Agricultural Misallocation: Mismeasurement, Misspecification, or Market Frictions?

Job Market Paper

Sara Esfahani*
Department of Economics
University of Notre Dame

November, 2017

Abstract

In developing countries, most people work in agriculture and agricultural incomes are low. Misallocation of agricultural resources has been shown to be an important detriment to agricultural productivity. Previous studies have focused on misallocation across households, which could arise due to frictions in credit or factor markets. I use detailed plot-level data from Ethiopia and Malawi that allows me to measure misallocation both within and across households. I find that the within-household misallocation across plots is substantially larger than between-household misallocation: approximately 70% of the overall misallocation is associated to within-household misallocation, even after controlling for many observable characteristics. Critically, I argue that market-level distortions generate within-household misallocation. In particular, the lack of land markets leads to households with spatially segregated plots, which are hard to optimize across. This accounts for half of the total share of within-household variations. A critical implication of agricultural market distortions is therefore their ability to distort within-household decision making. Not taking this effect on within-household misallocation into account, therefore, understates the gains from improving rural land markets.

*I am grateful to my adviser Joseph Kaboski, for his valuable comments and suggestions. I am indebted to the members of my committee, Kevin Donovan and Lakshmi Iyer for their patience and support. I also thank Abigail Wozniak, Wyatt Brooks, Douglas Gollin, Terence Johnson, and Nilesh Fernando for helpful conversations and comments. This study has benefited from presentations in the Macro and Development Seminars at the University of Notre Dame, H2D2 conference at the University of Michigan, and CEA meeting at Nova Scotia's St Francis Xavier University. Any errors are my own responsibility. Contact Sara Esfahani at Sara.Esfahani.1@nd.edu.

1 Introduction

Poor countries allocate a large fraction of their economic resources to the agricultural sector, despite the fact that agriculture is far less productive in these countries. Previous studies have documented an output per worker difference of a factor of 60 between the richest and the poorest 10 percent of countries.[1] A predominant set of theories and quantitative work suggest that a sizable share of this gap is due to misallocation in the agricultural sector in developing countries. A common feature of these theories is their reliance on market-based distortions, such as land, labor, credit, or insurance market frictions that generate misallocation across households but not necessarily within households across production units.[2] For example, a standard collateral constraint (i.e. the amount of capital available to an entrepreneur is limited by his personal assets) potentially distorts the capital intensity of poor entrepreneurs relative to rich, but does not necessarily imply misallocation across production units of the same entrepreneur.[3] On the other hand, a large microeconomics literature has long emphasized the importance of understanding within-household distortions (Udry (1996)).

In this paper, I use detailed micro-level panel data from Ethiopia and Malawi to quantify the relevant margins of agricultural misallocation in labor, intermediate inputs, and land. I adopt an "indirect approach", in which distortions are backed out from variation in factor ratios (Restuccia and Rogerson (2008), Hsieh and Klenow (2009)). In this setting the efficient resource allocation is achieved when all production units equalize the marginal products, which also implies equalized factor ratios across production units. My contribution is to explicitly consider within-household misallocation by studying variation in factor ratios across plots cultivated by the same household. This analysis is made possible by the fact that the data from Ethiopia and Malawi provide a detailed accounting of inputs and outputs at the plot level. Moreover, a large fraction of households cultivate more than one plot of land (roughly 80% of Ethiopian households and 65% of Malawian households), which allows me to study within-household resource allocation.

The first important result of this paper is that most variation in factor ratios is within households. Using the standard misallocation methodology, I find that 60-80% of the overall variation in the factor ratios is due to variation within households. I argue that the market-level frictions are at the core of misallocation within households, and not only across households, unlike what the previous literature shows. I provide a mechanism for which land market frictions can generate misallocation within households. I find that the lack of land markets leads to households with spatially segregated plots that are hard to optimize across, which then leads to within-household misallocation. I find that within-household variation in the distance of plots from home is 4.5 times larger for households who do not have marketed land and larger distances between plots makes it harder for the manager to optimize the allocation of resources across these distanced plots. Rela-

[1] See for example Gollin et al. (2002), Restuccia et al. (2008), and Gollin et al. (2014).

[2] See for example Restuccia et al. (2008), Buera et al. (2011), Mobarak and Rosenzweig (2012), Moll (2014), Karlan et al. (2014), Adamopoulos and Restuccia (2015), Restuccia and Rogerson (2016), Donovan (2016), Restuccia and Santaeulalia-Llopis (2017), and Cole et al. (2017).

[3] See for example Buera et al. (2011) and Moll (2014).

tive size of within-household variation in factor ratios decreases from 60-85% for households with distanced plots to 30-60% for households with plots that are close to one another. This finding implies that understanding the relative importance of within-household misallocation requires an understanding of how market frictions can generate within-household misallocation.

I also consider other possibilities, and show that they are significantly less important quantitatively. In particular, I consider two broad explanations for large within-household variations in factor ratios. First, all measured misallocation may not be true misallocation, rather it reflects measurement error or misspecification of the production function. I find evidence of both measurement error and misspecification due to non-convexities in fertilizer usage. These factors seem to lower both within- and between-household misallocation equally. Second, non-market frictions may lead to misallocation within the household. Differences in within-household bargaining power (e.g. husband vs. wife-cropped plots) do not seem to be a major force, but the location of plots is important.

The richness of the data allows me to examine measurement issues in several ways. First I control for a set of observable characteristics across plots such as land quality, land size, plot distance from household, variation in planted crops across plots, seed types, weather and other transitory shocks, and the characteristics of the decision maker on each plot. I also find that households misreport their land size and labor hours. Respondents substantially over-report labor hours for other household members compared to hours reported for own. Controlling for all these sources of variation decreases total inferred misallocation by about 20%. However, these features are not related exclusively to within-household misallocation. In fact, the variation implied by them translates into both within- and between-household inferred misallocation. Therefore, while total misallocation falls after isolating these features, the relative importance of within-household misallocation does not.

The second possibility is that variation in factor ratios may be driven by misspecification of the production function. Specifically, the assumption of diminishing returns in all factors, and all points of the production function may not be valid. This can be due to a fixed cost or a minimum requirement, in fertilizer, for example. Therefore, if a household does not have enough fertilizer for all the plots, they may apply it to some, and not all their plots. I find evidence of non-convexities in fertilizer. Specifically, I find that poor households don't use any fertilizer and richer households use fertilizer on all of their plots. The middle income households however, fertilize some, but not all their plots. The use of an input on some and not all plots generates larger within-household variations in factor ratios. Share of within-household variations are about 10% smaller among farmers who either do not fertilize their plots or fertilize all their plots compared to those who partially fertilize their plots.

Since both of these broad explanations provide substantially smaller explanatory power than land markets, in the last section of this paper, I return to market-based distortions and unlike the previous literature, I investigate how these market-based frictions generate within-household misallocation. I specifically consider land market frictions and show that the lack of land markets leads to households having plots that are spatially distanced and hard to optimize across. The data confirms this theory, and counteractions imply that this accounts for almost half of the total share of within-household variations. I also briefly investigate risk as another market friction that can manifest within households. Ogaki and Zhang (2001) and Donovan (2016) show that subsistence requirements can amplify risk concerns for households with low consumption levels. I find that within-household variance in intermediate input per square meter of land in Malawi is about 2 times larger for poorer households compared to wealthier households. This finding suggests that future studies need to consider risk as another source of within-household misallocation.

The remainder of the paper is organized as follows. Section 2 review the literature. The theoretical framework is presented in section 3, where the efficient allocation and variance decomposition are presented. Section 4 explains the data. Section 5 presents the findings on resource allocation and within versus between-household variations in factor ratios. In Section 6, I investigate possible explanations for observed variations within households. Since neither measurement or technological issues seem to be the lone driving force behind understanding the high share of within-household misallocation, in Section 7, I explain how market distortions such as the lack of land markets and risk manifest within households and provide evidence from the data to support the arguments. Section 8 concludes.

2 Related Literature

My paper contributes to several existing literatures. A large body of the literature studies misallocation of resources across production units within a sector through studying the variation in factor ratios, a method Restuccia and Rogerson (2016) refer to as the "indirect approach" in which distortions are backed out from a model (Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Bartelsman et al. (2013), Restuccia and Rogerson (2016), Restuccia and Santaeulalia-Llopis (2017), and Gollin and Udry (2017)). I also apply this indirect approach; however, unlike these studies in which the focus of the analysis is on misallocation across factories or farm-households, I explicitly consider resource allocation across plots within a household.

Misallocation across production units has been studied in both micro and macro literature and the use of farm household data has contributed significantly to understanding efficiency in this literature. Among micro development literature, Udry (1996)'s work associates within-household misallocation to the bargaining power between husband- and wife-cropped farms. In contrast, I show that the gender mix of the managers of plots within a household does not seem to be a major source in Ethiopia and Malawi. In the macro literature, this procedure has been used to motivate a number of distortions in developing countries, including financial frictions, land market and labor market distortions (Benjamin (1992), LaFave and Thomas (2016), and Shenoy (2017).). My paper complements these findings by showing that market frictions can manifest themselves into within-household distortions.

Several studies argue land market frictions are at the core of low agricultural productivity in poor countries (Restuccia et al. (2008), Adamopoulos and Restuccia (2015), Restuccia and Rogerson (2016), and Restuccia and Santaeulalia-Llopis (2017).), but the focus of all these studies is on misallocation across households. I find that the lack of land markets distorts the resource allocation even within the household by leading to households having plots that are distanced from one another and hard to optimize across. Another stream of the literature studies misallocation considering risk and the lack of insurance markets in developing countries (Mobarak and Rosenzweig (2012), Karlan et al. (2014), Donovan (2016), and Cole et al. (2017)). I do not find evidence of risk distorting within-household allocation of resources.

This paper also contributes to the literature considering measurement error in LSMS data. Since the use of detailed micro data in studying factor productivity and misallocation has become more popular, concerns about issues with measurement have raised. Several studies in the literature document measurement error in labor and land size in micro data (Goldstein and Udry (1999), Keita and Carfagna (2009), Garlick et al. (2016), Dillon et al. (2016), Arthi et al. (2016), and Bils et al. (2017)). Findings of my paper about land mismeasurement are consistent with the literature. With regards to mismeasurement in labor hours, my paper provides new findings indicating significant differences between hours reported by the respondent for self versus hours reported for other household members.

My paper is most closely related to a concurrent study by Gollin and Udry (2017), who relate dispersion in productivity across plots within a farm to three sources: idiosyncratic shocks to agriculture, measurement error, and heterogeneity in unobserved land quality. They assume that within a farm, the allocation of resources across plots is efficient. Then any dispersion across plots within a farm is measurement error or varying realization of risk. Given this assumption, they correct the data to study misallocation across farms and conclude that misallocation leads to substantial output loss, but it cannot explain much of the cross-country income differences. I show that not all within-household dispersion across plots is driven by measurement error, and market frictions can lead to misallocation of resources within a household. In the next section I provide the theoretical environment in which efficient allocation of resources is defined and explain the variance decomposition technique used in the analysis.

3 Theoretical Framework

In this section I present the model, drive the variance of factor ratios as a measure of misallocation, and show how these variances can be decomposed into within-household and between-household variances following Davis and Haltiwanger (1991)'s variance decomposition method. Lastly, I will show the mapping between the variance decomposition and the model.

I use a more general form of the production function used by Restuccia and Santaeulalia-Llopis (2017) and incorporate multi-plot households, and generalize production into a CES, where technology varies across plots.[4] Production at household i, plot j planted to crop c is defined by a

[4] Restuccia and Santaeulalia-Llopis (2017) have a farm-level Cobb-Douglas production function and focus their

Cobb-Douglas between land and a CES aggregate of labor and intermediate inputs:

$$Y_{ijc} = a_{ijc} \cdot T_{ij}^{\gamma_c} \cdot (\alpha_c \, L_{ij}{}^{\rho} + (1-\alpha_c) \, M_{ij}{}^{\rho})^{\frac{1-\gamma_c}{\rho}}$$

where Y_{ijc} is quantity of output in household i, plot j planted to crop c, a_{ijc} is plot-crop-level productivity, T_{ij} is plot size, M_{ij} is intermediate input expenditures, L_{ij} is total labor hours at household i, plot j planted to crop c, α_c is the share parameter and γ_c captures the extent of decreasing return from land, both specific to crop c.

3.1 Within-Household Optimal Allocation

Household i's total value of production; Y_i is the sum of production across all plots within the household evaluated at market prices P_c for crop c, where the output distortion is denoted by τ^Y:

$$Y_i = \sum_{j,c} (1-\tau_i^Y)(1-\tau_{ij}^Y) P_c Y_{ijc}$$

where τ_i^Y represents output distortions across households and τ_{ij}^Y captures output distortions within household i across plots. These distortions could all be folded into one distortion, but I separate them for the purpose of the analysis in this paper to highlight the relative importance of each. The household chooses the total amount of labor, land, and intermediate inputs in the market and after choosing what crop to be planted on each plot, household allocates these aggregate household resources across plots within the household to maximize profits at the household level:

$$\begin{aligned}
\max_{T_{ij}, L_{ij}, M_{ij}, \mathbb{I}_{ijc}} \quad & \sum_{j,c} \mathbb{I}_{ijc}(1-\tau_i^Y)(1-\tau_{ij}^Y) P_c Y_{ijc} - (1+\tau_i^T)(1+\tau_i^T) r T_i \\
& - (1+\tau_i^L)(1+\tau_{ij}^L) w L_i - (1+\tau_i^M)(1+\tau_{ij}^M) q M_i \\
\text{s.t.} \quad & T_i = \sum_j T_{ij} \\
& L_i = \sum_j L_{ij} \\
& M_i = \sum_j M_{ij} \\
& Y_{ijc} = a_{ijc} \cdot T_{ij}^{\gamma_c} \cdot (\alpha_c \, L_{ij}{}^{\rho} + (1-\alpha_c) \, M_{ij}{}^{\rho})^{\frac{1-\gamma_c}{\rho}}
\end{aligned}$$

where $\mathbb{I}_c$ is an indicator function that takes the value of 1 if household i's plot j is planted with crop c, and zero otherwise; T_i is total household i's land holding, L_i is total household i's labor, M_i is the total amount of intermediate inputs; r is the market rental rate of land, w is the market wage for labor, and q is the market price for intermediate inputs; and the distortions that increase the effective marginal cost of land, labor, and intermediate inputs are denoted by τ^T, τ^L, and τ^M,

analysis on the allocation of capital and land and abstract from labor differences by writing the inputs and output per hour.

respectively.[5] For example, τ_i^T represents land market frictions across households and τ_{ij}^T captures how land market frictions distort allocation of resources across plots within household i. Similar to the output distortions, these factor market distortions can all be folded into one distortion, but I separate them for the purpose of the analysis in this paper to highlight the distinct role of each.

Assuming μ as the multiplier for the land holding constraint, λ as the multiplier for the labor hours constraint, and η as the multiplier for the intermediate inputs constraint, from the first order conditions with respect to T_{ijt}, L_{ijt} and M_{ijt}, one can solve for the ratio of factor inputs across all j's in household i, planted to crop c:

$$\frac{L_{ij}}{M_{ij}} = \left(\frac{\eta q(1+\tau_i^M)(1+\tau_{ij}^M)}{\lambda w(1+\tau_i^L)(1+\tau_{ij}^L)} \times \frac{\alpha_c}{1-\alpha_c} \right)^{\frac{1}{1-\rho}} \tag{1}$$

$$\frac{T_{ij}}{L_{ij}} = \frac{\gamma_c}{1-\gamma_c} \frac{\lambda w(1+\tau_i^L)(1+\tau_{ij}^L)}{\mu r(1+\tau_i^T)(1+\tau_{ij}^T)} \left(1 + \left(\frac{\lambda w(1+\tau_i^L)(1+\tau_{ij}^L)}{\eta q(1+\tau_i^M)(1+\tau_{ij}^M)} \right)^{\frac{\rho}{1-\rho}} \left(\frac{1-\alpha_c}{\alpha_c} \right)^{\frac{1}{1-\rho}} \right) \tag{2}$$

$$\frac{T_{ij}}{M_{ij}} = \frac{\gamma_c}{1-\gamma_c} \frac{\eta q(1+\tau_i^M)(1+\tau_{ij}^M)}{\mu r(1+\tau_i^T)(1+\tau_{ij}^T)} \left(1 + \left(\frac{\eta q(1+\tau_i^M)(1+\tau_{ij}^M)}{\lambda w(1+\tau_i^L)(1+\tau_{ij}^L)} \right)^{\frac{\rho}{1-\rho}} \left(\frac{1-\alpha_c}{\alpha_c} \right)^{\frac{1}{1-\rho}} \right) \tag{3}$$

It is important to notice that the within-household frictions; τ_{ij}^T, τ_{ij}^L, and τ_{ij}^M, are important determinants of the optimal allocation of household resources across plots within the household, however between-household distortions; τ_i^T, τ_i^L, and τ_i^M, are constant with household i across all plots, so they cannot lead to within household variation in factor ratios across plots. In the next section, I show the efficient allocation of resources across plots within a household in the absence of within-household distortions.

Within-Household Efficient Allocation

Suppose no distortions exist within the household. Then the distortion terms; τ_{ij}^T, τ_{ij}^L, and τ_{ij}^M, can not affect the shadow value of factor inputs. Then all factor ratios will be constant within a household across plots. From equations (1), (2) and (3), where $\tau_{ij}^T = \tau_{ij}^L = \tau_{ij}^M = 0$, one can show the following is true at the efficient allocation across all j's in household i, planted to crop c:

$$\frac{L_{ij}}{M_{ij}} = \left(\frac{\eta q(1+\tau_i^M)}{\lambda w(1+\tau_i^L)} \times \frac{\alpha_c}{1-\alpha_c} \right)^{\frac{1}{1-\rho}}$$

$$\frac{T_{ij}}{L_{ij}} = \frac{\gamma_c}{1-\gamma_c} \frac{\lambda w(1+\tau_i^L)}{\mu r(1+\tau_i^T)} \left(1 + \left(\frac{\lambda w(1+\tau_i^L)}{\eta q(1+\tau_i^M)} \right)^{\frac{\rho}{1-\rho}} \left(\frac{1-\alpha_c}{\alpha_c} \right)^{\frac{1}{1-\rho}} \right)$$

[5] The signs of the distortions are defined to be analogous to taxes, but they may also be negative, representing effective subsidies.

$$\frac{T_{ij}}{M_{ij}} = \frac{\gamma_c}{1-\gamma_c}\frac{\eta q(1+\tau_i^M)}{\mu r(1+\tau_i^T)}\left(1+\left(\frac{\eta q(1+\tau_i^M)}{\lambda w(1+\tau_i^L)}\right)^{\frac{\rho}{1-\rho}}\left(\frac{1-\alpha_c}{\alpha_c}\right)^{\frac{1}{1-\rho}}\right)$$

where τ_i^T, τ_i^L, and τ_i^M are constant across plots within household i. In other words, at the efficient allocation, household allocates resources across plots such that the ratio of any two factor inputs in equalized across all plots within the household. Any deviations in data from the above efficient allocation criteria will be interpreted through the lens of the model as misallocation within the household. Therefore, the critical implication of the model is that the variance of factor ratios is zero across plots within a household at the efficient allocation.

Recall from equations (1), (2) and (3), variation in the factor ratios across all household-plots has two sources: (1) distortions within households and (2) market frictions distorting factors between households. The goal of this paper is to decompose this overall variance in the factor ratios into within- and between-household components. I show how to accomplish this using a simple additive variance decomposition in the next section.

3.2 Variance Decomposition Method

Following Davis and Haltiwanger (1991)'s variance decomposition, I decompose total variance in the factor ratios for all household-plots across all years into a within-household and a between-household variance component. Let R_{ijt} be the ratio of any two factors under consideration. Variance of R_{ijt} across all household-plots and across years is

$$V = \frac{1}{N}\sum_i \sum_j (R_{ij} - \bar{R})^2 \tag{4}$$

where N is the total number of household-plot, and $\bar{R}$ is the average factor ratio across all household-plot and is defined as

$$\bar{R} = \frac{1}{N}\sum_i \sum_j R_{ij}$$

I define R_i as within household i average of R_{ij}'s such that

$$R_i = \frac{1}{N_i}\sum_j R_{ij}$$

where N_i is the total number of plots within household i. Appendix B shows how one can decompose the total variance across all household-plot into within-household and between-household, so that:

$$V = V_{WH} + V_{BH} \tag{5}$$

Let V_i be the within-household variance of R_{ij} defined as:

$$V_i = \frac{\sum_j (R_{ij} - R_i)^2}{N_i} \tag{6}$$

Then the within-household variance V_{WH} is the average of within household variances weighted by the number of plots within household:

$$V_{WH} = \frac{1}{N} \sum_i N_i V_i \tag{7}$$

Also,

$$V_{BH} = \frac{1}{N} \sum_i N_i (R_i - \bar{R})^2 \tag{8}$$

is the weighted average of the variance between households.

To measure the relative size of within-household variance to the total household-plot variance in data, I define S as the ratio of equation (7) to equation (5), so that

$$S = \frac{V_{WH}}{V} = \frac{\frac{1}{N} \sum_i N_i V_i}{\frac{1}{N} \sum_i \sum_j (R_{ij} - \bar{R})^2} \tag{9}$$

The statistics S is used throughout the paper and indicates the share of the variation in the factor ratios that comes from within-household variations.

3.3 Mapping Variance Measures to the Model

In this section, I show how the variances discussed in Section 3.2 map into the factor ratios in the model. As an example, lets define V^* as the total variance in the ratio of labor to intermediate inputs from equation (1), across all household-plots:

$$\begin{aligned} V^* = Var(\frac{L_{ij}}{M_{ij}}) &= Var\left[\left(\frac{\eta q(1+\tau_i^M)(1+\tau_{ij}^M)}{\lambda w(1+\tau_i^L)(1+\tau_{ij}^L)} \times \frac{\alpha_c}{1-\alpha_c}\right)^{\frac{1}{1+\rho}}\right] \\ &= C \times Var\left[\left(\frac{(1+\tau_i^M)(1+\tau_{ij}^M)}{(1+\tau_i^L)(1+\tau_{ij}^L)}\right)^{\frac{1}{1-\rho}}\right] \end{aligned} \tag{10}$$

where $C = \left(\frac{\eta q}{\lambda w} \times \frac{\alpha_c}{1-\alpha_c}\right)^{\frac{2}{1-\rho}}$ is a constant term. Assuming perfect factor markets and no within-household distortions, total variance would be equal to zero; however, because of the distortions, total variance in the factor ratios will be greater than zero. Also, the variance of the the ratio of labor to intermediate inputs across plots within household i is defined as:

$$V_i^* = Var(\frac{L_{ij}}{M_{ij}}) = B \times Var\left[\left(\frac{(1+\tau_{ij}^M)}{(1+\tau_{ij}^L)}\right)^{\frac{1}{1-\rho}}\right] \tag{11}$$

where $B = \left(\frac{(1+\tau_i^M)}{(1+\tau_i^l)} \times \frac{\eta q}{\lambda w} \times \frac{\alpha_c}{1-\alpha_c}\right)^{\frac{2}{1-\rho}}$ is a constant term across plots within household i. If there are no distortion within the household, then $\tau_{ij}^L = \tau_{ij}^M = 0$, and therefore, $V_i^* = 0$. Within-household variance in the the ratio of labor to intermediate inputs V_{WH}^* is then an average of all V_i^*'s, weighted by the number of plots within household i; N_i :

$$V_{WH}^* = \frac{1}{N}\sum_i\sum_t N_i V_i^* = B \times \sum_i \frac{N_i}{N} Var\left[\left(\frac{1+\tau_{ij}^M}{1+\tau_{ij}^L}\right)^{\frac{1}{1-\rho}}\right] \tag{12}$$

The share of the overall variance in the ratio of labor to intermediate inputs that is due to within-household variations is then defined as the ratio of equation (12) to equation (10):

$$S^* = \frac{B \times \sum_i \frac{N_i}{N} Var\left[\left(\frac{1+\tau_{ij}^M}{1+\tau_{ij}^L}\right)^{\frac{1}{1-\rho}}\right]}{C \times Var\left[\left(\frac{(1+\tau_i^M)(1+\tau_{ij}^M)}{(1+\tau_i^L)(1+\tau_{ij}^L)}\right)^{\frac{1}{1-\rho}}\right]} \tag{13}$$

Notice that the distortions affect both numerator (within-household variance) and the denominator (overall variance) of the share S^*. If no distortion exist within households ($\tau_{ij}^L = \tau_{ij}^M = 0$), the allocation of resources across plots within the households would be efficient, and the numerator of equation (13) equals zero and all variation is due to market frictions and the variation in factor ratios across households. Also, as the distortions within households become relatively larger and more significant, the share of the overall variation due to within-household variance in the factor ratios increases. Therefore, a larger S^* implies that the within-household distortions are relatively more significant than between-household distortions. In the next section, I describe the important elements of the micro data used for the variance analysis in this study.

4 Data

The exercise laid out theoretically in the last section requires detailed information on plot-level inputs and outputs. I therefore use two sets of data from Ethiopia and Malawi, collected by the World Bank in collaboration with the National Central Statistical Agencies as part of the Living Standard Measurement Survey (LSMS) project.[6] LSMS is a household survey program implementing nationally representative panel household surveys with a strong focus on agriculture. Ethiopian data is from the 2011, 2013, and 2015 Ethiopian Socioeconomic Surveys (ESS1, ESS2,

[6]LSMS data is used in studies such as Adamopoulos et al. (2016),

and ESS3 respectively).[7] The data for Malawi is from the 2010 and 2013 Integrated Household Panel Survey (IHPS).[8]

The feature of the data that makes it suitable for the analysis in this paper, is that it provides very detailed information at plot level and more importantly, in both countries, there are several households cultivating the same crop on multiple plots within the same year. This provides a framework in which within-household allocation of resources can be studied.

Both datasets collect data on labor (household, free, and hired labor hours), intermediate inputs (seeds and fertilizer expenditure), and land (size and quality) at household-plot level during the agricultural season. The data also have very detailed information on the characteristics of the household individuals involved in agricultural activities, such as gender, age, education, health, relationship to the household head, and whether individual has spend time away from household.

The agricultural data at plot-level provides information on the crops grown, land size, land ownership status, soil type and soil quality, irrigation, land wetness, elevation, and slope, and shocks to harvest. The data also distinguishes the land owner from the primary decision maker on the plot, and the survey respondent. I use this feature of the data to show larger measurement error in hours reported by the respondent for other household members. See Section 6.1 for more details.

I also use the consumption expenditure survey data to construct a household welfare measure to classify households based on their poverty level. I use this measure to study risk and misallocation. The survey includes questions about expenditure on food and nonfood items for the past week, month, and year. I use food share of household expenditure when studying risk as a potential source of misallocation within households.

Summary statistics of the data can be found in Table 1 and Table 2. Households in Malawi cultivate fewer plots of land than households in Ethiopia (plots are larger in size in Malawi than in Ethiopia). The median number of plots per household in Malawi is 2 compared to 6 in Ethiopia. Ethiopia and Malawi are quite different in terms of crops grown. There are 123 types of crops grown in Ethiopia, but only 27 types of crops in Malawi. This could be one of the reasons for fewer plots of land within Malawian households compared to Ethiopia. Figure 1 and Figure 2 show the distribution of the number of plots in household for Ethiopia and Malawi respectively.

I measure labor input as the total plot level labor hours used during planting and harvesting seasons. This includes mostly household labor for both countries. A small share of labor comes from free/exchange labor and hired labor. This information comes from the agricultural questionnaires. Surveys collect data on total weeks, average days per week, average hours per day per individual per plot by activity type (planting and harvesting for Ethiopia, and planting, other non-harvest, and harvesting for Malawi). Household-plot level labor hours is the sum of all hours from all individuals

[7] ESS began as ERSS (Ethiopian Rural Socioeconomic Survey) in 2011/12. The first wave of data included rural and small town areas. The same households that were interviewed in ESS1 were tracked and re-interviewed in ESS2 and ESS3. In the second and third waves, the sample was expanded to all urban areas as well. ESS1, ESS2, and ESS3 together create a panel dataset of households from rural and small town areas. ESS2 and ESS3 together represent a panel of individuals for rural and all urban areas.

[8] The first round of the panel comprises 3,246 households and the second round has a sample of 4,000 households.

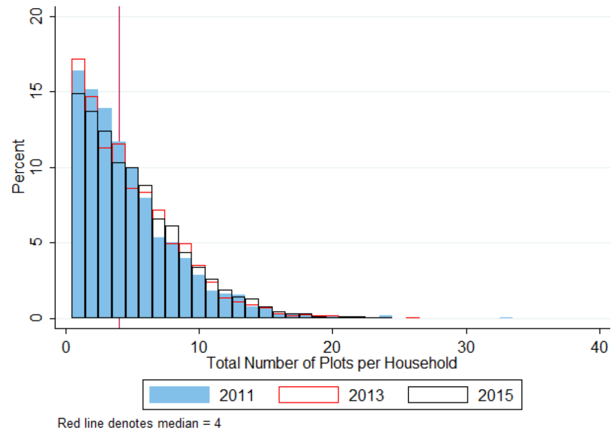


Figure 1: Ethiopia - Distribution of the Total Number of Plots Cultivated by Households

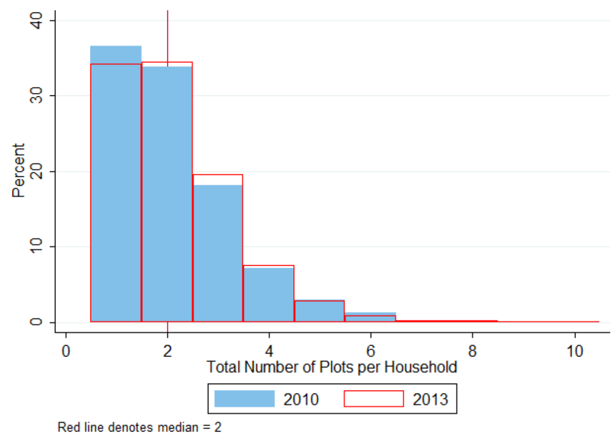


Figure 2: Malawi - Distribution of the Total Number of Plots Cultivated by Households

(household and non-household) for all activities during the season.

Total intermediate input expenditure is the sum of household seed value, fertilizer, herbicide, and pesticide used on each plot. I use the median price from those households who purchase their seeds or fertilizer in the market to evaluate seed and fertilizer values for households who don't purchase their intermediate inputs. Median seed prices are measured based on the seed type and crop type. Median fertilizer prices are calculated based on type of the fertilizer and the network source of the fertilizer. Prices vary across suppliers, and the household network is assumed to determine the prices they are facing.

Finally, the measure of land is the GPS measures or rope and compass measures of land. Data also provides household reported measures of land, but the differences between the two measures are substantial.[9] Also, GPS measures are less accurate for smaller plots due to the error bound on the GPS coordinates of the corners of smaller plots. For these plots rope and compass measures

[9]Figure 4 and Figure 5 show the distribution over the percent difference between measured and self-reported land size for Ethiopia and Malawi respectively.

are used.

The plot size distribution differs slightly across the two countries as well. More than 50% of plots in Ethiopia are less than 0.1 Hectares and about 40% are between 0.1 and 0.5 Hectares. In Malawi however, about 66% of plots have an area between 0.1 and 0.5 Hectares and only about 14% of plots are less than 0.1 Hectares. Table A.2 summarizes the distribution of land size in the sample.

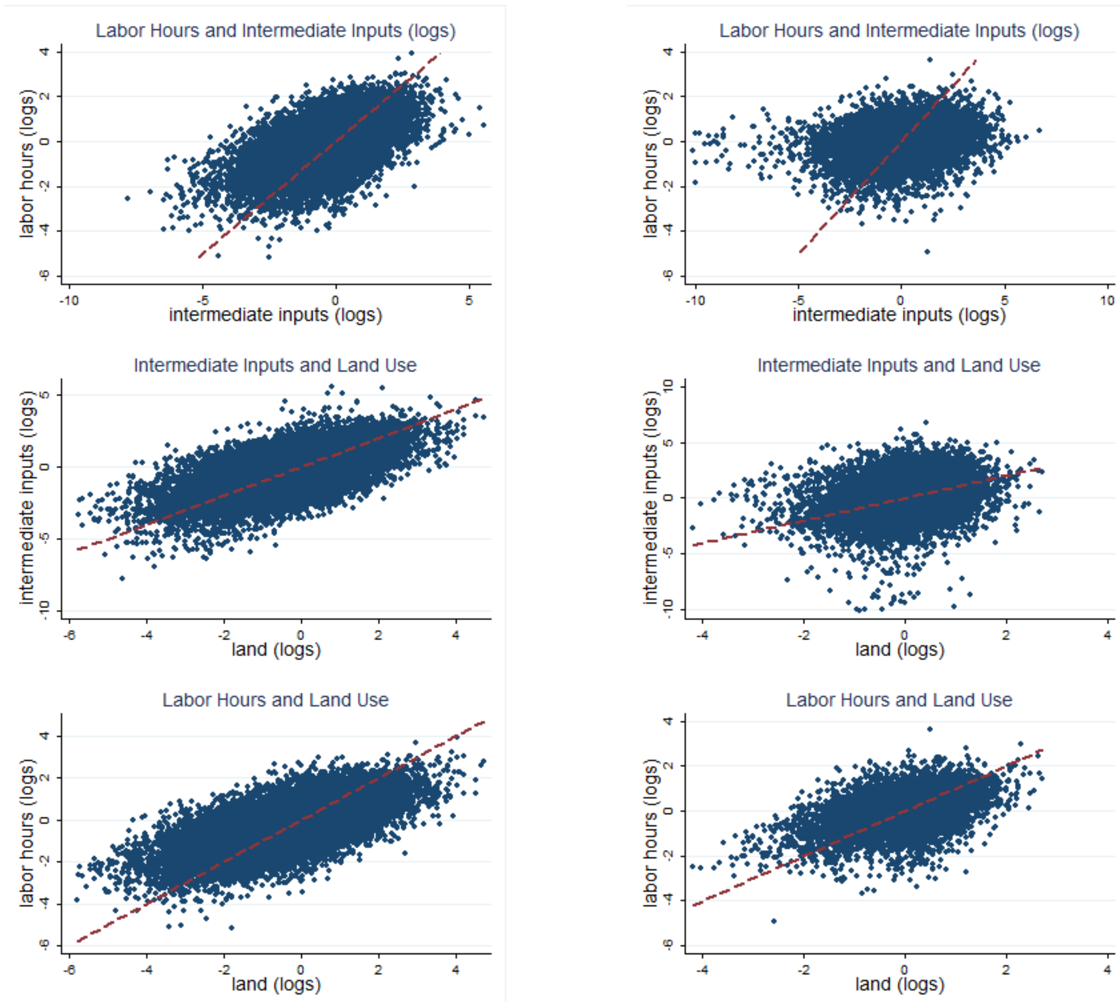
5 Variances in the Factor Ratios

Using measures of land, labor hours and intermediate input expenditure at household-plot level, I construct measures of factor ratio, R_{ijt} for household i plot j at year t. Three factor ratios of this study are (1) the ratio of labor to intermediate inputs, (2) the ratio of intermediate inputs to land, and (3) the ratio of labor to land. As discussed in Section 3, if household resources are allocated efficiently across plots, factor ratios should equalize across all plots within the household. Figure 3 shows household-plot level data on the three factors of production: labor, intermediate inputs, and land.

Top pictures in Figure 3 show labor hours versus intermediate inputs, the middle panel shows intermediate inputs versus land and the bottom panel illustrates labor hours and land. All variables are in log scale and demeaned, and the household-year fixed effects are removed, so that any variation left is within-household variations. Given the efficient allocation criteria, we would expect to see data points fall on a 45 degree line (dashed red line), because the data is demeaned. However, as figures reveal, large within-household variations exist in the data.

Using the variance decomposition method explained in Section 3, I quantify the extent of misallocation of resources within- and between- households from the data. The efficient allocation implies a within-household variance of zero in the factor ratios. I decompose total variation in the factor ratios across all household-plots into within-household and between-household variances and compare the two measures of variance. Table 3 and Table 4 show the variance decomposition of the three factor ratios of the study for Ethiopia and Malawi, respectively. For all three factor ratios, across all years and both countries, there are large within-household variations in the factor ratios. For example, in Ethiopia in 2013 from Panel A of Table 3, total variance in the ratio of labor to intermediate inputs is 4.57, which decomposes into 3.07 for within-household variance and 1.50 for between-household variance (all variances are normalized by mean). Therefore, 67% of the total variation is due to within-household variances. Relative size of within-household variance to total variances across the three factor ratios varies between 60 to 80 percent in Ethiopia and between 40 to 80 in Malawi.

I then test to see how much of the variances can be explained by observable characteristics of the plots and households. Several features exist that cause variations in the factor ratios, which do not reflect misallocation. For example, different crops have different factor intensities, so variations across crops exist, which does not indicate misallocation. Also, weather or other temporary shocks



(a) Ethiopia

(b) Malawi

Figure 3: Household-plot level data on three factors of production; labor, intermediate inputs, and land. All variables are in log scale for the purpose of illustration. Dashed red line is a 45 degree line indicating the efficient allocation. Variables are demeaned and household and year fixed effects are removed, so that the variation observed is all from within-household variations across plots.

to agriculture leads to ex post variations that does not imply ex ante misallocation. Similarly, different districts may reflect other unobserved differences in weather, land quality, etc. Therefore, I study variances after controlling for observable characteristics of the households and plots. I remove year and district fixed effects from the data and control for the characteristics of the plot manager, land quality, weather and other temporary agricultural shocks, and crop technology differences; so, any variation left is due to within-household variations unexplained by these observable characteristics. Specifically, I run the following regression:

$$R_{dijct} = X_{ijct}\beta + \theta_d + \zeta_t + \kappa_c + \epsilon_{dijct} \tag{14}$$

where R_{dijct} is the factor ratio on plot j planted to crop c at time t in household i of district d and

X_{ijct} is a plot-level vector of characteristics of plot j planted to crop c at time t in household i. X_{ijct} includes land size, soil type and soil quality, land characteristics (slope, elevation, and wetness), distance of the plot from household, seed type (traditional vs. improved), characteristics of the plot manager (age, gender, education, and whether has been away from household for longer than 6 months) and shocks to harvest; θ_d, ζ_t, and κ_c are district, year, and crop fixed effects respectively. And ϵ_{dijct} is the error term.

Variation in ϵ_{dijct} is variation in the factor ratios within a district on plots planted to the same crop in the same year, conditional on land size, land quality, weather and other transitory shocks and plot manager's characteristics. I then decompose total variation in ϵ_{dijct} into within- and between-household variations and calculate the share of total variation that is coming from within-household. Results of the variance decomposition after controlling for observable characteristics is also presented in Table 3 and Table 4. For Ethiopia in 2013 from Panel A of Table 3, total variance in the ratio of labor hours to intermediate inputs is 4.57 before controls and it decreases to 3.58 after controlling for observable characteristics (a 22% drop). After including the observable characteristics, within-household variance is the ratio of labor to intermediate inputs for Ethiopia in 2013, decreases from 3.07 to 2.46, which is a 20% decrease. After controlling for all these features, total variance in the factor ratios decreases by about 20% across the three ratios and years for the two countries. However, the variations by all these features translates into both within- and between-household variances equally. Therefore, even after controlling for observable differences across households and plots, most of the variation in the factor ratios is due to within-household variances.

The remaining unexplained variation in the factor ratios within a household can reflect true misallocation, mismeasurement, variation from misspecification, or unobservable plot-level characteristics. In the next section, I investigate several potential explanations for the large observed within-household variances such as measurement error, within-household bargaining power differences across plot managers, non-convexities in production function, plot distances from household, and managerial constraints.

6 Potential Explanations for Within-Household Variations

In the previous section, I showed that the relative size of within-household to total variance in the factor ratios are large. I also showed that controlling for several observable characteristics of households and plots explains only about 20% of the variations and in fact all these features contribute to both within- and between-household variations equally, such that the relative size of within-household variations to overall variation remains large.

In this section, I investigate several potential sources of variation in factor ratios within a household. I first study within-household bargaining power differences between plot managers within a household and show that more than 96% of households in the two countries in this study have a single manager across all plots and for the remaining 4% of the household, within-household variances are not larger. I then study measurement error in the factor inputs and find evidence in the data revealing mismeasurement of labor and land. I will also show that measurement error seems to matter to both within- and between-household variations, and it is not exclusively related to within-household variations in the factor ratios. I then provide evidence of indivisibility in fertilizer in the data. This can be due to a fixed cost or non-convexities in the production function. I find that households who have the financial resources to use fertilizer, but not enough to fertilize all their plots, end up fertilizing some of their plots and this group of households have the largest within-household variations. Finally, I explore the data for evidence of managerial constrains to efficiently allocate recurses across a large set of plots, or plots that are distanced from household. I find that within-household variations are significantly larger if the household has plots that are located at different distances from home. I find that the manager spends relatively less time on plots that are far from home, which can imply a time cost to distance.

6.1 Measurement Error

In this section, I provide evidence of measurement error in land size and labor hours in the data. I find substantial differences between reported and measured land size (based on GPS coordinates or using rope & compass) in the data in both countries. Figure 4 and Figure 5 show the distribution over the percent difference between household reported land size and measured land size by enumerators for Ethiopia and Malawi, respectively. Although GPS measures of land are not prone to measurement error, literature have documented that the measured land size is more reliable than reported land size.[10] Therefore, for the analysis in this paper, I use measured land size. However, in order to study the impact of mismeasurement of land size on the variance analysis, in Appendix E, I show the variances are much larger both within and between households if one uses reported land size, which contains more measurement error.

I also find that survey respondent seems to overstate the hours for other household members compared to what he/she reports for him/herself. The survey respondent is the most informed individual about household's agricultural activities. The respondent is asked to report hours for all household members (including him/herself) involved in agricultural activities throughout the season for each plot for each activity type (Land preparation and planting, other planting, and harvesting activities).

For each plot of land and for each activity type, respondent reports the identifier number of the household members who have worked on the plot (which is then used to back-out the characteristics of the worker from household survey), total number of weeks during the season he/she has worked as well as average number of days per week, and average number of hours per day. For each household member, I count the total number of plots the individual has worked on as well as the average hours he/she has worked per plot. Controlling for individual characteristics as well as land size and quality, and crop technology differences. Figure 6 presents the estimated coefficients on the number of plots per individual (omitted number is 1, so that each coefficient is being compared

[10] See for example Goldstein and Udry (1999) and Dillon et al. (2016).

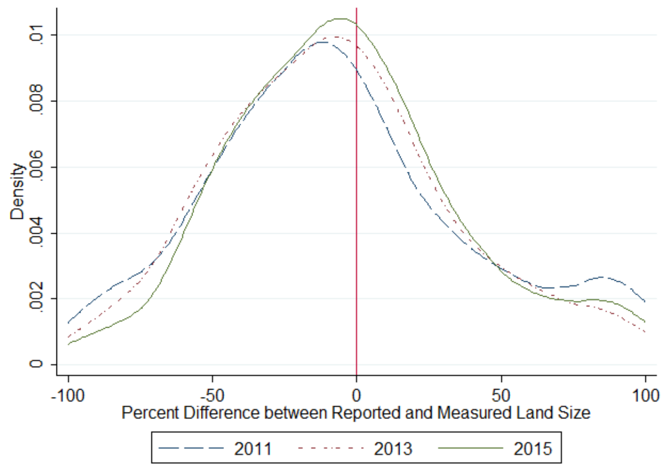


Figure 4: Ethiopia - Distribution of the Percent Difference Between Household-reported and Measured Land Size by year.

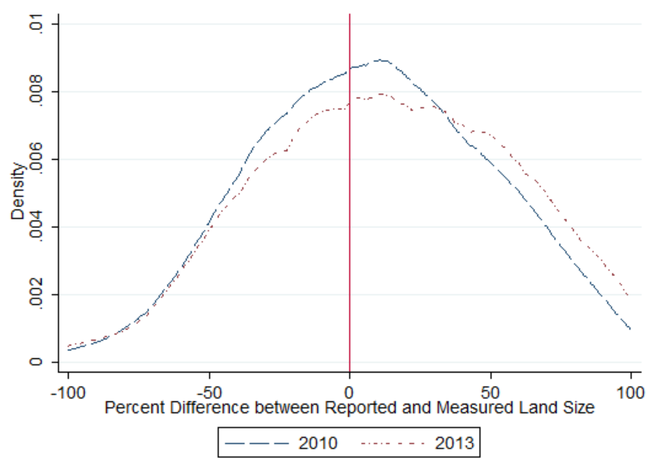
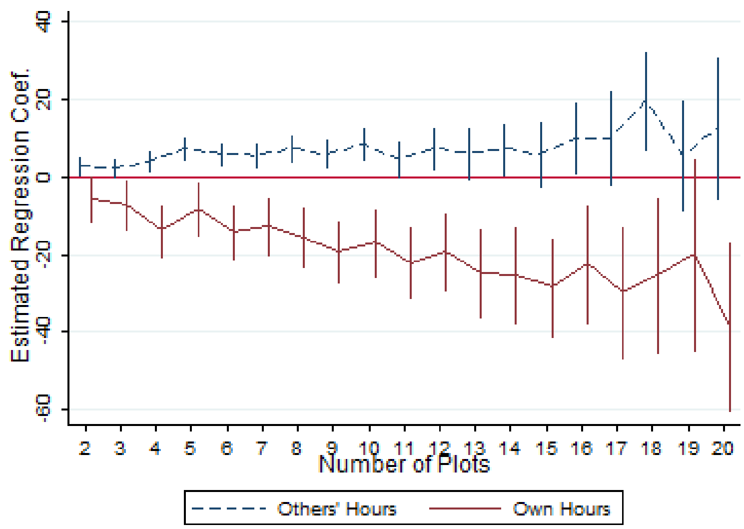


Figure 5: Malawi - Distribution of the Percent Difference Between Household-reported and Measured Land Size by year.

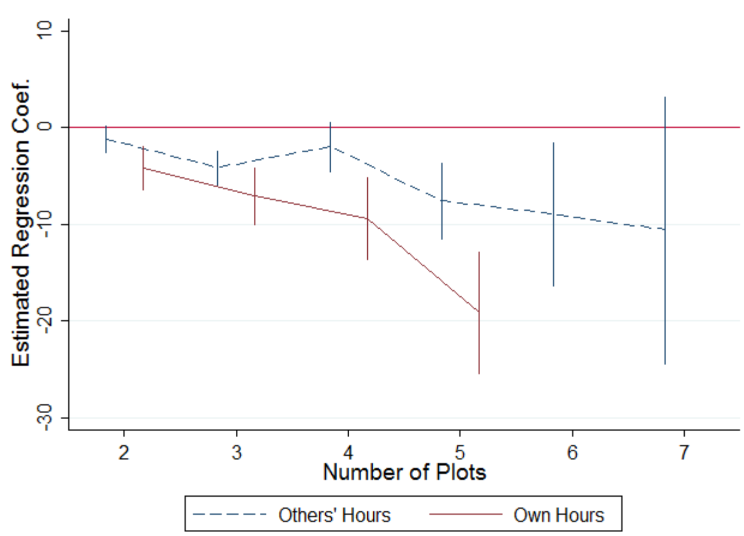
with a person within the household who only works on one plot of land).

The differences between own hours and others' hours are major, especially for Ethiopia. From Panel (a) of Figure 6, we can see that the coefficients are always positive for others' hours and always negative for own hours, and the differences are significant and get larger as the number of plots per individual increases. For own hours, an individual, who works on 20 plots of land, spends an average of 40 hours less than someone within the household who only works on one plot of land. For others' hours however, someone who works on 20 plots, spends an average of 11 hours more on each plot compared to a person within the same household who work son plot only.

Results for Malawi are not as pronounced as in Ethiopia, but the differences between own hours and others' hours are still statistically significant. For example, for own hours in Malawi ,an individual, who works on 5 plots of land, spends an average of 18 hours less than someone within the household who only works on one plot of land. For others' hours however, someone who works on 5 plots, spends an average of 8 hours less on each plot compared to a person within the same household who work son plot only.



(a) Ethiopia



(b) Malawi

Figure 6: Estimated coefficients of average individual hours on the number of plots per individual, controlling for age, gender, land size, land quality, harvest shocks, and including household, year, season, crop, and activity fixed effects. Confidence intervals are at 5% significance level.

Table 5 and Table 6 illustrate the the differences between hours reported for own versus hours reported for others in Ethiopia and Malawi respectively. Hours for others are over-reported by about 2 hours on average per individual per plot per agricultural activity type compared to hours reported for self. This means, for example for an individual who works on 20 plots, hours are over-reported by 40 hours per plot by activity type.

It is important to note that this overstating of hours can lead to both within-household and between-household variations in the factor ratios. The measurement error happens at individual, plot and activity level. For example within a household if more individuals work on one plot than the other, hours will be overstated by more on one plot than the other. Also, across households, one with more plots compared to another with fewer plots is overstating hours by more, or the margin of error due to mismeasurement in different in a household with more individuals compared to another with fewer individuals.

To study the impact of mismeasurement on the variance analysis, I construct measures of factor ratios (i.e. labor to intermediate inputs ratio and labor to land ratio), using actual total household labor hours and predicted total labor hours. To construct measures of predicted labor hours, I run two separates regressions for own hours and others ours. This is a regression of hours on the number of plots and all other controls (individual, land, and crop characteristics, and shocks) at individual level. Specifically, I run the following regression:

$$l_{kijt} = n_{kit} + X_{kijt}\beta + \theta_d + \zeta_t + \tau_c + \eta_a + \epsilon_{kijt} \tag{15}$$

where l_{kijt} is the average individual k's hours per plot in household i on plot j at time t; n_{kit} is total number of plots within household i that individual k has worked on at time t; X_{kijt} includes individual's age and gender, indicator for whether individual has been away from the household for more than 6 months, relationship to the household head, plot size, and land quality (slope, elevation, wetness, distance from household, soil quality); and θ_d, ζ_t, τ_c, and η_a are district, year, crop, and activity type (planting vs. harvesting) fixed effects, respectively.

I run the above regression separately for own hours and others' hours, and I take the error term ϵ_{kijt} (unexplained residuals) from others' hours regression. I then estimate mean hours for others using estimated betas from own hours regression and add in the error term from others' hours regression for the variation that comes from unobservables. This constructs a measure of predicted hours for others' that is corrected for measurement error.

After constructing the factor ratios using predicted hours, I decompose the variation in the factor ratios with each measure of labor to within- and between-household variances. Table 7 summarizes the findings. From the variance columns, we can see that correcting for labor hours, can lead to higher or lower variances in the factor ratios both within- and between-households, so the relative size of within-household variance to total may increase, decrease, or remain constant. For example, from Panel A for Ethiopia, average labor hours per square meter of land decreases from 1.78 to 0.74 after correcting for the measurement error. Total variance in the ratio of labor hours to land decreases from 1585 to 55, and within-household variance decrease from 1196 to 41 after correcting for mismeasurement in hours. Since, the impact is on both within-household and overall variance in the factor ratios, the share of the variation due to within-household variances remains at 75%.

On the other hand, correcting for mismeasurement in labor hours increases variances both within and between households for the ratio of hours to intermediate inputs in Ethiopia.

6.2 Within-Household Bargaining Power

One of the sources of within-household misallocation studied by Udry (1996) is within-household differences in the bargaining powers of the individuals managing different household plots, specifically the differences between male and female managers. Udry (1996) studied within-household allocation of agricultural resources across plots in Burkina Faso and found that plots controlled by women are farmed less intensively than similar plots in the household controlled by men.

In both Ethiopia and Malawi a very large fraction of households have only one individual managing all household plots (98% in Ethiopia and 96% in Malawi). Table 8 summarizes the number of managers and their gender for households in the two samples. The fact that only less than 4% of the households have multiple managers across plots suggests that within-household bargaining power differences cannot explain the large within-household variations observed in the factor ratios in this data. However, I examine to see how much larger are within-household variances in the factor ratios for households in which multiple people manage different household plots. For each country, I divide the sample into three subsample: 1. households with one manager for all plots, 2. households with multiple managers, and 3. households with multiple managers from the opposite sex. Then I study the variances in the factor ratios for each subsample and decompose the variances into within- and between-household variances. Table 9 summarizes the findings. I do not find a significant difference between single-manager and multi-manager households in Ethiopia. For Malawi however, for multi-manager households, as opposed to what Udry (1996)'s theory predicts, both levels of within-household and the relative size of within household to total variation in the ratio of labor to intermediate inputs is significantly lower than single-manager households.

Based on the analysis in this section, I conclude within-household bargaining power between male- and female-cropped plots is not the source of within-household variations in the factor ratios in either of the two countries in this study. In fact, results from Malawi contradict with the predictions of Udry (1996)'s theory. In the next section, I investigate the role of measurement error in the analysis of variations in the factor ratios.

6.3 Non-Convexities in Production

In this section, I investigate the importance of factor indivisibility in the observed variations in the factor ratios across plot within a household. This can be due to a fixed cost or technological differences, such as a minimum fertilizer requirement. Therefore, if a household does not have enough fertilizer for all the plots, they may apply it to some, and not all their plots. In fact, there are many households in the data who apply fertilizer to some of their plots and not to all. Among households who have fertilizer, roughly two thirds of the households in Ethiopia and half of the households in Malawi use fertilizer on some of their plots, but not on all.[11] Then the question is that if a household has fertilizer, why would they allocate it all to one plot and not allocate any to their other plot? Perhaps there is a non-convexity in production that requires a minimum amount of fertilizer to be effective, a fixed cost of applying the fertilizer, or some other cause of a region of local increasing returns to fertilizer use. Then if a household does not have enough fertilizer for all the plots, they allocate what they have to one or some of the plots.

I use the food share of household consumption as a measure of the household's wellbeing. I find that in Ethiopia, households who don't use any fertilizer on their plots are the poorest, and the households who apply fertilizer to some of their plots are poorer than those who apply fertilizer to all of their plots. In Malawi, although those with no fertilizer are the poorest, household who use fertilizer on some or all of their plots are not significantly different in terms of their wellbeing.

To study the impact of partially using fertilizer on household plots on the variance analysis, I divide the sample into two groups of households: (Group 1) those who apply fertilizer to none or all of their plots and Group 2. those who apply fertilizer to some of their plots, but not to all. Table 10 presents the results. (Group 2) households have much larger within-household variances in the factor ratios relative to total. In both countries, households who apply fertilizer to some of their plots and not to all have about 1.5 times more plots, therefore I adjust the relative share of within-household variance for the sample bias using the technique discussed in section 6.4. Also, notice that labor hours per square meter of land is not different in terms of variances across the two groups in Malawi. In Ethiopia, however Group 2 has slightly larger relative within-household variances in labor hours per square meter of land, which can potentially be due to larger mismeasurement for households who have more plots of land.[12]

Among all features studies in Section 3, the only one that is exclusively related to within-household misallocation is indivisibility of fertilizer. The variation implied by the rest of these features translates into both within- and between-household inferred misallocation, so that the relative importance of within-household misallocation does not change much. Even after taking into account these features, within-household variation in the factor ratios still accounts for most of the variation across household-plots. Therefore, theories of misallocation need to (1) either focus on within-household distortions, or (2) provide mechanisms by which market distortions manifest themselves as within-household distortions. In the next section of this paper, I provide two theories that could generate this type of link and test them using the data. First, I consider risk. The underlying idea would be that the household maximizes utility taking risk into account. That is with a risk-sharing mechanism across plots within the household, utility is maximized. One can assume there is a certain combination of a given factor input with on a given plot that interacts with marginal utilities such that the allocations observed in data become efficient. I will discuss in Section 7.2 that the evidence from data does not support this prediction. Second, in Section 7.1, I

[11] In Ethiopia, on average households apply fertilizer to 60 percent of their plots whereas in Malawi, they apply their fertilizer to 80 percent of their plots.

[12] Mismeasurement in labor is larger for households with more plots, and Group 2 households in this analysis have 1.5 times more plots than Group 1.

consider the fact that the lack of land markets leads to households having plots that are specially segregated and hard to optimize across.

6.4 Managerial Constraints

As mentioned before, the households in the sample cultivate multiple plots of land, and the number of cultivated plots per household varies a lot from one household to another (See Figure 1 and Figure 2). A relevant question to ask is whether within-household misallocation (variation in factor ratios) are larger among households who cultivate more plots. In other words, we would like to know whether it is harder for the manager to allocate resources efficiently across plots if the household has more plots.

In this section, by restricting the sample to all households with a certain number of plots at a time, I study the share of the total variance in the factor ratios that comes from within-household variances. One thing to note is that due to the nature of the variance decomposition technique explained earlier, the relative size of within-household to between-household variance gets larger as the number of plots in household increases. This is driven by the fact that within-household variances has a N_{it} term in the denominator, which is the total number of plots at household i at time t (See equation (6)).

This is not an issue when we measure variances for the entire sample. Because within-household variance for the entire sample; V_{WHH} is the average of all household variances weighted with the total number of cultivated plots per household. However, this becomes an issue when I restrict the sample to only households with a fixed number of plots. Imagine 100 observations are divided between 10 household, so that each household has 10 plots, versus a case in which the same 100 observations are divided between 50 households, so that each household has 2 plots of land. The share of within-household variance from the total variance is larger in the former case, due to the way the within-household variances are constructed.

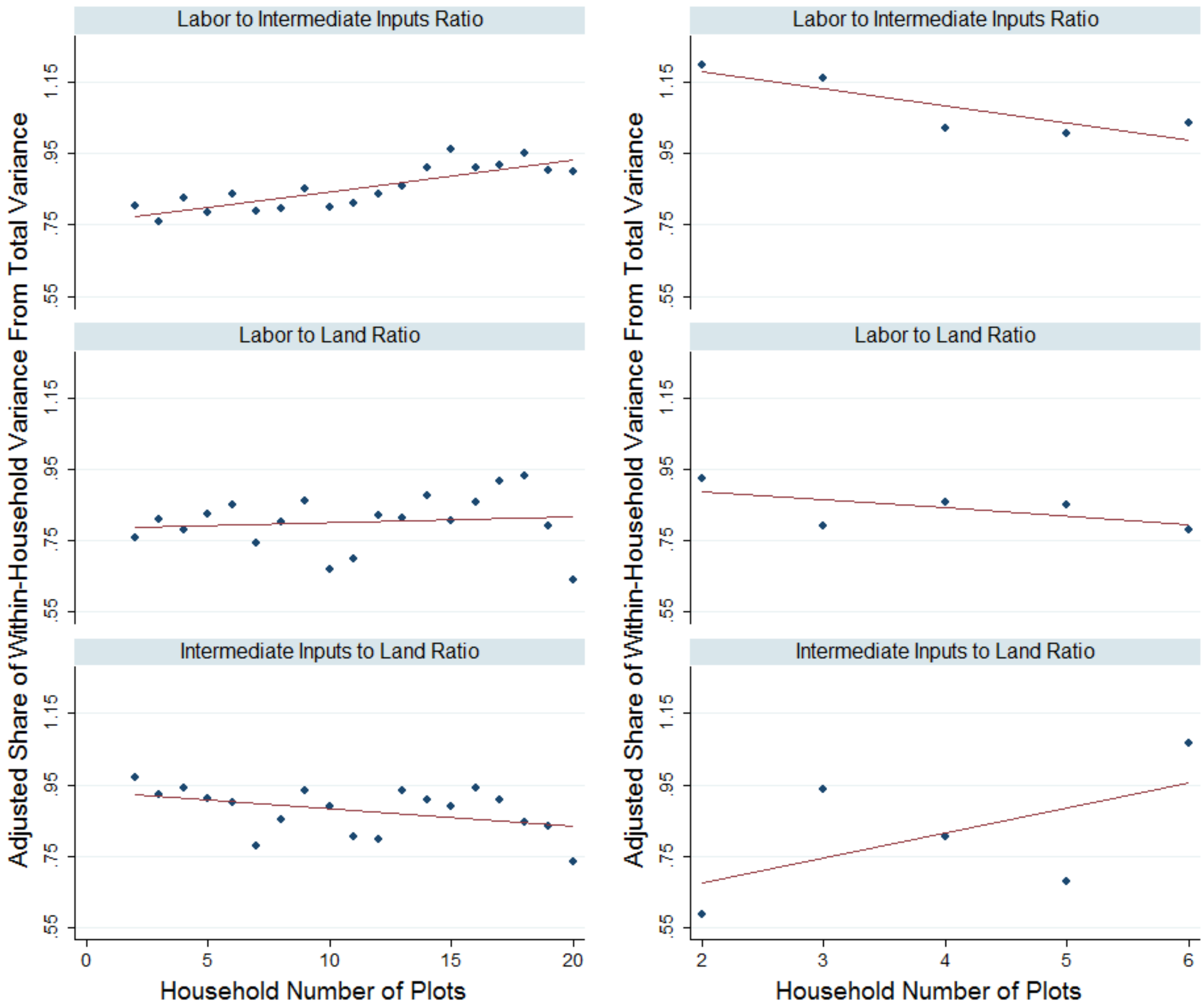
To solve the above problem, I adjust within-household variances (V_{it}) for the sample size bias by multiplying it by ($\frac{N_{it}}{N_{it}-1}$), such that

$$V_{it}^{adj} = \frac{\sum_{j}(R_{ijt} - R_{it})^2}{N_{it} - 1} \tag{16}$$

This adjustment changes the interpretation of the share of within-household from total variance (S). New S (*adjusted* S hereafter), will measure the amount of misallocation relative to a case where plots are all randomly assigned to households (the ceiling for S, which is equal to 1, meaning all of the variation comes from within-household). The higher the adjusted S is, the more within-household misallocation there is. More explanation on the adjustment on the variances can be found in appendix D.[13]

The left panel of Figure 7 for Ethiopia, shows the adjusted S on the y-axis and number of plots

[13] I perform a Monte Carlo simulation exercise to evaluate the variance analysis I use for subsamples. See appendix D for details about the simulation exercise.



(a) Ethiopia (b) Malawi

Figure 7: Adjusted relative share of within-household variance in the factor ratios from total by the number of plots per household. The analysis controls for crop technology, the characteristics of the primary decision maker, plot distance from household, and land quality(slope, elevation, and wetness index), harvest shocks, rainfall, and region fixed effects.

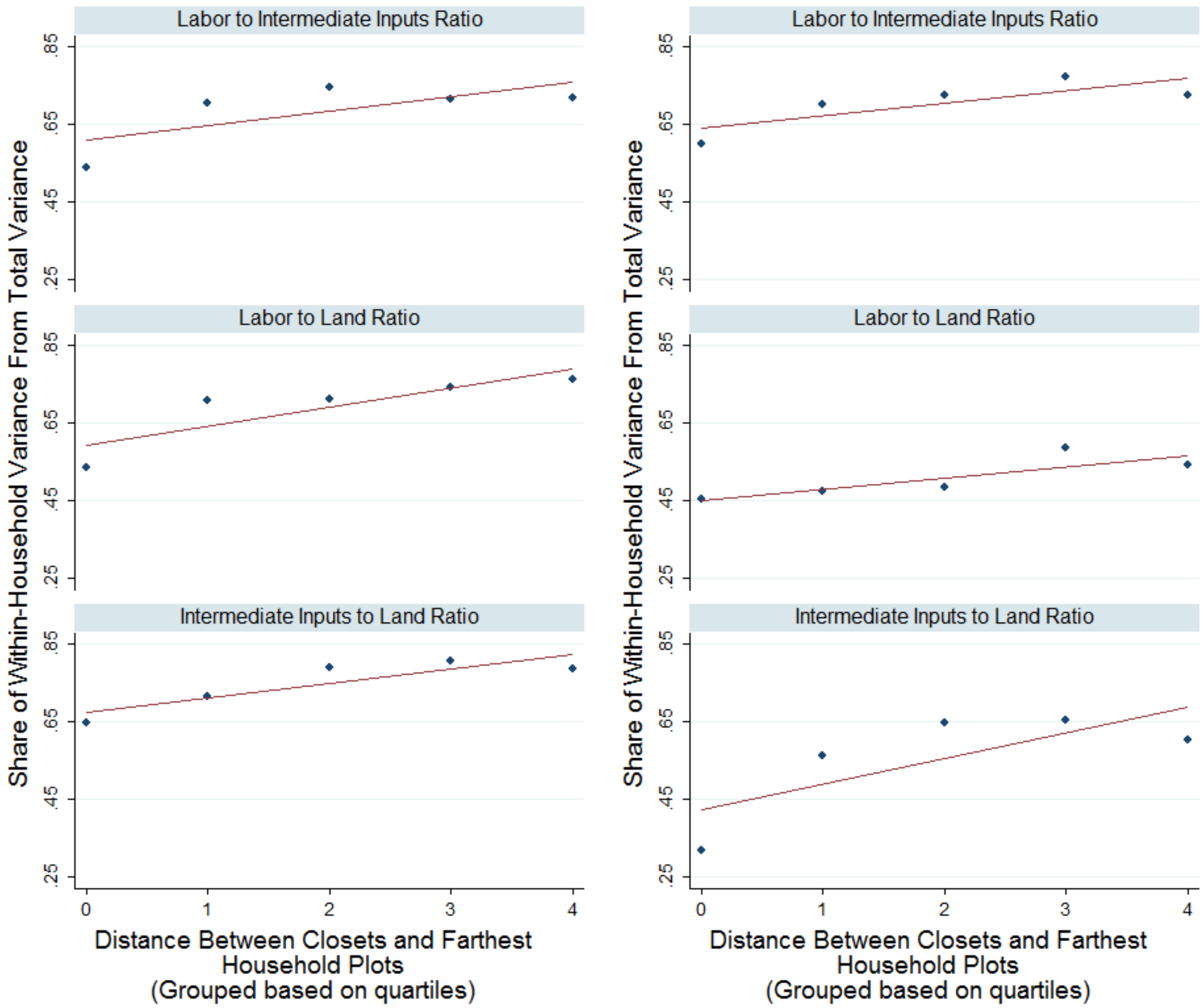
per household on the x-axis. Each Panel in the Figure shows one of the three factor ratios. I find a positive relationship between adjusted shares of within-household variance and the number of plots for the ratio of labor to intermediate inputs in Ethiopia, and for intermediate inputs per square meter of land in Malawi. In other words, households who cultivate a larger number of plots are further away from their efficient allocation of resources in the household for these specific cases. Again note that the increasing trend in the share in N_{it} is no longer driven by the way variance formulas work, because they are adjusted for the sample bias.

Two of potential explanations for observing any trends in this exercise can be (1) that the allocating across more plots is harder for the manager and/or (2) misreporting and therefore measurement error is more likely when household has more plots. First explanation means we would expect to see as the number of plots a household gets larger, the within-household variance in the factor ratios becomes more important in magnitude. Direction of the relationship for explanation two is not clear though. As discussed in details previously, measurement error may increase or decrease the relative size of within-household variance to total variance depending on how it impact the variances within- and between-households.

6.5 Plot Distances from Household

In this section I study the role that plot distance from household plays in within-household misallocation of resources. In both countries, some households have plots that are located at about the same distances from the household and some other households have plots that are located in very different distances from the household. The difference between plot distances from household are as large as 70 kilometers within the households in some cases. I find that within-household variances in the factor ratios are relatively smaller for households who have plots that are distanced equally from the household. Figure 8 presents the relationship between distances between plots within a household and the share of variances in the factor ratios that comes from within the household. Moving from left to right on the x-axis, variation in plot distances from home within a household increases, and y-axis shows the relative size of the variance in the factor ratio that is due to within-household variations. In both countries and for all factor ratios, for households who have plots located at very different distances from home, within-household misallocation is relatively larger. For the ratio of intermediate inputs per square meter of land in Malawi, within-household variance accounts for about 30% of the overall variation for households who have plots that are very close to one another. Whereas for households that have distanced plots, within-household variances account for about 65% of the overall variations. Among households, in which all plots are located near each other, within-household variations in the factor ratios account for 30-60 percent of the total variations. For households who have very distanced plots, within-household variances in the factor ratios account for 60-85% of the total variations.

There are several reasons for why households with segregated plots have larger within-household misallocation. This can be due to a time cost. Traveling to plots that are further away from home is more costly. I find that the managers spend less time on plots that are far from the household and therefore it is harder for them to monitor agricultural activities on these plots. Table 12 shows results of a fixed effects regression of the amount of time plot manager spends on the plot during the season as a function of plot distance from household, controlling for land size, the amount of time other household members work on the plot, and activity type. In both Ethiopia and Malawi, if the plot's distance from household is larger than the median distance, the manager spends less time on the plot. There may also be a ice-berg cost associated with transporting inputs to plots that are far from household. I find that plots that are far from home in Malawi are less likely to be fertilized in Malawi. In Ethiopia, however plots that are far from home are more likely to be fertilized and use of hired labor on far plots is more likely.



(a) Ethiopia

(b) Malawi

Figure 8: Distances Between Household Plots and Share of Within-Household Variance - On the horizontal axis, households are grouped into 5 groups: $x = 0$ representing all households, in which all plots are near one another, and $x = 1$, $x = 2$, $x = 3$, and $x = 4$ representing quartiles based of distance between plots within the household. These results are also presented in Table 11.

In the next section, I will argue that the source of distanced plots within a household is the lack of land markets. Without proper land markets, households will not be able to optimize their plot distances from home, which is then associated with having plots that are spatially segregated and hard to optimize across.

7 Market Distortions

In this section, I study two types of market distortions that could manifest themselves in within-household distortions. First, I study the lack of land markets and show that households with marketed land have plots that are located at roughly equal distances from home and therefore their within-household misallocation is smaller. I also study risk. The subsistence requirements can amplify risk concerns for households with low consumption levels. I use household food share of consumption expenditure as an indicator of households risk aversion and test to see whether households closer to substance requirements have larger within-household variation in factor ratios

7.1 Access to Land Markets

A large body of the literature on agricultural misallocation focuses on the imperfections in the land markets as the core of the productivity problem in agriculture in developing countries (Restuccia and Rogerson (2016), Restuccia and Santaeulalia-Llopis (2017), Chen et al. (2017)). In both Ethiopia and Malawi, there is very limited market for land. Only 12.4% of plots in Ethiopia and 14.5% of plots in Malawi are marketed land.[14] In a recent study, Chen et al. (2017) show that land rentals in Ethiopia are associated with a significant reduction in misallocation and an increase in productivity. Existence of proper land markets lead to reallocation of land from households with limited non-land resources to households rich in non-land resources. Restuccia and Rogerson (2016) also mention that the reallocation of factor inputs from the actual to the efficient allocation is done through rental markets.

Current literature well documents the impact of the lack of land markets on misallocation across households. In this section, I study how these land market frictions can show up within a household and lead to misallocation of resources across plots within a household. Table 14 shows that within-household variation in factor ratios are relatively smaller for households farming marketed plots of land. I classify households into two groups: (1) households whose cultivated area of land is mostly non-marketed land, and (2) households whose more than half of their cultivated area of land is marketed land. Results show that within household variances in the factor ratios are smaller among households who farm marketed land.

Without markets for land, households have plots that are spatially segregated and hard to optimize across. Table 15 summarizes plot distances from home for households who have no marketed land versus households who have marketed land. In Ethiopia, plots of households with marketed land are 0.29 kilometers closer to the household on average. In Malawi, however average plot distance from household is 1.25 kilometers larger for those who have marketed land.[15]

Last column of Table 15 shows within-household variance in plot distances from home. In both Ethiopia and Malawi, households who do not have marketed plots have 4.5 times more variation in

[14] I define marketed land as a parcel of land that is rented, borrowed for free, sharecropped, or purchased. Table 13 shows the ownership status of plots across years of samples in Ethiopia and Malawi. About 25% of households in Ethiopia and 20% of households in Malawi have some marketed land.

[15] I find than households seem to treat their marketed plots differently. Table A.8 summarizes the differences. In Ethiopia, marketed plots are significantly larger in size than non-marketed plots. In Malawi, unlike Ethiopia, marketed plots are smaller is size. In both countries, the use of hired labor is more common on marketed plots. In Ethiopia, labor and intermediate inputs use relative to land size is smaller on marketed plots. The differences in Malawi are not as significant as what I find in Ethiopia (See Table 16). For neither of the countries there seem to be a significant difference between marketed land and non-marketed land in terms of average labor to intermediate inputs ratio.

their plots' distances from home. In other words, plots of the households with marketed land are located in roughly same distances from home. Although households in Malawi are renting/purchasing plots that are on average further away from their homes, but all their plots are located in same distances from home. In Section 6.5, I find that when household plots are located in different distances from home, within-household misallocation in larger. Therefore, the lack of land markets is associated with households having spatially distanced plots, which are hard to optimize across. In the next section, I investigate the potential for risk to distort within-household allocation of resources.

7.2 Risk and Insurance

In this section, I investigate the data for evidence of risk-sharing within a household across plots. A large body of the macro literature studies misallocation considering risk and the lack of insurance markets in developing countries.[16] Literature argues that one possible cause of agricultural productivity differences across countries is that farmers in developing countries use fewer intermediate inputs. Donovan (2016) explains that the intermediate decisions are made before the realization of productivity shocks and because of the lack of well-functioning insurance markets, the farmers need to internalize the impact their intermediate input choice will have on their ex post consumption. Therefore, poorer farmers do not use intermediate inputs, because in case of a low shock, their consumption drops a lot.[17] More generally, equalizing input ratios across plots may maximize expected output because inputs are complements in this fashion, but they may be substitutes in maximizing risk-weighted output.[18]

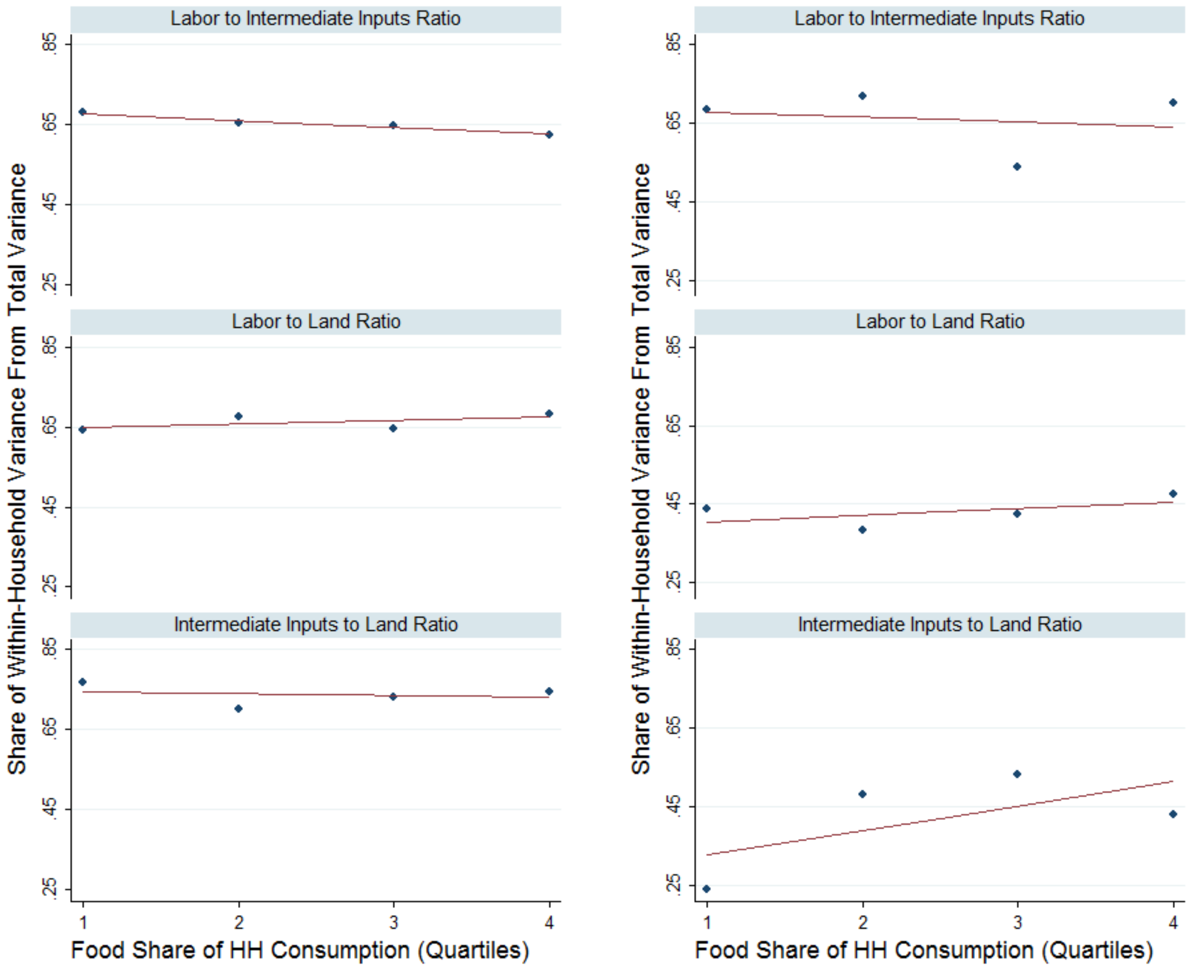
Ogaki and Zhang (2001) and Donovan (2016) show that subsistence requirements can amplify risk concerns for households with low consumption levels. To examine the importance of risk, I therefore use household's food share of consumption as an indicator of household's income, and therefore risk-aversion. Higher food share of consumption implies a poorer and therefore a more risk-averse household. I then sort households based on their food share of consumption, and for each quartile of income distribution, I calculate the relative size of within-household variations in the factor ratios to total variance. If households are taking a risk factor into account in their optimization problem, we would expect poorer households to have relatively larger within-household variances in the factor ratios. Results of this analysis are presented in Figure 9. For intermediate inputs per square meter of land in Malawi, I find some evidence of higher within-household variances among poorer households. For first decile (the quartile with the lowest food share of consumption or

[16]See for example Mobarak and Rosenzweig (2012), Karlan et al. (2014), Donovan (2016), and Cole et al. (2017).

[17]To understand how risk impacts resource allocation, imagine a simple household with two plots. Household gets output $Y_1 = zl_1^\alpha$ from plots 1 and $Y_2 = al_2^\alpha$ from plot 2, where $l_1 = L_1/T_1$ and $l_2 = L_2/T_2$ are labor per unit of land. Without loss of generality, assume $T_1 = T_2 = T$, then $l_2 = 1 - l_1$. Also, consumption is $C = Y_1 + Y_2$.
If $z \sim G(z)$ and $E(z) > a$, then plot 2 has higher risk and higher outcome. Household maximizes utility $U(C)$, where $U'(C) > 0$ and $U''(C) < 0$. Then solution implies all household labor to be allocated on plot 2, and we will have $l_1 = 1$ and $l_2 = 0$

[18]For example, fertilizer on land that generally receives good rainfall may increase output but also risk. The additional risk may increase the benefit of applying labor to less well-irrigated land.

the richest households) of household in Malawi, within-household variations in intermediate inputs per square meter of land only accounts for 24% of the total variations, and this statistics is 53% for larger quartiles of the households.



(a) Ethiopia

(b) Malawi

Figure 9: Food Share of Consumption and Relative Size of Within-household Variances - Horizontal axis is the decile of household food share of consumption, with larger deciles implying poorer households.

For Ethiopia, relative importance of within-household variations remains fairly constant across deciles of food share of household consumption across all factor ratios. Other than the ratio of intermediate inputs to land in Malawi, I cannot find strong evidence in the data to support the hypothesis that poorer (more risk-averse) households allocate their resources differently compared to richer households. However, the finding concerning intermediate inputs in Malawi suggests that future studies can investigate mechanism by which risk can distort the allocation of resources across plots within households.

In this section, I studied two types of market distortions that could manifest themselves in within-household distortions. First, I studied risk. I used household consumption expenditure data to construct an index of households risk aversion and tested to see whether there is evidence in the data that more risk-averse households allocate their resources differently across their plots. I did not find strong evidence to support this hypothesis. Secondly, I showed how inaccessibility of land markets can lead to within-household misallocation of resources. In an economy lacking proper land markets, households cannot sell the plots of land that are far from their household and it is hard for them to allocate their resources efficiently across plots that are segregated.

8 Summary and Conclusion

Using household-plot level data from Ethiopia and Malawi, I study the variations in three factor ratios: (1) labor to intermediate inputs ratio, (2) labor to land ratio, and (3) intermediate inputs to land ratio. By decomposing the total variation in each factor ratio into within- and between-household variations, I show that within-household variation accounts for 70% of the overall variation in the data. This finding signifies the importance of understanding the sources of within-household variation in factor ratios.

Controlling for the several observable characteristics of the households and plots, such as differences in crop technologies across plots, land ownership status, land quality, irrigation, weather and other temporary shocks to agriculture, measurement error, within-household bargaining power between plot managers, and other characteristics of the plot primary decision maker account for about 20% of the overall variations in the factor ratios. Among the controls, crop technology differences seem to be the most important determinant. Since different crops have different factor intensities, most of the explained variations are from controlling for crop-specific technologies. However, even after controlling for all these observable differences across plots, within-household variations remain large relative to between-household variation.

I also show that the household reported land size as well as labor hours contain large measurement error. The measurement issue, however is not exclusively related to the within-household variations in the factor ratios. I find that the survey respondent overstates hours for other household members by a factor of 2 per person, per plot, and per agricultural activities. Therefore, for example for a household with more individuals involved in agricultural activities, the margin of error is larger. Also, a household with more plots have a larger margin of measurement error in labor hours. I aim to correct for the measurement error in labor hours and show that the measurement error impacts variations in the factor ratios both within and between households.

This study suggests that the theories of misallocation need to provide mechanisms by which market distortions manifest themselves as within-household distortions. In the last section of the paper, I conduct an examination of two alternative explanations that provide mechanisms that could generate this type of link and test them using the data. First, I consider the fact that the lack of land markets leads to households having plots that are spatially segregated and hard to optimize across. I find distances between plots of households who have marketed land are smaller and households allocate their resources more efficiently across plots if their plots are located within the same distance from home. Second, I consider risk. The subsistence requirements can amplify risk concerns for households with low consumption. If risk distorts the within-household resource allocation, poorer households, who are also more risk-averse should have larger within-household misallocation. The evidence from the data does not support this prediction for Ethiopia. In Malawi, however, I find relatively larger within-household variations in the ratio of intermediate inputs to land among poorer households. This finding suggests that future studies may need to consider mechanism by which risk can distort within-household allocation of agricultural resources across plots.

References

T. Adamopoulos and D. Restuccia. Geography and Agricultural Productivity: Cross-Country Evidence from Micro Plot-Level Data. Technical report, University of Toronto Working Paper, 2015.

T. Adamopoulos, L. Brandt, J. Leight, D. Restuccia, et al. Misallocation, Selection and Productivity: a Quantitative Analysis with Panel Data from China. Technical report, Working Paper, 2016.

V. Arthi, K. Beegle, J. De Weerdt, and A. Palacios-Lopez. Measuring Household Labor on Tanzanian Farms. *Unpublished working paper, World Bank, Washington, DC*, 2016.

E. Bartelsman, J. Haltiwanger, and S. Scarpetta. Cross-Country Differences in Productivity: the Role of Allocation and Selection. *The American Economic Review*, 103(1):305–334, 2013.

D. Benjamin. Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models. *Econometrica: Journal of the Econometric Society*, pages 287–322, 1992.

M. Bils, P. J. Klenow, and C. Ruane. Misallocation or Mismeasurement? Technical report, Working Paper, 2017.

F. J. Buera, J. P. Kaboski, and Y. Shin. Finance and Development: a Tale of Two Sectors. *The American Economic Review*, 101(5):1964–2002, 2011.

C. Carletto, S. Gourlay, and P. Winters. From Guesstimates to Gpstimates: Land Area Measurement and Implications for Agricultural Analysis. *Journal of African Economies*, 24(5):593–628, 2015.

C. Chen, D. Restuccia, and R. Santaeulalia-Llopis. Land Markets, Resource Allocation, and Agricultural Productivity. Technical report, Working Paper, 2017.

S. Cole, X. Giné, and J. Vickery. How Does Risk Management Influence Production Decisions? Evidence From a Field Experiment. *The Review of Financial Studies*, 30(6):1935–1970, 2017.

S. J. Davis and J. Haltiwanger. Wage Dispersion Between and Within Us Manufacturing Plants, 1963-1986. Technical report, National Bureau of Economic Research, 1991.

A. Dillon, S. Gourlay, K. McGee, and G. Oseni. Land Measurement Bias and Its Empirical Implications: Evidence from a Validation Exercise. 2016.

K. Donovan. Agricultural Risk, Intermediate Inputs, and Cross-country Productivity Differences. 2016.

R. Garlick, K. Orkin, and S. Quinn. Call Me Maybe: Experimental Evidence on Using Mobile Phones to Survey Microenterprises. 2016.

M. Goldstein and C. Udry. Agricultural Innovation and Resource Management in Ghana. *Final Report to IFPRI under MP17, Mimeo. Yale University*, 1999.

D. Gollin and C. Udry. Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture. 2017.

D. Gollin, S. Parente, and R. Rogerson. The Role of Agriculture in Development. *American Economic Review*, pages 160–164, 2002.

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

C.-T. Hsieh and P. J. Klenow. Misallocation and Manufacturing TFP in China and India. *The quarterly journal of economics*, 124(4):1403–1448, 2009.

D. Karlan, R. Osei, I. Osei-Akoto, and C. Udry. Agricultural Decisions After Relaxing Credit and Risk Constraints. *The Quarterly Journal of Economics*, 129(2):597–652, 2014.

N. Keita and E. Carfagna. Use of Modern Geo-Positioning Devices in Agricultural Censuses and Surveys: Use of GDP for Crop Area Measurement. In *Bulletin of the International Statistical Institute, the 57th Session, 2009, Proceedings, Special Topics Contributed Paper Meetings (STCPM22), Durban*, 2009.

D. LaFave and D. Thomas. Farms, Families, and Markets: New Evidence on Completeness of Markets in Agricultural Settings. *Econometrica*, 84(5):1917–1960, 2016.

A. M. Mobarak and M. R. Rosenzweig. Selling Formal Insurance to the Informally Insured. 2012.

B. Moll. Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation? *The American Economic Review*, 104(10):3186–3221, 2014.

M. Ogaki and Q. Zhang. Decreasing Relative Risk Aversion and Tests of Risk Sharing. *Econometrica*, 69(2):515–526, 2001.

D. Restuccia and R. Rogerson. Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Review of Economic dynamics*, 11(4):707–720, 2008.

D. Restuccia and R. Rogerson. The Causes and Costs of Misallocation. *Unpublished manuscript, University of Toronto*, 2016.

D. Restuccia and R. Santaeulalia-Llopis. Land Misallocation and Productivity. Technical report, National Bureau of Economic Research, 2017.

D. Restuccia, D. T. Yang, and X. Zhu. Agriculture and Aggregate Productivity: a Quantitative Cross-Country Analysis. *Journal of Monetary Economics*, 55(2):234–250, 2008.

A. Shenoy. Market Failures and Misallocation. *Journal of Development Economics*, 128:65–80, 2017.

C. Udry. Gender, Agricultural Production, and the Theory of the Household. *Journal of political Economy*, 104(5):1010–1046, 1996.

Tables

		2011			2013			2015		
		Mean	**Median**	**N**	**Mean**	**Median**	**N**	**Mean**	**Median**	**N**
Panel A: Labor Hours										
Total plot hours		169	103	12,059	167	104	13,712	159	98	14,271
	Total	196	134		205	138		197	131	
Households with non-family labor	Family labor	176	116	3,942	183	119	3,829	172	112	3,465
	Non-family labor	20	10		22	10		25	15	
Households with family labor only		156	90	8,117	152	92	9,883	147	88	10,806
Panel B: Intermediate Inputs										
Total intermediate inputs					24	8	13,965	50	21	14,645
Seeds	Traditional				19	7	10,899	29	11	10,878
	Improved				21	12	673	34	19	695
Fertilizer					28	17	4,524	43	20	4,923
Panel C: Land										
Plot Size (1000 Square meters)		1.7	1	12,059	1.7	0.9	13,712	1.8	0.9	14,271
Plot distance in kms to household		3.62	0.31	10,703	1.53	0.3	13,917	0.86	0.2	14,472
Plot slope (percent)		13.36	10.13	10,703	13.77	10.2	13,917	12.86	9.8	14,472
Plot elevation (m)		1942	1916	10,703	1978	1969	13,917	1976	1965	14,472
Plot potential wetness index		12.71	12	10,703	12.67	12	13,917	12.68	12	14,472
Number of plots per household		4.8	4	2,310*	4.9	4	2,813*	5.2	4	2,776*

Table 1: Ethiopia - Table of Summary Statistics: Labor Hours, Intermediate Inputs, and Land Size. Labor hours is total planting and harvesting labor hours for the entire agricultural season on each plot. Intermediate input expenditure is total expenditure on seeds and fertilizer in constant 2010 USD. * represents total number of households.

		2010			2013		
		Mean	**Median**	**N**	**Mean**	**Median**	**N**
Panel A: Labor Hours							
Total plot hours		282	216	4433	299	236	5571
Households with non-family labor	Total	267	203		286	222	
	Family labor	267	203	3,928	286	222	3,567
	Non-family labor	40	20		41	24	
Households with family labor only		268	204	10037	282	216	3476
Panel B: Intermediate Inputs							
Total intermediate inputs		205	33	4433	454	179	5,571
Seeds		145	5	4433	106	40	5,571
Fertilizer		60	18	4433	348	129	5,571
Panel C: Land							
Plot Size (1000 Square meters)		3.6	2.9	4433	3.5	2.8	5571
Plot distance in kms to household		0.81	1.17	4438	1.86	8.32	6012
Fraction of Wetland Plots		0.15		4438	0.16		6012
Number of plots per household		2.11	1.18	2218*	2.15	1.17	2800*
		Freq.	Perc.		Freq.	Perc.	
Plot Slope	Flat	2,502	0.56		3,308	0.59	
	Slight Slope	1,417	0.32		1,708	0.31	
	Moderate Slope	355	0.08		393	0.07	
	Steep	158	0.04		160	0.03	

Table 2: Malawi - Table of Summary Statistics: Labor Hours, Intermediate Inputs, and Land Size. Labor hours is total planting and harvesting labor hours for the entire agricultural season on each plot. Intermediate input expenditure is total expenditure on seeds and fertilizer in constant 2010 USD. * represents total number of households.

		N	Variances			Share of Within-household Variance From Total Variance
			Within-household	Between-household	Total	
Panel A: Labor to Intermediate Inputs						
2013	No Controls	12784	3.07	1.50	4.57	0.67
	All Controls	12784	2.46	1.12	3.58	0.69
2015	No Controls	13331	2.84	1.83	4.67	0.61
	All Controls	13331	2.36	1.50	3.86	0.61
All Years	No Controls	26115	3.48	1.12	4.60	0.75
	All Controls	26115	2.90	0.89	3.79	0.76
Panel B: Labor to Land						
2011	No Controls	9227	2.13	1.39	3.52	0.60
	All Controls	9227	1.78	1.14	2.92	0.62
2013	No Controls	12784	2.25	1.33	3.58	0.62
	All Controls	12784	1.75	1.07	2.82	0.64
2015	No Controls	13331	1.88	1.28	3.16	0.61
	All Controls	13331	1.49	0.93	2.42	0.63
All Years	No Controls	35342	2.74	0.81	3.55	0.77
	All Controls	35342	2.25	0.61	2.86	0.80
Panel C: Intermediate Inputs to Land						
2013	No Controls	12784	2.25	0.94	3.19	0.70
	All Controls	12784	1.91	0.71	2.62	0.72
2015	No Controls	13331	3.80	1.41	5.21	0.73
	All Controls	13331	2.95	1.12	4.07	0.74
All Years	No Controls	26115	3.84	0.90	4.74	0.81
	All Controls	26115	3.25	0.67	3.92	0.82

Table 3: Ethiopia - Variance decomposition of the factor ratios. Variances are normalized by the means. Controls include year, region, and crop fixed effects, characteristics of the primary decision maker of the plot (age, gender, education, whether manager has been away from household for more than 6 months, and relationship to the household head), shocks to harvest (due to rain, crop disease, or bad soil), total annual precipitation based on GPS coordinates of the plot, land quality(land size, soil quality, slope, elevation, and wetness), land ownership status, and plot distance from household.

		N	Variances			Share of Within-household Variance From Total Variance
			Within-household	Between-household	Total	
Panel A: Labor to Intermediate Inputs						
2010	No Controls	3479	65.07	38.24	103.31	0.63
	All Controls	3479	65.07	38.00	103.07	0.63
2013	No Controls	5018	3.00	2.62	5.62	0.53
	All Controls	5018	2.98	2.62	5.60	0.53
All Years	No Controls	8497	169.61	54.07	223.68	0.76
	All Controls	8497	167.47	53.47	220.94	0.76
Panel B: Labor to Land						
2010	No Controls	3479	0.44	0.63	1.07	0.41
	All Controls	3479	0.44	0.63	1.07	0.41
2013	No Controls	4930	0.65	0.76	1.41	0.47
	All Controls	4930	0.65	0.76	1.41	0.47
All Years	No Controls	8409	0.9	0.58	1.48	0.60
	All Controls	8409	0.76	0.66	1.42	0.61
Panel C: Intermediate Inputs to Land						
2010	No Controls	3479	9.92	6.89	16.81	0.59
	All Controls	3479	9.92	6.89	16.81	0.59
2013	No Controls	4930	6.91	7.59	14.5	0.48
	All Controls	4930	6.91	7.59	14.5	0.48
All Years	No Controls	8409	10.77	7.72	18.49	0.58
	All Controls	8409	8.44	7.99	16.43	0.58

Table 4: Malawi - Variance decomposition of the factor ratios. Variances are normalized by the means. Controls include year, season, and crop fixed effects, characteristics of the primary decision maker of the plot (age, gender, education, whether manager has been away from household for more than 6 months, and relationship to the household head), household size, shocks to harvest (due to rain, crop disease, or bad soil), total annual precipitation based on GPS coordinates of the plot, land quality(land size, soil quality, slope, elevation, and wetness), land ownership status, and plot distance from household.

	(1) All	(2) All	(3) Others' Hours	(4) Own Hours
No. of Plots	0.20*	0.19*	0.68***	-1.32***
	(0.10)	(0.10)	(0.11)	(0.26)
Plot Size (logs)	2.35***	2.33***	2.49***	1.87**
	(0.27)	(0.27)	(0.32)	(0.60)
Gender (Female=1)	-9.63***	-9.28***	-6.81***	-9.22
	(0.81)	(0.81)	(0.94)	(6.09)
Age	0.02	0.01	0.11*	-0.28
	(0.04)	(0.04)	(0.05)	(0.20)
Relationship to HH Head				
Spouse	-7.36***	-0.04	-7.09*	5.91
	(1.04)	(2.10)	(3.50)	(9.42)
Son/Daughter	-9.38***	-2.23	-6.49	10.76
	(1.40)	(2.27)	(3.70)	(8.96)
Other	-3.88*	3.22	-2.41	-6.31
	(1.72)	(2.47)	(3.81)	(10.67)
Away from HH for 6+ months (Yes=1)	0.71	0.71	1.62	-3.78
	(1.15)	(1.15)	(1.32)	(3.26)
Plot Distance to HH (KMs)	0.01	0.01	-0.10	0.02
	(0.09)	(0.09)	(0.12)	(0.18)
Own Hours (Yes=1)		8.09***		
		(2.02)		
Mean of Dep. Var.	30.13***	22.78***	16.37**	62.98***
	(4.23)	(4.61)	(6.06)	(12.77)
R^2	0.44	0.45	0.50	0.58
N	24,444	24,444	15,205	9,239
Household FE	Yes	Yes	Yes	Yes
Crop FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Activity Type	Yes	Yes	Yes	Yes

Table 5: Ethiopia - Self vs. Non-self Reported Hours - Dependent variable is average individual hours per plot. Observation units are household individuals. Regressions include household, year, and crop fixed effects, as well as controls for land quality (soil quality, slope, elevation, wetness), plot distance from household, land ownership status, harvest shocks, and individual's relationship to the household head. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	(1) All	(2) All	(3) Others' Hours	(4) Own Hours
No. of Plots	-2.00***	-2.00***	-1.58***	-2.79***
	(0.31)	(0.31)	(0.38)	(0.64)
Plot Size (logs)	2.58***	2.58***	2.35***	3.13***
	(0.33)	(0.33)	(0.43)	(0.61)
HH Size	0.64*	0.65*	1.55***	-0.40
	(0.30)	(0.30)	(0.39)	(0.56)
Gender (Female=1)	-1.49**	-1.66**	-1.69*	4.96
	(0.57)	(0.57)	(0.74)	(2.91)
Age	0.15***	0.15***	0.17***	0.26*
	(0.03)	(0.03)	(0.04)	(0.13)
Relationship to HH Head				
Spouse	0.51	0.95	0.17	-4.00
	(0.70)	(0.72)	(1.19)	(2.96)
Child	-10.17***	-9.14***	-9.88***	-0.57
	(0.99)	(1.03)	(1.43)	(5.64)
Other	-8.56***	-7.50***	-9.00***	3.90
	(1.29)	(1.32)	(1.72)	(8.38)
Away for 6+ months	-0.96	-0.76	-1.53	-1.18
	(2.11)	(2.11)	(2.70)	(7.07)
Plot Distance to HH (KMs)	-0.08	-0.08	-0.10	-0.07
	(0.07)	(0.07)	(0.10)	(0.12)
Own Hours (Yes=1)		1.67***		
		(0.47)		
Mean of Dep. Var.	38.04***	36.95***	30.66***	37.88***
	(3.69)	(3.70)	(4.89)	(8.44)
R^2	0.40	0.40	0.41	0.48
N	38,137	38,137	23,383	14,754
Household FE	Yes	Yes	Yes	Yes
Crop FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Season FE	Yes	Yes	Yes	Yes
Activity Type	Yes	Yes	Yes	Yes

Table 6: Malawi - Self vs. Non-self Reported Hours - Dependent variable is average individual hours per plot. Observation units are household individuals. Regressions include household, year, season, and crop fixed effects, as well as controls for land quality (soil quality, slope, elevation, wetness), plot distance from household, land ownership status, harvest shocks, and individual's relationship to the household head. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Mean	N	Variances			Share of Within-household Variance From Total
			Within-Household	Between-Household	Total	
Ethiopia (2011, 2013, and 2015)						
Panel A: Labor Hours to Land Size						
Actual Hours	1.78	52,989	1196	888	1585	0.75
Corrected Hours	0.74	52,989	41	15	55	0.75
Panel B: Labor Hours to Intermediate Inputs Ratio						
Actual Hours	0.389	52,989	2737	111	3446	0.79
Corrected Hours	0.23	52,989	12192	1255	19255	0.63
Malawi (2010 and 2013)						
Panel A: Labor Hours to Land Size						
Actual Hours	0.74	6,974	276	253	606	0.46
Corrected Hours	0.18	6,974	13922	2256	29114	0.48
Panel B: Labor Hours to Intermediate Inputs Ratio						
Actual Hours	0.19	7,010	778	212	1448	0.54
Corrected Hours	0.19	7,010	587	42	862	0.68

Table 7: Household-Reported Labor Hours versus Predicted Hours and the Impact on the Variance Analysis - When predicted hours are used in the factor ratios, variances in the factor ratios are much smaller and a larger fraction of the variations is explained by controls. This Reveals measurement error in hours size leads to larger variances both within and between households. Hours in this analysis are only hours for household members for planting and harvesting activities, and it does not take into account hired or free labor. Variances are normalized by means. Controls include year, and crop, and region fixed effects, age and gender of the manager, whether manager has been away for more than 6 months, land ownership status, plot distance from household, land quality (slope, elevation, and wetness index), seed type, indicator for fertilizer use, and for hired or free labor on the plot, and whether there has been any crop damage on the plot during the season.

Number of Holders per Household	**Ethiopia**				**Malawi**			
	Male	Female	Mix	Total	Male	Female	Mix	Total
1	6,267	1,541	0	7,808	3,461	1,544	0	5,005
	(0.78)	(0.19)	(0.00)	(0.98)	(0.67)	(0.30)	(0.00)	(0.96)
2	97	9	57	163	15	12	153	180
	(0.01)	(0.00)	(0.01)	(0.02)	(0.00)	(0.00)	(0.03)	(0.03)
3	8	1	9	18	0	0	4	4
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
4	0	0	1	1	0	0	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total	6,372	1,551	67	7,990	3476	1,556	157	5,189

Table 8: Plot Manager's Gender: Numbers in parenthesis indicate the percent frequency. In Ethiopia, among 98% of the households who have one individual managing all household plots, 80% are male holders and 20% are female holders. In Malawi, 70% of single-manager households are male and 30% are female. In total, there are only 67 household out of 7990 households in Ethiopia and 157 out of 5189 in Malawi, in which plot managers are a mix of males and females.

	N	Variances			Share of Within-HH Variance from Total Variance
		Within-household	Between-household	Total	
Ethiopia					
Panel A: Labor to Intermediate Inputs					
One-Manager Households	24,871	2.89	0.91	3.8	0.76
Multi-Manager Households	1,035	1.75	0.72	2.47	0.71
Multi-Manager/Mixed-Gender Households	348	0.9	0.39	1.29	0.70
Panel B: Labor Hours to Land					
One-Manage Households r	24,871	2.09	0.68	2.77	0.75
Multi-Manager Households	1,035	2.42	0.47	2.89	0.84
Multi-Manager/Mixed-Gender Households	348	2.2	0.53	2.73	0.80
Panel C: Intermediate Inputs to Land					
One-Manager Households	24,871	3.21	0.72	3.93	0.82
Multi-Manager Households	1,035	3.75	0.89	4.64	0.81
Multi-Manager/Mixed-Gender Households	348	5.03	1.41	6.44	0.78
Malawi					
Panel A: Labor to Intermediate Inputs					
One-Manager Households	8,036	254.29	69.49	323.78	0.79
Multi-Manager Households	461	31.64	129.7	161.34	0.20
Multi-Manager/Mixed-Gender Households	405	28.34	116.2	144.54	0.20
Panel B: Labor Hours to Land					
One-Manager Households	7,952	0.85	0.55	1.4	0.61
Multi-Manager Households	457	0.84	0.61	1.45	0.58
Multi-Manager/Mixed-Gender Households	401	0.87	0.62	1.49	0.58
Panel C: Intermediate Inputs to Land					
One-Manager Households	7,952	7.82	6.73	14.55	0.54
Multi-Manager Households	457	8.66	5.07	13.73	0.63
Multi-Manager/Mixed-Gender Households	401	8.89	5.11	14.00	0.64

Table 9: Variance decomposition of the factor ratios by the number of managers and manager's gender. The variances are average across the years of the panel, and they are normalized by the means. The results are after controlling for observable characteristics of the decision maker, land, region, and crop differences. The purpose of this analysis is to see whether inefficiencies in the allocation of factors is larger within households in which multiple people make primary decisions about agricultural activities on different household plots as well as shocks to harvest.

	Mean	N	Variances			Adjusted Share of Within-household Variance From Total
			Within-Household	Between-Household	Total	
Ethiopia (2011, 2013, and 2015)						
Panel A: Labor Hours to Intermediate Inputs Ratio						
Group 1	2.04	11615	2.32	1.24	3.56	0.77
Group 2	1.74	14291	3.23	0.7	3.93	0.91
Panel B: Labor Hours per Sq. Meter of Land						
Group 1	0.26	11615	2.03	0.97	3.00	0.79
Group 2	0.28	14291	1.94	0.53	2.47	0.85
Panel C: Intermediate Inputs per Sq. Meter of Land						
Group 1	0.43	11615	3.13	1.06	4.19	0.88
Group 2	0.58	14291	2.98	0.62	3.6	0.92
Malawi (2010 and 2013)						
Panel A: Labor Hours to Intermediate Inputs Ratio						
Group 1	16.842	3836	0.68	0.55	1.23	0.89
Group 2	27.294	4771	2.55	0.8	3.35	1.05
Panel B: Labor Hours per Sq. Meter of Land						
Group 1	0.15	3836	0.64	0.64	1.28	0.80
Group 2	0.12	4771	0.84	0.52	1.36	0.80
Panel C: Intermediate Inputs per Sq. Meter of Land						
Group 1	0.21	3836	5.9	11.26	17.16	0.55
Group 2	0.14	4771	8.16	4.09	12.25	0.93

Table 10: Use of Fertilizer on Some and Not All Plots Within a Household and the Impact on the Variance Analysis - Group 1 includes all households in which fertilizer is applied to none or all household plots, and Group 2 includes all households who apply fertilizer to some of their plots and not to all. Variances are normalized by means. Controls include year, and crop, and region fixed effects, age and gender of the manager, whether manager has been away for more than 6 months, land ownership status, plot distance from household, land quality (slope, elevation, and wetness index), seed type, indicator for fertilizer use, and for hired or free labor on the plot, and whether there has been any crop damage on the plot during the season.

	Mean	N	Variances: Within-household	Between-household	Total	Share of Within-HH Variance from Total Variance
Ethiopia						
Panel A: Labor to Intermediate Inputs						
0	2.42	1,954	1.6	1.37	2.97	0.54
1	2.01	4,647	2.51	1.05	3.56	0.70
2	1.92	7,041	3.08	1.07	4.15	0.74
3	1.70	5,874	2.87	1.15	4.02	0.71
4	1.73	6,390	2.39	0.93	3.32	0.72
Panel B: Labor Hours to Land						
0	0.32	1,954	1.49	1.36	2.85	0.53
1	0.31	4,647	1.83	0.77	2.6	0.71
2	0.27	7,041	2.19	0.97	3.16	0.71
3	0.26	5,874	1.71	0.66	2.37	0.74
4	0.23	6,390	2.06	0.67	2.73	0.76
Panel C: Intermediate Inputs to Land						
0	0.43	1,954	3.82	2.03	5.85	0.65
1	0.57	4,647	2.72	1.14	3.86	0.71
2	0.49	7,041	3.45	0.97	4.42	0.79
3	0.52	5,874	2.8	0.68	3.48	0.81
4	0.50	6,390	2.75	0.78	3.53	0.78
Malawi						
Panel A: Labor to Intermediate Inputs						
0	391.72	1,691	90.57	61.04	151.61	0.60
1	288.55	1,197	195.05	83.52	278.57	0.70
2	271.57	1,867	236.85	89.66	326.51	0.73
3	97.26	1,843	258.93	76.68	335.61	0.77
4	120.09	1,993	370.85	140.95	511.8	0.72
Panel B: Labor Hours to Land						
0	0.14	1,674	0.51	0.8	1.31	0.45
1	0.14	1,188	0.51	0.64	1.15	0.47
2	0.13	1,837	0.59	0.57	1.16	0.48
3	0.13	1,825	0.85	0.48	1.33	0.58
4	0.13	1,967	0.85	0.66	1.51	0.54
Panel C: Intermediate Inputs to Land						
0	0.15	1,674	7.47	16.21	23.68	0.32
1	0.15	1,188	5.14	3.86	9.00	0.56
2	0.15	1,837	9.82	5.13	14.95	0.65
3	0.19	1,825	11.70	6.47	18.17	0.65
4	0.18	1,967	7.41	5.23	12.64	0.60

Table 11: Variance Decomposition of the Factor Ratios by Plot Distance from Household - Households are grouped into 5 groups: $x = 0$ representing all households, in which all plots are near one another, and $x = 1$, $x = 2$, $x = 3$, and $x = 4$ representing quartiles based of distance between plots within the household.

Dependent Variable: Manager's Total Labor Hours on Plot (logs)	(1) Ethiopia		(2) Malawi	
Plot Distance from HH larger than Median (Yes=1)	-0.20^{***}	(0.02)	-0.05^{**}	(0.02)
Total Hours Worked by Others (log)	0.58^{***}	(0.00)	0.81^{***}	(0.00)
Land Area ($\leq$ 0.1 Ha omitted)				
0.1-0.5 Ha	0.34^{***}	(0.01)	0.08^{***}	(0.01)
0.5-1 Ha	0.50^{***}	(0.02)	0.11^{***}	(0.01)
1-2 Ha	0.61^{***}	(0.04)	0.15^{***}	(0.02)
2-5 Ha	0.36^{***}	(0.10)	0.15^{*}	(0.06)
5-10 Ha	0.53^{**}	(0.19)	0.12	(0.11)
10 > Ha	0.22^{**}	(0.08)		
Major Crop (Yes=1)	0.23^{***}	(0.01)	-0.01	(0.01)
Agricultural Activity Type (Planting is omitted)				
Other Planting	-	-	-0.03^{***}	(0.01)
Harvesting	-0.36^{***}	(0.01)	-0.19^{***}	(0.01)
Mean of Dep. Variable	1.29^{***}	(0.02)	0.47^{***}	(0.02)
R^2	0.50		0.78	
N	54,149		24,315	
Household FE	Yes		Yes	
Year FE	Yes		Yes	

Table 12: Fixed Effect Determinants of Total Manager's Labor Hours on Plot - Dependent variable is the total amount of labor hours spent by the manager working on the plot in log terms. The unit of observation is a plot of land. Regression also includes year-distance interaction terms. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Ethiopia					Malawi			
	2011	2013	2015	Total	Perc.	2010	2013	Total	Perc.
Granted by Local Leaders	5165	5916	5492	16573	0.42	524	569	1093	0.10
Inherited	4574	5834	6353	16761	0.42	3562	4654	8216	0.74
Rent	1166	1295	651	3112	0.08	397	580	977	0.09
Borrowed for Free	384	108	120	612	0.02	127	174	301	0.03
Moved in Without Permission	293	154	181	628	0.02	31	36	67	0.01
Purchased	0	0	409	409	0.01	132	165	297	0.03
Other	418	303	871	1592	0.04	88	89	177	0.02

Table 13: Distribution of Land Ownership Status: 84% of the plots in both countries are either granted by the local leaders or inherited. Also, about 10-12% of the plots are rented or borrowed for free.

	N	Variances			Share of Within-HH Variance from Total Variance
		Within-household	Between-household	Total	
Ethiopia					
Panel A: Labor to Intermediate Inputs					
HHs without Marketed Land	23,125	2.87	0.87	3.74	0.77
HHs with Marketed Land	3,257	2.42	1.35	3.77	0.64
Panel B: Labor Hours to Land					
HHs without Marketed Land	31,260	2.18	0.58	2.76	0.79
HHs with Marketed Land	4,401	2.19	0.86	3.05	0.72
Panel C: Intermediate Inputs to Land					
HHs without Marketed Land	23,125	3.17	0.68	3.85	0.82
HHs with Marketed Land	3,257	3.53	1.07	4.60	0.77
Malawi					
Panel A: Labor to Intermediate Inputs					
HHs without Marketed Land	7,538	206.25	55.02	261.27	0.79
HHs with Marketed Land	1,053	247.87	123.87	371.74	0.67
Panel B: Labor Hours to Land					
HHs without Marketed Land	7,451	0.76	0.48	1.24	0.61
HHs w/ Marketed Land	1,040	1.02	0.82	1.84	0.55
Panel C: Intermediate Inputs to Land					
HHs without Marketed Land	7,451	10.20	6.89	17.09	0.60
HHs with Marketed Land	1,040	11.51	5.91	17.42	0.66

Table 14: Variance Decomposition of the Factor Ratios by Land Ownership Status of the Household - If more than half of household land area is marketed land, I classify the household as a household with marketed land.

	Mean of Plot Distance from Home (KMs)	N	Average Number of Plots per Household	Within-Household Variance in Plot Distances from Home
Panel A: Ethiopia (2011, 2013, and 2015)				
Households with No Marketed Land	1.87	24,998	10.4	436.3
Households with Marketed Land	1.58	12,060	11.0	88.2
Panel B: Malawi (2010 and 2013)				
Households with No Marketed Land	1.01	6,747	3.6	5.0
Households with Marketed Land	2.26	1,954	3.3	1.2

Table 15: Within-Household Variance in Plot Distances from Home by Household's Land Ownership Status. The plot distances are in kilometers.

	(1) Hours to Int. Inputs Ratio		(2) Hours to Land Ratio		(3) Int. Inputs to Land Ratio	
Panel A: Ethiopia						
Marketed Land	**0.00**	**(0.09)**	**-0.20***	**(0.09)**	**0.20***	**(0.08)**
Manager Gender (Female=1)	-0.33	(0.22)	-0.22	(0.22)	-0.37	(0.20)
Crop Damage (Yes=1)	-0.04	(0.07)	0.03	(0.07)	-0.04	(0.06)
Land Area in Sq. Meters (logs)	-0.35***	(0.03)				
Labor Hours (logs)			0.77***	(0.03)		
Int. Inputs Expenditure (logs)					-1.32***	(0.02)
Mean of Dep. Variable	5.15***	(0.71)	-1.82**	(0.68)	9.78***	(0.61)
R^2	0.08		0.11		0.26	
N	16338		16338		16338	
Year Fixed Effects	Yes		Yes		Yes	
Region Fixed Effects	Yes		Yes		Yes	
Crop Fixed Effects	Yes		Yes		Yes	
Household Fixed Effects	Yes		Yes		Yes	
Panel B: Malawi						
Marketed Land (Yes=1)	**-0.71**	**(1.84)**	**-0.62**	**(1.82)**	**-0.09**	**(1.70)**
Manager Gender (Female=1)	1.42	(2.33)	1.58	(2.31)	-0.69	(2.16)
Crop Damage (Yes=1)	-1.11	(1.19)	-1.02	(1.18)	-2.43*	(1.10)
Land Area in Sq. Meters (logs)	-0.16	(0.66)				
Labor Hours (logs)			0.65	(0.61)		
Int. Inputs Expenditure (logs)					-6.84***	(0.29)
Mean of Dep. Variable	3.20	(5.96)	-1.85	(4.18)	59.50***	(3.33)
R^2	0.02		0.02		0.15	
N	6400		6443		6443	
Year Fixed Effects	Yes		Yes		Yes	
Season Fixed Effects	Yes		Yes		Yes	
Crop Fixed Effects	Yes		Yes		Yes	
Household Fixed Effects	Yes		Yes		Yes	

Table 16: OLS Estimates of the Determinants of Plot-level Factor Ratio. Regressions include soil type, soil quality, slope, and wetness index. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A Data Appendix

I use plot level panel data from Ethiopia and Malawi. For Ethiopia, data is from the 2011/12, 2013/14, and 2015/16 Ethiopian Socioeconomic Surveys (ESS1, ESS2, and ESS3 respectively) collected ESS1, ESS2, and ESS3 together create a panel data set of households from rural and small town areas. ESS2 and ESS3 together represent a panel of households and individuals for rural and all urban areas. ESS1 does not provide enough information on seed and fertilizer expenditure. Thus, in the analysis whenever intermediate inputs data is needed, I exclude ESS1 from the panel data. The analysis of labor or land across households uses the panel of all three years of data. The analysis at individual level only includes ESS2 and ESS3.

Information on basic household demographics, education, health, and household location come from the household questionnaire. This data is used for the characteristics of plot holder as well as labor characteristics. A holder is a person exercising management control over the operations of the agricultural holdings and makes primary decisions regarding the utilization of the available resources. The holder may operate the holding as the owner or as a manager. It is possible to have more than one holder in a single sampled household; however it is a lot more common that one holder manages all household plots. Each holder manages one or multiple parcels of land and each parcel has one or several sub-parcels called plots. [19] The analysis in this study is at the plot level. And the focus is on the allocation of resources across plots within a household versus between households.

Ethiopian data does not differentiate between dry and rainy seasons; however due to the timing of the survey, majority of plots reflect the rainy season (main agricultural season). Data for Malawi provides all the data for two agricultural seasons (dry and rainy). In Malawi as well most of the sample is plot cultivated during the rainy season.

I mainly use the two post-planting and post-harvest agricultural questionnaires to construct measures of land, labor and intermediate inputs from the data. The agriculture questionnaires focus on farming activities and information on land ownership and use, farm labor, input use, land size, land location, agriculture capital, irrigation and harvest. Land quality variables are from the geospacial data files provided with ESS. The field-crop roaster provides information on damages occurred to the crop during the agricultural season. Analysis is restricted to all plots that have been cultivated during the agricultural season and have information on all factors of interest: land, labor hours, and intermediate inputs.

Total household-plot labor hours consists of total planting and harvesting labor hours. This includes both household labor and non-household labor. Non-household labor includes hired and free labor. Data provides information on weeks, days per week, and hours per day worked by individual household members for planting and harvesting activities. To compute household-plot level labor hours, I aggregate the hours of all activities and individuals on each plot. Total household

[19] A parcel is a piece of land with fixed boundaries, which is owned by an owner. Household parcels may be located near one another or far apart; but all plots on a parcel will be connected to one another.

labor hours is measured directly from the data. [20]

For hired and free (exchange) labor, the data only provide total number of individuals, total number of days, and average daily wages per person per day by gender, age group, and activity type (planting and harvesting) for each plot, but not their hours. To construct a measure of labor hours for non-household labor, I calculate the median hours worked for household members by gender, age group, and activity type, and estimate hours for hired and free labor at plot level using the measured median daily hours. The measures of total family hours, hired labor and free labor hours adds up to total household-plot hours.

Intermediate inputs expenditure is calculated at household-plot level and includes the total seeds and fertilizer used on plot evaluated at median purchasing price. The median seed and fertilizer purchasing price is calculated for each crop and seed type (traditional and improved) for those farmers who bought their intermediate inputs. Then these median prices are used to evaluate the value of intermediate inputs for farmers who acquired their seed and fertilizer without purchasing. Fertilizer expenditure includes expenses on urea, DAP, and other inorganic fertilizer. Data does not provide information on the amount of organic fertilizer used on plots; thus, organic fertilizer expenditure is not included in my measure of plot-level intermediate inputs. In Ethiopia. traditional seed is used on more than 95% of the plots. Improved seeds are more common for maize, wheat, and teff plots.

I only study the cultivated plots in this paper. In order to study variations in the factor ratios across plots within households, I need to limit the analysis to the households that cultivate more than one plot of land. In both Ethiopian and Malawi, a large fraction of households cultivate more than one plot of land, which allows to study within-household allocation of agricultural resources across plots. Ethiopian households on average cultivate 5.5 plots of land. Median number of plots per household is 5 in Ethiopia. Figure 1 and Figure 2 show the distribution of total number of cultivated plots per household for Ethiopia and Malawi respectively. About 80% of the Ethiopian households and 65% of Malawian households cultivate more than one plot of land. Also, about 20% of the observations represent a plot cultivated with multiple crops. For the plots with a mix of crops, I define new crop code/names. For example, a plot with maize and teff has a different crop code/name from a plot with maize and sorghum, or from a plot with maize, teff and sorghum.

The ESS data provides two measures of land. One is the standard unit measures of land from GPS and/or rope and compass measures, and the other one is self-reported values, which may be reported in local units. I use conversion factors provided with the data to convert local measures of land to square meters. The two measures of land allows to compare self-reported and measured values and test the accuracy of household reported measures of land. I calculate the percent difference between self-reported land size and the enumerator measure of land size and I find large differences between measured and reported land sizes. In Appendix E, I will provide evidence that the measured land size is a more consistent measure of land.

[20] If for an individual on the plot, household has reported more than 7 days per week, or more than 12 hours of work per day per individual, that plot is entirely excluded from the analysis. This accounts for less than half a percent of the plots.

The geospacial data files provided with ESS contain information on plot's distance from household, plot slope, elevation and a wetness index. The post-planting agricultural questionnaire provides data on plot irrigation and erosion. This information is only provided for the plots that have been visited for GPS-based land-area measurement. All these plot characteristics has been used to evaluate the quality of land. Table 1 and Table 2 summarize plot characteristics. The household-level geo file contains more information on elevation, soil type, nutrients availability, nutrient retention capacity, rooting conditions, oxygen availability to roots, excess salts, toxicity, workability, and average rainfall. However, this data is at household level and not plot level. Therefore, it can be used when analyzing resource allocation across households, but cannot be used to investigate the within-household variations in the resource allocation across plots.

Data is cleaned for plots with zero or missing seeds expenditure. Also, the top and bottom 1% of the observations (plots) with extreme factor ratios are dropped from the analysis. This data cleaning stage is performed for all three factor ratios in the analysis: the ratios of labor to intermediate inputs, labor to land, and intermediate inputs to land. This results in dropping 4.8% of the plots from the analysis. Given that only about 2% of the outliers appear to be within a household, it is expected that by removing the outliers the within-household variations decrease by more than the decline in between-household variations in the factor ratios.

Table A.1 summarizes the results of the variance decomposition of the three factor ratios with and without outliers. Excluding the outliers from the analysis results in smaller variances both within and between households. However, the decrease in within-household variances in the factor ratios is larger than the decrease in between-household variations (with one exception of 2015 variations in the ratio of labor to intermediate inputs). Therefore, when i exclude the outliers from the analysis, the share of within-household variance from total variance in the factor ratios drops.

B Variance Decomposition Appendix

Following Davis and Haltiwanger (1991) variance decomposition, I decompose total observed variance in factor ratios for all household-plots across years into within-household and between-household variances. R_{ij} for household i plot j is defined as the ratio of two factors. For example ratio of labor to L_{ij} to total expenditure on intermediate inputs, M_{ij} at household i plot j is defined as:

$$R_{ij} = \frac{L_{ij}}{M_{ij}}$$

Total variance of R_{ij} across all household-plots is

$$V = \frac{1}{N} \sum_i \sum_j (R_{ij} - \bar{R})^2$$

where N is the total number of household-plots, and $\bar{R}$ is the average factor ratio across all household-plots and is defined as

$$\bar{R} = \frac{1}{N}\sum_i\sum_j R_{ij}$$

I define R_i as within household i average of R_{ij}'s such that

$$R_i = \frac{1}{N_i}\sum_j R_{ij}$$

where N_i is the number of plots within household i. The numerator of the total variance equation can be rewritten in the following form:

$$\begin{aligned}\sum_i\sum_j(R_{ij}-\bar{R})^2 =&\sum_i\sum_j\big((R_{ij}-R_i)+(R_i-\bar{R})\big)^2 = \\ &\sum_i\sum_j(R_{ij}-R_i)^2+ \\ &\sum_i\sum_j(R_i-\bar{R})^2+ \\ &\sum_i\sum_j 2(R_{ij}-R_i)(R_i-\bar{R})\end{aligned}$$

Notice that the last term in the above equation is equal to zero, because

$$\begin{aligned}\sum_i\sum_j 2(R_{ij}-R_i)(R_i-\bar{R}) =&2\sum_i\left[(R_i-\bar{R})\sum_j(R_{ij}-R_i)\right] = \\ &2\sum_i\left[(R_i-\bar{R})(\sum_j R_{ij}-\sum_j R_i)\right] = \\ &2\sum_i\big[(R_i-\bar{R})(N_iR_i-N_iR_i)\big] = 0\end{aligned}$$

Therefore, the the following holds:

$$\sum_i\sum_j(R_{ij}-\bar{R})^2 = \sum_i\sum_j(R_{ij}-R_i)^2+\sum_i\sum_j(R_i-\bar{R})^2$$

Then multiply and divide the first term on the right-hand side by N_i:

$$\sum_i\sum_j(R_{ij}-\bar{R})^2 = \sum_i N_i\frac{\sum_j(R_{ij}-R_i)^2}{N_i}+\sum_i\sum_j(R_i-\bar{R})^2$$

Define

$$V_i = \frac{\sum_j (R_{ij} - R_i)^2}{N_i}$$

as within household variance of R_{ij}.

Also, we can write

$$\sum_i \sum_j (R_i - \bar{R})^2 = \sum_i N_i (R_i - \bar{R})^2$$

Given the above simplifications, I get

$$\sum_i \sum_j (R_{ij} - \bar{R})^2 = \sum_i N_i V_i + \sum_i N_i (R_i - \bar{R})^2$$

Divining the above equation by N, we get the total variance on the left hand-side:

$$\frac{1}{N} \sum_i \sum_j (R_{ij} - \bar{R})^2 = \frac{1}{N} \sum_i N_i V_i + \frac{1}{N} \sum_i N_i (R_i - \bar{R})^2$$

Notice that the left hand side is now the total variance across all household-plots and across all years and it is decomposed into two variances: first term on the right hand-side of the equation, which is within-household variance in the factor ratios and the second term, which is the between-household component, such that

$$V = V_{WH} + V_{BH}$$

where

$$V_{WH} = \frac{1}{N} \sum_i N_i V_i$$

is the weighted average of within household variances with weights being the number of plots within household at time t, and

$$V_{BH} = \frac{1}{N} \sum_i N_i (R_i - \bar{R})^2$$

is the weighted average of the variance between households.

C Within-Household Bargaining Power Differences

In Section 6.2, I explain that a very small fraction of households in Ethiopia and Malawi (less than 4%) have multiple people managing different household plots, therefore Udry (1996)'s explanation for within-household misallocation does not apply in this dataset. However, I find some differences between the characteristics of the plots managed by men versus plots managed by women. For example, in both countries, men's plots are significantly larger than women's plots and a larger share of men's plots are marketed land (rental or purchased land).[21] Therefore, using a similar approach

[21] Table A.11 presents summary statistics of inputs used on male-cropped and female-cropped plots as well as some plot characteristics.

to Udry (1996), I examine the variation in the factor ratios as a function of plot characteristics and the characteristics of the plot's primary decision maker within a group of plots planted to the same crop by the members of a household (household and crop fixed effects).

Table A.12 and Table A.13 report estimates of a fixed effects regression the factor ratios on plot characteristics for Ethiopia and Malawi respectively. In Ethiopia, within a household, there does not seem to be a significant difference between plots managed by men versus plots managed by women in terms of factor ratios. For Malawi, however plots managed by women within a household have fewer labor relative to both land and intermediate inputs (See Column (1) and Column(2) of Table A.13). But, there is no significant difference between male- versus female-cropped plots within a household in terms of intermediate input per square meter of land.

D Monte Carlo Simulation

As the number of plots in the households increases, the relative size of the within-household to total variance in the factor ratios increases. It is important to note that some of this increase is due to the variance decomposition method used in the analysis. Imagine 100 observations divided between 2 groups of 50 observations versus between 50 groups of 2 observations. We would expect the size of within-group variance from total to be much larger in the former case. If I divide the sample into groups of households based on the number of plots household cultivates and decompose the variances in the factor ratios for each group of households, I see the relative size of within-household variance to total becomes more significant as the number of plots within the households increase. Table A.5 summarizes these results. For example, from Panel A for Ethiopia, among households who cultivate two plots of land, only 40% of the variation in the ratio of labor hours to intermediate inputs comes from within the households. This number for households who cultivate 20 plots of land increases to 85%. In this appendix, I perform a Monte Carlo simulation exercise to investigate how much of the increase in the factor ratios is due to technicality of the variance analysis used.

Recall that the variable of interest is the relative size of within-household to total variance in the factor ratios. Let's call this variable within-household variance share. In this simulation exercise, I build a probability distribution over a range of possible outcomes for the within-household variance share. I build separate probability distributions based on the total number of plots per household. For example, to study households who cultivate a total of 2 plots of land, I randomly select a set of observation pairs and assign each pair to a household. Then, I measure the within-household and total variance in the factor ratios for this newly generated sample of households and calculate the within-household variance share. I repeat this exercise 100 times and build a probability distribution over the outcomes. Table **??** summarizes the results of the Monte Carlo simulation. Each column represents the results for each of the three factor ratio. Results from the Monte Carlo simulation show

E Measurement Error in Household Reported Land Size

The ESS data provides two measures of land. One is the standard unit measures of land from GPS and/or rope and compass measures, and the other one is self-reported values, which may be reported in local units. In Ethiopia for example, one of the most common non-standard units is the timad, traditionally defined as the amount of land a pair of oxen can plough in one day. This measure varies significantly by region, even by farmer. I use conversion factors provided with the data to convert local measures of land to square meters. All the analysis presented throughout the paper use measured land size. In this section, however I show how measurement error in a factor input impacts the variance analysis.

The measurement error is household reported measures of land has been documented by several studies. Goldstein and Udry (1999) show that the correlation coefficient between reported and measured land size is very small in their study of land size in Ghana. Dillon et al. (2016) using plot-level data from the second wave of Nigerian LSMS, showed that the GPS measures are more reliable than farmers estimates of land size. Even the GPS measures of land are not prone to error due to the error bound on devices determining the GPS coordinates. It is also important to note that the GPS measures of land size are not prone to measurement error. GPS devices have a margin of error that is dependent on environmental factors. Also, the LSMS team at the Word Bank claim that the GPS measures of land for smaller plots are less reliable, so the enumerators are instructed to use the rope and compass method to measure the size of smaller plots. Keita and Carfagna (2009) find that the difference between GPS measures and rope and compass measures are only statistically significant for smaller plots but not for larger plots. The measured land size in this section is the rope and compass measures when available (mainly for small plots) and GPS coordinate measures otherwise (for larger plots).

Having two measures of land allows to compare household-reported and measured values and test the reliability of household's responses. I calculate the percent difference between self-reported land size and the enumerator measure of land size. Figure 4 and Figure 5 show the distribution over the percent difference between household reported land size and measured land size by enumerators for Ethiopia and Malawi, respectively. The differences between measured and reported land sizes are substantial. In fact, the correlation coefficient between reported and measured land size is 0.007. These results are consistent with what Goldstein and Udry (1999) found in their study of land size in Ghana. They showed that the correlation between self-reported and measured land size is only 0.15 for Eastern Ghana. Across all three years of data, reported land size is about 5.3% smaller than measured land size on average at the plot level. In 2011, reported land on average is 5.6% smaller than measured land size; reported land on average is 6.4% in 2013 and 3.8% in 2015 smaller than measured land size.

The difference between reported and measured land size are larger for larger plots. Data in table A.6 shows that the area of smaller plots (those with an area less than 0.1 hectares) is overstated, and the area of larger plots is understated. This finding is consistent with the results from the studies by Dillon et al. (2016) from Nigeria and Carletto et al. (2015) in Malawi, Uganda, Tanzania, and

Niger. Results also show that households tend to understate their plot size by more, when their plots are larger, indicating larger measurement error in reported land size for larger plots of land. This evidence suggests that the measurement error in self-reported land size can not be taken as classical measurement error.

I also find that the area of plots with ownership certificates are understated by less than the plots with no ownership certificate (-4% for plots with certificate versus -7.35 for plots without certificate). The least difference between reported and measured land size is observed for purchased plots. For purchased parcels, the reported plot size is 1.07% smaller than measured land size (compared to -5.62% for the entire sample).

To evaluate the impact of the measurement error in land size on the variance analysis in the factor ratios, I construct two measure for the ratio of labor to land and the ratio of intermediate inputs to land; one with the measured land size and another with the reported land size. I then decompose the variations in each factor ratio into within and between household variances, and investigate how much of the variation in each measure will be explained by the observable characteristics discussed in section 3.

Table A.7 summarizes the variance decomposition for the measures of factor ratios with measured land size as well as reported land size. Panel A shows the variances in the ratio of hours to land. Both within-household and between-household variances are much higher when the reported land size is used versus when measured land is used. Results also show that the observable characteristics of household and plot explain a lot more of the variances when measured land size is used, indicating that there is better relationship between observable characteristics and the measured land size. These findings also hold for the ratio of intermediate inputs to land presented in panel B of the table.

F Tables & Figures

	Year	N	Standard Deviations			Share of Within-household Variance From Total Variance
			Within-household	Between-household	Total	
Panel A: Labor to Intermediate Inputs						
With Outliers	2013	14702	220.92	111.84	247.62	0.80
	2015	14979	93.01	94.43	132.55	0.49
Without Outliers	2013	13997	2.37	2.10	3.17	0.56
	2015	14260	2.30	1.85	2.95	0.61
Panel A: Labor to Land						
With Outliers	2011	10177	24.71	16.66	29.81	0.69
	2013	14095	27.32	9.82	29.03	0.89
	2015	14861	7.79	3.62	8.59	0.82
Without Outliers	2011	9951	0.36	0.32	0.48	0.56
	2013	13415	0.52	0.37	0.64	0.66
	2015	14145	0.47	0.35	0.59	0.64
Panel A: Intermediate Inputs to Land						
With Outliers	2013	14506	319.82	131.33	345.74	0.86
	2015	15192	293.75	128.81	320.75	0.84
Without Outliers	2013	13415	1.49	0.92	1.76	0.72
	2015	14461	2.41	1.43	2.80	0.74

Table A.1: Variance decomposition of the factor ratios with and without outliers. For the analysis without outliers, the top and bottom 1% of the observations with extreme factor ratios are dropped from the data. By excluding outliers from the analysis, within-household variation in the factor ratios decreases by more than the between-household variations.

Panel A: Ethiopia

	2011			2013			2015		
	Freq.	**Perc.**	**Cum.**	**Freq.**	**Perc.**	**Cum.**	**Freq.**	**Perc.**	**Cum.**
$\leq$ 0.1 Ha	6104	0.51	0.51	7155	0.52	0.52	8404	0.54	0.54
0.1-0.5 Ha	5140	0.43	0.94	5725	0.42	0.94	6162	0.40	0.94
0.5-1 Ha	736	0.06	1.00	666	0.05	0.99	796	0.05	0.99
1-2 Ha	68	0.01	1.00	141	0.00	1.00	150	0.01	1.00
2-5 Ha	10	0.00	1.00	54	0.00	1.00	37	0.00	1.00
5-10 Ha	1	0.00	1.00	4	0.00	1.00	7	0.00	1.00
10 $>$ Ha	0	0.00	1.00	1	0.00	1.00	3	0.00	1.00

Panel B: Malawi

	2010			2013		
	Freq.	**Perc.**	**Cum.**	**Freq.**	**Perc.**	**Cum.**
$\leq$ 0.1 Ha	536	0.12	0.12	776	0.14	0.14
0.1-0.5 Ha	2937	0.66	0.78	3660	0.66	0.80
0.5-1 Ha	802	0.18	0.96	932	0.17	0.97
1-2 Ha	144	0.03	0.99	181	0.03	1.00
2-5 Ha	14	0.00	0.99	21	0.00	1.00
5-10 Ha	0	0.00	0.99	0	0.00	1.00
10 $>$ Ha	0	0.00	1.00	1.00	0.00	1.00

Table A.2: Distribution of plot size (Hectares) - Plots are average are smaller in Ethiopia than they are in Malawi. In both countries almost all plots are less than 1 hectares.

	(1) Hours to Int. Inputs Ratio		(2) Hours to Land Ratio		(3) Int. Inputs to Land Ratio	
Plot Area Category						
0.1-0.5 Ha	0.08*	(0.04)	-0.23***	(0.01)	-0.36***	(0.03)
0.5-1 Ha	0.28***	(0.08)	-0.28***	(0.01)	-0.40***	(0.07)
1-2 Ha	0.65***	(0.18)	-0.28***	(0.03)	-0.37*	(0.15)
2-5 Ha	0.94**	(0.33)	-0.29***	(0.05)	-0.38	(0.28)
5-10 Ha	0.34	(0.82)	-0.27*	(0.14)	-0.38	(0.71)
10 > Ha	1.82	(1.29)	-0.27	(0.22)	0.02	(1.11)
Hired Labor (Yes=1)	0.00	(0.04)	-0.01	(0.01)	-0.07*	(0.03)
Used Fertilizers (Yes=1)	0.47***	(0.04)	0.02*	(0.01)	0.37***	(0.04)
Manager Age	-0.00	(0.00)	0.00***	(0.00)	0.00	(0.00)
Manager Gender (Female=1)	0.11*	(0.05)	-0.01	(0.01)	0.02	(0.04)
Manager has ever attended school (No=1)	-0.08*	(0.04)	0.01	(0.01)	-0.05	(0.03)
Manager has been away from household?						
Less than 6 months	-0.06	(0.07)	0.01	(0.01)	0.06	(0.06)
More than 6 months	0.20	(0.53)	-0.15	(0.09)	-0.25	(0.45)
Manager Relationship to the Head						
Spouse	0.08	(0.19)	-0.01	(0.03)	-0.11	(0.16)
Child	-0.16	(0.14)	-0.01	(0.02)	-0.10	(0.12)
Other	-0.20	(0.25)	0.05	(0.04)	-0.15	(0.22)
Land Ownership Status						
Inherited	-0.06	(0.04)	0.00	(0.01)	-0.02	(0.04)
Rent	-0.01	(0.07)	-0.01	(0.01)	-0.13*	(0.06)
Borrowed for Free	0.13	(0.18)	0.05	(0.03)	-0.06	(0.16)
Moved in Without Permission	-0.33*	(0.16)	-0.04	(0.03)	-0.09	(0.13)
Shared Crop in	-0.12	(0.10)	-0.01	(0.02)	0.37***	(0.08)
Purchased	0.10	(0.15)	0.02	(0.02)	-0.98***	(0.13)
Other	-0.20	(0.16)	-0.02	(0.03)	0.39**	(0.13)
Avg 12-month total rainfall(mm)	-0.00*	(0.00)	-0.00***	(0.00)	-0.00***	(0.00)
Constraint to work on the plot?						
Moderate	0.05	(0.05)	0.00	(0.01)	0.02	(0.05)
Severe	0.06	(0.06)	-0.02**	(0.01)	0.05	(0.05)
Very Severe	-0.07	(0.06)	-0.03**	(0.01)	-0.17***	(0.05)
Mainly Non-Soil Plot	0.46***	(0.10)	0.06***	(0.02)	-0.20*	(0.09)
Water	-0.36	(2.61)	0.01	(0.44)	-0.73	(2.24)
Plot Distance from Household (Quartiles)						
Q2	-0.03	(0.05)	-0.06***	(0.01)	-0.03	(0.04)

Continued on next page

	(1) Hours to Int. Inputs Ratio		(2) Hours to Land Ratio		(3) Int. Inputs to Land Ratio	
Q3	-0.03	(0.06)	-0.07***	(0.01)	0.02	(0.05)
Q4	-0.03	(0.05)	-0.06***	(0.01)	0.04	(0.05)
Plot Soil Quality						
Fair	-0.04	(0.04)	0.01	(0.01)	-0.01	(0.03)
Poor	-0.12*	(0.05)	-0.01	(0.01)	-0.04	(0.04)
Plot Slope						
Slight slope	-0.05	(0.05)	-0.01	(0.01)	-0.06	(0.04)
Moderate slope	-0.11*	(0.05)	-0.00	(0.01)	-0.05	(0.05)
Steep	-0.07	(0.06)	-0.01	(0.01)	-0.05	(0.05)
Plot Elevation (km)	-0.10*	(0.05)	0.06***	(0.01)	0.20***	(0.04)
Plot Potential Wetness Index	-0.01	(0.01)	0.00	(0.00)	0.00	(0.01)
Harvest Shocks (No=1)	-0.00	(0.04)	-0.00	(0.01)	0.03	(0.03)
Mean of Dep. Variable	0.76***	(0.22)	0.35***	(0.04)	1.21***	(0.19)
R^2	0.03		0.08		0.07	
N	26577		26577		26577	
Year Fixed Effects	Yes		Yes		Yes	
Region Fixed Effects	Yes		Yes		Yes	
Crop Fixed Effects	Yes		Yes		Yes	

Table A.3: Ethiopia - OLS Estimates of the Determinants of Plot-level Factor Ratio. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1) Hours to Int. Inputs Ratio		(2) Hours to Land Ratio		(3) Int. Inputs to Land Ratio	
Plot Area Category						
0.1-0.5 Ha	0.01	(0.79)	-0.19***	(0.00)	-12.36***	(1.24)
0.5-1 Ha	0.42	(1.00)	-0.25***	(0.01)	-16.32***	(1.55)
1-2 Ha	0.26	(1.66)	-0.27***	(0.01)	-20.17***	(2.58)
2-5 Ha	-0.41	(4.55)	-0.29***	(0.02)	-21.09**	(7.04)
Manager Gender (Female=1)	0.17	(0.68)	-0.01**	(0.00)	0.41	(1.05)
Manager Age	-0.01	(0.02)	0.00**	(0.00)	0.02	(0.03)
Household Size	-0.01	(0.11)	0.01***	(0.00)	0.06	(0.18)
Manager Relationship to the Head						
Spouse	1.65	(1.12)	-0.01	(0.01)	5.31**	(1.74)
Child	-1.01	(4.09)	-0.01	(0.02)	-1.86	(6.32)
Other	0.07	(3.74)	-0.04	(0.02)	-3.39	(5.77)
Manager's Education Level						
Some Schooling	0.15	(0.76)	-0.02***	(0.00)	0.79	(1.18)
Non-University Diploma	0.07	(3.21)	-0.04*	(0.02)	13.03**	(5.00)
University Diploma	0.49	(3.92)	-0.06*	(0.02)	16.11**	(6.21)
Post-Grad	-2.11	(9.08)	-0.03	(0.05)	9.74	(14.02)
Land Ownership Status						
Inherited	-1.17	(0.90)	0.00	(0.00)	1.16	(1.39)
Bride Price	-1.27	(3.38)	-0.04*	(0.02)	-1.08	(5.22)
Purchased w/ Title	-2.95	(2.53)	-0.00	(0.01)	0.58	(3.94)
Purchased w/o Title	-1.93	(2.11)	-0.00	(0.01)	0.26	(3.26)
Leasehold	6.13	(5.03)	-0.01	(0.03)	17.03*	(7.77)
Short-term Rent	-2.04	(1.26)	-0.02**	(0.01)	1.76	(1.95)
Farming as Tenant	-1.99	(4.92)	-0.04	(0.03)	-1.56	(7.59)
Borrowed for Free	-1.32	(1.83)	-0.01	(0.01)	-0.23	(2.86)
Moved in w/o Permission	-2.37	(3.52)	-0.01	(0.02)	11.61*	(5.43)
Other	-3.80	(4.14)	-0.02	(0.02)	-1.94	(6.40)
Plot Distance from Household (Quartiles)						
Q2	0.34	(0.72)	-0.01	(0.00)	-0.45	(1.11)
Q3	-0.14	(0.76)	-0.00	(0.00)	0.53	(1.18)
Q4	-0.48	(0.74)	-0.02***	(0.00)	-0.43	(1.15)
Plot Soil Quality						
Fair	0.47	(0.56)	0.00	(0.00)	0.59	(0.87)
Poor	-0.76	(0.86)	-0.00	(0.00)	0.49	(1.34)
Plot Slope						

Continued on next page

	(1) Hours to Int. Inputs Ratio		(2) Hours to Land Ratio		(3) Int. Inputs to Land Ratio	
Slight slope	-0.29	(0.58)	0.00	(0.00)	2.51**	(0.90)
Moderate slope	-1.06	(1.02)	0.00	(0.01)	1.77	(1.58)
Steep	-1.07	(1.53)	-0.00	(0.01)	-2.62	(2.38)
Wetland (No=1)	-0.16	(0.74)	-0.02***	(0.00)	0.97	(1.16)
Irrigation						
Bucket	0.05	(2.93)	-0.01	(0.02)	8.90	(4.71)
Hand pump	-1.26	(20.85)	-0.13	(0.11)	-14.91	(32.30)
Treadle pump	1.37	(8.14)	0.01	(0.05)	-8.76	(13.31)
Motor pump	-0.03	(25.26)	0.01	(0.14)	-11.98	(39.04)
Gravity	-0.93	(10.62)	0.05	(0.06)	10.53	(16.42)
Rainfed/No irrigation	0.56	(3.27)	-0.02	(0.02)	-2.96	(5.11)
Other	0.03	(5.00)	0.06*	(0.03)	-23.31**	(8.13)
Harvest Shock (Yes=1)	-0.73	(0.57)	-0.00	(0.00)	-0.91	(0.88)
Mean of Dep. Variable	4.86	(3.63)	0.29***	(0.02)	21.05***	(5.67)
R^2	0.01		0.24		0.03	
N	10221.00		10120.00		10120.00	
Year Fixed Effects	Yes		Yes		Yes	
Region Fixed Effects	Yes		Yes		Yes	
Crop Fixed Effects	Yes		Yes		Yes	

Table A.4: Malawi - OLS Estimates of the Determinants of Plot-level Factor Ratio. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel A: Ethiopia

Household Number of Plots	Relative Size of Within-HH Variance from Total		
	Ratio of Labor Hours to Intermediate Inputs	Labor Hours per Sq. Meter of land	Intermediate Inputs per Sq. Meter of land
2	0.40	0.38	0.49
3	0.51	0.54	0.63
4	0.62	0.59	0.70
5	0.65	0.65	0.73
6	0.72	0.72	0.74
7	0.67	0.67	0.66
8	0.73	0.72	0.73
9	0.84	0.76	0.84
10	0.81	0.60	0.81
11	0.74	0.64	0.74
12	0.77	0.76	0.73
13	0.79	0.74	0.85
14	0.84	0.84	0.85
15	0.90	0.74	0.84
16	0.85	0.81	0.89
17	0.86	0.91	0.85
18	0.90	0.88	0.81
19	0.85	0.78	0.80
20	0.85	0.61	0.71

Panel B: Malawi

Household Number of Plots	Relative Size of Within-HH Variance from Total		
	Ratio of Labor Hours to Intermediate Inputs	Labor Hours per Sq. Meter of land	Intermediate Inputs per Sq. Meter of land
2	0.6	0.44	0.3
3	0.77	0.54	0.63
4	0.76	0.62	0.6
5	0.72	0.72	0.52
6	0.66	0.66	0.94
7	0.67	0.67	0.81

Table A.5: Relative Size of Within-Household Variance to Total Variance in the Factor Ratios and Household Number of Plots - As total number of cultivated plots within a household increases, the share of the total variance in the factor ratios that comes from within the household increases.

Plot Area (Ha)	Summary of the percentage difference between reported and measured land Size					
	Ethiopia			**Malawi**		
	Mean	Stdv.	Freq.	Mean	Stdv.	Freq.
<= 0.1 Ha	3.76	47.04	7,708	53.63	37.09	1,798
0.1-0.5 Ha	-9.35	33.72	11,418	7.90	36.60	6,743
0.5-1.0 Ha	-21.68	28.53	1,517	-17.75	31.83	1,703
1.0-2.0 Ha	-21.59	30.71	254	-27.00	32.03	292
2.0-5.0 Ha	-17.11	31.33	74	-41.18	35.63	23
5.0-10.0 Ha	-30.05	44.39	11			
10.0 > Ha	-6.03	4.70	4			
Total	-5.61	39.60	20,986	10.47	42.17	10,559

Table A.6: Area categories are based on measured land size. Ethiopian households tend to understate land size on average and Malawian households tend to overstate land size. For both countries, households tend to overstate the size of smaller plots and understate the size of the larger plots.

	Mean	N	Variances			Share of Within-household Variance From Total
			Within-Household	Between-Household	Total	
Ethiopia (2011, 2013, and 2015)						
Panel A: Labor Hours to Land Size						
Measured Land Size	0.2	14,796	1.66	1.75	2.41	0.69
Reported Land Size	0.43	14,796	73	75	105	0.70
Panel B: Intermediate Inputs to Land Size						
Measured Land Size	0.38	14,796	2.99	2.52	3.91	0.76
Reported Land Size	1.15	14,796	414	359	548	0.76
Malawi (2010 and 2013)						
Panel A: Labor Hours to Land Size						
Measured Land Size	0.13	8,295	0.87	1.1	1.4	0.62
Reported Land Size	2.47	8,375	1968	1566	2516	0.78
Panel B: Intermediate Inputs to Land Size						
Measured Land Size	0.17	8,295	9.56	12.15	15.46	0.62
Reported Land Size	9.17	8,375	1454	1225	1901	0.76

Table A.7: Household-Reported versus Measured Land Size and its Impact on the Variance Analysis - When household-reported land size is used in the factor ratios, variances in the factor ratios are much higher and a smaller fraction of the variations is explained by controls. This Reveals measurement error in land size leads to larger variances both within and between households.
Controls include year, crop, and region fixed effects, age and gender of the manager, land ownership status, plot distance from household, land quality (slope, elevation, and wetness index), seed type, indicator for fertilizer use, and for hired or free labor on the plot, and whether there has been any crop damage on the plot during the season.

Variable	Year	Non-Marketed Plots		Marketed Plots		Difference
Panel A: Ethiopia						
Measured land size (1000 sq meters)	2011	1.57	(2.08)	2.34	(2.84)	-0.77***
	2013	1.6	(2.94)	2.77	(3.56)	-1.17***
	2015	1.62	(3.38)	2.7	(4.58)	-1.08***
Total Labor Hours per Sq. Meter of Land	2011	0.29	(0.54)	0.18	(0.37)	0.11***
	2013	0.29	(0.56)	0.14	(0.26)	0.15***
	2015	0.28	(0.49)	0.21	(0.44)	0.07***
Intermediate Inputs Expenditure per Sq. Meter of Land	2013	0.42	(0.76)	0.35	(0.62)	0.07***
	2015	0.61	(1.4)	0.48	(1.21)	0.13***
Labor Hours to Intermediate Inputs Expenditure Ratio	2011	2.14	(4.53)	1.35	(2.8)	0.79***
	2013	1.78	(3.83)	1.86	(4.08)	-0.08
Percent with hired labor	2011	0.32	(0.47)	0.36	(0.48)	-0.04***
	2013	0.27	(0.45)	0.32	(0.47)	-0.05***
	2015	0.24	(0.43)	0.28	(0.45)	-0.04***
Plot Distance from Household	2011	3.46	(83.81)	4.72	(51.76)	-1.26
	2013	1.53	(6.88)	1.58	(3.64)	-0.05
	2015	0.78	(1.85)	1.30	(3.12)	-0.52***
	2011	10,509		1,550		
N	2013	12,309		1,403		
	2015	12,256		2,015		
Panel B: Malawi						
Measured land size (1000 sq meters)	2010	3.51	(3.01)	3.35	(3.09)	0.16
	2013	3.35	(3.13)	3.03	(2.89)	0.32**
Total Labor Hours per Sq. Meter of Land	2010	0.12	(0.13)	0.1	(0.12)	0.02***
	2013	0.15	(0.18)	0.13	(0.18)	0.02*
Intermediate Inputs Expenditure per Sq. Meter of Land	2010	11.04	(46.44)	10.19	(41.15)	0.85
	2013	9.84	(38.3)	13.15	(39.72)	-3.31*
Labor Hours to Intermediate Inputs Expenditure Ratio	2010	3.46	(37.89)	1.92	(21.94)	1.54
	2013	0.12	(0.27)	0.09	(0.25)	0.03***
Percent with hired labor	2010	0.34	(0.47)	0.38	(0.49)	-0.04
	2013	0.34	(0.47)	0.4	(0.49)	-0.06**
Plot Distance from Household	2010	0.74	(1.06)	1.25	(1.65)	-0.51***
	2013	1.62	(7.42)	3.21	(12.22)	-1.59***
N	2013	3,997		660		
	2013	5,140		890		

Table A.8: Summary Statistics by Land Ownership Status (Marketed versus Non-marketed) - I defined marketed land as a parcel of land that is rented, borrowed for free, sharecropped, or purchased. Standard Deviations in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

		Mean	N	Standard Deviations			Share of
				Within-Household	Between-Household	Total	Within-household Variance From Total
Panel A: Labor Hours to Land Size							
No Controls	Variable Prices - Value						
	2013	2.04	13082	3.19	2.12	3.83	0.69
	2015	1.76	13737	2.83	2.14	3.55	0.64
	Constant Prices - Quantity						
	2013	2.07	13082	3.46	2.20	4.10	0.71
	2015	1.82	13737	3.38	2.51	4.21	0.64
With Controls	Variable Prices - Value						
	2013	2.03	13082	3.23	2.18	3.89	0.69
	2015	1.77	13737	2.83	2.19	3.58	0.63
	Constant Prices - Quantity						
	2013	2.07	13082	3.50	2.26	4.17	0.71
	2015	1.82	13737	3.38	2.54	4.23	0.64
Panel B: Intermediate Inputs to Land							
No Controls	Variable Prices - Value						
	2013	0.47	13418	0.77	0.42	0.88	0.77
	2015	0.62	14565	1.19	0.66	1.36	0.77
	Constant Prices - Quantity						
	2013	0.59	13082	9.55	4.89	10.73	0.79
	2015	0.61	13899	1.58	0.84	1.79	0.78
With Controls	Variable Prices - Value						
	2013	0.47	13082	0.76	0.44	0.88	0.75
	2015	0.62	13899	1.19	0.67	1.37	0.76
	Constant Prices - Quantity						
	2013	0.60	13082	9.54	4.90	10.73	0.79
	2015	0.60	13899	1.63	0.92	1.87	0.76

Table A.9: Comparison between the variances with two different measures of intermediate inputs. In one measure, seed expenditure is household-reported value of all seeds used on the plot and in cases where this value is missing, I use median prices by crop and seed type to evaluate the expenditure on seeds. The other measure on the other hand uses constant median prices by crop and seed type to evaluate seed quantity used on the plot. Measure one is more of a value measure whereas measure two is a quantity measure in some sense.

	Number of Holders per Household	Number of Households 2011	2013	2015	All Years	Percent	Cumulative Percent
Ethiopia	1	2,449	2,716	2,696	7,861	0.98	0.98
	2	55	60	49	164	0.02	1.00
	3	4	7	7	18	0.00	1.00
	4	0	0	1	1	0.00	1.00
	Total	2,508	2,783	2,753	8,044	1.00	
		2010	**2013**		**All Years**		
Malawi	1	2,212	2,821		5,061	0.98	0.98
	2	89	91		112	0.02	1.00
	3	2	2		3	0.00	1.00
	Total	2,303	2,914		5,176	1.00	

Table A.10: Number of holders/managers per household: 98% of the households have one individual managing all household plots in both countries. The maximum number of managers/holders in a household is 4 for Ethiopia and 3 for Malawi.

Year	Variable	N	Mean		Difference
			Male	Female	
Panel A: Ethiopia					
2011	Measured Land Size (1000 sq. meters)	10215	1.72	1.38	0.33***
	Plot Distance from HH (KMs)	8922	3.05	6.93	-3.88*
	Fraction of Marketed Plots	10215	0.14	0.07	0.07***
	Fraction of Plots w/ Hired Labor	10215	0.3	0.48	-0.18***
	Hours per Square Meter of Land	10215	0.27	0.31	-0.04***
2013	Measured Land Size (1000 sq. meters)	11448	1.74	1.33	0.41***
	Plot Distance from HH (KMs)	11411	1.48	1.77	-0.29*
	Fraction of Marketed Plots	11448	0.11	0.04	0.08***
	Fraction of Plots w/ Hired Labor	11448	0.25	0.41	-0.16***
	Fraction of Plots w/ Fertilizer	11448	0.33	0.34	-0.01
	Hours per Square Meter of Land	11448	0.27	0.31	-0.04***
	Intermediate Inputs per Square Meter of Land	11448	0.41	0.44	-0.03*
2015	Measured Land Size (1000 sq. meters)	12055	1.86	1.26	0.6***
	Plot Distance from HH (KMs)	11942	0.9	0.6	0.3***
	Fraction of Marketed Plots	12055	0.13	0.04	0.08***
	Fraction of Plots w/ Hired Labor	12055	0.22	0.35	-0.13***
	Fraction of Plots w/ Fertilizer	12055	0.35	0.35	0
	Hours per Square Meter of Land	12055	0.26	0.32	-0.06***
	Intermediate Inputs per Square Meter of Land	12055	0.57	0.7	-0.13***
Panel B: Malawi					
2010	Measured Land Size (1000 sq. meters)	3370	3.57	3.22	0.35***
	Plot Distance from HH (KMs)	3189	0.81	0.78	0.03
	Fraction of Marketed Plots	3370	0.15	0.11	0.04***
	Fraction of Plots w/ Hired Labor	3370	0.35	0.35	0
	Fraction of Plots w/ Fertilizer	3370	0.64	0.63	0.01
	Hours per Square Meter of Land	3361	0.12	0.12	0
	Intermediate Inputs per Square Meter of Land	3361	0.08	0.05	0.04***
2013	Measured Land Size (1000 sq. meters)	4320	3.47	2.89	0.58***
	Plot Distance from HH (KMs)	4311	1.93	1.69	0.24
	Fraction of Marketed Plots	4320	0.17	0.1	0.07***
	Fraction of Plots w/ Hired Labor	4320	0.35	0.36	-0.01
	Fraction of Plots w/ Fertilizer	4320	0.62	0.63	-0.01
	Hours per Square Meter of Land	4255	0.14	0.14	0
	Intermediate Inputs per Square Meter of Land	4255	0.22	0.33	-0.11***

Table A.11: Summary Statistics by Plot Manager's Gender

	(1) Hours to Int. Inputs		(2) Hours to Land		(3) Int. Inputs to Land	
Manager's Gender (Female=1)	-0.01	(0.07)	-0.04	(0.06)	-0.03	(0.08)
Manager's Age	-0.00	(0.00)	-0.00	(0.00)	0.00	(0.00)
Manager's been away For 6- months	0.06	(0.04)	-0.00	(0.03)	-0.06	(0.04)
For 6+ months	-0.28	(0.32)	-0.41	(0.26)	-0.12	(0.35)
Manager's Relationship to the HH Head						
Spouse	-0.22	(0.12)	-0.23*	(0.10)	0.00	(0.14)
Child	-0.00	(0.11)	-0.13	(0.09)	-0.12	(0.12)
Other	-0.43**	(0.15)	-0.27*	(0.12)	0.17	(0.16)
Land Size Category (Ha)						
0.1-0.5 Ha	-0.70***	(0.02)	-0.87***	(0.01)	-0.12***	(0.03)
0.5-1 Ha	-1.07***	(0.04)	-1.52***	(0.03)	-0.36***	(0.05)
1-2 Ha	-1.32***	(0.07)	-1.84***	(0.06)	-0.43***	(0.09)
2-5 Ha	-1.49***	(0.17)	-2.11***	(0.14)	-0.50**	(0.19)
5-10 Ha	-2.40***	(0.41)	-2.33***	(0.33)	0.21	(0.45)
10 > Ha	-3.60***	(0.67)	-2.53***	(0.55)	1.23	(0.74)
Plot Distance from HH (Quartiles)						
Q2	0.04	(0.02)	-0.10***	(0.02)	-0.14***	(0.02)
Q3	0.08**	(0.03)	-0.15***	(0.02)	-0.21***	(0.03)
Q4	0.01	(0.03)	-0.20***	(0.02)	-0.20***	(0.03)
Soil Quality						
Fair	0.03	(0.02)	-0.00	(0.02)	-0.03	(0.02)
Poor	0.02	(0.03)	-0.06**	(0.02)	-0.08**	(0.03)
Plot Slope						
Slight slope	-0.04	(0.03)	-0.00	(0.02)	0.04	(0.03)
Moderate slope	-0.05	(0.03)	0.01	(0.03)	0.06	(0.03)
Steep	-0.03	(0.04)	0.01	(0.03)	0.04	(0.04)
Plot Elevation (km)	0.45***	(0.08)	0.07	(0.06)	-0.39***	(0.08)
Plot Potential Wetness Index	-0.00	(0.01)	0.00	(0.00)	0.00	(0.01)
Harvest Shock (No=1)	0.05**	(0.02)	0.03	(0.01)	-0.02	(0.02)
Labor Hours (logs)	0.03**	(0.01)				
Intermediate Input Expenditure (logs)			-0.00	(0.01)		
Land Size (logs)					-0.04***	(0.01)
Mean of Dep. Var.	-2.15***	(0.21)	-1.96***	(0.17)	0.34	(0.24)
R^2	0.33		0.39		0.23	
N	26,759		26,759		26,759	
Year Fixed Effects	Yes		Yes		Yes	
Region Fixed Effects	Yes		Yes		Yes	
Household Fixed Effects	Yes		Yes		Yes	
Crop Fixed Effects	Yes		Yes		Yes	

Table A.12: Ethiopia - OLS Fixed Effects Estimates of the Determinants of Plot-level Factor Ratio. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1) Hours to Int. Inputs		(2) Hours to Land		(3) Int. Inputs to Land	
Manager's Gender (Female=1)	-0.37**	(0.14)	-0.15*	(0.08)	0.23	(0.15)
Manager's Age	-0.00	(0.01)	0.00	(0.00)	0.00	(0.01)
Household Size	0.01	(0.02)	0.06***	(0.01)	0.05	(0.03)
Manager's Relationship to the HH Head						
Spouse	0.51***	(0.15)	0.10	(0.08)	-0.41*	(0.16)
Child	0.02	(0.37)	-0.24	(0.20)	-0.25	(0.40)
Other	-1.11**	(0.39)	-0.77***	(0.22)	0.37	(0.42)
Manager's Education Level						
Some Schooling	0.04	(0.08)	-0.18***	(0.04)	-0.20*	(0.09)
Non-University Diploma	0.06	(0.38)	0.10	(0.21)	0.03	(0.41)
University Diploma	0.18	(0.46)	-0.40	(0.26)	-0.55	(0.50)
Post-Grad	2.19	(1.35)	0.46	(0.75)	-1.73	(1.46)
Land Size Category (Ha)						
0.1-0.5 Ha	-0.89***	(0.06)	-0.94***	(0.03)	0.01	(0.09)
0.5-1 Ha	-1.39***	(0.08)	-1.54***	(0.04)	-0.06	(0.13)
1-2 Ha	-1.77***	(0.13)	-2.06***	(0.07)	-0.16	(0.19)
2-5 Ha	-2.47***	(0.33)	-2.73***	(0.18)	-0.10	(0.39)
Plot Distance from HH (Quartiles)						
Q2	0.07	(0.06)	-0.05	(0.03)	-0.11	(0.06)
Q3	0.09	(0.07)	0.04	(0.04)	-0.05	(0.07)
Q4	0.00	(0.07)	-0.06	(0.04)	-0.06	(0.07)
Soil Quality						
Fair	-0.01	(0.05)	0.01	(0.03)	0.01	(0.05)
Poor	-0.09	(0.07)	-0.13**	(0.04)	-0.04	(0.08)
Plot Slope						
Slight slope	0.06	(0.05)	-0.01	(0.03)	-0.07	(0.05)
Moderate slope	0.08	(0.09)	-0.09	(0.05)	-0.16	(0.09)
Steep	-0.17	(0.13)	-0.07	(0.07)	0.10	(0.14)
Harvest Shock (Yes=1)	-0.16***	(0.05)	-0.06*	(0.03)	0.10	(0.05)
Labor Hours (logs)	0.06*	(0.02)				
Intermediate Input Expenditure (logs)			0.01	(0.01)		
Land Size (logs)					-0.06	(0.05)
Mean of Dep. Var.	-5.65**	(2.00)	-3.06**	(1.11)	2.76	(2.18)
R^2	0.54		0.33		0.47	
N	10,127		10,127		10,127	
Year Fixed Effects	Yes		Yes		Yes	
Season Fixed Effects	Yes		Yes		Yes	
Region Fixed Effects	Yes		Yes		Yes	
Household Fixed Effects	Yes		Yes		Yes	
Crop Fixed Effects	Yes		Yes		Yes	

Table A.13: Malawi - OLS Fixed Effects Estimates of the Determinants of Plot-level Factor Ratio. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$