Land Fragmentation with Double Bonuses

-- The Case of Tanzanian Agriculture

Xudong Rao

University of Minnesota

raoxx107@umn.edu

Selected Poster/Paper prepared for presentation at the Agricultural & Applied Economics Association's 2014 AAEA Annual Meeting, Minneapolis, Minnesota, 27-29 June 2014

Abstract:

Land fragmentation, also known as scattered land holdings, is a common phenomenon in agriculture around the world. In some cases, it has even persisted through government-supported land consolidation programs that aim to improve agricultural productivity. This study evaluates the effect of land fragmentation on agricultural production and hypothesizes that it may be beneficial to farmers by diversifying risk onto separate land plots that usually have heterogeneous growing conditions. Applying a stochastic frontier model to the Tanzania Living Standards Measurement Study (LSMS) data, we find evidence to support the risk-reduction hypothesis and indications that land fragmentation may be conducive to efficiency. This second finding may seem counterintuitive but is also supported by similar studies. We further argue that accounting for risk preferences that are absent from current framework in future research may help explain the double bonuses of land fragmentation.

Key words: Agricultural productivity, land fragmentation, risk management, stochastic production frontier

JEL codes: Q12, Q15, Q18

Introduction

Land fragmentation, which refers to a single farm consisting of numerous discrete plots scattered over a wide area (Binns, 1950), has long been deemed as an impediment to agricultural production and rural development. Policymakers describe it as "the blackest of evils" (Farmer, 1960), and researchers believe that it undermines efficiency and lowers profitability (e.g. Jabarin and Epplin, 1994; Nguyen et al, 1996; Wan and Cheng, 2001; Fan and Chan-Kang, 2005; and Tan et al, 2008). Until most recently, however, land fragmentation has remained a common phenomenon in both developed and developing countries. For example, Japanese rice growers operated more than four plots on average during the period 1985-2005 (Kawasaki, 2010); Albanian farmers owned four plots in 2005 (Deininger et al, 2012); and Tanzanian farms in the Mount Kilimanjaro regions cultivated an average of 2.5 plots per family in 2000 (Soini, 2005). This raises the question -- why has land fragmentation been so prevalent and persistent?

Scholars have come up with various explanations, including demographic, cultural and institutional reasons, to justify the prevalence and persistence of land fragmentation (For more discussions, see Heston and Kumar, 1983; Bentley, 1987; Blarel et al, 1992; Niroula and Thapa, 2005). Meanwhile, economists have attempted to re-interpret the role of land fragmentation in agricultural production from the perspective of risk management. McCloskey(1976) is among the first to formally hypothesize that cultivation on scattered plots with different soil and location can reduce risk, even though it incurs travel costs and other inconveniences. Such a risk-reducing function of land fragmentation has been corroborated by several other empirical studies such as Blarel et al (1992), Goland (1993), and Di Falco et al (2010).

In practice, voluntary land exchanges among farmers have been extremely rare (Heston and Kumar, 1983). Governments in many places have thus been advised to launch consolidation programs in the hope that farmers will benefit from more concentrated land holdings. Some of those programs have proven to be generally successful, while some have failed with resistance from farmers (See Heston and Kumar, 1983 for the failure cases in India; see Niroula and Thapa, 2005 for the failure cases in India, Pakistan and Thailand). Therefore, whether the existence of land fragmentation is economically justifiable is still largely inconclusive.

The fluctuation in agricultural income as a consequence of risk in agricultural production has profound implications on the well-being of farmers in developing countries. Unlike their counterparts in the developed world who may have access to government subsidies and crop insurance to protect themselves from adversity, those farmers have to resort to only primitive farming strategies, such as crop diversification and land fragmentation, to secure their production as the major income source. Further, as observed in many studies, farmers' aversion to risk may prohibit them from adopting new technologies and improved crop varieties even though they will be paid back with a higher expected return. This will lead to a stagnant growth in agricultural productivity, leaving farmers more vulnerable in the long run.

To investigate the role of land fragmentation in agricultural production, this study will discuss the economic implications of land fragmentation and evaluate its effects on both efficiency and risk. Applying the recent development in the stochastic frontier model to the analysis of land fragmentation, we expect to derive an improved characterization of this phenomenon through a careful discussion of determinants of production efficiency and production risk. The results from our model will be compared with those from similar studies to shed light on future land tenure reforms that aim to secure agricultural production and improve farmers' well-being.

Land fragmentation and plot heterogeneity

There is no unique measurement of land fragmentation whose economic implications extend beyond the scatterings of land. King and Burton (1982) propose a six-parameter characterization: farm size, plot number, plot size, plot shape, plot spatial distribution, and the size distribution of the fields, while Bentley (1987) further argues that methods of quantifying land fragmentation without a measure of distance are flawed. Among economists, the predominant measure has been the Simpson Index (SI), which may be used along with other dimension(s) of land fragmentation (e.g. Blarel et al, 1992; Hung et al, 2007; Tan et al, 2007; and Kawasaki, 2010). For a farm household cultivating a total of J plots, denote the area for plot j (j=1,2...J) by A_j, the Simpson Index is then defined as:

$$(1) \qquad SI = 1 - \sum_j^J \left(\frac{A_j}{\sum_j^J A_j} \right)^2 = 1 - \frac{1}{\left(\sum_j^J A_j\right)^2} \sum_j^J {A_j}^2 = 1 - \frac{1}{A^2} \sum_j^J {A_j}^2$$

where $A = \sum_j^J A_j$ is the total farm area. This index returns a value lying within the unit interval, and it goes up as farm becomes more fragmented. SI=1 refers to the infinite fragmentation scenario while SI=0 refers to the one-plot farm scenario. This value is jointly determined by the number of plots, the farm size, plot size and the plot size distribution.

One common phenomenon usually found associated with land fragmentation has been the heterogeneous soil quality and growing conditions across plots, or plot heterogeneity for short. It is sometimes believed to be a cause of land fragmentation or a restricting condition for land consolidation to be implemented (Mearns, 1999; Niroula and Thapa, 2005). What is significant about plot heterogeneity is its risk-management role discussed in the literature. By cultivating plots with varying micro-environments, farmers are able to reduce the variation in output or income because the risk caused by drought, flood and diseases is spread out for the same crop (Hung et al, 2007). Bentley (1987) collects a few studies from this perspective, covering both grain crops and cash crops, and argues that the risk management advantage of fragmented farms is applicable in many contexts.

Another value of plot heterogeneity is that it may encourage crop diversification (Bellon and Taylor, 1993; and Hung, 2006), a popular strategy for risk reduction. By matching the proper crop portfolio with the agro-ecological conditions across the whole farm, farmers are induced to increase crop diversity and stabilize the total farm output. Di Falco et al (2010) present an empirical analysis which finds that land fragmentation fosters crop diversification.

To summarize, the literature has spent a great deal of attention on land fragmentation's impacts on either productivity or profitability, and land fragmentation has been found to be detrimental in general. Meanwhile, the risk-management hypothesis of land fragmentation has not received sufficient empirical scrutiny, even though it was put forward in the literature a long time ago. The few existing studies that examine the risk effect of land fragmentation have focused solely on the dispersion of fields without ever considering plot heterogeneity. Considering the curious observation that land consolidation programs have succeeded mostly in places with uniform soils but failed in places with heterogeneous soils (Heston and Kumar, 1983; Mearns, 1999; Niroula and Thapa, 2005), it is reasonable to conjecture that the risk-reduction benefit of land fragmentation may be jointly determined by both plot dispersion and plot heterogeneity.

Conceptual Framework

In this section, we will provide a formal framework to characterize how land fragmentation affects both production efficiency and production risk, which is often measured by the variation in yield. The dominant approach to production efficiency analysis has been the stochastic frontier model, which was simultaneously developed by Aigner, Lovell and Schmidt (1977) and Meeusen and Van den Broeck (1977). To start, write the yield y_i (in its original unit) of farmer i (i=1, 2...N) as:

(2) $$y_i = F(\boldsymbol{X_i};\ \boldsymbol{\beta}) * exp(-u_i) * exp(v_i).$$

In (2), $F(\boldsymbol{X_i};\ \boldsymbol{\beta})$ is the deterministic production function where $\boldsymbol{X_i}$ is the input vector, including a constant term, and $\boldsymbol{\beta}$ is the corresponding parameter vector. The inefficiency term, u_i, is assumed to be greater than or equal to zero (hence it is also known as the one-sided error term) such that $exp(-u_i)$ lies within the unit interval, representing the proportion of $F(\boldsymbol{X_i};\ \boldsymbol{\beta})$ that is actually produced. When $exp(-u_i) = 1$, the production is completely efficient and lies right on the production frontier; otherwise, inefficiency exists and production lies below the frontier. Lastly, the term $exp(v_i)$ contains the regular error term v_i (also known as the two-sided error term), which captures all random factors such as noise and model misspecification. By having two separate error terms, the stochastic frontier model, which is also called the compound error model, allows the estimation of a stochastic production frontier with individual-specific inefficiency.

Empirical studies often focus on inputs and output in the logarithmic form and assume the deterministic production function after the logarithmic transformation, $f(\cdot)$, to take either the Cobb-Douglas form or the transcendental logarithmic (translog) form. This study will take the translog assumption as the more general case. This transformation allows us to see the three components of y more clearly:

(3) $\quad lny_i = f(ln\boldsymbol{X_i};\ \boldsymbol{\beta}) + v_i - u_i.$

The primary interest of stochastic production frontier analysis falls on the inefficiency term u_i, and more specific assumptions have been made about its distribution. With a truncated normal distribution for u_i, Kumbhakar et al. (1991) and Huang and Liu (1994) propose a model to parameterize the mean of the pre-truncated inefficiency distribution, μ_i, such that inefficiency could be explained by a group of exogenous variables $\boldsymbol{Z_i}$, including a constant term, through a linear function. That is:

(4) $\quad u_i \sim N^+(\mu_i, \sigma_u^2)$

where

(5) $\quad \mu_i = \boldsymbol{Z_i\gamma}$

The parameter vector $\boldsymbol{\gamma}$ in (5), or the so-called inefficiency effects, is left to be estimated. We will adopt the truncated normal assumption on u_i for the purpose of this study. Further, the two-sided error v_i is always assumed to follow the normal distribution $N(0, \sigma_v^2)$. Both v_i and u_i are often assumed to be independent of each other and $i.i.d.$ across observations.

In the traditional single-error model, heteroscedasticity usually does not cause too much trouble. In case of its presence, the coefficient estimates are still consistent although not efficient, and this problem could be easily fixed by more robust estimation procedures. However, heteroscedasticity is a much more serious problem in stochastic frontier models and may lead to inconsistent estimates of the inefficiency effects, the parameters of primary interest. This is because estimation of the inefficiency term is based upon residuals derived from the estimation of a frontier (Caudill et al, 1995; Hadri, 1999). Even worse, heteroscedasticity could be present in either or both of the one-sided error term u_i and the two-sided error term v_i, and mis-specification of either variance term, σ_v^2 or σ_u^2, will result in inconsistent estimates (Hadri, 1999). Therefore, a reliable stochastic frontier model demands a careful analysis of its two variance terms.

As reviewed in the section before, land fragmentation has been suspected of being related to production risk. In this study, we make the formal hypothesis that land fragmentation can diversify production risk onto separate land plots such that it reduces the risk on the entire farm. To see this, we follow a similar decomposition as the one in Blarel et al. (1992) and write the actual yield (in its original unit) on the jth plot of the ith farm by y_{ij} such that

(6) $$y_{ij} \equiv \bar{y}_i + d_{ij} + \theta_{ij} + e_{ij}$$

In (6), $\bar{y}_i$ is the expected farm-level yield. The term d_{ij} captures the plot-specific fixed effects that cause y_{ij} to deviate from $\bar{y}_i$, such as soil attributes. For example, if certain plot is more fertile than the other plots on the same farm, the yield on this plot will tend to be higher than the average yield on the whole farm. As opposed to d_{ij}, θ_{ij} is also plot-specific but stochastic, and it may be associated with precipitation, insolation, wind, and other random factors that define the microclimatic environment on each plot (Bentley, 1987). In general the distribution of θ_{ij} should vary from plot to plot and hence we assume $E(\theta_{ij}) = 0$ and $Var(\theta_{ij}) = \sigma_{\theta ij}^2$ for any j. Finally, e_{ij} captures all stochastic effects that are uniquely distributed for any plot on any farm, such as measurement errors, and it is assumed that $E(e_{ij}) = 0$ and $Var(e_{ij}) = \sigma_e^2$, for any i and j.

With such a decomposition, we are taking the production on the farm level as a portfolio of production on all individual plots, each of which has its own distribution of returns. To aggregate into the farm-level yield y_i, we have

$$(7) \qquad y_i = \frac{1}{A_i}\sum_j^J y_{ij}\,A_{ij} = \frac{1}{A_i}\sum_j^J\left[\left(\bar{y}_i + d_{ij}\right) * A_{ij} + \left(\theta_{ij} + e_{ij}\right) * A_{ij}\right]$$

Since we are concerned with the farm-level risk, take variance of y_i to get

$$(8) \qquad Var(y_i) = Var\left[\frac{1}{A_i}\sum_j^J\left[\left(\theta_{ij} + e_{ij}\right) * A_{ij}\right]\right]$$

$$= \frac{1}{A_i^2}Var\left[\sum_j^J\left(\theta_{ij} * A_{ij}\right)\right] + \frac{1}{A_i^2}\sum_j^J \sigma_e^2 A_{ij}^2$$

$$\equiv \sigma_{\theta i}^2 + (I - SI) * \sigma_e^2$$

Firstly, the second term on the right-hand side of (8), $(I - SI) * \sigma_e^2$, shows clearly that land fragmentation, measured by the Simpson Index, is negatively related to the yield variability on the whole farm by spreading out the common stochastic effects σ_e^2 across the plots. What is less obvious is the first term, $\sigma_{\theta i}^2$, which is the aggregation of stochastic effects that are specific to each plot and whose effect on yield variability is generally unknown unless the distribution (or at least the variance) of each θ_{ij} is given. In general, we should expect $\sigma_{\theta i}^2$ to be related to soil heterogeneity for reasons argued in Hung et al (2007). Moreover, if we believe that farmers can match the growing conditions on all plots with the proper crop portfolio as suggested by the high correlation between the two (Bellon and Taylor, 1993; and Hung, 2006), we should expect $\sigma_{\theta i}^2$ to be negatively associated with crop diversification given the latter's evident role for risk reduction.

In this way, we see that yield variability is not identical among all farms but is determined by several farm-specific factors, echoing our concern of heteroscedasticity. To be more specific, the variance of the common error term v_i should have its own explanatory variables; that is

$$(9) \qquad \sigma_{vi}^2 = \exp(\boldsymbol{h_i \alpha})$$

where $\boldsymbol{h_i}$ will include a constant term, the Simpson Index and variables for plot heterogeneity and crop diversification. Further, some factors of production have been found to affect either or both variance terms, such as labor (Hadri et al 2003). To avoid potential bias in the coefficient estimates, we retain the most general specification of σ_{ui}^2 at this step by allowing its own vector of determinants, $\boldsymbol{k_i}$, with the coefficient vector $\boldsymbol{\varphi}$:

$$(10) \qquad \sigma_{ui}^2 = \exp(\boldsymbol{k_i \varphi})$$

If heteroscedasticity is found to be absent from σ_{ui}^2 by the empirical estimation, $\boldsymbol{k_i}$ will contain only a constant term as in the homoscedastic case.

Data and Context

The data to be used for the empirical analysis come from the Tanzania National Panel Survey 2008-2009 as part of the Living Standards Measurement Study (LSMS) -- Integrated Surveys on Agriculture project conducted by the World Bank. This survey adopted a stratified, multi-stage cluster design to obtain nationally-representative sample. Rural family members were interviewed by team enumerators regarding their family socioeconomics and agricultural activities. Information such as location, ownership, soil conditions, crop varieties, input uses and harvest was collected for each cultivated plot.

For the purpose of this study, we will focus on plots that were grown either partially or fully with annual crops in the long rainy season (March, April and May) by realizing that the production of annual crops differs tremendously from that of perennial crops and trees. In this way, our sample contains 1,503 households with 2,756 plots; nearly half of the households cultivated only one plot and around 95 percent of households cultivated less than 4 plots (Table A-1). Maize is the predominant crop in terms of either frequency or planting area, and other popular annual crops include beans, groundnuts, paddy rice, and sorghum. More background information and descriptive statistics for key variables will be presented below.

In Tanzania, smallholder farming has been the predominant form of agriculture, which accommodated about 75 percent of the national population and accounted for about 45 percent of the GDP in 2008. Although Tanzania has vast areas of cropland that is suitable for intensive cultivation, the use of inputs is limited and productivity is generally low. In 2008, 37 percent of the rural population, i.e. more than one fourth of the total population, lived below the poverty line. Therefore, an efficient and secured food production has significance for Tanzania's millions of impoverished rural citizens as well as its national economy.

There is one particular issue of Tanzania's agriculture that is highly pertinent to the topic of this study -- land fragmentation. At the beginning of its independence, Tanzania adopted a communist approach and promoted collective land cultivation and shared labor for its agricultural production. An estimated 75% of the population were relocated from scattered homesteads and smallholdings to live in communal villages of 2,000-4,000 residents (Dondeyne et al, 2003; Maoulidi, 2004), even though there was a strong preference of farmers for individually allocated and individually cultivated farmland (USAID, 2011).

This approach was quickly abandoned by the following administration in the 1980s and a new legal framework was gradually installed to support private property rights and individualized control of farming. The law recognizes the rights to land and encourages productive and sustainable use of land. In principle, farmers have the rights to buy, sell, lease and mortgage their plots and decide on matters such as their crop choices and land use. By 2008, each rural household owned or cultivated an average of 2.5 plots. The shifts in Tanzania's land tenure system in the past several decades may better address the underlying economic motivations of land fragmentation as investigated in this research.

Empirical Model

Dependent Variable

Among the households in our sample, nearly 70 percent of them grew more than one crop and the crop portfolio varied from farm to farm, making it difficult to compare production efficiency across farms using a yield frontier. As an alternative, we focus on a revenue frontier by implicitly assuming revenue-maximizing farmers. A cost frontier has also been utilized in the literature, such as Kawasaki (2010) for Japanese rice growers and Tan et al (2007) for Chinese farmers. We could have tried either a cost frontier or a profit frontier[1] provided that the price for hired labor becomes available in our data set.

Therefore, the dependent variable of our empirical model is the logarithmic form of revenue per acre, which equals the aggregated value (in Tanzania shillings) of all crops grown on each farm divided by the farm area. In this survey, farmers were asked to estimate the value of their crops and the proportion of harvest finished by the time of the survey. Crop prices reported by village leaders are not adopted because of the apparent anomalies and missing observations.

Explanatory Variables of the Revenue Frontier

As stated earlier, farm area is calculated as the aggregated area for all annual crops and is included in the revenue function as an input. Besides land, labor has been the utmost important input in Tanzanian agriculture. The LSMS survey documents labor days spent by family members and, if any, hired workers on each plot at three stages of production, i.e. land preparation and planting, weeding, and harvesting, making it possible to differentiate labor spent on these activities as different inputs. For this study we add hired labor onto family labor for each activity and include in the inefficiency term (to be discussed below) the ratio of total hired labor to total family labor in order to control for the impact of labor heterogeneity on efficiency.

Inputs other than labor and land, such as fertilizers, irrigation, herbicides and pesticides, have been rare in Tanzania (Panel 1, Table A-2). Even fewer farmers have access, through either rental or possession, to draft animals (e.g., oxen) or farm machinery (e.g., tractor and thresher) although they may increase revenue significantly (Panel 2, Table A-2). Instead, the most common farm implement in Tanzania are hand hoes with all the households in our sample having at least one. In the empirical model, we will include the number of hand hoes per acre and a dummy variable for the use of any draft animal or machinery to control for their probable contribution to revenue.

Variables for average temperature and precipitation of the wettest quarter rather than those of the whole year are included as inputs to account for weather's impact on the agricultural production undertaken in the long rainy season[2]. Finally, our revenue frontier model contains a price index which equals the average price of all annual crops harvested on the farm weighted by their quantities (all in kilograms)[3].

Explanatory Variables of Inefficiency

Land Fragmentation. Variables from this category are of primary interest in this study regarding the determination of production efficiency. Table A-3 lists the descriptive statistics of the various dimensions of land fragmentation. It shows that the majority of the farms in our sample have a relatively small size with 95 percent of them less than 15 acres. The average plot size, with an average of 1.83 acres, is even smaller owing to the fragmentation of land on over half of the farms. Land fragmentation measured by the Simpson Index presents a clear bimodal distribution as a result of the large percentage of single-plot farms, while there exists only weak correlation between farm size and the Simpson Index. In terms of distance, about three fourths of the plots are located within 3 kilometers (approximately 2 miles) from either home or road. Meanwhile, less than 40 percent of the plots are within that distance from a nearby market.

To estimate the inefficiency term in the model, we will include farm size, the Simpson Index, an interaction term between the two as well as the three distance variables (from plot to home, road and market, respectively). To account for the varying effects of land fragmentation on plots with different sizes, we calculate the average plot area and average distance variables weighted by plot size. It turns out that the weighted average plot area, a somewhat obscure concept, equals farm area minus its interaction term with the Simpson Index; hence there is no need to add it to the model. To see this connection, recall that the area for the jth plot is denoted as $A_j$then the weighted average plot area is by our definition derived as

$$(11) \quad \frac{1}{A} * \sum_j^J A_j \, A_j = A * \sum_j^J \left(\frac{A_j}{A}\right)^2 = A * (1 - SI)$$

Finally, the number of plots on each farm will be excluded from the model since it is already captured by the Simpson Index[4].

Household Characteristics. In a cross-section analysis like this one, household characteristics, especially those related to labor, usually help to explain the variation in efficiency across households. Here we adopt the average age and average education[5] (measured in school years) of family workers who actually worked in the fields instead of those of all family workers, some of whom may work in non-agricultural sectors. Labor days by male workers and labor days by hired workers as the respective proportion of the total labor days will also be included.

Further, households will allocate their resources to activities other than the growing of annual crops, such as housework and perennial crops or fruit trees. With the information available, we will include the ratio of the number of kids under the age of five to the number of family field workers and the ratio of farm area used for perennial crops/fruit trees to farm area used for annual crops to control for their potential negative impacts on the efficiency. Table A-4 in the appendix lists the descriptive statistics of these household characteristics variables.

Soil Conditions. Using the geo-referenced homestead location data, the LSMS survey has imported soil and terrain data from the Harmonized World Soil Database at a resolution of 0.083degree (about 10 kilometer grids). The measures we choose to explain production efficiency are: nutrient availability, oxygen availability to roots and workability for field management (Table A-5). To include each of the measures in the estimation, we use "severe constraints" as the reference and create respective dummy variables for the other two categories, “Moderate constraints” and “No or slight constraint”, both of which expect a negative coefficient.

Explanatory Variables for Heteroscedasticity

Plot heterogeneity and Crop Diversification. As argued in the conceptual framework, the variance of yield is related to plot heterogeneity, crop diversification, and land fragmentation measured by the Simpson Index. In the LSMS survey, Tanzanian farmers are asked to report the soil type (sandy, loam, clay and others), erosion type (existent or not) and steepness of slope (flat bottom, flat top, slightly sloped and very steep) for each plot. Assuming that soil conditions can be jointly characterized by these three dimensions, we use the number of different soil profiles normalized by the number of plots to compare plot heterogeneity across farms (Table A-6).

Nearly 70 percent of farms in our sample have diversified their crop portfolio by either growing more than one crop on single plot and/or growing different crops on different plots (Table A-7). In this study, we simply use the number of different crops on the whole farm to account for its influence on yield variance.

Labor Inputs. Researchers have for long emphasized the effects of various inputs on risk, and a convenient specification has been the Just-Pope production function, which incorporates inputs into both the mean and variance functions of output. Evidence regarding the role of certain input, especially labor, has been mixed. For example, Antle and Crissman (1990) find labor to be risk reducing while Villano and Fleming (2006) argue that labor increases output variability. Further, the variance of either or both the one-sided error and two-sided error in a stochastic frontier model may be associated with producers' input use (Schmidt, 1986; Hadri, 1999; Hadri et al, 2003). Hadri et al (2003) report that expenditure on labor and machinery by farms will increase variability in efficiency, whereas land area and fertilizer cost have the opposite effect. In this paper, we will divide the aggregated labor days for all three activities by farm area and put the ratio in the variance function.

Estimation and Results

We are estimating a stochastic production frontier with a group of exogenous explanatory variables for the inefficiency term. Moreover, heteroscedasticity may be present in either or both variance terms. Instead of using the common two-step estimation approach which will generate biased estimates, this study uses the simultaneous estimation method proposed by Wang and Schmit (2002)[6].

Variance Structure

The main challenge to the empirical estimation stems from the indeterminate effects of labor on the two variance terms. Kumbhakar and Lovell (2003) propose a procedure that starts with a model that incorporates heteroscedasticity in both error components and then test the homoscedasticity restriction that respective coefficient(s) equals to zero. For this study, we start with a model, named HUV, where labor inputs appear in both variance terms with the Simpson Index, and the measures of plot heterogeneity and crop diversification in the variance of the two-sided error term. Then we move on to the two single-heteroscedasticity specifications, denoted as HU and HV respectively, where either the one-sided-error variance (U) or the two-sided-error variance (V) has its own determinant(s). Since labor input may affect the two variance terms differently from the other three variables, estimates from two alternative specifications (HU_1 and HUV_1) are also derived for model comparisons. Finally, the homoscedasticity model is estimated with only a constant term for each variance, and it is denoted as HO hereafter.

Table 1 lists the variance coefficient estimates for the six models above. Since model HUV could be seen as the unrestricted model for the other five, the likelihood ratio test can be applied to make pairwise comparisons between HUV and each of the other five. It shows that HUV is preferred to HU, HV_1 and HO but not HV and HUV_1, the likelihood ratios of which are close

Table 1 Comparison of various variance structures

	HUV	HV	HU	HO	HUV_1	HV_1
			One-sided error (*U*) variance			
Labor intensity	-0.000429		-8.90E-05		-0.000338	
	(0.001)		(0.000)		(0.000)	
Constant	-0.882***	-0.904***	-0.1	-0.104	-0.934***	-0.0293
	(0.277)	(0.302)	(0.283)	(0.286)	(0.304)	(0.306)
			Two-sided error (*V*) variance			
Simpson index	-0.512*	-0.497*			-0.535*	
	(0.295)	(0.295)			(0.295)	
Labor intensity	0.000178*	0.000187*				0.000222**
	(0.000)	(0.000)				(0.000)
Plot heterogeneity	0.16	0.161			0.18	
	(0.323)	(0.323)			(0.322)	
Crop diversification	-0.231*	-0.223*			-0.227*	
	(0.125)	(0.132)			(0.129)	
Constant	-0.305	-0.325	-0.852***	-0.852***	-0.28	-0.891***
	(0.356)	(0.352)	(0.101)	(0.101)	(0.354)	(0.101)
No. of observations	1,503	1,503	1,503	1,503	1,503	1,503
Log likelihood	-1,877.202	-1,877.750	-1,891.831	-1,892.058	-1,878.574	-1,889.501
Degree of freedom	N.A.	1	4	5	1	4
2*(LR1-LR2)	N.A.	1.096	29.258	29.712	2.745	24.598
Critical value (10%)	N.A.	2.71	7.78	9.24	2.71	7.78
Critical value (5%)	N.A.	3.84	9.49	11.07	3.84	9.49

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Estimates of the revenue frontier and the mean inefficiency function are omitted from here for presentation clarity.
All the statistics for the Likelihood Ratio tests are calculated from the pairwise comparisons between the corresponding models with model HUV.

enough to that of HUV to reject the specification of HUV. Further, both the significance test of coefficient estimates and the likelihood dominance criterion (Pollak and Wales, 1991), an approach to non-nested model selection, suggest that HV is preferable to HUV_1.

To conclude this section, as far as heteroscedasticity is concerned, HV is the statistically preferred model where heteroscedasticity appears only in the two-sided error term with four explanatory variables: Simpson Index, labor input, plot heterogeneity, and crop diversification. Discussions in the next section will be based on the HV model unless otherwise noted.

Hypothesis Tests

Following from the previous section, we can see that the Simpson Index is negatively correlated with the two-sided error variance as predicted by the conceptual framework, and its coefficient estimate is significant at the 10% level. A similar result also holds for crop diversification, measured by the number of different crop types on the whole farm. In contrast, plot heterogeneity is found to have a positive impact on the variance although the estimated coefficient is statistically insignificant. This finding is not completely surprising given the close connection between plot heterogeneity and crop diversification. With a better characterization of plot heterogeneity and its relationship with crop diversification, we may be able to derive its "net effect" on the variance in future work. Lastly, yield variance increases with the labor input, a result in accordance with the risk-increasing role of labor found by many studies.

Regarding the determinants of efficiency (Column 2, Table A-8), we find that average education of family workers and proportion of male labor have the expected positive effects on efficiency, and the ratio of farm land devoted to perennial crops and fruit trees and average age of workers have the expected negative effects, and all these effects are statistically significant at the 5% level. Meanwhile, the ratio of kids under the age of five to the number of family workers does not seem to affect efficiency. Leaving out this variable will not impact the overall performance of the model as shown by the comparison between Column 3 and 2 in Table A-8. This may be because over three fourths of families in our sample have only one young child or no child at all such that they place no big burden on family workers.

What turns out to be puzzling is the effect of hired labor, and the results suggest that the higher the ratio of hired labor to the overall labor is, the more efficient the production will be. This contradicts the common belief that hired labor is less efficient than family labor. A potential explanation is that we do not control for hired labor's age and education in the empirical model owing to a lack of information.

Among the variables that are associated with soil conditions, the two for nutrient availability report positive coefficient estimates while neither of the estimates is statistically significant. An exclusion test (Column 4, Table 11) on the two variables shows that leaving them out from the model will not significantly change the estimates of other variables or the overall model fit. As for oxygen availability, the dummy representing the category of "Moderate constraint" is found to be negative at the 10% significance level, whereas the one for "No or slight constraint" is not significant, implying that soil of this type has the same effects on production efficiency as that of "Severe constraint". This unusual estimate may be caused by the lack of variation in our sample, as 90% observations report no or slight constraint (Table A-5). At last, both the two dummies for "Workability" report significantly negative coefficient estimates, and the difference in magnitude between the two estimates suggests that the less constraining the workability is, the more efficient the production would be, a conclusion that is consistent with our expectation.

Our primary interest falls on the variables related to land fragmentation. The Simpson Index, the most popular measure in the literature, is found to have a significantly negative impact on inefficiency; in other words, the more fragmented the farm is, the more efficient the production would be. This relationship seems counterintuitive and contradicts with the results in many other studies, although it is robust to various model specifications in this research. As for other dimensions of land fragmentation, neither of them reports a statistically significant coefficient estimate on its own; however, they are jointly significant as can be seen from the results in Column 5 of Table A-8. This finding echoes previous call for a complete characterization of land fragmentation to measure its economic effects.

Finally for the production frontier, the coefficient estimates of various inputs are less relevant to our topic and are thus waived from discussion. The only thing worth noting here is that the use of ox or machinery in the production shows a significantly positive effect on revenue as expected.

Efficiency Estimates and Marginal Effects

Given the results from the significance tests, we estimate a parsimonious and also statistically preferable model of HV, HV_P, to derive estimates for mean inefficiency term or its opposite, the mean efficiency, for each farm. Since our production frontier is defined for the logarithms of revenue and inputs, those mean efficiency estimates are subject to a proper transformation before comprehensible economic interpretations could be reached. The estimator proposed by Jondrow et al. (1982) is used to facilitate the calculation of marginal effects in the next step, although the results turn out to be very close to those using the alternative Battese and Coelli (1988)'s estimator (Table A-9 and Figure A-1). It can be seen that the average revenue efficiency across the 1,503 farms is 0.42, implying that these farms realize, on average, 42 percent of the revenue of a fully efficient farm, i.e. one that has zero inefficiency. Table A-9 also shows the wide gap between the most efficient farms and those least efficient ones.

Using the convenient estimates of efficiency from the last step, we are able to derive the farm-specific marginal effects as presented in Table A-10. For example, if the average education of labor is increased by one year, it can add 0.75 percentage points on average to the existing efficiency; if farmers can update the workability of his land from "Severe Constraints", the reference category for the regression, to "No or Slight Constraints", they can expect the efficiency to grow by 10.13 percentage points. As for the Simpson Index, the estimated mean marginal effect suggests that if all the plots are consolidated into one, i.e. the Index goes from one to zero, the efficiency will be reduced by 12.20 percentage points.

Finally, we adopt more specifications of the empirical model to test the robustness of the results, such as using aggregated labor instead of three separate labor inputs or using alternative measure of crop diversification, and find no substantial changes to our major findings.

Discussion and Conclusions

To investigate the role of land fragmentation in agricultural production, this study applies a stochastic frontier model with heteroscedasticity to the Tanzania LSMS data and finds robust evidence to support the hypothesis that land fragmentation may reduce production risk as measured by revenue variability. This finding is consistent with the few empirical studies that have addressed the risk-reduction effect of land fragmentation, such as McCloskey(1976), Blarel et al (1992), and Goland (1993). Moreover, we emphasize the necessity of including plot heterogeneity in characterizing land fragmentation and more importantly, quantitatively measuring its effects on revenue by showing how revenue variability is jointly determined by the two factors and the closely associated crop diversification. This may help explain the curious observations made by Heston and Kumar (1983) and Niroula and Thapa (2005) that land consolidation programs have succeeded mostly in places with uniform soils but failed in places with heterogeneous soils.

Meanwhile, our analysis suggests that land fragmentation is efficiency enhancing by increasing the revenue on unit land, leaving it instrumental to farmers in terms of both efficiency and risk management (we dub this result “double bonuses”), a finding that contradicts those of many studies in the literature but not all. For example, a few studies have found either a statistically insignificant (e.g. Blarel et al 1992; and Di Falco et al 2010) or economically insignificant (e.g. Wan and Cheng 2001) effect of land fragmentation. On the other hand, our result is not without companions in the literature. Deininger et al. (2012) apply the stochastic frontier model to the LSMS survey data of Albania and find land fragmentation measured by number of plots has a statistically significant positive effect on efficiency although the authors suggest that this positive economic impact is small (Page 13)[7]. An even more interesting observation has been made by Niroula and Thapa (2007), who report that in Nepal parcels with smaller size resulted from land fragmentation see more labor inputs and a higher yield. They further argued that “land fragmentation has a rather positive impact on production… However … the higher crop yield from small parcels is attributed to the application of considerably higher amount of labor, fertilizers and compost.” Yet they did not give any clue on whether or how input intensity is connected with land fragmentation.

To provide one possible explanation to Niroula and Thapa’s unanswered question and the puzzling positive relationship between land fragmentation and production efficiency found in this paper and Deininger et al (2012)’s, we argue that an important component has been absent from this study and the entire literature – risk preference, which could play a pivotal role in interpreting land fragmentation and its effects. As observed by most studies, farmers generally show aversion toward risk in agricultural production, a preference which can preclude them from using as many inputs as they would under risk neutrality and thus leads to a reduced yield or revenue. It can be anticipated that a shift in production risk, such as the one caused by land fragmentation as corroborated by this study, would result in changes in input use decisions, which will ultimately affect economic performance. An improved analytical framework that accommodates production, risk and risk preference should improve our understanding of land fragmentation's role in agricultural production.

In spite of the pending impact of land fragmentation on efficiency, this study still generates sufficient implications for future land reforms. First and foremost, land fragmentation as a tool for farmers to manage risk should be recognized. By utilizing the heterogeneous growing conditions, land fragmentation can spread out risk onto separate plots and reduce the revenue variability on the whole farm. This aspect is of special significance to farmers with no or limited access to crop insurance to secure their agricultural income. Second, the vast differences in farm structure, agricultural productivity and farming traditions warn against any hasty generalization on fragmentation and once-and-for-all consolidation propositions. In a smallholding and traditional agriculture like the Tanzanian case, the small plot size and rare use of machinery can minimize the potential negative effects of land fragmentation, while it may become a more serious issue for places with a more mechanized agriculture such as Japan[8].

References

Aigner, D., C.A. K. Lovell, and P. Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6: 21-37.

Antle, J.M. and C.C. Crissman. 1990. "Risk, Efficiency, and the Adoption of Modern Crop Varieties: Evidence from the Philippines." *Economic Development and Cultural Change* 38(3): 517-537.

Battese, G., and T. Coelli. 1988. "Prediction of Firm Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data." *Journal of Econometrics* 38: 387-399.

Bellon, M.R., and J.E. Taylor. 1993. "Folk Soil Taxonomy and the Partial Adoption of New Seed Varieties." *Economic Development and Cultural Change* 41(4): 763-786.

Belotti, F., S. Daidone, G. Ilardi, and V. Atella. 2012. "Stochastic Frontier Analysis Using Stata." *Centre for Economic and International Studies (CEIS) Research paper series* Volume 10, Issue 12.

Bentley, J.W. 1987. "Economic and Ecological Approaches to Land Fragmentation: In Defense of a Much-Maligned Phenomenon." *Annual Reviews of Anthropology* 16: 31-67.

Blarel, B., P. Hazell, F. Place, and J. Quiggin. 1992. "The Economics of Farm Fragmentation: Evidence from Ghana and Rwanda." *World Bank Economics Review* 6: 233-254.

Binns, B.O. 1950. "The Consolidation of Fragmented Agricultural Holdings." *FAO Agricultural Studies 11*. Food and Agriculture Organization, Rome.

Caudill, S., J. Ford, and D. Gropper. 1995. "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroskedasticity." *Journal of Business Economics and Statistics* 13: 105-111.

Deininger, K., S. Savastano, and C. Carletto. 2012. "Land Fragmentation, Cropland Abandonment, and Land Market Operation in Albania." *Policy Research Working Paper* 6032, the World Bank.

Di Falco, S., I. Penov, A. Aleksiev, and T.M. van Rensburg. 2010. "Agrobiodiversity, Farm Profits and Land Fragmentation: Evidence from Bulgaria." *Land Use Policy* 27: 763-771.

Fan, S., and C. Chan-Kang. 2005. "Is Small Beautiful? Farm Size, Productivity, and Poverty in Asian Agriculture." *Agricultural Economics* 32: 135-146.

Farmer, B.H. 1960. "On Not Controlling Subdivision in Paddy Lands." *Transactions of the British Geographers* 28: 225-35.

Fuwa, N. 2000. "A Note on the Analysis of Female Headed Households in Developing Countries." *Technical Bulletin of the Faculty of Horticulture of Chiba University*, pp.125-138.

Goland, C. 1993. "Field Scattering as Agricultural Risk Management: A Case Study from Cuyo Cuyo, Department of Puno, Peru." *Mountain Research and Development* 13: 317-338.

Greene, W. 2008. "The Econometric Approach to Efficiency Analysis in the Measurement of Efficiency." In *The Measurement of Productive Efficiency and Productivity Growth*, H.O. Fried, C.A. Knox Lovell and S.S. Schmidt (editors), Oxford University Press, Oxford, Chapter 2, 92-250.

Hadri, K. 1999. "Estimation of a Doubly Heteroskedasticity Stochastic Frontier Cost Function." *Journal of Business Economics and Statistics* 17(3): 359-363.

Hadri, K., C. Guermat, and J. Whittaker. 2003. "Estimating Farm Efficiency in the Presence of Double Heteroskedasticity Using Panel Data." *Journal of Applied Economics* VI(2): 255-268.

Hays, J. "Rice Farming in Japan: History, Paddies, Planting, Harvesting and Mechanization." Personal Website at http://factsanddetails.com/japan.php?itemid=939&catid=24&subcatid=159. Accessed on September 3rd, 2013.

Heston, A., and D. Kumar. 1983. "The Persistence of Land Fragmentation in Peasant Agriculture: An Analysis of South Asian Cases." *Explorations in Economic His*tory 20: 199-220.

Huang, C. J., and J.T. Liu. 1994. "Estimation of a Non-Neutral Stochastic Frontier Production Function." *The Journal of Productivity Analysis* 5: 171-180.

Hung, P.V. 2006. "Fragmentation and Economies of Size in Multi-Plot Farms in Vietname." Unpublished doctoral thesis, University of Sydney.

Hung, P.V., T.G. MacAulay, and S.P. Marsh. 2007. "The Economics of Land Fragmentation in the North of Vietnam." *Australian Journal of Agricultural and Resource Economics* 51: 195-211.

Jabarin, A.S., and F.M. Epplin. 1994. "Impact of Land Fragmentation on the Cost of Production Wheat in the Rain-fed Region of Northern Jordan. *Agricultural Economics* 11: 191-196.

Jondrow, J., C. Lovell, I. Materov, and P. Schmidt. 1982. "On the Estimation of Technical Efficiency in the Stochastic Production Function Model." *Journal of Econometrics* 19: 233-238.

Kawasaki, K. 2010. "The Costs and Benefits of Land Fragmentation of Rice Farms in Japan." *Australian Journal of Agricultural and Resource Economics* 54: 509-526.

King, R.L., and S.P. Burton. 1982. "Land Fragmentation, a Fundamental Rural Spatial Problem." *Progress in Human Geography* 6: 475-494.

Kumbhakar, S.C., S. Ghosh, and J.T. McGuckin. 1991. "A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms." *Journal of Business & Economic Statistics* 9: 279-286.

Kumbhakar, S.C., and K.C.A. Lovell. 2003. *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge.

McCloskey, D.N. 1976. "English Open Fields as Behavior towards Risk." In Paul J. Uselding, ed. *Research in Economic History: An Annual Compilation*. Greenwich, Connecticut: Jai Press.

Mearns, R. 1999. "Access to Land in Rural India: Policy Issues and Options." World Bank, Washington D.C.

Nguyen, T., E.J. Cheng, and C. Findley. 1996. "Land Fragmentation and Farm Productivities in China in the 1990s." *China Economic Review* 7: 169-180.

Niroula, G.S., and G.B. Thapa. 2005. "Impacts and Causes of Land Fragmentation, and Lessons Learned from Land Consolidation in South Asia." *Land Use Policy* 22: 358-372.

Niroula, G.S., and G.B. Thapa. 2007. "Impacts of Land Fragmentation on Input Use, Crop Yield and Production Efficiency in the Mountains of Nepal." *Land Degradation & Development* 18: 237-248.

Pollak, R.A., and T.J. Wales. 1991. "The Likelihood Dominance Criterion: A New Approach to Model Selection." *Journal of Econometrics* 47: 227-242.

Schmidt, P. 1986. "Frontier Production Functions." *Econometric Reviews* 4: 289-328.

Soini, E. 2005. "Land Use Change Patterns and Livelihood dynamics on the Slopes of Mt. Kilimanjaro, Tanzania." *Agricultural Systems* 85: 306-323.

Tan, S., N. Heerink, G. Kruseman, and F. Qu. 2008. "Do Fragmented Landholdings Have Higher Production Costs? Evidence from Rice Farmers in Northeastern Jiangxi Province, P.R. China." *China Economic Review* 19: 347-358.

Villano, R. and E. Fleming. 2006. "Technical Inefficiency and Production Risk in Rice Farming: Evidence from Central Luzon Philippines." *Asian Economic Journal* 20(1): 29-46.

Wan, G.H., and E. Cheng. 2001. "Effects of Land Fragmentation and Returns to Scale in the Chinese Farming Sector." *Applied Economics* 33: 183-194.

Wang, H., and P. Schmidt. 2002. "One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels." *Journal of Productivity Analysis* 18(2): 289-296.

Appendix

Table A-1 Households by Number of Plots

No. of Plots per Household	Frequency	Percent	Cumulative Percent
1	687	45.71	45.71
2	514	34.20	79.91
3	215	14.30	94.21
4	55	3.66	97.87
5	25	1.66	99.53
6	4	0.27	99.80
8	1	0.07	99.87
9	1	0.07	99.93
10	1	0.07	100.00
Total	1,503	100.00	

Table A-2 Use of Advanced Inputs

Panel 1: Other inputs (N=2,756)

Inputs	No. of Plots	Percent
Irrigation	83	3.01
Organic Fertilizer	332	12.05
Inorganic Fertilizer	416	15.09
Herbicide/Pesticide	308	11.18

Panel 2: Draft animals and machinery (N=1,503)

Inputs	No. of Households	Percent
Hand Hoe	1,503	100.00
Ox Plough	128	8.52
Ox Seeder	143	9.51
Ox Cart	1	0.07
Tractor	42	2.79
Mechanical Plough	3	0.20
Mechanical Harrow	6	0.40
Thresher	1	0.07

Table A-3 Descriptive Statistics of Dimensions of Land Fragmentation

	No. of Obs.	Mean	Median	S.D.
Farm Area	1,503	4.96	2.5	11.88
Number of Plots	2,756	1.83	2	1.01
Plot Area	2,756	2.70	1	12.78
Simpson Index	1,503	0.52	0.63	0.33
Distance, plot to home	2,755	3.12	1.5	6.44
Distance, plot to road	2,755	1.91	1	3.02
Distance, plot to market	2,773	7.78	5	9.03

Notes:

- Area in acres and distance in kilometers.
- One acre≈0.405 hectares or 0.0015625 square miles; one kilometer≈0.621 miles.

Table A-4 Descriptive Statistics of Household Characteristics (N=1,503)

	Central Tendency			Range			
	Mean	Median	S.D.	Minimum	Maximum	5th Percentile	95th Percentile
Area ratio	0.050	0	0.245	0	5	0	0.375
Average age	36.499	32.667	13.577	0	97	21.5	67
Average education	4.740	5	2.665	0	12	0	8.333
Male labor proportion	0.470	0.50	0.255	0	1	0	1
Kids ratio	0.368	0.25	0.456	0	3	0	1
Hired labor proportion	0.092	0	0.174	0	1	0	0.5

Notes:

- Average age and average education are measured in year; the other four variables are measured on a scale of zero to one.
- Average age and average education are for family workers only. If certain family use only hired labor, the average age and average education are reported with a value of zero.

Table A-5 Soil Variables

	Nutrient Availability		Oxygen Availability to Roots		Workability	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
No or Slight Constraint	498	33.13	1,344	89.42	850	56.55
Moderate Constraint	838	55.76	124	8.25	421	28.01
Severe Constraint	167	11.11	35	2.33	232	15.44
Total	1,503	100.00	1,503	100.00	1,503	100.00

Notes: The following definitions of variables are adapted from the Harmonized World Soil Database accessible at: http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/SoilQuality.html?sb=10

- Nutrient availability is decisive for successful low level input farming and to some extent also for intermediate input levels.
- Oxygen availability in soils is largely defined by drainage characteristics of soils.
- Workability or ease of tillage depends on interrelated soil characteristics such as texture, structure, organic matter content, soil consistence/bulk density, the occurrence of gravel or stones in the profile or at the soil surface, and the presence of continuous hard rock at shallow depth as well as rock outcrops. For the variable of workability, we combine “Severe Constraint”, “Very Severe Constraint” and “Mainly Non-Soil” into one category called “Severe Constraint”.

Table A-6 Plot heterogeneity

No. of plots	No. of different soil profiles					Total
	1	2	3	4	5	
1	687	0	0	0	0	687
2	166	348	0	0	0	514
3	51	98	66	0	0	215
4	11	22	17	5	0	55
5	3	4	10	7	1	25
6	0	2	2	0	0	4
8	0	1	0	0	0	1
9	0	0	1	0	0	1
10	0	0	0	1	0	1
Total	918	475	96	13	1	1,503

Table A-7 Crop Diversification

No. of plots	No. of Crop Varieties							Total
	1	2	3	4	5	6	7	
1	359	215	76	25	9	3	0	687
2	90	234	120	52	13	5	0	514
3	21	65	79	36	8	3	3	215
4	2	20	15	12	5	1	0	55
5	1	7	9	4	4	0	0	25
6	0	2	1	1	0	0	0	4
8	0	0	0	1	0	0	0	1
9	0	1	0	0	0	0	0	1
10	0	1	0	0	0	0	0	1
Total	473	545	300	131	39	12	3	1,503

Table A-8 Hypothesis Tests

Part 1: Revenue Frontier Function

Variables	HV	HV_1	HV_2	HV_P
Labor1*Labor1	-0.00951	-0.00985	-0.00958	-0.00515
	(0.025)	(0.025)	(0.025)	(0.026)
Labor1*Labor2	-0.0686	-0.0662	-0.0672	-0.0818*
	(0.045)	(0.045)	(0.045)	(0.045)
Labor1*Labor3	0.0731**	0.0702*	0.0702*	0.0713*
	(0.037)	(0.037)	(0.037)	(0.037)
Labor1*Area	-0.085	-0.0845	-0.0833	-0.0741
	(0.065)	(0.065)	(0.065)	(0.065)
Labor1*Price	-0.0275	-0.0284	-0.0298	-0.0233
	(0.046)	(0.046)	(0.046)	(0.046)
Labor1*Precipitation	0.0114	0.0122	0.0118	0.0036
	(0.105)	(0.105)	(0.105)	(0.105)
Labor1*Temperature	0.0454	0.0453	0.047	0.0518
	(0.130)	(0.129)	(0.129)	(0.130)
Labor1*Hoes	-0.104	-0.102	-0.102	-0.106
	(0.068)	(0.068)	(0.068)	(0.068)
Labor2*Labor2	0.00537	0.00303	0.00292	0.00808
	(0.033)	(0.033)	(0.033)	(0.034)
Labor2*Labor3	0.0441	0.0454	0.0461	0.0464
	(0.036)	(0.036)	(0.036)	(0.036)
Labor2*Area	-0.033	-0.0378	-0.0348	-0.0234
	(0.071)	(0.071)	(0.071)	(0.071)
Labor2*Price	-0.0696	-0.0681	-0.0664	-0.0692
	(0.049)	(0.049)	(0.049)	(0.049)
Labor2*Precipitation	0.00577	0.00373	0.00607	0.0148
	(0.106)	(0.106)	(0.106)	(0.107)
Labor2*Temperature	0.076	0.0776	0.0731	0.0658
	(0.131)	(0.131)	(0.131)	(0.131)
Labor2*Hoes	0.0241	0.0218	0.0252	0.0352
	(0.074)	(0.074)	(0.074)	(0.074)
Labor3*Labor3	-0.0893***	-0.0875***	-0.0879***	-0.0900***
	(0.020)	(0.020)	(0.020)	(0.020)
Labor3*Area	0.0205	0.0219	0.0216	0.0218
	(0.051)	(0.051)	(0.051)	(0.051)
Labor3*Price	-0.112***	-0.112***	-0.113***	-0.113***
	(0.035)	(0.035)	(0.035)	(0.035)
Labor3*Precipitation	-0.0138	-0.0132	-0.0128	-0.0151
	(0.082)	(0.082)	(0.082)	(0.083)

Labor3*Temperature	0.187*	0.186*	0.186*	0.190*
	(0.101)	(0.102)	(0.102)	(0.102)
Labor3*Hoes	0.0679	0.0658	0.0653	0.0725
	(0.056)	(0.056)	(0.056)	(0.056)
Area*Area	0.113*	0.110*	0.109*	0.0653
	(0.063)	(0.063)	(0.063)	(0.060)
Area*Price	-0.0863	-0.0855	-0.0848	-0.0821
	(0.060)	(0.060)	(0.060)	(0.060)
Area*Precipitation	0.317**	0.321**	0.314**	0.314**
	(0.130)	(0.129)	(0.129)	(0.129)
Area*Temperature	-0.285*	-0.289*	-0.284*	-0.283*
	(0.167)	(0.167)	(0.167)	(0.167)
Area*Hoes	0.281**	0.280**	0.275**	0.227*
	(0.121)	(0.121)	(0.121)	(0.120)
Price*Price	0.0326**	0.0331**	0.0331**	0.0347**
	(0.015)	(0.015)	(0.015)	(0.015)
Price*Precipitation	-0.284***	-0.285***	-0.286***	-0.288***
	(0.106)	(0.106)	(0.106)	(0.107)
Price*Temperature	0.479***	0.479***	0.480***	0.477***
	(0.123)	(0.123)	(0.123)	(0.125)
Price*Hoes	0.0029	0.00375	0.00384	0.0139
	(0.070)	(0.070)	(0.070)	(0.069)
Precipitation*Precipitation	0.118	0.107	0.122	0.107
	(0.168)	(0.167)	(0.163)	(0.163)
Precipitation*Temperature	-0.0499	-0.0233	-0.0576	-0.019
	(0.402)	(0.400)	(0.390)	(0.391)
Precipitation*Hoes	0.306**	0.309**	0.305**	0.302**
	(0.141)	(0.141)	(0.141)	(0.141)
Temperature*Temperature	-0.322	-0.337	-0.318	-0.343
	(0.267)	(0.265)	(0.261)	(0.262)
Temperature*Hoes	-0.363**	-0.367**	-0.364**	-0.368**
	(0.180)	(0.179)	(0.179)	(0.180)
Hoes*Hoes	0.142*	0.143**	0.140*	0.117
	(0.073)	(0.073)	(0.073)	(0.072)
Dummy	0.339***	0.340***	0.345***	0.351***
	(0.077)	(0.076)	(0.076)	(0.076)
Constant	12.02***	11.94***	11.97***	12.10***
	(0.741)	(0.721)	(0.715)	(0.712)

Part 2: Mean Inefficiency Function

Variables	HV	HV_1	HV_P	HV_2
Area Ratio	0.359***	0.379***	0.379***	0.387**
	(0.135)	(0.141)	(0.144)	(0.150)
Average Age	0.00528**	0.00616**	0.00607**	0.00626**
	(0.003)	(0.003)	(0.003)	(0.003)
Average Education	-0.0265**	-0.0278*	-0.0283*	-0.0323**
	(0.013)	(0.014)	(0.015)	(0.016)
Male Labor Ratio	-0.401***	-0.420***	-0.425***	-0.395**
	(0.142)	(0.153)	(0.156)	(0.160)
Kids Ratio	-0.0821			
	(0.071)			
Hired Labor Ratio	-1.500***	-1.625***	-1.673***	-1.779***
	(0.527)	(0.565)	(0.586)	(0.688)
Nutrient Availability	0.00338	-0.00461		
-- No Constraint	(0.114)	(0.122)		
Nutrient Availability	0.0679	0.0643		
-- Moderate Constraint	(0.101)	(0.108)		
O2 Availability to Roots	-0.316	-0.342	-0.357	-0.370
--No Constraint	(0.205)	(0.217)	(0.220)	(0.238)
O2 Availability to Roots	-0.415*	-0.445*	-0.475*	-0.459*
--Moderate Constraint	(0.235)	(0.252)	(0.256)	(0.269)
Field Workability	-0.367***	-0.395***	-0.380***	-0.377***
--No Constraint	(0.126)	(0.135)	(0.131)	(0.141)
Field Workability	-0.252**	-0.277**	-0.261**	-0.252**
--Moderate Constraint	(0.116)	(0.126)	(0.124)	(0.128)
Farm Area	0.0047	0.00522	0.006	
	(0.007)	(0.007)	(0.007)	
Farm Area* SI	0.00409	0.00382	0.003	
	(0.009)	(0.009)	(0.009)	
SI	-0.425**	-0.454**	-0.457**	-0.350**
	(0.166)	(0.182)	(0.183)	(0.174)
Distance to Home	-0.0189	-0.0218	-0.023	
	(0.014)	(0.015)	(0.016)	
Distance to Road	0.00507	0.00699	0.008	
	(0.017)	(0.018)	(0.018)	
Distance to Market	-0.00901	-0.0103	-0.011	
	(0.006)	(0.007)	(0.007)	
Constant	2.026***	1.953***	1.978***	1.847***
	(0.295)	(0.288)	(0.272)	(0.266)

Part 3: One-sided Error Variance Function

Variables	HV	HV_1	HV_P	HV_2
Constant	-0.904***	-0.848***	-0.844***	-0.825**
	(0.302)	(0.317)	(0.324)	(0.372)

Part 4: Two-sided Error Variance Function

Variables	HV	HV_1	HV_P	HV_2
SI	-0.497*	-0.481*	-0.476*	-0.468
	(0.295)	(0.292)	(0.289)	(0.292)
Labor Intensity	0.000187*	0.000190**	0.000189**	0.000244***
	(0.000)	(0.000)	(0.000)	(0.000)
Plot Heterogeneity	0.161	0.138	0.126	0.148
	(0.323)	(0.309)	(0.304)	(0.305)
Crop Diversification	-0.223*	-0.208*	-0.206*	-0.203*
	(0.132)	(0.117)	(0.114)	(0.110)
Constant	-0.325	-0.324	-0.306	-0.331
	(0.352)	(0.344)	(0.339)	(0.342)

Part 5: Statistics and Tests

Variables	HV	HV_1	HV_P	HV_2
Observations	1,503	1,503	1,503	1,503
Log Likelihood	-1,877.750	-1,878.399	-1,878.886	-1,885.590
Degree of freedom	0	1	2	5
2*(LR1-LR2)	0	1.298	0.9734	13.4088
Critical value (10%)	2.71	2.71	4.61	6.25
Critical value (5%)	3.84	3.84	5.99	7.81

Notes:

- Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
- Labor1: Labor days used for land preparation and planting per acre, in the log form;
  Labor2: Labor days used for weeding per acre, in the log form;
  Labor3: Labor days used for harvest per acre, in the log form;
  Area: Total area planted with annual crops, in the log form;
  Price: Crop price index weighted by quantity (in kilograms), in the log form;
  Precipitation: Precipitation of the wettest quarter, in the log form;
  Temperature: Average temperature of the wettest quarter, in the log form;
  Hoes: Average number of hand hoes used per acre, in the log form;
  Dummy=1 if any ox or machinery ever used, =0 otherwise;
  Area Ratio: Ratio of farm area planted with perennial crops/trees to area planted with annual crops;

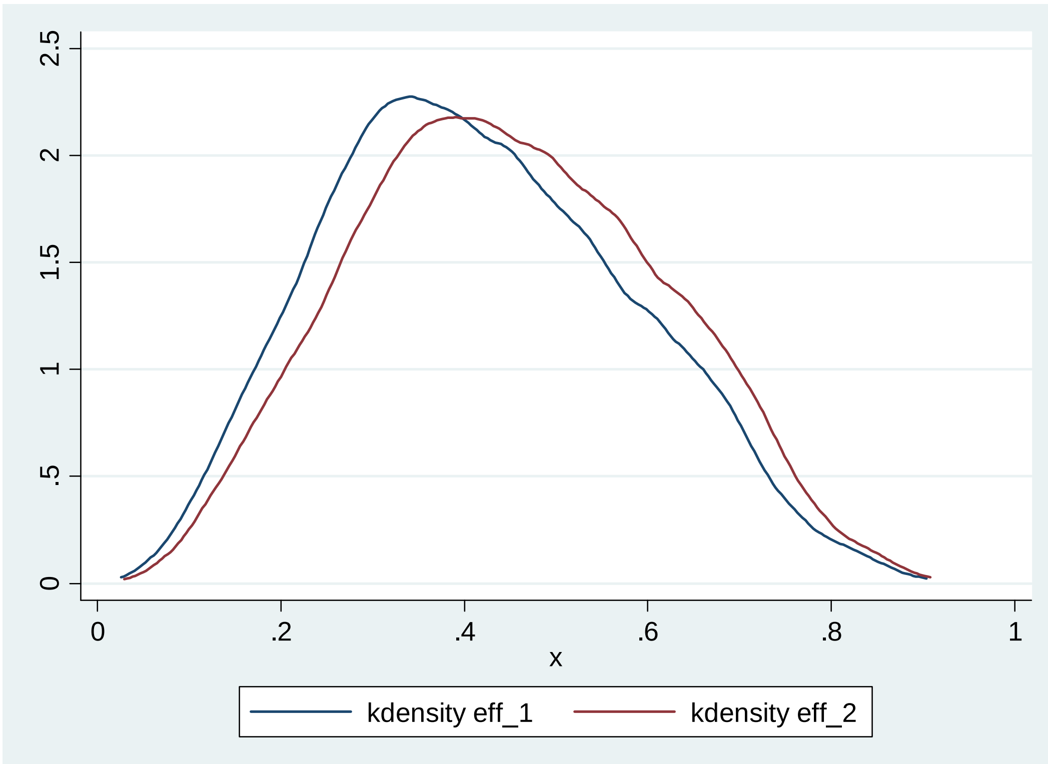
Age: Average age of family workers in the fields;
Education: Average number of years in school of family workers in the fields;
Male Labor Ratio: Ratio of labor days by male workers to labor days by both genders;
Hoes Ratio: Ratio of number of hoes to number of family workers in the fields;
Kids Ratio: Ratio of number of kids under the age of 5 to the number of family workers in the fields;
Hired Labor Ratio: Ratio of labor days by hired workers to days by both family and hired workers;
SI: the Simpson Index for land fragmentation;
Distance to Home/Road/Market: weighted by plot area;
Labor Intensity: Total labor days for all three activities per acre, i.e. Labor1+Labor2+Labor3;
Plot heterogeneity: Number of different soil profiles across the farm, normalized by number of plots;
Crop Diversification: Number of different annual crop varieties grown on the entire farm.

- All the statistics for the Likelihood Ratio tests are calculated from the pairwise comparison of the corresponding model with the preceding model.

Table A-9 Descriptive Statistics of Efficiency Estimates

	No. of Obs.	Mean	S.D.	Minimum	1st Quartile	Median	3rd Quartile	Maximum
Jondrow et al. (1982) estimator	1,503	0.42	0.16	0.03	0.29	0.40	0.53	0.90
Battese & Coelli (1988) estimator	1,503	0.45	0.16	0.03	0.32	0.44	0.57	0.91

Figure A-1 Distributions of Inefficiency Estimates



Kernel density *eff_1* is derived using the Jondrow et al. (1982) estimator, and kernel density *eff_2* is derived using the Battese & Coelli (1988) estimator.

Table A-10 Marginal Effects on Efficiency (N=1,503)

	Direction of Effect	Minimum	Maximum	Mean	Median
Area Ratio	Negative	0.0096	0.2223	0.1011	0.0997
Average Age	Negative	0.0002	0.0036	0.0016	0.0016
Average Education	Positive	0.0007	0.0166	0.0075	0.0074
Male Labor Proportion	Positive	0.0108	0.2494	0.1134	0.1118
Hired Labor Proportion	Positive	0.0426	0.9822	0.4464	0.4404
O2 Availability to Roots --No Constraint	Positive	0.0091	0.2097	0.0953	0.0940
O2 Availability to Roots --Moderate Constraint	Positive	0.0121	0.2788	0.1267	0.1250
Field Workability --No Constraint	Positive	0.0097	0.2229	0.1013	0.0999
Field Workability --Moderate Constraint	Positive	0.0066	0.1532	0.0696	0.0687
Farm Area	Negative	0.0001	0.0032	0.0015	0.0015
Farm Area * SI	Negative	0.0001	0.0020	0.0009	0.0009
SI	Positive	0.0116	0.2684	0.1220	0.1203
Distance to Home	Positive	0.0006	0.0133	0.0061	0.0060
Distance to Road	Negative	0.0002	0.0046	0.0021	0.0021
Distance to Market	Positive	0.0003	0.0064	0.0029	0.0029

See the notes of Table A-8 for variable definitions.

[1] Given the logarithmic transformation in our setting, households earning negative profits need to be dropped out from our sample in order to apply a profit frontier analysis.

[2] As a matter of fact, the average number for all year around is highly correlated with the average number for the wettest season. This is the case for both temperature and precipitation with the correlation coefficients equal to 0.98 and 0.92 respectively. Switching to the yearly statistics will not lead to any essential changes in our major findings as confirmed by our sensitivity test on this.

[3] We also tried generating an average crop price weighted by their contribution to total value and included it in the empirical model. All major findings remain the same except for the changes in the magnitude of coefficient estimates and therefore the inefficiency estimates and marginal effects.

[4] Also, it will be difficult to interpret the marginal effects if we include both the Simpson Index and number of plots.

[5] Many studies choose to use the age and education of household head as a proxy for experience. However, as argued in Fuwa (2000) and others, there have been various definitions of household headship (e.g. demographics-based or economics-based) and the household head elicited in the common household-level surveys may not necessarily be the one that is most relevant to the economic analysis under many circumstances. Therefore, we believe the average age and education of family laborers who actually worked in the fields to be a better proxy variable of farming experience in this study.

[6] This estimation procedure has been operationalized in Stata 12 by Belotti et al. (2012).

[7] Their study also investigates land fragmentation's impact on farmers' cropland abandonment decisions. They found that about 10 percent of Albania's productive land has been left idle mostly because of land market imperfections. In contrast, there are only a few cases of land abandonment where land fragmentation leads to plots too small for economically viable cultivation. Among those currently cultivated plots, land fragmentation is found to have a statistically significant positive effect on efficiency. Although their study does not give an overall appraisal of land fragmentation when both cultivation-related and abandonment-related productivity are considered, they conclude that their analysis does not support the argument of land fragmentation undermining productivity.

[8] According to Kawasaki (2010) who finds that land fragmentation reduces the cost efficiency of Japanese rice growing, the average farm size in his sample is about 6.8 acres, roughly comparable to the 6.1 acres among the Tanzanian farmers in our sample when area used for perennial crops and trees is also counted. In contrast to the Tanzanian case, in Japan the planting and harvesting is done mostly with small machines. Large machines are hardly used because they cannot maneuver around in small plots and need long tracts of uniform land to do the job efficiently (Hays, 2009).