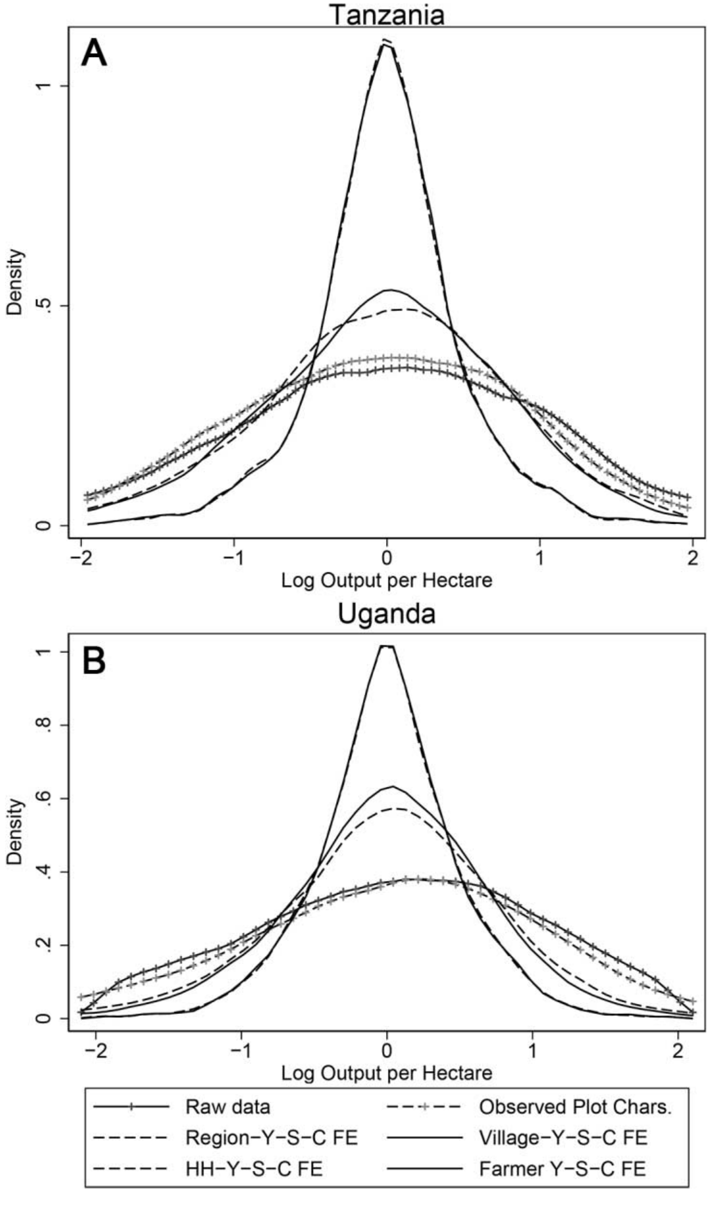


FIG. 1.—Dispersion of log output per hectare across plots. Plot Chars. = plot characteristics; HH = household; Y-S-C FE = year-season-crop fixed effects.

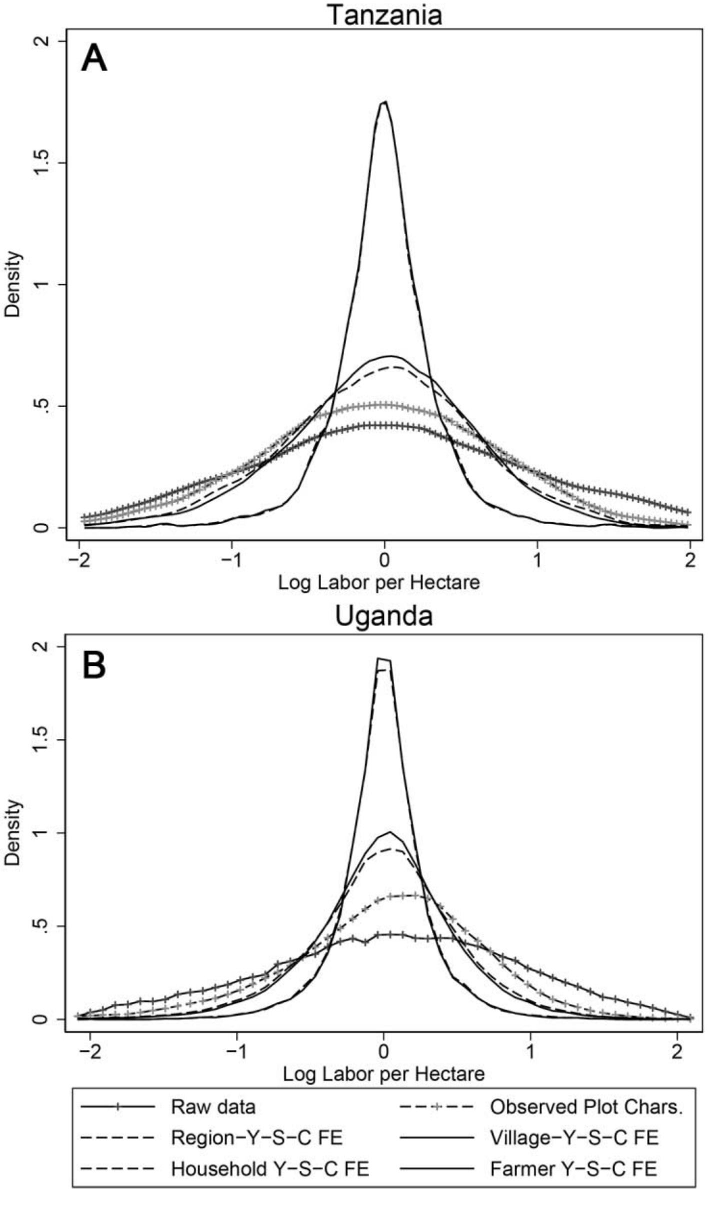


FIG. 2.—Dispersion of log labor per hectare across plots. Plot Chars. = plot characteristics; HH = household; Y-S-C FE = year-season-crop fixed effects.

we condition on measures of rainfall and their interactions with land characteristics as well. If these observed characteristics fully account for the variation in productivity, then in an efficient allocation output per hectare and labor per hectare would not vary across plots, once we control for observables.

To do so, land and labor inputs to production ($J \in \{L, X\}$) are modeled as the observed quantity of that factor ($J^{o}_{hit}$), observed as hectares or days of input $J$ on plot $i$ of household $h$ in season $t$, corrected for a factor-specific set of observables ($W_{J_{hit}}$) so that $J_{hit} = J^{o}_{hit} e^{W_{J_{hit}} \beta_J}$. (In essence, this is defining "effective labor units" and "effective land units" on the basis of observables.) Let $W_{Y_{hit}}$ be a set of observable determinants of TFP, including detailed measures of weather shocks realized after factor inputs are chosen, so that the production function now becomes

$$Y_{hit} = e^{W_{Y_{hit}} \beta_Y + \omega_{Y_h}} \left( L^{o}_{hit} e^{W_{L_{hit}} \beta_L} \right)^{\alpha_L} \left( X^{o}_{hit} e^{W_{X_{hit}} \beta_X} \right)^{\alpha_X}. \tag{1'}$$

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

$$\begin{aligned} x^{o}_{hit} + W_{X_{hit}} \beta_X - l^{o}_{hit} - W_{L_{hit}} \beta_L - (\ln(\alpha_X) - \ln(\alpha_L)) &= \ln\left(\frac{p_{L_{hit}}}{p_{X_{hit}}}\right) = \ln\left(\frac{p_L}{p_X}\right), \\ y_{hit} + W_{Y_{hit}} \beta_Y - l_{hit} - W_{L_{hit}} \beta_L + \ln(\alpha_L) &= \ln\left(\frac{p_{L_{hit}}}{p_{Y_{hit}}}\right) = \ln\left(\frac{p_L}{p_Y}\right), \end{aligned} \tag{2'}$$

$$\begin{aligned} y_{hit} - y_{kit} + (W_{Y_{hit}} - W_{Y_{kit}}) \beta_Y &= \frac{(\omega_h - \omega_k) + \alpha_L \ln(p_{Y_{hit}}/p_{L_{hit}}) + \alpha_X \ln(p_{Y_{hit}}/p_{X_{hit}})}{1 - \alpha_L - \alpha_X} \\ &= \frac{(\omega_h - \omega_k) + \alpha_L \ln(p_Y/p_L) + \alpha_X \ln(p_Y/p_X)}{1 - \alpha_L - \alpha_X}. \end{aligned}$$

Tables 2 and 3 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 data sets. The estimated density of the residuals from these regressions is illustrated as the curve labeled "Observed Plot Chars" in each of the panels 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.

TABLE 2
Log Output per Hectare in Tanzania

| | No Fixed Effects (1) | Year-Crop-Region (2) | Year-Crop-Village (3) | Year-Crop-Household (4) | Year-Crop-Farmer (5) |
|---|---|---|---|---|---|
| ln(ha) | −.49 | −.52 | −.48 | −.51 | −.5 |
| | (.009) | (.011) | (.012) | (.030) | (.031) |
| Female plot | −.16 | −.12 | −.15 | −.0014 | |
| | (.026) | (.028) | (.031) | (.200) | |
| Plot used free of charge | −.044 | −.021 | −.033 | .073 | .099 |
| | (.029) | (.032) | (.034) | (.095) | (.098) |
| Shared, rent | .015 | −.23 | −.02 | −.67 | −.62 |
| | (.140) | (.160) | (.160) | (.700) | (.710) |
| Shared, owned | .14 | .08 | .032 | −.13 | −.098 |
| | (.038) | (.048) | (.049) | (.140) | (.140) |
| Average quality | −.16 | −.17 | −.15 | −.054 | −.074 |
| | (.018) | (.020) | (.021) | (.070) | (.072) |
| Poor quality | −.28 | −.36 | −.35 | −.21 | −.21 |
| | (.038) | (.040) | (.046) | (.120) | (.120) |
| Loam | .17 | .071 | .14 | −.051 | −.063 |
| | (.024) | (.027) | (.029) | (.081) | (.084) |
| Clay | .2 | .09 | .16 | .012 | −.013 |
| | (.030) | (.033) | (.036) | (.095) | (.099) |
| Distance to market | .0023 | .0011 | .0011 | .0045 | .0046 |
| | (.001) | (.001) | (.001) | (.006) | (.006) |
| Irrigated | .46 | .17 | .23 | .69 | .49 |
| | (.061) | (.083) | (.085) | (.230) | (.240) |
| Erosion evident | −.018 | −.015 | −.044 | .026 | .022 |
| | (.026) | (.028) | (.031) | (.070) | (.072) |
| Sale value | −.047 | .021 | −.053 | .64 | .62 |
| | (.076) | (.087) | (.077) | (1.850) | (2.040) |
| Current morbidity ($z$) | −.043 | −.044 | −.035 | .22 | |
| | (.010) | (.010) | (.011) | (.140) | |
| Morbidity missing | 4.2 | 4.3 | 3.46 | −22.1 | |
| | (.950) | (1.020) | (1.140) | (14.000) | |
| $R^2$ | .29 | .52 | .53 | .71 | .71 |
| Log variance of residuals | 1.15 | .76 | .73 | .24 | .24 |
| $F$-statistic for plot characteristics | 122.8 | 70.8 | 56.1 | 10.5 | 10.3 |
| Corresponding $p$-value | 0 | 0 | 0 | 0 | 0 |

Note.—Standard errors are in parentheses. $R^2$ is from the OLS dummy variable specification. Variance of the dependent variable is 1.49. Regressions include 34 plot characteristics—e.g., soil type, toposequence, boundary markers, location, rainfall, and interactions with rain—and 14 household and farmer characteristics—e.g., housing, education, and age.

The first column of tables 4 and 5 reports the same subset of coefficients of the parallel regression of labor input per hectare on the observable exogenous shocks, farmer characteristics, and land characteristics. The estimated density of the residuals from these regressions is illustrated as the curve labeled “Observed Plot Chars” in each of the panels of figure 2.

TABLE 3
Log Output per Hectare in Uganda

| | No Fixed Effects (1) | Year-Crop-District (2) | Year-Crop-Village (3) | Year-Crop-Household (4) | Year-Crop-Farmer (5) |
|---|---|---|---|---|---|
| ln(ha) | −.56 | −.69 | −.7 | −.73 | −.73 |
| | (.006) | (.008) | (.008) | (.019) | (.020) |
| Male plot | .14 | .11 | .095 | .33 | |
| | (.024) | (.026) | (.029) | (.170) | |
| Leasehold plot | .0001 | .055 | .011 | −.10 | −.21 |
| | (.032) | (.036) | (.039) | (.110) | (.120) |
| Customary plot | −.024 | .04 | .012 | −.13 | −.17 |
| | (.045) | (.050) | (.052) | (.150) | (.150) |
| Mailo plot | −.060 | .022 | .011 | .0059 | .046 |
| | (.030) | (.036) | (.042) | (.16) | (.18) |
| Plot via occupancy | −.015 | .0030 | −.024 | −.030 | −.020 |
| | (.030) | (.035) | (.037) | (.110) | (.120) |
| Customary | −.30 | .0066 | .020 | −.052 | .029 |
| | (.016) | (.026) | (.028) | (.092) | (.099) |
| No document | .0046 | −.019 | −.029 | −.0037 | −.019 |
| | (.024) | (.027) | (.029) | (.093) | (.097) |
| Fair soil | −.033 | −.024 | −.015 | .0048 | .11 |
| | (.040) | (.044) | (.043) | (.12) | (.13) |
| Poor soil | −.23 | −.20 | −.16 | −.26 | −.52 |
| | (.097) | (.10) | (.10) | (.31) | (.36) |
| Sandy clay loam | .010 | −.0099 | −.0020 | −.015 | −.010 |
| | (.014) | (.016) | (.016) | (.048) | (.051) |
| Black clay | −.0042 | −.0001 | .020 | .075 | .088 |
| | (.018) | (.020) | (.022) | (.064) | (.067) |
| Fair soil × total rain | −.15 | −.15 | −.18 | .064 | .031 |
| | (.056) | (.067) | (.070) | (.200) | (.220) |
| Poor soil × total rain | .024 | .11 | .047 | −.64 | −.66 |
| | (.16) | (.17) | (.18) | (.49) | (.52) |
| Fair soil × flood duration | .016 | .042 | .033 | .19 | .25 |
| | (.029) | (.032) | (.033) | (.11) | (.15) |
| $R^2$ | .33 | .69 | .73 | .79 | .79 |
| Log variance of residuals | 1.42 | .66 | .52 | .26 | .26 |
| $F$-statistic for plot characteristics | 192.5 | 171.0 | 158.3 | 37.5 | 36.6 |
| Corresponding $p$-value | 0 | 0 | 0 | 0 | 0 |

Note.—Standard errors are in parentheses. $R^2$ is from the OLS dummy variable specification. Variance of the dependent variable is 1.82. Regressions include 37 plot characteristics, e.g., dummies for soil type, toposequence, boundary markers, and location; 11 interactions of plot characteristics with weather shocks; 12 household and farmer characteristics, e.g., education, age, and dummies for morbidity, and housing; 38 household-level self-reported shocks, e.g., drought, floods, job loss, and livestock disease; three community-level remote-sensing rainfall variables; and annual and seasonal rainfall totals.

Again, the set of observable characteristics is highly jointly significant in each regression, and the estimated variance of the residuals is significantly smaller than that of the raw data. The variation in observable characteristics, including shocks, is an important determinant of the variation

TABLE 4
LOG LABOR PER HECTARE IN TANZANIA

| | No Fixed Effects (1) | Year-Crop-Region (2) | Year-Crop-Village (3) | Year-Crop-Household (4) | Year-Crop-Farmer (5) |
|---|---|---|---|---|---|
| ln(ha) | −.63 | −.61 | −.61 | −.61 | −.6 |
| | (.006) | (.008) | (.008) | (.018) | (.019) |
| Female plot | −.12 | −.12 | −.12 | .12 | |
| | (.018) | (.019) | (.021) | (.110) | |
| Plot used free of charge | .066 | −.016 | .028 | .041 | .012 |
| | (.020) | (.022) | (.023) | (.055) | (.056) |
| Shared, rent | −.059 | −.018 | −.16 | −.96 | −.97 |
| | (.099) | (.110) | (.110) | (.470) | (.460) |
| Shared, owned | −.048 | −.022 | −.022 | −.018 | −.047 |
| | (.025) | (.032) | (.033) | (.083) | (.084) |
| Average quality | −.0053 | .0079 | .0071 | .043 | .038 |
| | (.013) | (.014) | (.015) | (.041) | (.042) |
| Poor quality | −.022 | −.0095 | −.029 | .079 | .1 |
| | (.025) | (.027) | (.031) | (.070) | (.072) |
| Loam | .011 | .025 | .037 | .027 | .03 |
| | (.016) | (.018) | (.020) | (.047) | (.048) |
| Clay | .053 | .059 | .074 | −.015 | −.0019 |
| | (.021) | (.023) | (.025) | (.057) | (.058) |
| Distance to market | −.00015 | −.00011 | −.00033 | .02 | .021 |
| | (.000) | (.001) | (.001) | (.004) | (.004) |
| Irrigated | −.075 | .018 | −.12 | .096 | .067 |
| | (.042) | (.059) | (.059) | (.140) | (.150) |
| Erosion evident | −.019 | −.066 | −.056 | −.19 | −.18 |
| | (.018) | (.020) | (.021) | (.042) | (.043) |
| Sale value | −.14 | −.19 | −.14 | .95 | 1.11 |
| | (.057) | (.062) | (.059) | (1.140) | (1.230) |
| Current morbidity ($z$) | −.0052 | −.0083 | −.011 | .14 | |
| | (.007) | (.007) | (.008) | (.078) | |
| Morbidity missing | .56 | .86 | 1.21 | −13.3 | |
| | (.650) | (.700) | (.780) | −7.76 | |
| $R^2$ | .34 | .46 | .46 | .61 | .61 |
| Log variance of residuals | .65 | .45 | .44 | .11 | .11 |
| $F$-statistic for plot characteristics | 333.8 | 202.7 | 181.4 | 37.6 | 37.6 |
| Corresponding $p$-value | 0 | 0 | 0 | 0 | 0 |

NOTE.—Standard errors are in parentheses. $R^2$ is from the OLS dummy variable specification. Variance of the dependent variable is 1.06. Regressions include 34 plot characteristics—e.g., soil type, toposequence, boundary markers, location, rainfall, and interactions with rain—and 14 household and farmer characteristics—e.g., housing, education, and age.

in both output and labor input per hectare across plots in each of these samples. But again, these observables do not generate much difference in the pattern of residuals, as shown in figure 2.

Differences in technology across farming systems and crops will presumably affect yield and input intensity, even in an efficient allocation. Similarly, variation over time in the shadow costs of factors of production

TABLE 5
Log Labor per Hectare in Uganda

| | No Fixed Effects (1) | Year-Crop-District (2) | Year-Crop-Village (3) | Year-Crop-Household (4) | Year-Crop-Farmer (5) |
|---|---|---|---|---|---|
| ln(ha) | −.72 | −.77 | −.78 | −.79 | −.79 |
| | (.004) | (.005) | (.006) | (.012) | (.012) |
| Male plot | −.014 | .012 | −.015 | −.077 | |
| | (.013) | (.016) | (.018) | (.096) | |
| Leasehold plot | −.088 | −.094 | −.11 | .065 | .15 |
| | (.019) | (.024) | (.027) | (.066) | (.075) |
| Customary plot | −.20 | −.17 | −.15 | −.13 | −.15 |
| | (.026) | (.032) | (.035) | (.087) | (.090) |
| Mailo Plot | −.10 | −.12 | −.081 | .21 | .13 |
| | (.017) | (.023) | (.028) | (.091) | (.10) |
| Plot via occupancy | −.14 | −.16 | −.12 | .0096 | .023 |
| | (.018) | (.023) | (.026) | (.068) | (.073) |
| Customary | −.12 | −.12 | −.12 | .15 | .18 |
| | (.0093) | (.017) | (.019) | (.055) | (.059) |
| No document | −.066 | −.041 | −.010 | .0027 | .0017 |
| | (.014) | (.018) | (.020) | (.056) | (.059) |
| Fair soil | .060 | .041 | .040 | −.19 | −.19 |
| | (.023) | (.028) | (.029) | (.071) | (.076) |
| Poor soil | −.023 | −.027 | −.013 | .11 | .15 |
| | (.057) | (.067) | (.068) | (.18) | (.22) |
| Sandy clay loam | .00054 | −.0090 | −.00055 | .023 | .033 |
| | (.0083) | (.010) | (.011) | (.028) | (.029) |
| Black clay | −.0040 | −.014 | −.0013 | .019 | .030 |
| | (.010) | (.013) | (.015) | (.036) | (.038) |
| Fair soil × total rain | −.028 | −.038 | −.048 | .039 | .13 |
| | (.033) | (.044) | (.048) | (.12) | (.13) |
| Poor soil × total rain | −.036 | −.10 | −.17 | −.43 | −.37 |
| | (.090) | (.11) | (.12) | (.29) | (.30) |
| Fair soil × flood duration | .019 | .048 | .033 | .075 | .063 |
| | (.016) | (.019) | (.021) | (.059) | (.076) |
| Poor soil × flood duration | .066 | .10 | .024 | | |
| | (.11) | (.14) | (.15) | | |
| $R^2$ | .41 | .51 | .53 | .60 | .60 |
| Log variance of residuals | .48 | .28 | .25 | .09 | .09 |
| $F$-statistic for plot characteristics | 719.7 | 487.0 | 402.7 | 110.7 | 107.7 |
| Corresponding $p$-value | 0 | 0 | 0 | 0 | 0 |

Note.—Standard errors are in parentheses. $R^2$ is from the OLS dummy variable specification. Variance of the dependent variable is 1.00. Regressions include 37 plot characteristics—e.g., dummies for soil type, toposequence, boundary markers, location; 11 interactions of plot characteristics with weather shocks; 12 household and farmer characteristics, e.g., education, age, and dummies for morbidity and housing; 38 household-level self-reported shocks, e.g., drought, floods, job loss, and livestock disease; three community-level remote-sensing rainfall variables; and annual and seasonal rainfall totals.

or the shadow value of output could generate time variation in output or labor per hectare. Therefore, column 2 in each of tables 2–5 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 shown as the curves labeled "Region-Y-S-C FE" in each of the panels of figures 1 and 2.

Qualitatively speaking, these tables provide evidence that observable characteristics of plots and shocks have a statistically meaningful effect on input intensity and yield. However, we note that, quantitatively speaking, these observables do not account for a very large fraction of the total dispersion.

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

It is apparent that substantial variation in output per hectare and labor per hectare remains after we account for a rich set of observable characteristics of land, including detailed measures of rainfall variation. The variance remains large even when we add year-season-crop-region fixed effects (col. 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 TFP (Hsieh and Klenow 2009; Adamopoulos et al. 2017; Restuccia and Santaeulalia-Llopis 2017). For this reason, we refer to the estimated residuals from the regressions based on equations (2′) and reported in column 2 of tables 2–5 as our baseline measures of dispersion in productivity across plots.[17]

The variation remains substantial as we move from the baseline specification to tighter specifications, adding fixed effects at progressively narrower geographic units. In tables 2–5, column 3 adds village fixed effects. The dispersion falls with successively narrower fixed effects, but it remains nontrivial.

[16] Claassen and Just (2011) report that for more than 500,000 observations in their US data, the 95th percentile corn yield is 190% higher than the 5th percentile yield, a difference they view as "quite wide" (148). By contrast, we find 95-5 ratios of 9,304% for Uganda and 2,558% for Tanzania. This reinforces our perception that the dispersion of yield across plots is quite high.

[17] An alternative baseline could be provided by examining the residuals from a similar regression with village-crop-year fixed effects. This would absorb the effects of unobserved village-level shocks that might otherwise be misinterpreted as misallocation, but it would also absorb any real misallocation of resources across villages. As can be seen in figs. 1 and 2, the estimated dispersion of the residuals from these two specifications is similar.

This baseline dispersion could be attributed to misallocation. However, even if factors of production are allocated with full efficiency, output shocks and measurement error generate output dispersion across farms. Consider three components of TFP. The first is the set of observable characteristics of the plot, farmer, or community, $W_{Y_{hit}}$, 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, $\omega_{Y_{hit}}$. Finally, there is a component that is unobserved in the data and unknown to the farmer at the time of input application, $\epsilon_{Y_{hit}}$. 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.)

$$\omega_{hit} = W_{Y_{hit}}\beta_Y + \omega_{Y_{hit}} + \epsilon_{Y_{hit}}. \tag{3}$$

Baseline dispersion in factor ratios could reflect unobserved characteristics of land or labor, or measurement error in either input. Land and labor inputs to production ($J \in \{L, X\}$) are modeled as the observed quantity of that factor ($J^o_{hit}$), 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 ($W_{J_{hit}}$) and subject to classical measurement error $\epsilon_{J_{hit}}$:

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

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

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

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

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

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

$$\begin{aligned} y_{hit} - y_{kit} + (W_{Y_{hit}} - W_{Y_{kit}})\beta_Y &= \frac{(\omega_h - \omega_k) + \alpha_L \ln(p_{Y_{hit}}/p_{L_{hit}}) + \alpha_X \ln(p_{Y_{hit}}/p_{X_{hit}})}{1 - \alpha_L - \alpha_X} \\ &\quad + \epsilon_{Y_{hit}} - \epsilon_{Y_{kit}} \\ &= \frac{(\omega_h - \omega_k) + \alpha_L \ln(p_Y/p_L) + \alpha_X \ln(p_Y/p_X)}{1 - \alpha_L - \alpha_X} + \epsilon_{Y_{hit}} \\ &\quad - \epsilon_{Y_{kit}}. \end{aligned} \tag{2''c}$$

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

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

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

If the allocation of factors across plots within a farm (during a single season) is efficient, then the dispersion of factor intensities and output-factor ratios across plots within a farm is generated by (1) imperfect measurement of factor inputs, (2) imperfect measurement of output, or (3) varying realizations of risk.[18]

The final two columns of tables 2–5 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

[18] It is of course possible that some farmers are systematically worse than others at allocating efficiently across plots. But it is not obvious that this should have a strong correlation with productivity levels. A bad farmer is arguably one who realizes equally poor yields across all plots, on the basis of allocating inputs with the same (improper) intensity across all plots.

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 panels of figures 1 and 2.[19]

In each country, when we consider either the yield regressions of tables 2 and 3 or the labor intensity regressions of tables 4 and 5, approximately one-fifth of the baseline dispersion reported in column 2 remains after we focus attention on variation in output per hectare across plots within a farm.

Given our assumption of efficient within-farm allocation, we conclude that this residual variation is evidence for significant heterogeneity or measurement error in factors of production or output. Alternatively, it could reflect unobserved shocks to output that do not affect the marginal product of factor inputs or that occur after the application of inputs to different plots within a farm. If the variance of these errors of measurement, or of shocks to output, is at least as large across farmers as it is across plots of a given farmer, then interpreting the residuals of the equations estimated in columns (2) of tables 2–5 as misallocation would result in an overestimate of the importance of misallocation.

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

### B. *Plot Size, Yield, and Factor Intensity*

We observe a strong and consistent negative relationship between output per hectare and plot size. While this pattern is reminiscent of the long-standing discussion of an inverse farm size–yield correlation, we find in the final column of tables 2–5 that this pattern holds across plots (planted with the same crop in the same season) within a farm. Across farms, factor market imperfections might explain an inverse relationship between land area and yield, but these market imperfections cannot explain this relationship across plots within a farm. Both Tanzania and Uganda exhibit fairly extreme negative relationships between log yield and log plot size within a farm; the estimated elasticity is $-0.73$ (SE $=$ 0.02) for Uganda and $-0.50$ (SE $=$ 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, Gourlay, and Winters 2015; Bevis and Barrett 2017; Carletto et al. 2017). One source of this estimated inverse relationship might be measurement error in plot size. Kilic et al. (2016) provide a careful

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

account of the role of this kind of measurement error, using these same Uganda and Tanzania data sets. They show that while measurement error does contribute to the estimated inverse plot size relationship, the relationship remains strong after using objective GPS measures of plot area and correcting for selection bias in the subset of plots measured with GPS. A number of other studies consider other reasons why yield might be negatively related to plot size, for example, Bevis and Barrett (2017). Most relevant for us, Gourlay, Kilic, and Lobell (2017) report the results of a methodological experiment, using different data from Uganda, that carefully examined the reporting of output data from farmers. Their findings suggest that self-reported yields are biased upward compared to measurement of crop cuts at harvest—and that this effect is stronger on smaller plots, perhaps fully explaining the observed relationship.

We find, however, that labor per hectare is also strongly declining in plot size within a farm (col. 5 of tables 4 and 5). The decline in both yield and labor per hectare with plot size suggests a different interpretation: namely, that smaller plots have higher unmeasured land quality.[20]

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

## III. Theoretical Framework

Our central argument is that heterogeneity in land quality and growing conditions plays an important role in explaining the dispersion of productivity at the level of plots and farms. This heterogeneity is unobservable to the econometrician but may be well recognized by farmers. Some of the unobservables involve intrinsic properties of the soil or land, such as the physical and chemical properties of the soil or the slope and topography. Other unobservables relate to highly localized shocks—such as hail that strikes one plot on a farm but spares another. Still others may involve complex interactions between shocks and plot characteristics: a

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

heavy early-season rain makes one low-lying plot unworkable at the start of the season because of mud, but the same rainstorm is actually beneficial for another plot that is well drained.

The importance of this kind of heterogeneity, often at a very fine-grained spatial level, is well documented in agronomic and economic studies.[21] Farmers can and do modify their practices to reflect heterogeneity of this kind; for example, 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 boundaries. This proves to be a good description of the realities of agriculture in our two countries. In the Uganda data, 62% of plots are subdivided from larger parcels of contiguous land controlled by an individual farmer, implying that these plots are literally or approximately adjacent.[22]

### A. *Agricultural Production with Continuous Variation in Land Quality*

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

At a location $k$, the quality of the land in effective units is denoted by $\zeta(k)$. For simplicity, assume that the function $\zeta(\cdot)$ is continuous and integrable. Land is used for producing an agricultural good. The production

[21] The importance of heterogeneity in agricultural systems at highly localized spatial scale has been shown previously in numerous contexts, e.g., by Hanna, Mullainathan, and Schwartzstein (2014). A data collection experiment conducted in Uganda (Lobell et al. 2018, 13) found that the yields on randomly chosen $8 \times 8$-m squares within typical maize plots correlated only weakly with the yields on the entire plots (which had a median size of 0.11 ha) because of "substantial intra-plot heterogeneity of yields in these systems." For African crop agriculture, see the work on agronomy by Tittonell et al. (2005, 2007, 2008), Vanlauwe, Tittonell, and Mukalama (2006), and Vanlauwe et al. (2015), along with papers in economics such as those by Suri (2011) and Tjernstrom, Carter, and Lybbert (2015). For US crop agriculture, Hurley, Malzer, and Kilian (2004) document high levels of agronomic heterogeneity within farmers' fields. This indeed is the premise for the emergence of "precision agriculture" technologies, as discussed by Stoorvogel, Kooistra, and Bouma (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.

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

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

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

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

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

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

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

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

### B. *Plot-Level Production with Continuous Variation in Land Quality*

Production could, in principle, be fine-tuned in this fashion to match the precise characteristics of each location, with inputs varying continuously across space. However, farming takes place at the level of a plot. A gardener nurtures each plant; a farmer has a goal of routinizing operations across the plot, transforming artisanal attention to each plant into systematic tasks that can be performed across the mass of plants growing together. Therefore, a plot is characterized by synchronicity of a sequence of tasks: land preparation, sowing, thinning, applying inputs, successive rounds of weeding, harvesting, processing, and so on. We define a plot as a set of contiguous locations that are managed uniformly.

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

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

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

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

Defining average land quality as $\zeta_{hi}(s) = (1/L_{hi})\int_0^{L_{hi}} \gamma_{hi}(k, s)\zeta_{hi}(k)\, dk$ and substituting into equation (9) gives the production function

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

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

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

Note that in equation (11), 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 farmer were maximizing at each location separately. Output will differ correspondingly. The lone exception is the case in which the fixed cost $c \to 0$, in which case $(\bar{k} - \underline{k}) \to 0$ and $Y^*_{hi} - \int_{\underline{k}}^{\bar{k}} q^*_h(k)\, dk \to 0$. With $c > 0$, the farmer chooses to divide the land into a finite number of plots. Section 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.

[23] We note that as a simple extension of the analysis, we can let the input vector $\xi$ 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.

### C. *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 section A1 that the maximum number of plots that could be cultivated profitably by a farmer depends inversely on the fixed cost and is also positively related to the average land quality across the farm. A farmer with very poor average land quality will ceteris paribus have a smaller maximum number of plots than a farmer with the same total land area but better-quality land. This does not necessarily give rise to an empirical prediction, because farms will not in general cultivate the maximum possible number of plots. But it does point to an underlying pattern that holds more generally: everything else equal, poor-quality plots must be sufficiently large that they will earn positive profits.

Consider the profit maximization for the $i$th plot cultivated by farmer $h$. The size of this plot is $\tilde{L}_{hi}$, with its boundaries at $L_{hi-1}$ and $L_{hi}$. The average productivity of this plot, conditional on the realization of the shock $\gamma_h(k, s)$, is $\zeta_{hi} = (1/\tilde{L}_{hi}) \int_{L_{hi-1}}^{L_{hi}} \gamma_h(k, s)\zeta_h(k)\, dk$. We can solve the profit maximization problem and then ask, for a given value of $\zeta_{hi}$, what is the smallest plot size that will yield nonnegative profits—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 $\zeta_{hi}$. The formulation of this is straightforward. Substituting the optimized value of $X^* = \tilde{L}_{hi}(\theta\zeta_{hi}/w_h)^{1/(1-\theta)}$ into the zero-profit condition and setting $\zeta_{hi}(s)\tilde{L}_{hi}^{1-\theta}(X_{hi})^{\theta} - w_h X_{hi} = c$, we get a relationship between the threshold plot size and land quality that will sustain nonnegative profits:

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

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

A farm of given area and heterogeneity with sufficiently high average land quality will be divided into more plots. Table 6 shows direct evidence of this pattern in Uganda, where we can identify the specific parcel of land to which each cultivated plot belongs. Labor intensity and output per acre over the entire parcel are higher on parcels that have been divided into more plots, as shown in column 1. In column 2, we condition on all of the measures of land quality, farmer characteristics,

TABLE 6
INTENSITY OF CULTIVATION AND PARCEL SUBDIVISION IN UGANDA

| | No Plot Characteristics (1) | No Fixed Effects (2) | Year-District (3) | Year-Village (4) | Year-Household (5) | Year-Farmer (6) |
|---|---|---|---|---|---|---|
| | A. Log Output per Hectare | | | | | |
| Log parcel area | −.60 | −.75 | −.75 | −.74 | −.80 | −.81 |
| | (.0068) | (.010) | (.011) | (.011) | (.017) | (.018) |
| No. of plots in parcel | .30 | .36 | .36 | .36 | .42 | .41 |
| | (.0066) | (.0075) | (.0078) | (.0080) | (.014) | (.015) |
| $R^2$ | .24 | .34 | .45 | .49 | .65 | .65 |
| Log variance of residuals | 1.34 | 1.20 | 1.01 | .93 | .57 | .55 |
| $F$-statistic for plot characteristics | | 23.4 | 13.4 | 12.3 | 8.22 | 7.18 |
| Corresponding $p$-value | | 0 | 0 | 0 | 0 | 0 |
| | B. Log Labor per Hectare | | | | | |
| Log parcel area | −.87 | −.83 | −.84 | −.83 | −.85 | −.86 |
| | (.0044) | (.0062) | (.0064) | (.0066) | (.0084) | (.0087) |
| No. of plots in parcel | .43 | .41 | .39 | .39 | .46 | .46 |
| | (.0043) | (.0046) | (.0047) | (.0048) | (.0067) | (.0072) |
| $R^2$ | .51 | .57 | .61 | .62 | .69 | .69 |
| Log variance of residuals | .59 | .45 | .37 | .34 | .14 | .14 |
| $F$-statistic for plot characteristics | | 17.0 | 11.4 | 10.6 | 11.4 | 9.97 |
| Corresponding $p$-value | | 0 | 0 | 0 | 0 | 0 |

NOTE.—Standard errors are in parentheses. $R^2$ is from the OLS dummy variable specification. Regressions include 37 plot characteristics, e.g., dummies for soil type, toposequence, boundary markers, location; 12 interactions of plot characteristics with weather shocks; 12 household and farmer characteristics, e.g., education, age, dummies for morbidity, housing; 38 household-level self-reported shocks, e.g., drought, floods, job loss, livestock disease; three community-level remote-sensing rainfall variables; and annual and seasonal rainfall totals.

and weather shocks used in tables 3 and 5; again we find that parcels that have been divided into more plots are cultivated with higher intensity and achieve greater yields. Columns 3–6 repeat this exercise with successively finer fixed effects, at the district-season, village-season, household-season, and farmer-season levels, and in each case we observe the same strong positive correlation between the number of plots into which a parcel has been divided and both input intensity and yield at the parcel level. This is consistent with the model that, conditional on parcel area and our rich set of observed characteristics of parcel land quality, there is important unobserved variation in land productivity across parcels, such that farmers subdivide higher-quality parcels into more plots.

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

[24] Our reason for abstracting from crop choice is not simply convenience. In our data, because of the ubiquity of intercropping, it is not sensible to think of farmers making a discrete crop choice. Instead, they are frequently choosing complex intercropped mixtures. In the Uganda data, we observe 428 distinct crops and crop mixes across farmers' plots; in Tanzania, the intercropping permutations give us an even higher number, with 853 distinct crops and crop mixes. Many of these crop mixtures are derived from variations on a much smaller set of distinct commodities. For instance, a farmer might grow maize by itself on one plot and a maize-cassava mix on another plot, with maize-beans-cassava on a third plot. In Tanzania, 43% of plots have some maize on them. In Uganda, 63% of plots are planted with one or more of maize, beans, cassava, and cooking bananas. Even beyond that, a maize-cassava intercrop could be 90% maize and 10% cassava, or the opposite. Given this complexity, we find it simpler to carry out our analysis in terms of a composite agricultural output, which then corresponds very closely to our model.

Accepting these slight departures from our model, we rewrite the production function (10), adding an index $t$ to denote the season and year in which production takes place, because we are 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 $\zeta_{hit} = (1/L_{hit})\int_0^{L_{hit}} \gamma_{hit}(k, s(t))\zeta_{hi}(k)\, dk$ incorporates the plot average of land quality at each point, denoted $\zeta_{hi}(k)$, and the plot average effect of shocks received in season $t$ at each point $\gamma_{hi}(k, s(t))$. Both land quality and the effects of shocks may vary across plots and across households in ways that we do not observe but that our model implies will potentially be correlated with plot size. To account for the possible dependence of average plot productivity on plot size, we parameterize productivity as $(1/L_{hit})\int_0^{L_{hit}} \gamma_{hi}(k, s(t))\zeta_{hi}(k)\, dk = e^{\omega_{hit}}(L_{hit})^{\varphi_{L_{hit}}}$.

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

$$\begin{aligned} Y_{hit} &= e^{\omega_{hit}}(L_{hit})^{\varphi_{L_{hit}}}(L_{hit})^{1-\theta}(X_{hit})^{\theta} \\ &= e^{\omega_{hit}}(L_{hit})^{\alpha_{L_{hit}}}(X_{hit})^{\alpha_{X_{hit}}}. \end{aligned} \tag{13}$$

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

$$\begin{aligned} y_{hit} = {} & \alpha_{L_{hit}} l_{hit}^{o} + \alpha_{X_{hit}} x_{hit}^{o} + W_{Y_{hit}}\beta_Y \\ & + \sum_{J\in\{L,X\}} \alpha_{J_{hit}}\left(W_{J_{hit}}\beta_J - \epsilon_{J_{hit}}\right) + \omega_{Y_{hit}} + \epsilon_{Y_{hit}}. \end{aligned} \tag{14}$$

The vector of observable determinants of TFP $(W_{Y_{hit}} = (W_{E_{hit}}, W_{H_{hit}}))$ 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 $W_{E_{hit}}$ the subset of those shock

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

indicators that occur before the early harvest season begins, sufficiently early that farmers may be able to adjust factor inputs in response. Similarly, $W_{H_{hit}}$ is the subset of those shock indicators that occur at harvest season, too late for farmers to adjust factor inputs in response.

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

### *E. Identification of the Production Function*

Our strategy for identification of the farm production function rests on the shadow price framework of Singh, Squire, and Strauss (1986) and the subsequent approach of Benjamin (1992). In our setting, we observe many firms (farms) producing essentially identical outputs in close geographic proximity, subject to numerous idiosyncratic production shocks. Because of spatial frictions and other market imperfections, the production shock of one farmer, on one plot, will affect shadow factor prices on that farmer's other plots and can feed through to the factor prices faced by other farmers in the same community. However, because output is undifferentiated and typically tradable beyond the community, output prices are not affected. These characteristics of the setting allow us to use observed shocks as instruments in a way that might not be appropriate for, say, manufacturing firms producing differentiated products in rich countries.

We use variation in input demand at the plot level, driven by variation across farmers in the shadow prices of inputs, to estimate the plot-level production function. Observed shocks to factor productivity—and hence factor demand—on one plot affect the shadow cost of factors on other plots of the same farmer and on the plots of other farmers in the same community. The substantive assumption we require is that, conditional on the observed shocks affecting a given plot $i$, the observed shocks affecting a different plot $j$ are not correlated with unobserved determinants of factor demand on plot $i$.

To be more precise, if factor markets are imperfect, then conditional on the realization of $W_{E_{hit}}$ on plot $i$, the realizations of $W_{E_{h,-i,t}}$ 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 intervillage 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, $W_{E_{h,-i,t}}$ and $W_{E_{-h,j,t}}$, along with more conventional measures of household wealth and demographics, constitute a set of potential instruments $G_{hit}$ for factor inputs on plot $i$.

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

### *F. 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 $\hat{\alpha}_L$ and $\hat{\alpha}_X$ of the expected values of the random factor productivities $\alpha_{L_{hit}}$ and $\alpha_{X_{hit}}$, a first approximation to the distribution of log TFP across plots might simply be the residual

$$\begin{aligned} \widehat{\ln \mathrm{TFP}}^{\mathrm{A}}_{hit} = {} & y_{hit} - \hat{\alpha}_L l^{\mathrm{o}}_{hit} - \hat{\alpha}_X x^{\mathrm{o}}_{hit} - W_{Y_{hit}}\hat{\beta}_Y \\ & - \hat{\alpha}_L W_{L_{hit}}\hat{\beta}_L - \hat{\alpha}_X W_{X_{hit}}\hat{\beta}_X. \end{aligned} \tag{15}$$

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

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

$$\begin{aligned}\widehat{\ln \mathrm{TFP}}^{\mathrm{A}}_{hit} = {} & \underbrace{\omega_{Y_{hit}}}_{\text{unobserved TFP}} + \underbrace{\sum_{J\in\{L,X\}} \left(\alpha_{J_{hit}} - \hat{\alpha}_J\right)\left(W_{J_{hit}}\hat{\beta}_J\right)}_{\substack{\text{unobserved productivity of}\\ \text{observed characteristics}}} \\ & + \underbrace{\sum_{J\in\{L,X\}} \left(\alpha_{J_{hit}} - \hat{\alpha}_J\right)\ln\left(J^{\mathrm{o}}_{hit}\right)}_{\substack{\text{unobserved productivity}\\ \text{of factors}}} \\ & - \underbrace{\sum_{J\in\{L,X\}} \left(\alpha_{J_{hit}} - \hat{\alpha}_J\right)\epsilon_{J_{hit}}}_{\substack{\text{unobserved productivity of}\\ \text{factor measurement error}}} - \underbrace{\sum_{J\in\{L,X\}} \hat{\alpha}_J\epsilon_{J_{hit}}}_{\substack{\text{factor}\\ \text{measurement error}}} \\ & + \underbrace{\epsilon_{Y_{hit}}}_{\substack{\text{postinput shocks and}\\ \text{measurement error in } y}}.\end{aligned} \tag{16}$$

The final three terms are sources of variation in measured productivity (i.e., $\widehat{\ln \mathrm{TFP}}^{\mathrm{A}}_{hit}$), but they do not give rise to actual productivity variation. The dispersion in measured productivity arising from these three terms cannot be remedied through reallocation. Reallocation cannot "solve" measurement error, nor can reallocation "solve" late-season idiosyncratic shocks that affect yield.

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

$$\begin{aligned}\mathrm{Var}\left(\widehat{\ln \mathrm{TFP}}^{\mathrm{A}}_{hit}\right) = {} & \mathrm{Var}\left(\omega_{Y_{hit}} + \sum_{J\in\{L,X\}} \left(\alpha_{J_{hit}} - \hat{\alpha}_J\right)\left(W_{J_{hit}}\hat{\beta}_J + \ln\left(J^{\mathrm{o}}_{hit}\right)\right)\right) \\ & + \hat{\alpha}^2_L\mathrm{Var}\left(\epsilon_{L_{hit}}\right) + \hat{\alpha}^2_X\mathrm{Var}\left(\epsilon_{x_{hit}}\right) + \mathrm{Var}\left(\epsilon_{Y_{hit}}\right).\end{aligned} \tag{17}$$

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

In what follows, we maintain the assumption that the allocation of factors within a farm is efficient, at least in the narrow sense that market failures do not affect farmers' ability to make allocative decisions across their plots. Uninsured risk, imperfect financial markets, labor market frictions, and missing markets for land all assuredly influence a farmer's overall choice set—but not their allocation decisions across plots. Within a farm, shadow factor prices are constant across plots.[26] Within-farm variation in observed output and observed inputs depends only on (1) unobserved dimensions of risk or measurement error in output ($\epsilon_{Y_{hit}}$), (2) measurement error in factor inputs ($\epsilon_{L_{hit}}, \epsilon_{X_{hit}}$), or (3) unobserved heterogeneity in factor-specific productivity ($\alpha_{L_{hit}}, \alpha_{X_{hit}}$) or TFP ($\omega_{Y_{hit}}$).

We show in section 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 TFP ($\sigma_z^2$), factor-specific productivity and their covariance ($\sigma_L^2$, $\sigma_X^2$, $\sigma_{LX}$), factor measurement error ($\sigma_{\epsilon L}^2$, $\sigma_{\epsilon X}^2$) and output measurement error and postinput risk ($\sigma_{\epsilon Y}^2$), as well as the covariance of plot-level TFP and factor-specific productivity ($\sigma_{zL}$, $\sigma_{zX}$).

We do not separately identify variation in all three types of unobserved heterogeneity in factor-specific productivity ($\omega_{L_{hit}}, \omega_{X_{hit}}$) and TFP ($\omega_{Y_{hit}}$): a parallel increase in $\omega_{L_{hit}}$ and $\omega_{X_{hit}}$ is equivalent to an increase in $\omega_{Y_{hit}}$. Hence, we normalize $\omega_{L_{hit}} + \omega_{X_{hit}} = 0$. Intuitively, a change in $\omega_{L_{hit}}$ relative to $\omega_{X_{hit}}$ is a change in the slope of an isoquant; a change in $\omega_{Y_{hit}}$ is a shift in or out of an isoquant. This normalization implies $\sigma_L^2 = \sigma_X^2$, $\sigma_{LX} = -\sigma_L^2$, and $\sigma_{zL} = -\sigma_{zX}$.

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

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 independently and identically distributed 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 measurement error, then we can expect (as shown in sec. A3) that our estimate of the variance of unexplained output attributable to late-season risk and measurement error in output and factor

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

inputs ($\hat{\sigma}^2_{\epsilon Y} + \hat{\alpha}^2_L \hat{\sigma}^2_{\epsilon L} + \hat{\alpha}^2_X \hat{\sigma}^2_{\epsilon X}$) is a lower-bound estimate of the variance in the final three terms of equation (15). Subtracting $\hat{\sigma}^2_{\epsilon Y} + \hat{\alpha}^2_L \hat{\sigma}^2_{\epsilon L} + \hat{\alpha}^2_X \hat{\sigma}^2_{\epsilon X}$ from $\mathrm{Var}(\ln \mathrm{TFP}^{\mathrm{A}}_{hit})$, therefore, provides an upper-bound estimate of $\mathrm{Var}(\omega_{Y_{hit}} + \Sigma_{J \in \{L,X\}} (\alpha_{J_{hit}} - \hat{\alpha}_J)(W_{J_{hit}} \hat{\beta}_J + \ln(J^{\mathrm{o}}_{hit})))$. 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, $\ln \mathrm{TFP}^{\mathrm{A}}_{hit}$, and shrinking it toward its mean $\mu^{\mathrm{A}}$ to account for the variances of the measurement errors and late-season shocks we can measure within the farm. Our revised estimate of unobserved productivity is

$$\widehat{\ln \mathrm{TFP}}^{\mathrm{B}}_{hit} = \mu^{\mathrm{A}} + \left(\widehat{\ln \mathrm{TFP}}^{\mathrm{A}}_{hit} - \mu^{\mathrm{A}}\right) \times \left(\frac{\mathrm{Var}\left(\widehat{\ln \mathrm{TFP}}^{\mathrm{A}}_{hit}\right) - \sigma^2_{\epsilon Y} - \sum_{J \in \{L,X\}} \alpha^2_J \sigma^2_{\epsilon J}}{\mathrm{Var}\left(\widehat{\ln \mathrm{TFP}}^{\mathrm{A}}_{hit}\right)}\right)^{1/2}. \tag{18}$$

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

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

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

## IV. Empirical Analysis

### A. *Estimation Procedure*

We estimate the agricultural production functions, the implied residuals $\ln \mathrm{TFP}^{\mathrm{A}}_{hit}$, the associated variances of unobserved heterogeneity and measurement error, and the revised estimates of unobserved productivity $\ln \mathrm{TFP}^{\mathrm{B}}_{hit}$ both by 2SLS (two-stage least squares) and by using Masten and Torgovitsky's (2016) IVCRC (instrumental variables correlated random coefficients) estimator.

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

Building on Florens et al. (2008), Masten and Torgovitsky (2016) suggest an IVCRC estimator, which allows for heterogeneity in the first-stage regressions of the instruments on the endogenous factor demands. The IVCRC estimator is a control function approach and therefore relies on an assumption of single-dimensional heterogeneity in the first-stage equations. As in any control function estimator, the idea is that it is possible to invert the first-stage factor demands to reveal the problematic heterogeneity. In the second-stage production function, factor inputs are orthogonal to the heterogeneity in productivity, conditional on this control function. The IVCRC estimator uses a sequence of quantile regressions in the first stage to generate conditional (on the instruments) ranks for each observation; in the second stage, OLS (ordinary least squares) estimation of equation (14) conditional on the estimated rank equal to $r$ identifies $E(\alpha_{L_{hit}}|\text{Rank} = r)$ and $E(\alpha_{X_{hit}}|\text{Rank} = r)$, and then averaging over ranks provides $E(\alpha_{L_{hit}})$ and $E(\alpha_{X_{hit}})$.

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

### *B. Estimation Results*

Tables 7 and 8 present OLS and quantile regression estimates of the determinants of land and labor inputs into production in Tanzania and

TABLE 7
OLS and Quantile Regression Determinants of Log Land and Log Labor Inputs in Tanzania

| | OLS | | 25th percentile | | 50th percentile | | 75th percentile | | Interquartile Range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Log Land (1) | Log Labor (2) | Log Land (3) | Log Labor (4) | Log Land (5) | Log Labor (6) | Log Land (7) | Log Labor (8) | Log Land (9) | Log Labor (10) |
| Male manager | .35 | .23 | .37 | .28 | .32 | .22 | .33 | .22 | −.034 | −.069 |
| | (.02) | (.02) | (.02) | (.03) | (.02) | (.02) | (.02) | (.02) | (.03) | (.03) |
| Land value | .21 | .1 | .19 | .11 | .2 | .1 | .24 | .1 | .044 | −.006 |
| | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) |
| Drought/flood × good soil | .096 | .024 | .14 | .011 | .045 | −.03 | .045 | .0023 | −.091 | −.009 |
| | (.03) | (.03) | (.04) | (.04) | (.04) | (.04) | (.04) | (.04) | (.04) | (.05) |
| Drought/flood × average (avg.) soil | .073 | .035 | .12 | −.004 | −.004 | .055 | .025 | .036 | −.1 | .04 |
| | (.03) | (.03) | (.04) | (.04) | (.04) | (.04) | (.04) | (.04) | (.04) | (.07) |
| Drought/flood × poor soil | −.027 | −.035 | −.014 | −.083 | −.031 | −.082 | −.006 | −.016 | .0083 | .067 |
| | (.08) | (.07) | (.08) | (.09) | (.09) | (.10) | (.07) | (.08) | (.10) | (.11) |
| Illness/accident of HH member | .035 | −.039 | .042 | −.075 | .056 | −.010 | .0073 | .018 | −.035 | .093 |
| | (.03) | (.03) | (.05) | (.04) | (.03) | (.03) | (.04) | (.04) | (.06) | (.04) |
| GS rain × good soil in HH (×1,000)[a] | −.038 | −.004 | .042 | .037 | −.045 | −.007 | −.059 | −.03 | −.1 | −.067 |
| | (.03) | (.03) | (.03) | (.04) | (.03) | (.03) | (.02) | (.02) | (.04) | (.03) |
| GS rain × avg. soil in HH (×1,000)[a] | −.006 | .017 | .078 | .032 | −.027 | .011 | −.025 | .015 | −.1 | −.017 |
| | (.03) | (.02) | (.03) | (.04) | (.03) | (.02) | (.02) | (.02) | (.04) | (.04) |
| GS rain × poor soil in HH (×1,000)[a] | −.061 | −.05 | −.093 | −.081 | −.11 | −.065 | −.019 | .028 | .075 | .11 |
| | (.05) | (.05) | (.06) | (.07) | (.04) | (.06) | (.04) | (.04) | (.09) | (.08) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Drought/flood × good soil in HH[a] | .0053 | .038 | .02 | .046 | .011 | .051 | .0067 | .032 | −.014 | −.014 |
| | (.02) | (.02) | (.02) | (.03) | (.02) | (.01) | (.02) | (.02) | (.02) | (.03) |
| Drought/flood × avg. soil in HH[a] | −.009 | .02 | −.029 | .04 | .019 | .026 | .0011 | .011 | .03 | −.029 |
| | (.02) | (.02) | (.02) | (.03) | (.02) | (.02) | (.02) | (.02) | (.03) | (.03) |
| Drought/flood × poor soil in HH[a] | .11 | .1 | .11 | .097 | .08 | .17 | .072 | .025 | −.042 | −.072 |
| | (.05) | (.04) | (.04) | (.05) | (.08) | (.05) | (.03) | (.04) | (.07) | (.05) |
| Adverse shock to HH plots[a] | −.089 | −.032 | −.11 | −.046 | −.081 | −.032 | −.061 | −.022 | .052 | .023 |
| | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) | (.02) | (.01) |
| Pseudo $R^2$ | .13 | .09 | .07 | .05 | .06 | .05 | .08 | .05 | | |
| *F*-statistics: | | | | | | | | | | |
| For joint significance of instruments | 17.1 | 11.5 | 13.7 | 7.32 | 24.4 | 11.3 | 19.4 | 11.2 | | |
| *p*-value | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| For $h_0$[b] | | | | | | | | | 14.8 | 16.0 |
| *p*-value | | | | | | | | | 0 | 0 |

NOTE.—Bootstrapped (500 samples) standard errors, clustered at the household level, are in parentheses. The full sets of coefficients are reported in table B1. These include the sale value of land; distances of the plot from home and from the nearest road; three levels of soil quality; four soil types; the gender, health status, literacy, and age of the plot manager; indicator variables for household (HH)-level drought or floods, crop disease or pests, severe water shortage or other shocks that led to crop loss, and growing-season (GS) rainfall, interacted with soil type and soil quality dummies.

[a] Variable serves as instrument in table 9.

[b] $h_0$: coefficients are equal for 25th and 75th percentiles.

TABLE 8
OLS and Quantile Regression Determinants of Log Land and Log Labor Inputs in Uganda

| | OLS | | 25th percentile | | 50th percentile | | 75th percentile | | Interquartile Range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Log Land (1) | Log Labor (2) | Log Land (3) | Log Labor (4) | Log Land (5) | Log Labor (6) | Log Land (7) | Log Labor (8) | Log Land (9) | Log Labor (10) |
| Male plot | .14 | .062 | .13 | .044 | .14 | .066 | .14 | .083 | .014 | .039 |
| | (.009) | (.0071) | (.010) | (.0092) | (.009) | (.008) | (.009) | (.0074) | (.010) | (.011) |
| Fair soil | −.091 | −.078 | −.10 | −.079 | −.14 | −.096 | −.099 | −.12 | −.0031 | −.036 |
| | (.024) | (.021) | (.031) | (.028) | (.027) | (.018) | (.023) | (.023) | (.038) | (.034) |
| Poor soil | −.20 | −.11 | −.18 | −.19 | −.33 | −.18 | −.13 | −.020 | .056 | .17 |
| | (.067) | (.062) | (.076) | (.048) | (.044) | (.066) | (.062) | (.044) | (.12) | (.074) |
| Fair soil × drought duration | .026 | .024 | .032 | .021 | .032 | .024 | .025 | .035 | −.0077 | .013 |
| | (.005) | (.005) | (.006) | (.006) | (.006) | (.005) | (.005) | (.005) | (.0059) | (.007) |
| Poor soil × drought duration | .036 | .027 | .033 | .026 | .039 | .024 | .038 | .0050 | .0048 | −.021 |
| | (.012) | (.0093) | (.015) | (.010) | (.012) | (.011) | (.011) | (.0049) | (.024) | (.011) |
| Illness incidence in HH[a] | .0016 | −.051 | −.0018 | −.066 | .013 | −.066 | −.0082 | −.064 | −.0064 | .0026 |
| | (.013) | (.0098) | (.015) | (.013) | (.014) | (.011) | (.012) | (.010) | (.018) | (.016) |
| HH fair soil × peak rain (×1,000)[a] | .082 | −.074 | .1 | −.074 | .047 | −.07 | .046 | −.031 | −.055 | .042 |
| | (.023) | (.019) | (.023) | (.025) | (.027) | (.020) | (.027) | (.017) | (.032) | (.025) |
| HH poor soil × peak rain (×1,000)[a] | .210 | −.053 | .220 | −.035 | .250 | −.120 | .230 | −.035 | .011 | .000 |
| | (.073) | (.053) | (.088) | (.087) | (.064) | (.081) | (.055) | (.056) | (.092) | (.086) |
| HH fair soil × drought duration[a] | −.005 | −.007 | −.007 | −.008 | −.006 | −.008 | −.007 | −.008 | −.0004 | −.0007 |
| | (.002) | (.002) | (.002) | (.002) | (.002) | (.001) | (.002) | (.002) | (.003) | (.002) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| HH fair soil × total rain[a] | −.048 | .017 | −.053 | .018 | −.045 | .012 | −.042 | −.0003 | .011 | −.018 |
| | (.008) | (.007) | (.008) | (.009) | (.010) | (.007) | (.010) | (.006) | (.013) | (.008) |
| HH poor soil × total rain[a] | −.083 | .023 | −.069 | .018 | −.097 | .042 | −.093 | −.0036 | −.023 | −.022 |
| | (.025) | (.020) | (.032) | (.031) | (.022) | (.029) | (.018) | (.019) | (.029) | (.028) |
| Shocks on HH plots[a] | .021 | −.0005 | .017 | −.0012 | .062 | .0005 | .10 | −.0013 | .085 | −.0001 |
| | (.003) | (.001) | (.001) | (.003) | (.007) | (.001) | (.009) | (.001) | (.008) | (.003) |
| Pseudo $R^2$ | .08 | .02 | .038 | .013 | .048 | .013 | .052 | .015 | | |
| *F*-statistics: | | | | | | | | | | |
| For joint significance of instruments | 37.1 | 14.7 | 33.0 | 12.6 | 33.5 | 14.8 | 33.6 | 12.1 | | |
| *p*-value | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | | |
| For $h_0$[b] | | | | | | | | | 65.2 | 57.1 |
| *p*-value | | | | | | | | | .00 | .00 |

Note.—Bootstrapped (500 samples) standard errors, clustered at the household level, in parentheses. The full sets of coefficients are reported in table B2. These include four categories of soil quality; three sources of water; six indicators of plot toposequence; the level of erosion; indicators of the gender, literacy, and access to agricultural advice of the plot manager; household (HH)-level indicators of drought and flood and their interactions with plot level soil quality and village level seasonal and annual rain; and their interactions with plot-level soil quality.

[a] Variable serves as instrument in table 9.

[b] $h_0$: coefficients are equal for 25th and 75th percentiles.

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

The penultimate row of tables 7 and 8 shows that the instruments are strong predictors of plot-level land and labor demand. For example, in both Tanzania and Uganda, illness in the household reduces labor use in household plots. In Tanzania, when there is abundant growing-season rain, farmers reduce labor on a given plot when a higher fraction of that household's other plots have good soil. In Uganda, when peak-season rainfall is higher, farmers reduce labor on a given plot when a higher fraction of that household's other plots have fair or poor soil.

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

If the production function has random coefficients, then the associated input demand functions will have heterogeneous coefficients as well. The IVCRC estimator was developed to allow for heterogeneity in the first-stage regressions. The final pair of columns reports the difference in coefficients at the 75th and 25th percentiles of the factor demand quantile regressions and the $F$-test that these differences are jointly zero for the instrumental variables. We strongly reject that these differences are zero. For example, in Tanzania, at the 25th percentile of the demand for land, the effect of more growing-season rain when the household has more plots with good soil is positive, while at the 75th percentile, the effect is negative. Nevertheless, we present 2SLS estimates of the production function and their implications for estimates of the dispersion of unobserved productivities for comparison.

The 2SLS and IVCRC estimates of $E(\alpha_{L_{hit}})$ and $E(\alpha_{X_{hit}})$ in the production function for agricultural plots in Tanzania and Uganda are presented in panel A of table 9.

Columns 1 and 3 present 2SLS estimates of the Cobb-Douglas factor coefficients for Tanzania and Uganda, respectively; columns 2 and 4 provide the corresponding IVCRC estimates, with the log total value of crop output as the dependent variable. Crop-year-season-region fixed effects are included, as are a rich set of observable characteristics of land and

[27] Bootstrapped standard errors, clustered at the household-season level, are reported. The respective full sets of coefficient estimates are presented in tables B1 and B2.

TABLE 9
PRODUCTION FUNCTION AND VARIANCE COMPONENTS

| | Tanzania | | Uganda | |
|---|---|---|---|---|
| | 2SLS | IVCRC | 2SLS | IVCRC |
| | A. Cobb-Douglas Factor Coefficients | | | |
| Land | .73 | .61 | .69 | .53 |
| | (.17) | (.01) | (.05) | (.00) |
| Labor | .28 | .26 | .22 | .43 |
| | (.23) | (.02) | (.12) | (.01) |
| | B. Implied Plot-Level Variances of Productivity, Risk, and Measurement Errors | | | |
| Plot TFP | .35 | .38 | .18 | .19 |
| | (.01) | (.02) | (.00) | (.01) |
| Land/labor productivity | .07 | .10 | .07 | .05 |
| | (.01) | (.02) | (.01) | (.01) |
| Late-season risk and output measurement error | .65 | .70 | .67 | .84 |
| | (.02) | (.02) | (.01) | (.01) |
| Land measurement error | .08 | .03 | .13 | .16 |
| | (.02) | (.02) | (.01) | (.01) |
| Labor measurement error | .27 | .32 | .18 | .39 |
| | (.02) | (.02) | (.01) | (.01) |
| Covariance of TFP and land/labor productivity | .06 | .09 | .05 | .09 |
| | (.01) | (.02) | (.00) | (.01) |
| Observations | 14,535 | 14,535 | 43,187 | 43,187 |

NOTE.—Bootstrapped (500 samples) standard errors, clustered at the household level, are in parentheses. Full production function results shown in tables A1 and A2.

labor, as well as plot-level observable shocks.[28] Reflecting the simple technology of Tanzanian and Ugandan smallholder agriculture, these coefficients imply a much larger share of income for land than is observed in typical macroeconomic data and a much smaller share for labor.[29] The preferred IVCRC estimates imply a larger share for labor and a smaller share for land than the 2SLS estimates.

With estimates $\hat{\alpha}_L = E(\alpha_{L_{hit}})$, $\hat{\alpha}_X = E(\alpha_{X_{hit}})$, $\hat{\beta}_Y$, $\hat{\beta}_L$, and $\hat{\beta}_X$ in hand, we generate a first approximation, $\widehat{\ln \text{TFP}}^{\text{A}}_{hit}$, to the distribution of log TFP

[28] Bootstrapped standard errors are reported. The full sets of coefficients are reported in tables A1 and A2. For Tanzania, these include the sale value of land; distances of the plot from home and from the nearest road; three levels of soil quality; four soil types; the gender, health status, literacy, and age of the plot manager; indicator variables for household-level drought or floods, crop disease or pests, severe water shortage, or other shocks that led to crop loss; and growing-season rainfall, interacted with soil type and soil quality dummies. For Uganda, they include four categories of soil quality; three sources of water; six indicators of plot toposequence; the level of erosion; indicators of the gender, literacy, and access to agricultural advice of the plot manager; household-level indicators of drought and flood and their interactions with plot-level soil quality; and village-level seasonal and annual rain and their interactions with plot-level soil quality.

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

across plots. Our measure of TFP is the usual production function residual, as in equation (13). Figure 3 provides the empirical distribution of $\widehat{\ln \mathrm{TFP}}^{A}_{hit}$ in Tanzania and Uganda, using both the IVCRC and 2SLS estimates. The apparent dispersion is high: in Tanzania (Uganda) the

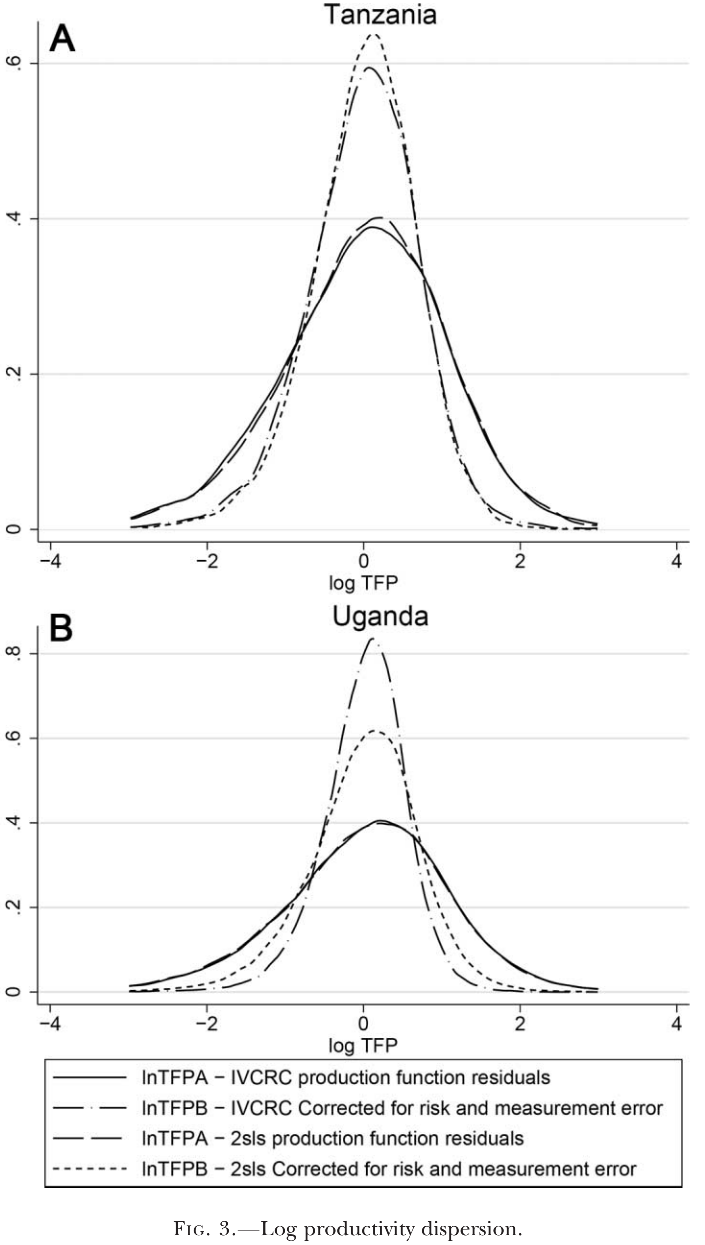


FIG. 3.—Log productivity dispersion.

variance based on the IVCRC estimates is 1.29 (1.25); the 90-10 log difference in TFP is 2.72 (2.67), corresponding to a 90-10 ratio of 15 (14) in TFP levels.[30]

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

Maintaining the assumption of within-farm efficient factor allocation, table 9 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 ($\sigma^2_z$) is also important, though less so in Uganda, where, as noted above, most plots of a given farmer are spatially contiguous or nearly so. Across-plot, within-farm variation in factor-specific productivity also exists but is relatively less important. The results are similar whether they are based on production function parameters estimated with 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 9, panel 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 TFP based on the assumption that the overall variances of $\sum_{J \in \{L,X\}} \alpha_J \epsilon_{J_{hit}}$ and $\epsilon_{Y_{hit}}$ across all farms are no smaller than these within-farm variances. The variance of $\widehat{\ln \mathrm{TFP}}^{\mathrm{B}}_{hit}$ is an upper bound on the dispersion of unobserved TFP across plots more generally.[31]

In both Tanzania and Uganda, accounting for measurement error in factors of production and output and for late-season shocks dramatically reduces the apparent dispersion of TFP across plots. In Tanzania, the variance of the naïve log residual of the production function is 1.29.

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

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

The variance of the estimate corrected for measurement error and late-season risk falls to 0.55. The 90-10 log difference in estimated TFP falls from 2.72 to 1.78, corresponding to a drop in the 90-10 ratio of TFP from 14.8 to 5.9. In Uganda, the correction is even more dramatic. The variance of the log production function residual is 1.25, falling to 0.29 when corrected. The 90-10 log difference in estimated TFP falls from 2.67 to 1.29, corresponding to a correction in the 90-10 ratio of TFPs from 14.4 to 3.6.[32]

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

We have examined three within-farmer frictions that could restrict the efficient reallocation of factors across plots within farms. First, a farmer cultivating widely dispersed plots could face varying input costs across these plots. When we repeat the estimation procedure restricting attention to close-by plots, however, our results change little, as reported in section A5.[33] The IVCRC estimate of the dispersion of $\widehat{\ln \mathrm{TFP}}^{B}_{hit}$ is actually larger when attention is restricted to nearby plots. Second, land tenure restrictions across physical space could interfere with the smooth reallocation of land across plots of a farmer. We repeat the procedure, using only plots contained within a single parcel of each farmer, and again find very similar results. Finally, some of the "farmers" are instead particular pairs of individuals farming jointly; there may be inefficiencies in the allocation of resources in such circumstances. The results, however, are robust to restricting attention to single-person "farmers," except again that the IVCRC estimate of the dispersion of $\widehat{\ln \mathrm{TFP}}^{B}_{hit}$ increases.

### C. *Implications for Characterizing Misallocation*

Late-season production shocks and measurement error in factors of production and output together account for about half to two-thirds of the variance in log productivity residuals. Since these are not susceptible to reallocation, the aggregate productivity gains that could be attained

[32] The corrections are similarly dramatic using 2SLS estimates of the production function. Using these estimates in Tanzania, the variance of unobserved log TFP falls from 1.27 to 0.47, and the 90-10 difference falls from 2.66 to 1.68, corresponding to a fall in the 90-10 ratio of TFP from 14.4 to 5.3. Using the 2SLS estimates in Uganda, the variance of unobserved log TFP falls from 1.27 to 0.53. The 90-10 log difference falls from 2.69 to 1.74, corresponding to a fall in the 90-10 ratio of TFP from 14.8 to 5.7.

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

from a hypothetical reallocation exercise are correspondingly smaller. The effect of our adjustments on estimates of the magnitude of misallocation in an economy depend, of course, on the specifics of the reallocation exercise. Any reallocation exercise must impose a great deal of structure on what is ultimately an artificial exercise. For example, results will be sensitive to whether land alone is reallocated to the best farmers or whether labor is allowed to move along with land. If land is to be reallocated, is it limited to within-village or within-region reallocation? The exact magnitude of the reallocation gains will depend on the extent to which misallocation gives rise to systematic rank reversals, as shown in Hopenhayn (2014). Any of these calculations can be carried out with our estimates in hand, but there is no real discipline on the exercise from either theory or practice. As an alternative approach, in what follows, we have preferred to impose less structure and simply to focus on the dispersion of productivity.

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

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

where $\omega_i$ is TFP, known to the producer, and $\epsilon_i$ is measurement error in output or output risk that is realized after factors are committed.

Relative to an existing baseline allocation, the gains to efficient reallocation are proportional to $\sigma_\omega^2 / e^{2(1-\alpha_L-\alpha_X)^2}$, as shown in section A6, and therefore depend on the dispersion of TFP and the concavity of the production function. An overestimate of the variance of TFP across producers, caused, for example, by misinterpreting measurement error or pure risk as variation in TFP, leads directly to an overestimate of the potential gain from reallocation.

Let $\sigma_A^2 = \mathrm{Var}(\widehat{\ln \mathrm{TFP}_{hit}^{A}}) > \mathrm{Var}(\widehat{\ln \mathrm{TFP}_{hit}^{B}}) = \sigma_B^2$. The overstatement of the potential gains from reallocating resources from an existing baseline to the efficient allocation from assuming that the variance of TFP is $\sigma_A^2$ rather than $\sigma_B^2$ is

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

In Tanzania, this ratio is 2.1. In Uganda, the overstatement of the gains to correcting misallocation is 2.6. The extent of misallocation is substantially overstated if the contributions of risk and measurement error to the apparent dispersion of TFP 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 TFP 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 equation (21) remains unchanged.

## V. Discussion

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

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

There are limits to our analysis. As noted in section I, 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 issues different from those we have addressed here.

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

This paper also suggests that within the literature on agriculture and development, there is a need to pay close attention to heterogeneity in unobservable characteristics of plots. These may be linked to soil and land quality, which vary in quantitatively significant ways at very fine geographic scale. But there may also be a high degree of spatial variation in shadow prices (reflecting, e.g., 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.

## VI. 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 artifact 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 that commonly used approaches in the literature overstate the dispersion of log TFP by about 100%. The gains from a hypothetical reallocation are thus correspondingly overstated by a factor of two or three. On the basis of our estimates, reallocation can generate nontrivial 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, too, for researchers to consider other sources of productivity dispersion, including heterogeneity and measurement error.

## Appendix A

## Model Detail and Extensions

### *A1. Endogenous Plot Selection*

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

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

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

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 modeling 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 $L_{h1}$, so that the two plots are $[0, L_{h1}]$ and $[L_{h1}, L_h]$. In this case, an interior solution for the size of the two plots must hold; expected total profits could not be increased by moving this location either to the left or to the right on the number line.

The profit maximization problem can be written as

$$\max_{L_{h1}} \int_{s \in S} \left[ \max_{X_{h1}, X_{h2}} \left( \left( \frac{X_{h1}}{L_{h1}} \right)^{\theta} \int_{0}^{L_{h1}} \gamma_h(k, s) \zeta_h(k) \, dk \right. \right.$$
$$+ \left( \frac{X_{h1}}{L_h - L_{h1}} \right)^{\theta} \int_{L_{h1}}^{L_h} \gamma_h(k, s) \zeta_h(k) \, dk \tag{A2}$$
$$\left. \left. - \; w_h X_{h1} - w_h X_{h2} - 2c \right) \right] d\Delta(s).$$

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

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

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

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

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

$$E\hat{\pi}(I) = \max_{\{L_{hi}\}_{i=1}^{I}} \int_{s \in S} s \left[ \max_{\{X_{hi}\}_{i=1}^{I}} \left( \zeta_{hi} \tilde{L}_{ij} X_{hi}^{\theta} - \sum_{i=1}^{I} w_h X_{hi} \tilde{L}_{hi} - cI \right) \right] d\Delta(s). \tag{A3}$$

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 $q_h^*(k, s) = \zeta_h(k) \gamma_h(k, s) (\theta \gamma_h(k, s) \zeta_h(k)/w_h)^{\theta/(1-\theta)}$, with corresponding profits of $\pi_h^*(k, s) = q_h^*(k, s) - w_h \xi_h^*(k, s)$. Across the entire landholding 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 landholdings 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, $I^* = (\pi_h^*/c) + 1$ 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 $\hat{I} = \operatorname{argmax}_j\{E\hat{\pi}(j)\}_{j=1}^{I^*}$.

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 $\zeta_{hi}^{\mathrm{A}} \neq \zeta_{hi}^{\mathrm{B}}$, 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 $\pi_{1i} = \zeta_{hi}(\zeta_{hi}\theta/w_h)^{\theta/(1-\theta)} - w_h(\zeta_{hi}\theta/w_h)^{1/(1-\theta)}$ as the profit from farming the parcel as a unit. Let $L_{hi}^{\mathrm{A}}$ and $L_{hi}^{\mathrm{B}} = 1 - L_{hi}^{\mathrm{A}}$ be the areas of the two plots that optimally divide parcel $i$ (the solution to [A2]). So $\pi_{1i}^{\mathrm{A}} = L_{hi}^{\mathrm{A}}\zeta_{hi}^{\mathrm{A}}(\zeta_{hi}^{\mathrm{A}}\theta/w_h)^{\theta/(1-\theta)} - w_h L_{hi}^{\mathrm{A}}(\zeta_{hi}^{\mathrm{A}}\theta/w_h)^{1/(1-\theta)}$ and $\pi_{1i}^{\mathrm{B}} = L_{hi}^{\mathrm{B}}\zeta_{hi}^{\mathrm{B}}(\zeta_{hi}^{\mathrm{B}}\theta/w_h)^{\theta/(1-\theta)} - w_h L_{hi}^{\mathrm{B}}(\zeta_{hi}^{\mathrm{B}}\theta/w_h)^{1/(1-\theta)}$. 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 $\pi_{zi}$, $\pi_{zi}^{\mathrm{A}}$, and $\pi_{zi}^{\mathrm{B}}$ as the profits from farming the full parcel and the optimally divided plots if the productivity process is $\zeta_{zi}(k)$. Finally, define $\tilde{\pi}_{zi}^{\mathrm{A}} = L_{hi}^{\mathrm{A}}\zeta_{zi}^{\mathrm{A}}(\zeta_{zi}^{\mathrm{A}}\theta/w_h)^{\theta/(1-\theta)} - w_h L_{hi}^{\mathrm{A}}(\zeta_{zi}^{\mathrm{A}}\theta/w_h)^{1/(1-\theta)}$ and similarly $\tilde{\pi}_{zi}^{\mathrm{B}}$ 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.

By construction, $\zeta_{hi}^{\mathrm{A}} \neq \zeta_{hi}^{\mathrm{B}}$, so that $X_{1i}^{\mathrm{A}}/L_{1i}^{\mathrm{A}} \neq X_{1i}^{\mathrm{B}}/L_{1i}^{\mathrm{B}}$, so for $\theta < 1$,

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

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

$$\begin{aligned} \pi_{zi}^{\mathrm{A}} + \pi_{zi}^{\mathrm{B}} - c \geq \tilde{\pi}_{zi}^{\mathrm{A}} + \tilde{\pi}_{zi}^{\mathrm{B}} - c &= z^{1/(1-\theta)}(\pi_{1i}^{\mathrm{A}} + \pi_{1i}^{\mathrm{B}}) - c \\ &> z^{1/(1-\theta)}\pi_{1i} = \pi_{zi}. \end{aligned} \tag{A4}$$

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 $z$, $\pi_{zi}^{\mathrm{A}} + \pi_{zi}^{\mathrm{B}} < c$, and it is not feasible to divide the parcel.

### 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

$$z_{hit} \equiv \frac{1}{1 - \sum_{J\in\{L,X\}}\alpha_{J_{hit}}}\left(W_{E_{hit}}\beta_E + \omega_{Y_{hit}} + \alpha_{L_{hit}}\ln\left(\frac{\alpha_{L_{hit}}}{p_{L_{ht}}}\right) + \alpha_{X_{hit}}\ln\left(\frac{\alpha_{X_{hit}}}{p_{X_{ht}}}\right)\right). \tag{A5}$$

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

$$\begin{aligned} y_{hit} &= W_{H_{hit}}\beta_H + \epsilon_{Y_{hit}} + z_{hit}, \\ l_{hit} &= \ln(\alpha_{L_{hit}}) - \ln(p_{L_{ht}}) + z_{hit}, \\ x_{hit} &= \ln(\alpha_{X_{hit}}) - \ln(p_{X_{ht}}) + z_{hit}. \end{aligned} \tag{A6}$$

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 work in terms of observable inputs, and output, adjusted for the estimated effects of observed characteristics

$$\begin{aligned} y_{hit} - W_{H_{hit}}\hat{\beta}_H &= \epsilon_{Y_{hit}} + z_{hit}, \\ l^{o}_{hit} + W_{L_{hit}}\hat{\beta}_L &= \hat{\alpha}_L + \omega_{L_{hit}} - \ln(p_{L_{ht}}) + \epsilon_{L_{hit}} + z_{hit}, \\ x^{o}_{hit} + W_{X_{hit}}\hat{\beta}_X &= \hat{\alpha}_X + \omega_{X_{hit}} - \ln(p_{X_{ht}}) + \epsilon_{X_{hit}} + z_{hit}; \end{aligned} \tag{A7}$$

$\omega_{L_{hit}}$ and $\omega_{X_{hit}}$ are plot-level productivities of land and labor, respectively, and $z_{hit}$ is plot-level total productivity. We examine deviations of log output, log land, and log labor from their within-farmer-season averages (e.g., $\bar{y}_{h.t}$ is mean log output over the plots cultivated by farmer $h$ in season $t$).

$$\begin{aligned} \tilde{y}_{hit} &\equiv y_{hit} - \bar{y}_{h.t} - (W_{H_{hit}} - \bar{W}_{H_{h.t}})\hat{\beta}_H = \epsilon_{Y_{hit}} - \bar{\epsilon}_{Y_{h.t}} + z_{hit} - \bar{z}_{h.t}, \\ \tilde{l}_{hit} &\equiv l^{o}_{hit} - \bar{l}^{o}_{h.t} + (W_{L_{hit}} - \bar{W}_{L_{h.t}})\hat{\beta}_L = \omega_{L_{hit}} - \bar{\omega}_{L_{h.t}} + \epsilon_{L_{hit}} - \bar{\epsilon}_{L_{h.t}} + z_{hit} - \bar{z}_{h.t}, \\ \tilde{x}_{hit} &\equiv x^{o}_{hit} - \bar{x}^{o}_{h.t} + (W_{X_{hit}} - \bar{W}_{X_{h.t}})\hat{\beta}_X = \omega_{X_{hit}} - \bar{\omega}_{X_{h.t}} + \epsilon_{X_{hit}} - \bar{\epsilon}_{X_{h.t}} + z_{hit} - \bar{z}_{h.t}. \end{aligned} \tag{A8}$$

The left-hand sides of these are observable. Their covariances (and a normalization discussed below) provide us with sufficient information to identify the within-farm variances of plot-level TFP ($\sigma^2_z$), factor-specific productivity and their covariance ($\sigma^2_L$, $\sigma^2_X$, $\sigma_{LX}$), factor measurement error ($\sigma^2_{\epsilon L}$, $\sigma^2_{\epsilon X}$), and output measurement error and postinput risk ($\sigma^2_{\epsilon y}$), as well as the covariance of plot-level TFP and factor-specific productivity ($\sigma_{zL}$, $\sigma_{zX}$):

$$\begin{aligned} \mathrm{Var}(\tilde{y}_{hit}) &= \sigma^2_z + \sigma^2_{\epsilon y_{hit}}, \\ \mathrm{Var}\big(\tilde{l}_{hit}\big) &= \sigma^2_L + \sigma^2_{\epsilon L} + \sigma^2_z + 2\sigma_{zL}, \\ \mathrm{Var}(\tilde{x}_{hit}) &= \sigma^2_X + \sigma^2_{\epsilon X} + \sigma^2_Q + 2\sigma_{zX}, \\ \mathrm{Cov}\big(\tilde{y}_{hit}, \tilde{l}_{hit}\big) &= \sigma_{zL} + \sigma^2_z, \\ \mathrm{Cov}(\tilde{y}_{hit}, \tilde{x}_{hit}) &= \sigma_{zX} + \sigma^2_z, \\ \mathrm{Cov}\big(\tilde{l}_{hit}, \tilde{x}_{hit}\big) &= \sigma_{LX} + \sigma_{zL} + \sigma_{zX} + \sigma^2_z. \end{aligned} \tag{A9}$$

We do not separately identify variation in all three types of unobserved heterogeneity in factor-specific productivity ($\omega_{L_{hit}}$, $\omega_{X_{hit}}$) or TFP ($z_{hit}$): a parallel increase in $\omega_{L_{hit}}$ and $\omega_{X_{hit}}$ is equivalent to an increase in $z_{hit}$. Hence, we normalize $\omega_{L_{hit}} + \omega_{X_{hit}} = 0$. Intuitively, a change in $\omega_{L_{hit}}$ relative to $\omega_{X_{hit}}$ 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

$$\begin{aligned} \sigma_L^2 &= \sigma_X^2, \\ \sigma_{LX} &= -\sigma_L^2, \\ \sigma_{zL} &= -\sigma_{zX}. \end{aligned} \tag{A10}$$

From equations (A9) and (A10) we calculate the parameters $(\hat{\sigma}_z^2, \hat{\sigma}_L^2, \hat{\sigma}_X^2, \hat{\sigma}_{\epsilon Y}^2, \hat{\sigma}_{\epsilon L}^2, \hat{\sigma}_{\epsilon X}^2, \hat{\sigma}_{LX}, \hat{\sigma}_{zL}, \hat{\sigma}_{zX})$ that are consistent with the observed covariance of plot-level output and inputs across plots within farms, given an estimate of the production function parameters and the assumption of efficient allocation across plots within a farm.

TABLE A1
Tanzania Production Function Estimates ($N = 14{,}535$)

| | 2SLS | IVCRC |
|---|---|---|
| Land | .73 | .61 |
| | (.17) | (.01) |
| Labor | .28 | .26 |
| | (.23) | (.02) |
| Land value | .04 | .07 |
| | (.03) | (.01) |
| Land value missing | .41 | .91 |
| | (.26) | (.33) |
| Distance home | .00 | .00 |
| | (.00) | (.00) |
| Distance to road | −.01 | −.01 |
| | (.00) | (.00) |
| Good soil | .35 | 1.56 |
| | (.12) | (3.41) |
| Average soil | .20 | 1.39 |
| | (.12) | (3.43) |
| Sandy soil | −.06 | −2.37 |
| | (.21) | (4.55) |
| Loamy soil | .02 | −2.27 |
| | (.21) | (4.53) |
| Clay soil | .06 | −2.28 |
| | (.21) | (4.54) |
| Single manager | .02 | .04 |
| | (.04) | (.02) |
| Poor health | −.03 | −.03 |
| | (.01) | (.00) |
| Missing health | −.12 | −.17 |
| | (.05) | (.09) |
| Illiterate | −.06 | −.09 |
| | (.04) | (.08) |
| Literacy missing | .07 | −6.23 |
| | (.33) | (6.34) |
| Male manager | .02 | .07 |
| | (.04) | (.02) |
| Manager age | .00 | .00 |
| | (.00) | (.00) |
| Age missing | 3.70 | 4.26 |
| | (1.45) | (6.75) |
| Crop shock | −.02 | −.02 |
| | (.04) | (.01) |

TABLE A1 (*Continued*)

| | 2SLS | IVCRC |
|---|---|---|
| Drought × good soil | −.07 | −.06 |
| | (.04) | (.04) |
| Drought × average soil | −.07 | −.07 |
| | (.04) | (.07) |
| Drought × poor soil | .01 | .34 |
| | (.09) | (3.39) |
| Crop disease × good soil | .02 | .01 |
| | (.04) | (.06) |
| Crop disease × average soil | −.05 | −.06 |
| | (.04) | (.03) |
| Crop disease × poor soil | .17 | −1.54 |
| | (.09) | (2.65) |
| GS rainfall × good soil | .00 | .00 |
| | (.00) | (.00) |
| GS rainfall × average soil | .00 | .00 |
| | (.00) | (.00) |
| GS rainfall × poor soil | .00 | .00 |
| | (.00) | (.00) |
| GS rainfall × loamy soil | .00 | .00 |
| | (.00) | (.00) |
| GS rainfall × clay soil | .00 | .00 |
| | (.00) | (.00) |
| GS rainfall × other soil | .00 | −.01 |
| | (.00) | (.01) |
| Water shortage | .00 | .01 |
| | (.03) | (.03) |
| Constant | .00 | .01 |
| | (.00) | (.24) |

NOTE.—Standard errors are in parentheses. GS = growing season.

TABLE A2
UGANDA PRODUCTION FUNCTION ESTIMATES ($N$ = 43,187)

| | 2SLS | IVCRC |
|---|---|---|
| Land | .69 | .53 |
| | (.05) | (.01) |
| Labor | .22 | .43 |
| | (.12) | (.01) |
| Fair soil × drought duration | −.03 | −.03 |
| | (.01) | (.00) |
| Poor soil × drought duration | −.03 | −.03 |
| | (.02) | (.01) |
| Missing soil × drought duration | .00 | −.17 |
| | (.03) | (.06) |
| Fair soil × flood duration | .02 | .07 |
| | (.02) | (.02) |
| Poor soil × flood duration | .22 | .43 |
| | (.17) | (.57) |
| Missing soil × flood duration | −.23 | −2.16 |
| | (.31) | (.68) |
| Fair soil × total rain (×10) | .00 | .00 |
| | (.00) | (.00) |

TABLE A2 (*Continued*)

| | 2SLS | IVCRC |
|---|---|---|
| Poor soil × total rain (×10) | .03 | .02 |
| | (.01) | (.01) |
| Missing soil × total rain (×10) | .04 | .03 |
| | (.03) | (.08) |
| Fair soil × peak rain (×10) | .00 | .01 |
| | (.01) | (.00) |
| Poor soil × peak rain (×10) | −.07 | −.05 |
| | (.02) | (.01) |
| Missing soil × peak rain (×10) | .06 | .01 |
| | (.02) | (.01) |
| Fair soil | −.09 | −.07 |
| | (.04) | (.01) |
| Poor soil | −.23 | −.17 |
| | (.10) | (.04) |
| Missing soil | −1.02 | .52 |
| | (.50) | (1.38) |
| Rainfed | .19 | .18 |
| | (.06) | (.02) |
| Wetland | .35 | .27 |
| | (.08) | (.04) |
| Missing water | .51 | 1.22 |
| | (.23) | (.77) |
| Flat land | .03 | .02 |
| | (.03) | (.00) |
| Gentle slope | .08 | .07 |
| | (.03) | (.01) |
| Steep slope | .10 | .08 |
| | (.04) | (.01) |
| Valley | −.01 | .02 |
| | (.05) | (.02) |
| Other slope | .28 | 1.43 |
| | (.28) | (1.33) |
| Missing toposequence | −.20 | −1.62 |
| | (.25) | (1.01) |
| No erosion | .07 | .07 |
| | (.02) | (.00) |
| Missing erosion | .53 | .34 |
| | (.17) | (.28) |
| Male plot | .10 | .11 |
| | (.01) | (.00) |
| Agricultural advice | .09 | .11 |
| | (.02) | (.00) |
| Constant | .00 | −.02 |
| | (.00) | (.01) |

NOTE.—Standard errors are in parentheses.

### *A3. Measurement Error/Shock Variances across All Plots and Average across Farmers of Within-Farmer Variances*

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

Denote $y_{fi}$ the realization of any of these errors/shocks $(\epsilon_{Y_{hit}}, \epsilon_{L_{hit}}, \epsilon_{X_{hit}})$.[34] Let $N$ be the total number of plots, $N^{\mathrm{f}}$ 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 $\sigma_{\mathrm{F}}^2 \equiv (1/N^{\mathrm{f}})\Sigma_{f=1}^{N^{\mathrm{f}}}(1/N_f^i)\Sigma_{i=1}^{N_f^i}(y_{fi} - \bar{y}_f)^2$. The variance of $y$ across plots in the sample is $\sigma^2 \equiv (1/N)\Sigma_{f=1}^{N^{\mathrm{f}}}\Sigma_{i=1}^{N_f^i}(y_{fi} - \bar{y})^2$. 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 $\sigma_{\mathrm{F}}^2 = \sigma^2$.

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 $\{M_1, M_2, \ldots, M_{\bar{k}}\}$ 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

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

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

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

$$\begin{aligned}\sigma^2 &= \frac{1}{N}\sum_{k=1}^{\bar{k}}\sum_{f\in M_k}\sum_{i=1}^{k}y_{fi}^2 - \left(\frac{1}{N}\sum_{k=1}^{\bar{k}}kM_k\overline{y_k}\right)^2\\ &\geq \frac{1}{N}\sum_{k=1}^{\bar{k}}\sum_{f\in M_k}\sum_{i=1}^{k}y_{fi}^2 - \frac{1}{N}\sum_{k=1}^{\bar{k}}M_k k(\overline{y_k})^2 \qquad \text{(A11)}\\ &= \frac{1}{N}\sum_{k=1}^{\bar{k}}M_k k\sigma_k^2,\end{aligned}$$

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

$$\sigma_{\mathrm{F}}^2 = \frac{1}{N_{\mathrm{F}}}\sum_{k=1}^{\bar{k}}M_k\sigma_{\mathrm{F}k}^2.$$

So

$$\sigma^2 - \sigma_{\mathrm{F}}^2 \geq \sum_{k=1}^{\bar{k}}\left(\frac{k}{N} - \frac{1}{N_{\mathrm{F}}}\right)M_k\sigma_{\mathrm{F}k}^2. \qquad \text{(A12)}$$

If each farmer has the same number of plots, then the weak inequality in equation (A12) is an equality, $\Sigma_{k=1}^{\bar{k}}[(k/N) - (1/N_{\mathrm{F}})]M_k\sigma_{\mathrm{F}k}^2 = 0$, and the average across farmers of the within-farmer variance of plot yield is the same as the overall variance of plot yields.

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

Note that $(k/N) - (1/N_\mathrm{F})$ is increasing in $k$. If the average number of plots per farmer is less than or equal to 2, then $(k/N) - (1/N_\mathrm{F}) \geq 0$ for all $k$ and $\sigma^2 - \sigma_\mathrm{F}^2 \geq 0$. 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 two 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 right-hand side of equation (A12) is negative. Given the observed number of plots ($N$), number of farmers ($N_\mathrm{F}$) and numbers of farmers cultivating $k$ plots ($M_k$), then we can calculate that if $\sigma_{\mathrm{F}2}^2 \leq 3.82 \times \sigma_{\mathrm{F}k}^2$ for $k > 2$, then $\Sigma_{k=1}^{\bar{k}}[(k/N) - (1/N_\mathrm{F})]M_k\sigma_{\mathrm{F}k}^2 > 0$. That is, as long as the average variance across plots of $y$ of farmers cultivating two plots is no more than about four times as large as the average variance across plots of $y$ of farmers cultivating more than two plots, then the overall variance of $y$ across plots is larger than the average across farmers of within-farmer cross-plot variance of $y$.

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

### *A4. Consistency of the IVCRC Estimator*

Assumption I2 of Masten and Torgovitsky (2016) requires that the unobserved heterogeneity in any of the endogenous variables be single dimensional. This is equivalent to the assumption of rank invariance, which "means that the ordinal ranking of any two agents in terms of . . . [land or labor demand] would be the same if both agents received the same realization of [the instrument set] $Z$, for any realization of $Z$" (Masten and Torgovitsky 2016, 1003).

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

$$\begin{aligned}\ln p_{X_{hi}} &= a_{X_i} + Z_i\beta_{pX} + \gamma_{pX}W_i,\\ \ln p_{L_{hi}} &= a_{L_i} + Z_i\beta_{pL} + \gamma_{pL}W_i.\end{aligned}$$

Substituting the factor demands (eq. [A7]) yields first-stage factor demand functions of the form

$$\begin{aligned}l_i^{\mathrm{o}} &= f_{l1}(\epsilon_{Y_i}, \epsilon_{L_i}, \epsilon_{X_i}, \omega_{Y_i}, \omega_{L_i}; W_i) + f_{l2}(Z_i) + \omega_{L_i}\frac{Z_i(\beta_{pX} - \beta_{pL})}{1 - \alpha_X - \alpha_L},\\ x_i^{\mathrm{o}} &= f_{x1}(\epsilon_{Y_i}, \epsilon_{L_i}, \epsilon_{X_i}, \omega_{Y_i}, \omega_{L_i}; W_i) + f_{x2}(Z_i) + \omega_{L_i}\frac{Z_i(\beta_{pX} - \beta_{pL})}{1 - \alpha_X - \alpha_L},\end{aligned} \tag{A13}$$

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

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

Define

$$l_i^o(Z) \equiv f_{l1}(\epsilon_{Y_i}, \epsilon_{L_i}, \epsilon_{X_i}, \omega_{Y_i}, \omega_{L_i}; W_i) + f_{l2}(Z) + \omega_{L_i}\frac{Z(\beta_{pX} - \beta_{pL})}{1 - \alpha_X - \alpha_L}$$

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

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

for any values of the instrument vector $\{Z', Z''\}$ and any values of $(\epsilon_{Y_i}, \epsilon_{L_i}, \epsilon_{X_i}, \omega_{Y_i}, \omega_{L_i}; W_i)$ and $(\epsilon_{Y_j}, \epsilon_{L_j}, \epsilon_{X_j}, \omega_{Y_j}, \omega_{L_j}; W_j)$.

Rank invariance does not hold for equation (A13) in general. However, given our estimates, we show that violations are rare. We proceed in two steps. First, we randomly assign values of the instrumental variables (drawn from the sample distribution) to randomly chosen pairs of plots (drawn from the sample distribution and estimated distribution of unobserved heterogeneity) to quantify the frequency of rank reversals. For Tanzania, rank reversals occur in less than 0.0004 of cases for labor and less than 0.0004 cases for land. For Uganda, they occur in less than 0.006 cases for either land or labor.

Note as well that if a rank reversal ever occurs for a given pair of plots, it always occurs at the extreme values of $Z_i(\beta_{pX} - \beta_{pL})$. So we couple this analysis with a calculation of the frequency of violations of rank invariance when pairs of plots are assigned values of the instruments that correspond to extreme values of $\beta_{pX} - \beta_{pL}$. Even in this case, rank reversals are rare. For Tanzania, they occur in 0.001 of cases; for Uganda, the reversal rate is 0.051 for land and 0.055 for labor.

### *A5. Robustness Checks on Restricted Samples*

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

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

Table A3 reports the consequences of these relaxations on our estimates of the variances and covariances of measurement error, late-season risk, and unobserved

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

heterogeneity. Results are reported for samples restricted to plots in Tanzania within 1 km of the respondent's home, within 100 m of the respondent's home, or cultivated by a single individual. The baseline estimates are in column 1 for comparison. Results are reported for Uganda restricting attention to plots within a 15-minute walk of the respondent's home, to plots within contiguous parcel cultivated by a specific farmer, or to plots that are cultivated by a single individual. The baseline estimates for the full sample are provided in the initial columns for comparison. The bottom panel of table A3 reports the consequences of the same set of relaxations for our estimates of the variances of $TFP^A$ and $TFP^B$. Figure A1 shows the implications for productivity dispersion of assumption 1 in Tanzania. Figure A2 shows the implications for productivity dispersion of assumptions 1 (*A*) and 2 (*B*) in Uganda. Figure A3 shows the implication of assumption 3 in both Tanzania and Uganda.

In all cases, comparison with figure 3 shows that the changes to the estimates are minor, with the exception that the IVCRC estimate of the variance of $TFP^B$ is lower in the base estimates than in the alternatives.

TABLE A3
VARIANCE COMPONENTS AND PRODUCTIVITY DISPERSION FOR RESTRICTED SAMPLES

| | Tanzania | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Base | | 1 km | | 100 m | | Single Manager | |
| | 2SLS | IVCRC | 2SLS | IVCRC | 2SLS | IVCRC | 2SLS | IVCRC |
| | A. Variance Components | | | | | | | |
| Plot TFP | .35 | .38 | .36 | .39 | .36 | .39 | .35 | .39 |
| Land/labor productivity | .07 | .10 | .07 | .10 | .08 | .10 | .06 | .10 |
| Late-season risk and output measurement error | .65 | .70 | .67 | .71 | .67 | .71 | .68 | .73 |
| Land measurement error | .08 | .03 | .08 | .03 | .08 | .03 | .13 | .07 |
| Labor measurement error | .27 | .32 | .27 | .34 | .27 | .33 | .31 | .31 |
| Covariance of TFP and land/labor productivity | .06 | .09 | .06 | .09 | .06 | .09 | .07 | .10 |
| | B. Productivity Dispersion | | | | | | | |
| Var($TFP^A$) | 1.27 | 1.29 | 1.13 | 1.29 | 1.13 | 1.29 | 1.17 | 1.29 |
| Var($TFP^B$) | .47 | .55 | .40 | .54 | .39 | .54 | .39 | .51 |
| Implied share of measurement error/unobserved heterogeneity in Var($TFP^A$) | .61 | .57 | .65 | .58 | .65 | .58 | .67 | .60 |
| | Uganda | | | | | | | |
| | Base | | 15-Minute Walk | | Within Parcel | | Single Manager | |
| | 2SLS | IVCRC | 2SLS | IVCRC | 2SLS | IVCRC | 2SLS | IVCRC |
| | A. Variance Components | | | | | | | |
| Plot TFP | .18 | .18 | .18 | .18 | .18 | .18 | .17 | .17 |
| Land/labor productivity | .07 | .07 | .08 | .08 | .08 | .08 | .06 | .06 |
| Late-season risk and output measurement error | .67 | .67 | .69 | .69 | .71 | .71 | .63 | .63 |

TABLE A3 (*Continued*)

| | Uganda | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Base | | 15-Minute Walk | | Within Parcel | | Single Manager | |
| | 2SLS | IVCRC | 2SLS | IVCRC | 2SLS | IVCRC | 2SLS | IVCRC |
| Land measurement error | .13 | .13 | .16 | .16 | .18 | .18 | .14 | .14 |
| Labor measurement error | .18 | .18 | .20 | .20 | .18 | .18 | .20 | .20 |
| Covariance of TFP and land/labor productivity | .05 | .05 | .06 | .06 | .05 | .05 | .06 | .06 |
| | B. Productivity Dispersion | | | | | | | |
| Var(TFP$^{A}$) | 1.26 | 1.25 | 1.34 | 1.31 | 1.38 | 1.35 | 1.21 | 1.22 |
| Var(TFP$^{B}$) | .52 | .29 | .56 | .54 | .58 | .56 | .51 | .52 |
| Implied share of measurement error/unobserved heterogeneity in Var(TFP$^{A}$) | .59 | .77 | .58 | .59 | .58 | .59 | .58 | .57 |

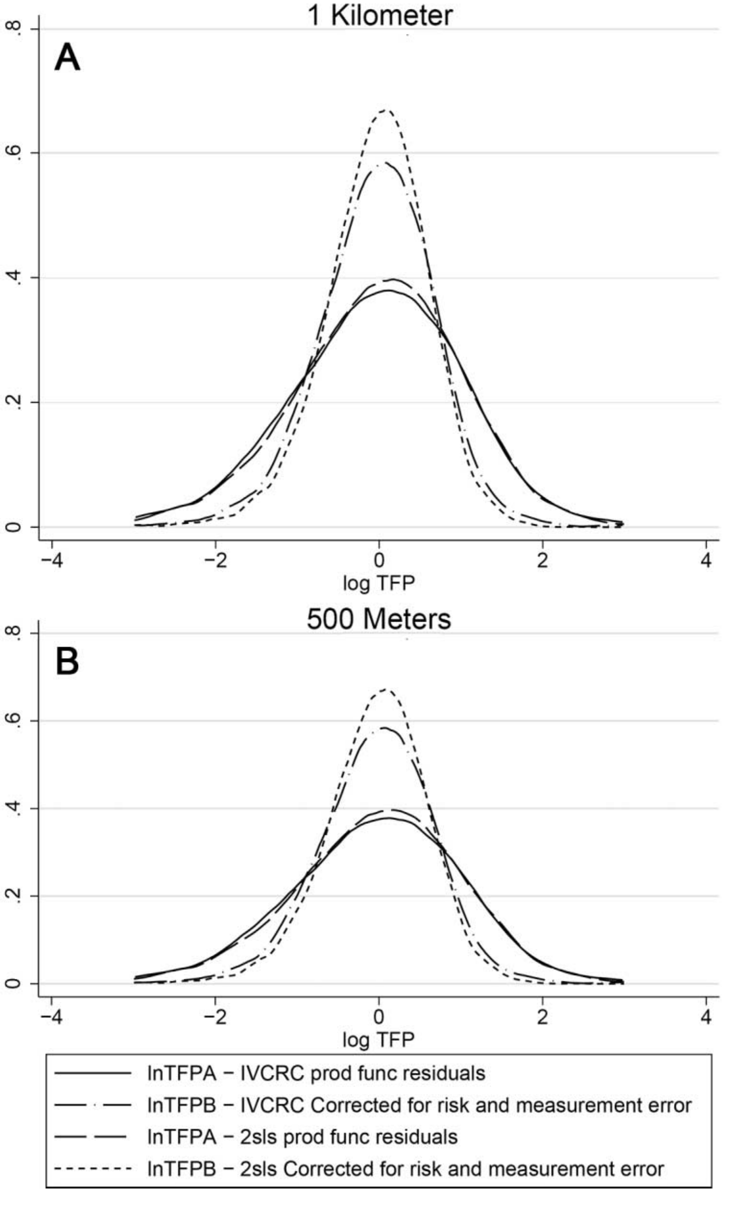


FIG. A1.—Tanzania distance effects on log productivity dispersion. prod func = production function.

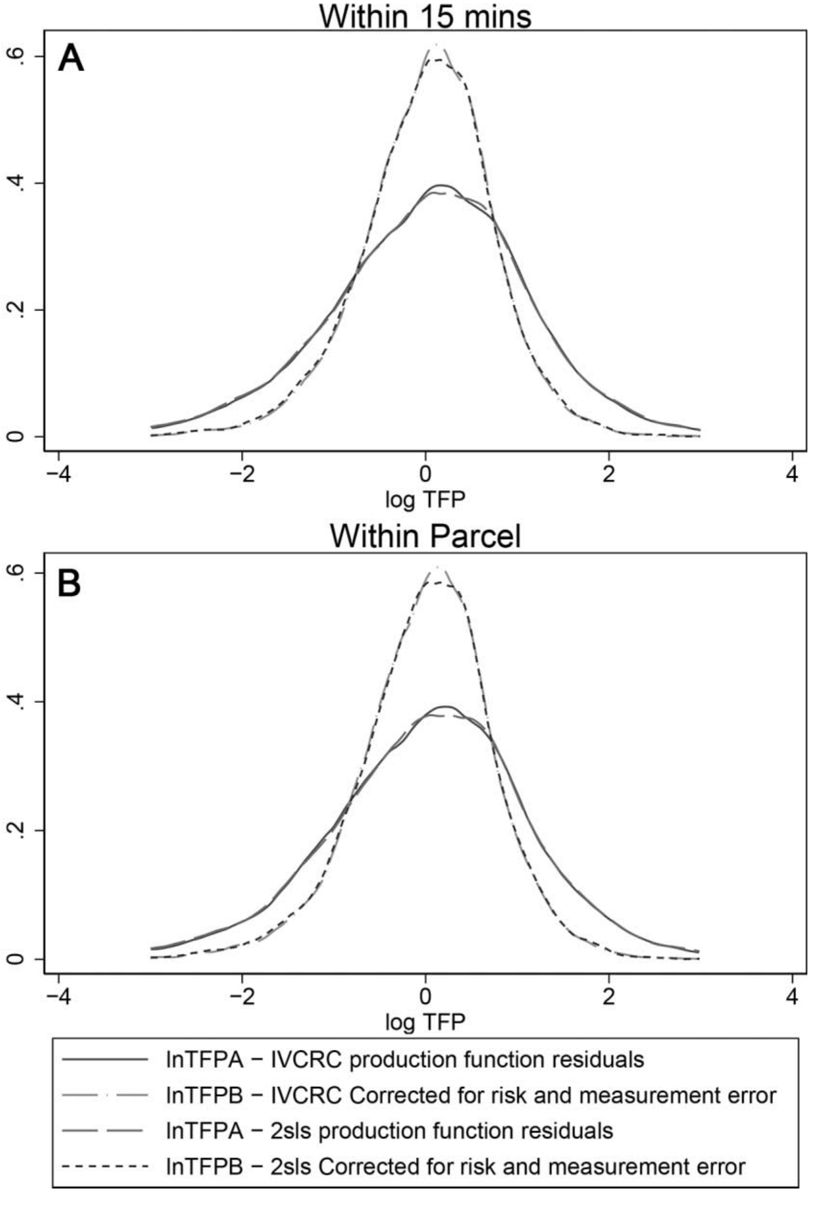


FIG. A2.—Uganda distance and parcel effects on log productivity dispersion.

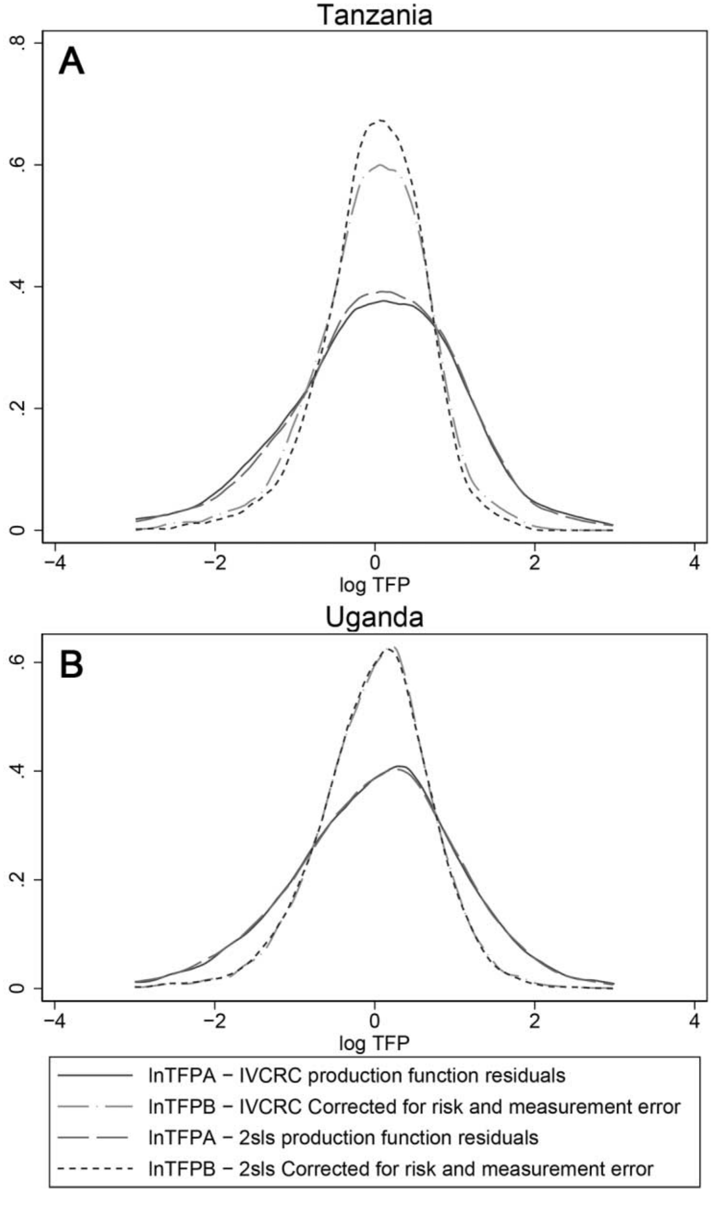


FIG. A3.—Single-managers-only log productivity dispersion.

### *A6. Gains from Reallocation*

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

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

$\bar{L}$ and $\bar{X}$ are, respectively, aggregate endowments of land and labor, and $\bar{\omega} = \Sigma_i e^{[1/(1-\alpha_L-\alpha_X)]\omega_i}$. Measured output of producer $i$ in an efficient allocation is

$$Y_i^e = \left(\frac{1}{\bar{\omega}}\right)^{\alpha_L+\alpha_X} e^{[\omega_i/(1-\alpha_L-\alpha_X)]+\epsilon_i} (\bar{L})^{\alpha_L} (\bar{X})^{\alpha_X}. \tag{A14}$$

If $\omega_i$ and $\epsilon_i$ are normally distributed and independent of each other, then expected output is

$$\begin{aligned} E(Y_i^e) &= \left(\frac{1}{\bar{\omega}}\right)^{\alpha_L+\alpha_X} (\bar{L})^{\alpha_L} (\bar{X})^{\alpha_X} E(e^{\epsilon_i}) E\left(e^{\omega_i/(1-\alpha_L-\alpha_X)}\right) \\ &= \left(\frac{1}{\bar{\omega}}\right)^{\alpha_L+\alpha_X} (\bar{L})^{\alpha_L} (\bar{X})^{\alpha_X} E(e^{\epsilon_i}) e^{E(\omega_i)/(1-\alpha_L-\alpha_X)} e^{\sigma_\omega^2/2(1-\alpha_L-\alpha_X)^2} \qquad \text{(A15)} \\ &\equiv \overline{Y^e}(\sigma_\omega^2), \end{aligned}$$

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

## Appendix B

### Coefficient Estimates: Full Sets

TABLE B1
OLS and Quantile Regression Determinants of Log Land and Log Labor Inputs in Tanzania

| | OLS | | 25th percentile | | 50th percentile | | 75th percentile | | Interquartile Range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Log Land (1) | Log Labor (2) | Log Land (3) | Log Labor (4) | Log Land (5) | Log Labor (6) | Log Land (7) | Log Labor (8) | Log Land (9) | Log Labor (10) |
| Male manager | .35 | .23 | .37 | .28 | .32 | .22 | .33 | .22 | −.034 | −.069 |
| | (.019) | (.018) | (.023) | (.025) | (.023) | (.021) | (.023) | (.020) | (.028) | (.031) |
| Land value | .21 | .1 | .19 | .11 | .2 | .1 | .24 | .1 | .044 | −.0056 |
| | .009) | (.008) | .010) | (.010) | (.010) | .009) | .010) | .009) | (.014) | (.011) |
| Drought/ flood × good soil | .096 | .024 | .14 | .011 | .045 | −.03 | .045 | .0023 | −.091 | −.0085 |
| | (.034) | .031) | (.040) | (.040) | (.041) | (.036) | (.044) | (.036) | (.044) | (.048) |
| Drought/ flood × average (avg.) soil | .073 | .035 | .12 | −.0043 | −.0038 | .055 | .025 | .036 | −.1 | .04 |
| | (.033) | (.031) | (.039) | (.043) | (.042) | (.036) | (.039) | (.036) | (.040) | (.067) |
| Drought/ flood × poor soil | −.027 | −.035 | −.014 | −.083 | −.031 | −.082 | −.0056 | −.016 | .0083 | .067 |
| | (.077) | (.072) | (.078) | (.085) | (.089) | (.098) | (.065) | (.080) | (.095) | (.110) |
| Illness/ accident of HH member | .035 | −.039 | .042 | −.075 | .056 | −.0095 | .0073 | .018 | −.035 | .093 |
| | (.031) | (.029) | (.046) | (.040) | (.031) | (.034) | (.036) | (.036) | (.056) | (.043) |
| GS rain × good soil in HH[a] | −4E−05 | −4E−06 | 4E−05 | 4E−05 | −5E−05 | −7E−06 | −6E−05 | −3E−05 | −.0001 | −7E−05 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| GS rain × avg soil in HH[a] | −6E−06 | 2E−05 | 8E−05 | 3E−05 | −3E−05 | 1E−05 | −3E−05 | 2E−05 | −.0001 | −2E−05 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| GS rain × poor soil in HH[a] | −6E−05 | −5E−05 | −9E−05 | −8E−05 | −.0001 | −7E−05 | −2E−05 | 3E−05 | 8E−05 | .0001 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| Drought/ flood × good soil in HH[a] | .0053 | .038 | .02 | .046 | .011 | .051 | .0067 | .032 | −.014 | −.014 |
| | (.017) | (.016) | (.017) | (.025) | (.018) | (.014) | (.016) | (.019) | (.022) | (.025) |
| Drought/ flood × avg. soil in HH[a] | −.009 | .02 | −.029 | .04 | .019 | .026 | .0011 | .011 | .03 | −.029 |
| | (.017) | (.016) | (.015) | (.025) | (.020) | (.019) | (.019) | (.016) | (.025) | (.033) |
| Drought/ flood × poor soil in HH[a] | .11 | .1 | .11 | .097 | .08 | .17 | .072 | .025 | −.042 | −.072 |
| | (.045) | (.042) | (.038) | (.047) | (.075) | (.051) | (.028) | (.040) | (.072) | (.045) |
| Adverse shock to HH plots[a] | −.089 | −.032 | −.11 | −.046 | −.081 | −.032 | −.061 | −.022 | .052 | .023 |
| | (.011) | (.010) | (.012) | (.013) | (.011) | (.010) | (.012) | (.011) | (.016) | (.013) |
| Land value missing | 1.87 | .87 | 1.72 | .93 | 1.82 | .86 | 2.2 | .93 | .48 | −.0055 |
| | (.099) | (.092) | (.110) | (.120) | (.110) | (.100) | (.120) | (.092) | (.130) | (.140) |
| Distance home | .0017 | 8E−05 | .0021 | .0001 | .0017 | 1E−04 | .0019 | −.0003 | −.0002 | −.0004 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.001) | (.000) |
| Distance to road | .021 | .018 | .024 | .021 | .023 | .02 | .024 | .018 | .0007 | −.0026 |
| | (.002) | (.002) | (.002) | (.002) | (.002) | (.001) | (.002) | (.001) | (.003) | (.002) |
| Good soil | −.093 | −.12 | −.14 | −.19 | −.067 | −.18 | −.17 | −.18 | −.029 | .0096 |
| | (.100) | (.097) | (.120) | (.120) | (.091) | (.120) | (.110) | (.094) | (.160) | (.180) |
| Avg. soil | −.19 | −.22 | −.28 | −.29 | −.23 | −.32 | −.24 | −.25 | .036 | .038 |
| | (.110) | (.098) | (.120) | (.120) | (.093) | (.120) | (.110) | (.095) | (.160) | (.160) |
| Sandy soil | −.38 | −.4 | −.43 | −.36 | −.38 | −.54 | −.5 | −.51 | −.069 | −.15 |
| | (.150) | (.140) | (.110) | (.130) | (.220) | (.230) | (.110) | (.100) | (.220) | (.200) |
| Loamy soil | −.3 | −.44 | −.34 | −.36 | −.33 | −.48 | −.26 | −.45 | .077 | −.091 |
| | (.140) | (.130) | (.089) | (.120) | (.210) | (.220) | (.110) | (.094) | (.230) | (.200) |
| Clay soil | −.27 | −.3 | −.26 | −.24 | −.35 | −.32 | −.23 | −.31 | .026 | −.068 |
| | (.150) | (.140) | (.100) | (.130) | (.210) | (.220) | (.120) | (.100) | (.240) | (.200) |
| Single manager | −.037 | −.13 | −.02 | −.13 | −.044 | −.15 | −.045 | −.15 | −.025 | −.029 |
| | (.018) | (.017) | (.023) | (.022) | (.022) | (.020) | (.022) | (.019) | (.026) | (.023) |
| Poor health[a] | −.017 | −.0056 | −.022 | .0017 | −.016 | −.0011 | −.019 | −.0047 | .0031 | −.0064 |
| | (.005) | (.005) | (.006) | (.007) | (.006) | (.006) | (.006) | (.005) | (.010) | (.009) |

TABLE B1 (*Continued*)

| | OLS | | 25th percentile | | 50th percentile | | 75th percentile | | Interquartile Range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Log Land (1) | Log Labor (2) | Log Land (3) | Log Labor (4) | Log Land (5) | Log Labor (6) | Log Land (7) | Log Labor (8) | Log Land (9) | Log Labor (10) |
| Missing health[a] | −.038 | .06 | −.049 | .079 | −.086 | .077 | −.031 | .027 | .018 | −.052 |
| | (.030) | (.028) | (.038) | (.032) | (.034) | (.033) | (.038) | (.033) | (.054) | (.055) |
| Illiterate[a] | .057 | .1 | .088 | .13 | .067 | .085 | .056 | .083 | −.032 | −.042 |
| | (.026) | (.024) | (.036) | (.034) | (.030) | (.028) | (.030) | (.027) | (.051) | (.036) |
| Literacy missing[a] | −.22 | .022 | .028 | .093 | −.33 | −.23 | −.11 | −.13 | −.14 | −.22 |
| | (.220) | (.200) | (.160) | (.089) | (.095) | (.190) | (.300) | (.110) | (.450) | (.320) |
| Manager age[a] | .0041 | .0077 | .0037 | .0077 | .0033 | .0084 | .0053 | .0083 | .0015 | .0005 |
| | (.001) | (.001) | (.001) | (.001) | (.001) | (.001) | (.001) | (.001) | (.001) | (.001) |
| Age missing[a] | −4.150 | −7.490 | −4.010 | −7.390 | −3.340 | −8.100 | −5.380 | −8.120 | −1.370 | −.730 |
| | (.560) | (.520) | (.690) | (.670) | (.660) | (.610) | (.700) | (.550) | (.800) | (.930) |
| Crop shock | .08 | .14 | .14 | .17 | .058 | .17 | .053 | .096 | −.083 | −.079 |
| | (.017) | (.015) | (.021) | (.021) | (.019) | (.018) | (.020) | (.017) | (.019) | (.021) |
| Crop disease × good soil | −.027 | .077 | −.033 | .067 | .034 | .096 | −.0063 | .065 | .026 | −.0014 |
| | (.036) | (.033) | (.052) | (.043) | (.043) | (.038) | (.043) | (.038) | (.055) | (.048) |
| Crop disease × avg. soil | .039 | .12 | .11 | .091 | .061 | .077 | .021 | .12 | −.086 | .032 |
| | (.035) | (.032) | (.048) | (.047) | (.042) | (.036) | (.040) | (.036) | (.051) | (.045) |
| Crop disease × poor soil | −.16 | −.054 | −.2 | −.036 | −.12 | −.03 | −.14 | −.095 | .054 | −.059 |
| | (.078) | (.072) | (.077) | (.100) | (.067) | (.087) | (.061) | (.074) | (.110) | (.130) |
| (soil_quality = 1) × wetQ | −.0002 | −.0002 | −.0002 | −.0003 | −.0002 | −3E−05 | .0002 | 2E−05 | .0004 | .0003 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| (soil_quality = 2) × wetQ | −1E−04 | −1E−04 | −.0002 | −.0001 | −2E−05 | .0001 | .0002 | 4E−05 | .0004 | .0001 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| (soil_quality = 3) × wetQ | −.0003 | −.0004 | −.0005 | −.0007 | −.0004 | −.0003 | −.0001 | −.0003 | .0003 | .0004 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| GS rainfall × loamy soil | 2E−06 | .0003 | .0001 | .0003 | −6E−06 | 7E−05 | −.0004 | 6E−05 | −.0005 | −.0002 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| GS rainfall × clay soil | −4E−05 | .0002 | 3E−05 | .0002 | −4E−07 | −4E−05 | −.0005 | −7E−05 | −.0005 | −.0003 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| GS rainfall × other soil | −.0006 | −.0006 | −.0006 | −.0005 | −.0008 | −.0009 | −.001 | −.0008 | −.0004 | −.0003 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| Water shortage | .12 | .049 | .099 | .078 | .14 | .045 | .14 | .065 | .043 | −.013 |
| | (.018) | (.017) | (.026) | (.024) | (.022) | (.019) | (.022) | (.019) | (.028) | (.034) |
| Livestock death or stolen[a] | .065 | .055 | .039 | .028 | .082 | .038 | .075 | .088 | .037 | .06 |
| | (.020) | (.018) | (.027) | (.023) | (.024) | (.021) | (.022) | (.023) | (.039) | (.037) |
| Death of HH member[a] | .1 | .099 | .09 | .12 | .075 | .12 | .061 | .091 | −.029 | −.029 |
| | (.026) | (.024) | (.024) | (.039) | (.031) | (.027) | (.032) | (.022) | (.034) | (.048) |
| Property crime in HH[a] | −.002 | −.055 | .008 | −.095 | −.019 | −.038 | −.013 | −.047 | −.021 | .048 |
| | (.032) | (.030) | (.034) | (.048) | (.040) | (.034) | (.029) | (.027) | (.036) | (.051) |
| HH good soil × shock on plots in HH[a] | .0015 | −.044 | −.031 | −.074 | −.022 | −.045 | .0072 | −.024 | .038 | .05 |
| | (.017) | (.015) | (.023) | (.025) | (.021) | (.014) | (.017) | (.020) | (.035) | (.020) |
| HH avg. soil × shock on plots in HH[a] | −.0052 | −.0072 | .0003 | −.0021 | −.011 | .011 | −.01 | −.0062 | −.011 | −.0041 |
| | (.017) | (.016) | (.019) | (.023) | (.015) | (.017) | (.019) | (.015) | (.025) | (.023) |
| HH poor soil × shock on plots in HH[a] | −.086 | −.045 | −.034 | −.017 | −.082 | −.05 | −.12 | −.07 | −.09 | −.052 |
| | (.046) | (.043) | (.043) | (.050) | (.053) | (.041) | (.030) | (.048) | (.090) | (.060) |
| HH loamy soil × GS rain[a] | −7E−05 | −.0001 | −.0001 | −.0001 | −6E−05 | −.0001 | −1E−04 | −.0001 | 2E−05 | −1E−05 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| HH clay soil × GS rain[a] | −6E−05 | −7E−05 | −.0001 | −.0001 | −6E−05 | −8E−05 | −6E−05 | −8E−05 | 4E−05 | 2E−05 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |
| HH other soil × GS rain[a] | −.0001 | −9E−05 | −.0002 | −8E−05 | −.0001 | −7E−05 | −.0002 | −.0001 | 2E−05 | −3E−05 |
| | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) | (.000) |

TABLE B1 (*Continued*)

| | OLS | | 25th percentile | | 50th percentile | | 75th percentile | | Interquartile Range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Log Land (1) | Log Labor (2) | Log Land (3) | Log Labor (4) | Log Land (5) | Log Labor (6) | Log Land (7) | Log Labor (8) | Log Land (9) | Log Labor (10) |
| Constant | −2E−08 | −9E−09 | −.58 | −.55 | .028 | .062 | .62 | .61 | 1.19 | 1.16 |
| | (.008) | (.007) | (.011) | (.011) | (.010) | (.009) | (.010) | (.009) | (.013) | (.011) |
| *F*-statistics: | | | | | | | | | | |
| For joint significance of instruments | 17.1 | 11.5 | 13.7 | 7.32 | 24.4 | 11.3 | 19.4 | 11.2 | | |
| *p*-value | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| For $h_0$[b] | | | | | | | | | 30.1 | 5.46 |
| *p*-value | | | | | | | | | 0 | 0 |

Note.—$N$ = 14,535. Bootstrapped (500 samples) standard errors, clustered at the household level, in parentheses. HH = household; GS = growing season.

[a] Variable serves as instrument in table 9.

[b] $h_0$: coefficients are equal for 25th and 75th percentiles.

TABLE B2
OLS and Quantile Regression Determinants of Log Land and Log Labor Inputs in Uganda

| | OLS | | 25th percentile | | 50th percentile | | 75th percentile | | Interquartile Range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Log Land (1) | Log Labor (2) | Log Land (3) | Log Labor (4) | Log Land (5) | Log Labor (6) | Log Land (7) | Log Labor (8) | Log Land (9) | Log Labor (10) |
| Male plot | .14 | .062 | .13 | .044 | .14 | .067 | .14 | .083 | .015 | .039 |
| | (.0086) | (.0071) | (.0097) | (.0093) | (.0093) | (.0075) | (.0093) | (.0075) | (.014) | (.011) |
| Fair soil | −.092 | −.078 | −.100 | −.080 | −.14 | −.098 | −.10 | −.12 | −.0045 | −.036 |
| | (.024) | (.021) | (.030) | (.029) | (.027) | (.018) | (.024) | (.023) | (.025) | (.024) |
| Poor soil | −.20 | −.11 | −.19 | −.19 | −.31 | −.18 | −.13 | −.022 | .064 | .17 |
| | (.067) | (.062) | (.082) | (.046) | (.056) | (.066) | (.060) | (.047) | (.14) | (.065) |
| Fair soil × drought duration | .026 | .024 | .032 | .021 | .032 | .024 | .025 | .035 | −.0071 | .013 |
| | (.0052) | (.0046) | (.0056) | (.0060) | (.0057) | (.0048) | (.0052) | (.0050) | (.0081) | (.0059) |
| Poor soil × drought duration | .035 | .027 | .035 | .027 | .036 | .021 | .040 | .0044 | .0049 | −.023 |
| | (.012) | (.0092) | (.015) | (.010) | (.012) | (.0097) | (.010) | (.0040) | (.018) | (.013) |
| Illness incidence in HH[a] | .0027 | −.051 | −.0019 | −.067 | .013 | −.066 | −.0078 | −.064 | −.0059 | .0027 |
| | (.013) | (.0098) | (.015) | (.013) | (.014) | (.011) | (.012) | (.010) | (.016) | (.013) |
| HH fair soil × peak rain[a] | .000082 | −.000074 | .00010 | −.000073 | .000045 | −.000070 | .000045 | −.000032 | −.000058 | .000042 |
| | (.000023) | (.000019) | (.000023) | (.000024) | (.000026) | (.000020) | (.000027) | (.000018) | (.000035) | (.000027) |
| HH poor soil × peak rain[a] | .00021 | −.000053 | .00022 | −.000037 | .00025 | −.00011 | .00022 | −.000036 | .0000057 | .000000075 |
| | (.000073) | (.000055) | (.000086) | (.000087) | (.000058) | (.000081) | (.000067) | (.000055) | (.00010) | (.00011) |
| HH fair soil × drought duration[a] | −.0053 | −.0069 | −.0067 | −.0076 | −.0061 | −.0078 | −.0074 | −.0083 | −.00062 | −.00073 |
| | (.0018) | (.0015) | (.0018) | (.0017) | (.0019) | (.0014) | (.0015) | (.0017) | (.0026) | (.0027) |

TABLE B2 (*Continued*)

| | OLS | | 25th percentile | | 50th percentile | | 75th percentile | | Interquartile Range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Log Land (1) | Log Labor (2) | Log Land (3) | Log Labor (4) | Log Land (5) | Log Labor (6) | Log Land (7) | Log Labor (8) | Log Land (9) | Log Labor (10) |
| HH fair soil × total rain[a] | −.048 | .017 | −.053 | .018 | −.045 | .012 | −.041 | −.00038 | .012 | −.018 |
| | (.0083) | (.0069) | (.0082) | (.0086) | (.0097) | (.0071) | (.0098) | (.0059) | (.011) | (.0092) |
| HH poor soil × total rain[a] | −.083 | .023 | −.070 | .018 | −.097 | .037 | −.091 | −.0038 | −.020 | −.022 |
| | (.025) | (.020) | (.031) | (.031) | (.019) | (.029) | (.022) | (.018) | (.027) | (.036) |
| Shocks on HH plots[a] | .023 | −.00034 | .017 | −.0011 | .069 | .00065 | .10 | −.00049 | .087 | .00065 |
| | (.0032) | (.0013) | (.0018) | (.0035) | (.0072) | (.0012) | (.0080) | (.0022) | (.0090) | (.0029) |
| HH non-agricultural shock[a] | .084 | −.0063 | .083 | −.0067 | .059 | −.020 | .093 | −.024 | .010 | −.018 |
| | (.026) | (.021) | (.019) | (.026) | (.030) | (.021) | (.032) | (.022) | (.031) | (.030) |
| Value of HH assets[a] | 1.6E−09 | 1.5E−10 | 1.7E−09 | 1.0E−10 | 1.6E−09 | 2.8E−10 | 1.5E−09 | 2.9E−10 | −2.4E−10 | 1.9E−10 |
| | (1.4E−10) | (1.3E−10) | (1.5E−10) | (1.5E−10) | (1.5E−10) | (1.2E−10) | (1.1E−10) | (1.5E−10) | (2.7E−10) | (1.7E−10) |
| Cement construction[a] | .068 | .023 | .059 | .019 | .052 | .0052 | .054 | .027 | −.0051 | .0077 |
| | (.012) | (.0097) | (.013) | (.013) | (.013) | (.010) | (.013) | (.010) | (.017) | (.017) |
| No. of HH members[a] | .025 | .010 | .023 | .014 | .022 | .011 | .022 | .010 | −.00035 | −.0041 |
| | (.0020) | (.0017) | (.0024) | (.0022) | (.0022) | (.0018) | (.0022) | (.0018) | (.0027) | (.0026) |
| No. of adults in HH[a] | .017 | .0081 | .018 | .0054 | .018 | .0078 | .016 | .0075 | −.0022 | .0022 |
| | (.0029) | (.0025) | (.0036) | (.0031) | (.0031) | (.0026) | (.0033) | (.0026) | (.0043) | (.0041) |
| Every member of HH has ≥1 pair of shoes[a] | −.0074 | −.070 | −.017 | −.094 | .0014 | −.058 | .018 | −.034 | .035 | .060 |
| | (.0092) | (.0078) | (.010) | (.010) | (.0099) | (.0080) | (.0098) | (.0081) | (.012) | (.013) |
| Does this house have electricity?[a] | .078 | .053 | .12 | .044 | .054 | .021 | .026 | .038 | −.094 | −.0062 |
| | (.020) | (.016) | (.022) | (.019) | (.021) | (.015) | (.023) | (.017) | (.033) | (.027) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Literacy of plot managers[a] | .062 | .042 | .054 | .049 | .069 | .039 | .063 | .025 | .0081 | −.024 |
| | (.012) | (.011) | (.015) | (.013) | (.013) | (.011) | (.014) | (.011) | (.017) | (.019) |
| Schooling level of plot managers[a] | −.0016 | −.00099 | −.021 | .0025 | .0016 | −.0050 | −.0061 | −.012 | .015 | −.014 |
| | (.013) | (.011) | (.014) | (.015) | (.014) | (.011) | (.012) | (.012) | (.019) | (.012) |
| Plot manager is recent resident[a] | .015 | .074 | .022 | .076 | .038 | .077 | .045 | .069 | .023 | −.0064 |
| | (.019) | (.016) | (.020) | (.023) | (.018) | (.017) | (.018) | (.016) | (.024) | (.022) |
| One or both managers serve on committee[a] | .022 | .013 | .0092 | −.0048 | .0047 | .029 | −.024 | .025 | −.034 | .029 |
| | (.014) | (.011) | (.016) | (.014) | (.015) | (.012) | (.015) | (.012) | (.019) | (.015) |
| HH adult equivalence scale[a] | .000025 | .000091 | .000012 | .000099 | −.000063 | .000065 | −.000014 | .000045 | −.000026 | −.000054 |
| | (.000048) | (.000038) | (.000044) | (.000065) | (.000058) | (.000058) | (.000081) | (.000024) | (.000074) | (.000045) |
| V fair soil × drought duration[a] | .0031 | .00047 | .0024 | .00067 | .0030 | .00053 | .0027 | .00030 | .00030 | −.00037 |
| | (.00037) | (.00030) | (.00041) | (.00040) | (.00039) | (.00033) | (.00036) | (.00034) | (.00051) | (.00031) |
| HH poor soil × drought duration[a] | −.0029 | −.0035 | −.0024 | −.0039 | −.0057 | −.0042 | −.0070 | .0033 | −.0046 | .0072 |
| | (.0043) | (.0037) | (.0053) | (.0041) | (.0038) | (.0043) | (.0049) | (.0035) | (.0068) | (.0056) |
| V poor soil × drought duration[a] | .0066 | −.0012 | .0050 | −.00063 | .0019 | −.0012 | .0031 | −.0024 | −.0019 | −.0018 |
| | (.0013) | (.00100) | (.0012) | (.0013) | (.0011) | (.00098) | (.0012) | (.0011) | (.0018) | (.0020) |
| HH missing soil × drought duration[a] | −.028 | .019 | −.036 | .0044 | −.023 | .013 | −.010 | .027 | .026 | .023 |
| | (.0099) | (.0061) | (.0094) | (.0063) | (.0081) | (.0058) | (.0078) | (.011) | (.018) | (.011) |

TABLE B2 (*Continued*)

| | OLS | | 25th percentile | | 50th percentile | | 75th percentile | | Interquartile Range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Log Land (1) | Log Labor (2) | Log Land (3) | Log Labor (4) | Log Land (5) | Log Labor (6) | Log Land (7) | Log Labor (8) | Log Land (9) | Log Labor (10) |
| V missing soil × drought duration[a] | −.0020 | .00061 | −.0023 | .0044 | −.0016 | −.00083 | −.0033 | −.00075 | −.0010 | −.0052 |
| | (.0022) | (.0017) | (.0019) | (.0019) | (.0020) | (.0015) | (.0013) | (.0015) | (.0034) | (.0022) |
| HH fair soil × flood duration[a] | .016 | .0088 | .025 | .0083 | .016 | .011 | .0067 | .0032 | −.019 | −.0052 |
| | (.0054) | (.0056) | (.012) | (.0091) | (.0056) | (.0031) | (.0060) | (.0068) | (.012) | (.012) |
| V fair soil × flood duration[a] | −.0037 | −.0017 | −.0035 | −.0025 | −.0052 | −.0017 | −.0038 | −.0010 | −.00034 | .0015 |
| | (.0013) | (.0011) | (.0022) | (.00041) | (.0022) | (.0012) | (.0018) | (.0013) | (.0021) | (.0020) |
| HH poor soil × flood duration[a] | .056 | .022 | .028 | .054 | .069 | .029 | .019 | .00091 | −.0090 | −.053 |
| | (.043) | (.027) | (.17) | (.023) | (.23) | (.020) | (.044) | (.031) | (.12) | (.029) |
| V poor soil × flood duration[a] | .0029 | .018 | .0031 | .026 | .011 | .022 | .016 | .013 | .013 | −.013 |
| | (.012) | (.0094) | (.0075) | (.022) | (.0054) | (.0099) | (.013) | (.013) | (.015) | (.019) |
| HH missing soil × flood duration[a] | .14 | .079 | .15 | .096 | .12 | .11 | .16 | .074 | .015 | −.022 |
| | (.037) | (.028) | (.083) | (.071) | (.28) | (.042) | (.074) | (.093) | (.079) | (.075) |
| V missing soil × flood duration[a] | −.077 | −.018 | −.035 | −.016 | −.068 | −.019 | −.058 | −.0079 | −.023 | .0085 |
| | (.020) | (.015) | (.0098) | (.017) | (.018) | (.016) | (.0078) | (.011) | (.030) | (.026) |
| V fair soil × total rain[a] | −.0047 | −.00077 | −.0054 | −.00017 | −.0065 | −.00019 | −.0055 | −.00051 | −.000097 | −.00034 |
| | (.0010) | (.00089) | (.0013) | (.0011) | (.0011) | (.00084) | (.0010) | (.00087) | (.0018) | (.0012) |
| V poor soil × total rain[a] | −.036 | −.0030 | −.030 | .0017 | −.030 | −.0076 | −.031 | −.0063 | −.00094 | −.0080 |
| | (.0059) | (.0044) | (.0069) | (.0060) | (.0061) | (.0053) | (.0068) | (.0046) | (.0093) | (.0083) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| HH missing soil × total rain[a] | −.00011 | −.0000067 | −.00012 | .000066 | −.000075 | −.000021 | −.00019 | −.000085 | −.000067 | −.00015 |
| | (.000056) | (.000040) | (.000068) | (.000063) | (.000081) | (.000036) | (.00006) | (.00004) | (.00009) | (.000078) |
| V missing soil × total rain[a] | .000025 | .0000059 | .000034 | −.000014 | .000036 | .000010 | .000028 | .000012 | −.000006 | .000026 |
| | (.000008) | (.000006) | (.000010) | (.000008) | (.000009) | (.000007) | (.00001) | (.00001) | (.00001) | (.000007) |
| V fair soil × peak rain[a] | .000004 | .0000040 | .000008 | .0000031 | .000010 | .000002 | .000002 | .000000 | −.000006 | −.0000030 |
| | (.000003) | (.000002) | (.000004) | (.0000030) | (.000003) | (.000002) | (.000003) | (.00000) | (.000004) | (.000003) |
| V poor soil × peak rain[a] | .000044 | −.000027 | .000032 | −.000042 | .000044 | −.000021 | .000047 | −.000012 | .000016 | .000031 |
| | (.000016) | (.000012) | (.000019) | (.000017) | (.000017) | (.000015) | (.000020) | (.000013) | (.000026) | (.000022) |
| HH missing soil × peak rain[a] | .00038 | −.00015 | .00045 | −.00028 | .00027 | −.00012 | .00041 | .000087 | −.000042 | .00036 |
| | (.00016) | (.00012) | (.00021) | (.00016) | (.00023) | (.00010) | (.00016) | (.00012) | (.00019) | (.00025) |
| V missing soil × peak rain[a] | −.000040 | .0000024 | −.000060 | .000018 | −.000070 | .0000082 | −.000039 | .0000042 | .000022 | −.000014 |
| | (.000017) | (.000013) | (.000020) | (.000016) | (.000017) | (.000013) | (.000018) | (.000013) | (.000017) | (.000018) |
| Shocks on village plots[a] | .0025 | .00067 | .0026 | .00091 | .0025 | .00075 | .0039 | .00084 | .0013 | −.000066 |
| | (.00044) | (.00029) | (.00055) | (.00045) | (.00029) | (.00027) | (.00054) | (.00029) | (.00071) | (.00045) |
| Missing soil × drought duration | .028 | −.017 | .013 | −.032 | .012 | .000006 | .026 | .0022 | .014 | .034 |
| | (.027) | (.020) | (.023) | (.018) | (.021) | (.015) | (.038) | (.027) | (.044) | (.031) |
| Fair soil × flood duration | −.021 | −.024 | −.057 | −.024 | −.0082 | −.023 | −.0023 | −.0095 | .055 | .015 |
| | (.019) | (.018) | (.026) | (.011) | (.029) | (.015) | (.012) | (.021) | (.034) | (.034) |
| Poor soil × flood duration | −.12 | .10 | .047 | .091 | −.14 | −.0021 | −.18 | .016 | −.23 | −.075 |
| | (.084) | (.078) | (.94) | (.22) | (.47) | (.025) | (.092) | (.061) | (.25) | (.19) |
| Missing soil × flood duration | −.46 | −.22 | −.63 | −.21 | −.31 | −.39 | −.59 | −.45 | .042 | −.25 |
| | (.17) | (.12) | (.48) | (.42) | (.53) | (.25) | (.48) | (.50) | (.52) | (.29) |

TABLE B2 (*Continued*)

| | OLS | | 25th percentile | | 50th percentile | | 75th percentile | | Interquartile Range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Log Land (1) | Log Labor (2) | Log Land (3) | Log Labor (4) | Log Land (5) | Log Labor (6) | Log Land (7) | Log Labor (8) | Log Land (9) | Log Labor (10) |
| Fair soil × total rain | .087 | −.0070 | .15 | −.013 | .12 | .037 | .10 | .083 | −.045 | .095 |
| | (.029) | (.027) | (.036) | (.035) | (.032) | (.026) | (.034) | (.028) | (.043) | (.037) |
| Poor soil × total rain | .087 | −.027 | .12 | .031 | .15 | .060 | .060 | −.015 | −.058 | −.046 |
| | (.079) | (.068) | (.089) | (.062) | (.071) | (.068) | (.085) | (.073) | (.13) | (.084) |
| Missing soil × total rain | −.0011 | .00012 | −.0014 | −.00016 | −.00064 | .000064 | −.00069 | .00047 | .00067 | .00063 |
| | (.00026) | (.00020) | (.00022) | (.00024) | (.00021) | (.00018) | (.00029) | (.00021) | (.00033) | (.00042) |
| Fair soil × peak rain | −.00011 | .00013 | −.00025 | .00016 | −.000075 | .000071 | −.000086 | −.000083 | .00016 | −.00024 |
| | (.00006) | (.00006) | (.00008) | (.00008) | (.00007) | (.00006) | (.00008) | (.00006) | (.00011) | (.000073) |
| Poor soil × peak rain | −.00020 | .000014 | −.00039 | −.000000070 | −.000059 | .000016 | −.00019 | −.00013 | .00020 | −.00013 |
| | (.00017) | (.00013) | (.00018) | (.00016) | (.00017) | (.00013) | (.00021) | (.00017) | (.00025) | (.00025) |
| Missing soil × peak rain | −.00029 | .00011 | −.000047 | .00018 | −.00020 | .00021 | −.00016 | .00017 | −.00011 | −.000014 |
| | (.00022) | (.00018) | (.00024) | (.00023) | (.00025) | (.00018) | (.00025) | (.00021) | (.00034) | (.00043) |
| Missing soil | 1.48 | .24 | 1.54 | .56 | .82 | .16 | 1.52 | −.55 | −.024 | −1.11 |
| | (.41) | (.26) | (.33) | (.22) | (1.34) | (.26) | (.41) | (.96) | (.69) | (.56) |
| Rainfed | −.15 | −.062 | −.14 | −.14 | −.14 | −.057 | −.23 | −.030 | −.083 | .11 |
| | (.038) | (.027) | (.050) | (.033) | (.040) | (.020) | (.049) | (.020) | (.063) | (.036) |
| Wetland | −.20 | −.047 | −.18 | −.13 | −.18 | −.083 | −.26 | −.035 | −.079 | .097 |
| | (.049) | (.038) | (.068) | (.047) | (.048) | (.038) | (.065) | (.036) | (.083) | (.062) |
| Missing water | −.55 | −.51 | −.75 | −.45 | −.80 | −.48 | −.89 | −.20 | −.14 | .25 |
| | (.22) | (.18) | (.055) | (.11) | (1.46) | (.072) | (.052) | (.95) | (.41) | (.27) |
| Flat land | −.025 | −.036 | −.020 | −.043 | −.0083 | −.012 | .027 | −.019 | .048 | .025 |
| | (.018) | (.013) | (.021) | (.019) | (.018) | (.014) | (.019) | (.014) | (.022) | (.026) |
| Gentle slope | −.066 | −.0087 | −.091 | −.020 | −.078 | .0035 | −.026 | −.010 | .065 | .010 |
| | (.017) | (.013) | (.020) | (.018) | (.017) | (.014) | (.019) | (.014) | (.023) | (.023) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Steep slope | −.072 | −.029 | −.100 | −.055 | −.051 | −.0031 | −.038 | .015 | .062 | .070 |
| | (.027) | (.022) | (.034) | (.029) | (.028) | (.021) | (.032) | (.023) | (.041) | (.034) |
| Valley | −.024 | −.0051 | −.064 | −.023 | −.058 | .00070 | −.032 | .0067 | .032 | .029 |
| | (.035) | (.025) | (.035) | (.034) | (.041) | (.029) | (.049) | (.026) | (.054) | (.047) |
| Other slope | −.33 | .29 | −.56 | .14 | −.15 | .53 | −.27 | .43 | .29 | .29 |
| | (.24) | (.23) | (.51) | (.95) | (.15) | (.051) | (.60) | (.17) | (.45) | (.42) |
| Missing toposequence | .31 | .035 | .54 | −.12 | .50 | .12 | .16 | .082 | −.38 | .20 |
| | (.22) | (.16) | (.25) | (.20) | (.22) | (.17) | (.047) | (.051) | (.43) | (.33) |
| No erosion | .013 | −.021 | .018 | −.031 | .026 | −.023 | −.022 | −.021 | −.041 | .0096 |
| | (.012) | (.0093) | (.013) | (.012) | (.012) | (.0095) | (.014) | (.010) | (.013) | (.010) |
| Missing erosion | −.12 | −.0047 | −.0054 | .00078 | −.0038 | .037 | −.25 | .078 | −.24 | .077 |
| | (.11) | (.088) | (.13) | (.086) | (.098) | (.17) | (.23) | (.034) | (.23) | (.13) |
| HH received advice on agricultural production (AGSEC9) | .13 | .043 | .14 | .051 | .13 | .041 | .090 | .027 | −.048 | −.024 |
| | (.0098) | (.0085) | (.012) | (.011) | (.011) | (.0088) | (.010) | (.0093) | (.014) | (.011) |
| Constant | −7.3E−09 | 1.0E−09 | −.52 | −.38 | .012 | .059 | .54 | .45 | 1.06 | .83 |
| | (.0041) | (.0034) | (.0051) | (.0047) | (.0048) | (.0037) | (.0051) | (.0038) | (.0052) | (.0044) |
| *F*-statistics: | | | | | | | | | | |
| For joint significance of instruments | 37.1 | 14.7 | 30.4 | 20.9 | 34.6 | 14.6 | 34.9 | 11.6 | | |
| *p*-value | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | | |
| For $h_0$[b] | | | | | | | | | 71.2 | 36 |
| *p*-value | | | | | | | | | .00 | .00 |

Note.—$N$ = 43,187. Bootstrapped (500 samples) standard errors, clustered at the household (HH) level, are in parentheses. V = village with; e.g., "V poor soil" means "village with poor soil."

[a] Variable serves as instrument in table 9.

[b] h0: coefficients are equal for 25th and 75th percentiles.

**References**


Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83 (6): 2411–51.

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

Adamopoulos, Tasso, and Diego Restuccia. 2014. "The Size Distribution of Farms and International Productivity Differences." *A.E.R.* 104 (6): 1667–97.

———. 2015. "Land Reform and Productivity: A Quantitative Analysis with Micro Data." Working Paper no. 540, Dept. Econ., Univ. Toronto.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic Inputs and Resource (Mis)Allocation." *J.P.E.* 122 (5): 1013–63.

Banerjee, Abhijit V., and Benjamin Moll. 2010. "Why Does Misallocation Persist?" *American Econ. J. 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. 2012. "Reliability of Recall in Agricultural Data." *J. Development Econ.* 98 (1): 34–41.

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

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

———. 1995. "Can Unobserved Land Quality Explain the Inverse Productivity Relationship?" *J. Development Econ.* 46 (1): 51–84.

Bento, Pedro, and Diego Restuccia. 2017. "Misallocation, Establishment Size, and Productivity." *American Econ. J. Macroeconomics* 9 (3): 267–303.

Bevis, Leah, and Christopher B. Barrett. 2017. "Close to the Edge: Do Behavioral Explanations Account for the Inverse Productivity Relationship?" Manuscript, Dept. Appl. Econ. and Management, Cornell Univ.

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

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

Carletto, Calogero, Sydney Gourlay, and Paul Winters. 2015. "From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis." *J. African Econ.* 24 (5): 593–628.

Caselli, Francesco. 2005. "Accounting for Cross-Country Income Differences." In *Handbook of Economic Growth*, vol. 1A, edited by Philippe Aghion and Steven Durlauf, 679–741. Amsterdam: North-Holland.

Chen, Chaoran. 2016. "Untitled Land, Occupational Choice, and Agricultural Productivity." Working Paper no. 573, Dept. Econ., Univ. Toronto.

Claassen, Roger, and Richard E. Just. 2011. "Heterogeneity and Distributional Form of Farm-Level Yields." *American J. Agricultural Econ.* 93 (1): 144–60.

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

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

De Nicola, Francesca, and Xavier Giné. 2014. "How Accurate Are Recall Data? Evidence from Coastal India." *J. Development Econ.* 106:52–65.

Emran, M. Shahe, and Forhad J. Shilpi. 2015. "Do Land Market Restrictions Hinder Structural Change in a Rural Economy? Evidence From Sri Lanka." Policy Res. Working Paper no. 7525, World Bank, Washington, DC.

Esfahani, Sara. 2018. "Agricultural Misallocation: Mismeasurement, Misspecification, or Market Frictions?" Manuscript, Dept. Econ. and Internat. Bus., Sam Houston State Univ.

Florens, Jean-Pierre, James J. Heckman, Costas Meghir, and Edward Vytlacil. 2008. "Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects." *Econometrica* 76 (5): 1191–1206.

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

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

García-Santana, Manuel, and Josep Pijoan-Mas. 2014. "The Reservation Laws in India and the Misallocation of Production Factors." *J. Monetary Econ.* 66:193–209.

Gottlieb, Charles, and Jan Grobovšek. 2019. "Communal Land and Agricultural Productivity." *J. Development Econ.* 138:135–52.

Gourlay, Sydney, Talip Kilic, and David Lobell. 2017. "Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for Inverse Scale-Productivity Relationship in Uganda." Manuscript, Living Standards Measurement Study Survey Unit, Development Data Group, World Bank, Rome.

Guner, Nezih, Gustavo Ventura, and Xu Yi. 2008. "Macroeconomic Implications of Size-Dependent Policies." *Rev. Econ. Dynamics* 11 (4): 721–44.

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

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

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning through Noticing: Theory and Evidence from a Field Experiment." *Q.J.E.* 129 (3): 1311–53.

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." *J. Human Resources* 33 (4): 974–87.

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

Hopenhayn, Hugo A. 2014. "Firms, Misallocation, and Aggregate Productivity: A Review." *Ann. Rev. Econ.* 6 (1): 735–70.

Hsieh, Chang-Tai, and Peter Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Q.J.E.* 124 (4): 1403–48.

Hurley, Terrance M., Gary L. Malzer, and Bernard Kilian. 2004. "Estimating Site-Specific Nitrogen Crop Response Functions." *Agronomy J.* 96 (5): 1331–43.

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

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

Kilic, Talip, Alberto Zezza, Calogero Carletto, and Sara Savastano. 2016. "Missing(ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements." *World Development* 92:143–57.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Rev. Econ. Studies* 70 (2): 317–41.

Lobell, David B., George Azzari, Marshall Burke, et al. 2018. "Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis in Uganda." Policy Res. Working Paper no. 8374, World Bank, Washington, DC.

Maksimovic, Vojislav, and Gordon Phillips. 2001. "The Market for Corporate Assets: Who Engages in Mergers and Asset Sales and Are There Efficiency Gains?" *J. Finance* 56 (6): 2019–65.

Masten, Matthew A., and Alexander Torgovitsky. 2016. "Identification of Instrumental Variable Correlated Random Coefficients Models." *Rev. Econ. and Statis.* 98 (5): 1001–05.

Matvos, Gregor, and Amit Seru. 2014. "Resource Allocation within Firms and Financial Market Dislocation: Evidence from Diversified Conglomerates." *Rev. Financial Studies* 27 (4): 1143–89.

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

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *A.E.R.* 104 (2): 422–58.

Olley, G. Steven, and Ariel Pakes. 1992. "The Dynamics of Productivity in the Telecommunications Equipment Industry." Working Paper no. 3977 (January), NBER, Cambridge, MA.

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

Petrin, Amil, and Jagadeesh Sivadasan. 2013. "Estimating Lost Output from Allocative Inefficiency, with an Application to Chile and Firing Costs." *Rev. Econ. and Statis.* 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." *Rev. Econ. Dynamics* 14 (1): 3–26.

Porzio, Tommaso. 2016. "Cross-Country Differences in the Optimal Allocation of Talent and Technology." Manuscript, Dept. Econ., Univ. California, San Diego.

Restuccia, Diego. 2018. "Misallocation and Aggregate Productivity across Time and Space." Working Paper no. 596, Dept. Econ., Univ. Toronto.

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

———. 2013. "Misallocation and Productivity." *Rev. Econ. Dynamics* 16 (1): 1–10.

———. 2017. "The Causes and Costs of Misallocation." *J. Econ. Perspectives* 31 (3): 151–74.

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

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

Rotemberg, Martin, and T. Kirk White. 2017. "Measuring Cross-Country Differences in Misallocation." Manuscript, Dept. Econ., New York Univ.

Schimmelpfennig, David. 2016. "Farm Profits and Adoption of Precision Agriculture." Report no. 217, Econ. Res. Service, US Dept. Agriculture.

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

Ševčík, Pavel. 2015. "Financial Frictions, Internal Capital Markets, and the Organization of Production." *Rev. Econ. Dynamics* 18 (3): 505–22.

Shenoy, Ajay. 2017. "Market Failures and Misallocation." *J. Development Econ.* 128:65–80.

Singh, Inderjit, Lyn Squire, and John Strauss, eds. 1986. *Agricultural Household Models: Extensions, Applications, and Policy.* Baltimore: Johns Hopkins Univ. Press (for the World Bank).

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,* edited by Rattan Lal and B. A. Stewart, 37–72. Boca Raton, FL: CRC.

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

Syverson, Chad. 2004. "Product Substitutability and Productivity Dispersion." *Rev. Econ. and Statis.* 86 (2): 534–50.

———. 2011. "What Determines Productivity?" *J. Econ. Literature* 49 (2): 326–65.

Tittonell, Pablo, Keith D. Shepherd, Bernard Vanlauwe, and Ken E. Giller. 2008. "Unravelling the Effects of Soil and Crop Management on Maize Productivity in Smallholder Agricultural Systems of Western Kenya—An Application of Classification and Regression Tree Analysis." *Agriculture, Ecosystems and Environment* 123 (1–3): 137–50.

Tittonell, Pablo, Bernard Vanlauwe, Nico 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): 376–90.

Tittonell, Pablo, Bernard Vanlauwe, Peter 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 (3–4): 166–84.

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

Udry, Christopher. 1996. "Gender, Agricultural Production, and the Theory of the Household." *J.P.E.* 104 (5): 1010–46.

Valentinyi, Akos, and Berthold Herrendorf. 2008. "Measuring Factor Income Shares at the Sectoral Level." *Rev. Econ. Dynamics* 11 (4): 820–35.

Vanlauwe, Bernard, Katrien Descheemaeker, Ken E. Giller, et al. 2015. "Integrated Soil Fertility Management in Sub-Saharan Africa: Unravelling Local Adaptation." *Soil* 1 (1): 491–508.

Vanlauwe, Bernard, Pablo Tittonell, and John Mukalama. 2006. "Within-Farm Soil Fertility Gradients Affect Response of Maize to Fertiliser Application in Western Kenya." *Nutrient Cycling Agroecosystems* 76 (2–3): 171–82.

White, T. Kirk, Jerome P. Reiter, and Amil Petrin. 2018. "Imputation in U.S. Manufacturing Data and Its Implications for Productivity Dispersion." *Rev. Econ. and Statis.* 100 (3): 502–9.

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