European Review of Agricultural Economics Vol **46** **(4)** (2019) pp. 609–635
doi:10.1093/erae/jby034
Advance Access Publication 17 October 2018

Land fragmentation with double dividends – the case of Tanzanian agriculture

Xudong Rao*

The Business Economics Group at Wageningen University, The Netherlands

Received April 2017; editorial decision August 2018; final version accepted September 2018

Review co-ordinated by Ada Wossink

Abstract

This study evaluates the economic effects of land fragmentation on agricultural production and hypothesises that fragmentation may benefit farmers by diversifying production risk among separate land plots with heterogeneous agronomic conditions. Applying a stochastic production frontier model to the Tanzania Living Standards Measurement Study data, we find robust evidence to support the risk-reduction hypothesis, as well as indications that fragmentation is positively associated with technical efficiency. We argue the low level of fragmentation in Tanzania may have limited its negative impact on efficiency, while crop diversification concurrent with fragmentation may have increased efficiency, leaving the net effect to be positive.

Keywords: agricultural productivity, land fragmentation, risk management, stochastic production frontier, Tanzania

JEL classification: Q12, Q15, Q18

1. Introduction

Land fragmentation – that is, a single farm consisting of numerous discrete plots scattered over a wide area (Binns, 1950) – has long been deemed an impediment to agricultural production and rural development. Policy makers describe it as 'the blackest of evils' (Farmer, 1960); researchers claim that it undermines efficiency and lowers profits (e.g. Jabarin and Epplin, 1994; Nguyen, Cheng and Findley, 1996; Wan and Cheng, 2001; Fan and Chan-Kang, 2005; Tan *et al.*, 2008). Until recently, however, land fragmentation 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 an average of four plots per farmer in 2005 (Deininger, Savastano and

*Corresponding author: E-mail: xudong.rao@wur.nl

Carletto, 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 provided various explanations for the prevalence and persistence of land fragmentation, including demographic, cultural and institutional reasons (e.g. Heston and Kumar, 1983; Bentley, 1987; Blarel *et al*., 1992; Niroula and Thapa, 2005). Meanwhile, economists have attempted to re-interpret land fragmentation's role in agricultural production from the perspective of risk management. McCloskey (1976) was among the first to argue that cultivation on scattered plots with different soil types and locations can reduce risk, even though it incurs additional travel costs and other inconveniences. Several empirical studies (e.g. Blarel *et al*., 1992; Goland, 1993; Di Falco *et al*., 2010) have corroborated the risk-reducing function of land fragmentation.

In practice, voluntary land exchanges among farmers have been extremely rare (Bentley, 1987). Many governments have been advised to launch consolidation programs in the hope that farmers will benefit from more concentrated land holdings. Some of those programs were successful at creating more consolidation across farms, while others have failed due to resistance from farmers (see Heston and Kumar, 1983 for the failure cases in India; Niroula and Thapa, 2005 for the failure cases in India, Pakistan and Thailand). Therefore, it remains largely inconclusive whether the existence of land fragmentation is economically justifiable.

Fluctuations in agricultural production and income have profound implications for the well-being of farmers in developing countries. Unlike their counterparts in developed countries, who can often access crop insurance and other resources (e.g. irrigation and pest control chemicals) to protect themselves from adversities, farmers in developing countries are faced with limited options for risk control and have to rely on their use of conventional inputs. Further, farmers' aversion to risk may discourage them from adopting new technologies and crop varieties despite the higher expected returns (Liu 2013). Understanding farmers' risk management strategies and the implications on productivity and income has been of keen interest to both researchers and policy makers.

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 technical efficiency and production risk. Applying a stochastic frontier model to analyse land fragmentation, we aim to derive an improved characterisation of this phenomenon through a careful discussion of determinants of technical efficiency and production risk. Our findings will shed light on future land tenure reforms that aim to secure agricultural production and improve farmers' well-being.

2. Land fragmentation and plot heterogeneity

There is no single metric of land fragmentation to capture the economic implications of its multiple aspects. King and Burton (1982) proposed a six-parameter characterisation: farm size, plot number, plot size, plot shape, plot spatial distribution and the size distribution of the fields, while Bentley (1987) argued that efforts to quantify the notion of land fragmentation will be flawed if they fail to account for measures of distance. Among economists, the predominant measure has been the Simpson Index (SI), which is often used along with other metrics of land fragmentation (e.g. Blarel *et al.*, 1992; Hung, MacAulay and Marsh, 2007; Tan *et al.*, 2008; Kawasaki, 2010). For a farm cultivating a total number of J plots, denoting the area for plot j ($j = 1, 2\ldots J$) by A_j, the SI is defined as:

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

where $A = \sum_j^J A_j$ is the total farm area. This index returns a value ranging from zero to one and increases as a farm becomes more fragmented. SI = 1 represents an infinite fragmentation scenario and SI = 0 represents one-plot farms. SI is jointly determined by the number of plots, the farm size, the plot size and the plot size distribution.

One confounding phenomenon in characterising land fragmentation is the concurrence of heterogeneous soil quality and growing conditions across plots, or plot heterogeneity for short. It is believed to be a cause of land fragmentation and a restricting condition for the implementation of land consolidation (Mearns, 1999; Niroula and Thapa, 2005). What is significant about plot heterogeneity is its risk-controlling role discussed in the literature. By cultivating plots with varying micro-environments, farmers are able to reduce output variations by spreading out the risk caused by drought, flood and diseases (Hung, MacAulay and Marsh, 2007). Bentley (1987) reviewed several studies from this perspective and concluded that the risk management advantage of fragmented farms is applicable to many contexts.

Another value of plot heterogeneity is that it may encourage crop diversification (Bellon and Taylor, 1993; Hung, 2006), a popular strategy for risk reduction. By matching the proper crop portfolio with the agro-ecological conditions across the entire farm, farmers tend to increase crop diversity and stabilise farm outputs. Di Falco *et al.* (2010) found empirical evidence that land fragmentation fosters crop diversification.

To summarise, previous studies have spent a great deal of attention on the impact of land fragmentation on efficiency and productivity. Meanwhile, the risk management hypothesis of land fragmentation has not received much empirical scrutiny, even though risk management plays an equally vital role in agriculture. The few existing studies examining the risk effect of land fragmentation have focused solely on the dispersion of fields without considering plot heterogeneity. Considering the 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 speculate that the risk-reducing benefit of land fragmentation may be jointly determined by both plot dispersion and plot heterogeneity.

3. Conceptual framework

This section will provide a formal framework to evaluate how land fragmentation affects both technical efficiency and production risk. The latter is often measured by variation in output. A common approach to efficiency analysis is the stochastic frontier model developed by Aigner, Lovell and Schmidt (1977) and Meeusen and Van den Broeck (1977). We write the yield y_i (in its level form) of farmer i ($i = 1, 2 \ldots N$) as:

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

In equation (2), $F(\boldsymbol{X_i};\ \boldsymbol{\beta})$ is the deterministic production frontier, 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, also known as the one-sided error term, is assumed to be greater than or equal to zero such that $\exp(-u_i)$ lies within the unit interval and represents the proportion of $F(\boldsymbol{X_i};\ \boldsymbol{\beta})$ that is actually produced. When $exp(-u_i) = 1$, the production is fully efficient and lies on the frontier; if not, 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 misspecifications. Having two separate error terms, the stochastic frontier model – also known as the compound error model – allows the estimation of a stochastic production frontier with individual-specific efficiency scores.

Empirical studies often assume the deterministic production function to take either the Cobb–Douglas form or the transcendental logarithmic (translog) form. This study will assume the translog functional form due to its flexibility.[1] This transformation allows us to see the three components of y more clearly:

$$lny_i = f(ln\boldsymbol{X}_i;\ \boldsymbol{\beta}) + v_i - u_i. \tag{3}$$

The primary interest of stochastic production frontier analysis usually falls on the inefficiency term u_i, which is often assumed to follow a certain statistical distribution. With a truncated normal distribution for u_i, Kumbhakar, Ghosh and McGuckin (1991) and Huang and Liu (1994) proposed a model to parameterise the mean of the pre-truncated inefficiency distribution, μ_i. In this way, inefficiency can accommodate a group of exogenous explanatory variables $\boldsymbol{Z_i}$, including a constant term through a linear function, that is:

1 Supplementary Material Section 2 (at *ERAE* online) reports the model estimates under the two alternative assumptions (i.e. Cobb–Douglas and translog) on the frontier functional form and conduct a likelihood-ratio (LR) test to choose between the two models. The statistics (LR = 65.65 and p-value = 0.0013) reject the null hypothesis at the 5 per cent significance level and suggest the translog form as the more appropriate option.

$$u_i \sim N^{+}(\mu_i,\ \sigma_u^2) \tag{4}$$

where

$$\mu_i = \mathbf{Z}_i \boldsymbol{\gamma} \tag{5}$$

The vector of parameters $\boldsymbol{\gamma}$, also called inefficiency effects, can be estimated. We will adopt the truncated normal assumption on u_i for the purpose of this study. Following the majority of the literature, we assume the two-sided error v_i to be normally distributed as $N(0,\ \sigma_v^2)$, and u_i and v_i to be independent of each other[2] and independently and identically distributed across observations.

In linear models with one error term, heteroskedasticity can easily be tackled with robust estimation procedures. In stochastic frontier models, however, heteroskedasticity is a more challenging problem and will lead to inconsistent estimates of the inefficiency effects and the frontier parameters (Caudill, Ford and Gropper, 1995; Hadri, 1999). Even worse, heteroskedasticity may appear in either or both the one-sided error term u_i and the two-sided error term v_i, and misspecification 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 previously, land fragmentation is likely related to production risk. In this study, we make the formal hypothesis that land fragmentation can diversify production risk among separate land plots such that it reduces the risk on the entire farm. Following a decomposition similar to that by Blarel *et al.* (1992: 250), we denote the actual yield (in its level form) on the jth plot of the ith farm as y_{ij} such that

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

In equation (6), $\bar{y}_i$ is the expected farm-level yield. The term d_{ij} captures plot-specific fixed effects (e.g. soil attributes) that cause y_{ij} to deviate from $\bar{y}_i$. For example, if one plot is more fertile than other plots on the same farm, the yield on this plot will be higher than the average farm yield. Compared to d_{ij}, θ_{ij} is also plot-specific but stochastic, and it is associated with precipitation, insolation, wind and other random factors that define the microclimatic environment on each plot. In general, the distribution of θ_{ij} varies from plot to plot and hence we assume $E(\theta_{ij}) = 0$ and $\mathrm{Var}(\theta_{ij}) = \sigma_{\theta ij}^2$ for any j. Finally, e_{ij} captures all stochastic effects (e.g. measurement errors) that are identically

2 The independence between the inefficiency term (u) and the error term (v) has been a ubiquitous assumption in the frontier literature mainly for the ease of econometrical estimation (Kumbhakar, Wang and Horncastle, 2015). Recently, researchers have started to challenge the orthogonality assumption based upon conceptual grounds and develop copula-function-based approaches to allow for correlated u and v. Given the peculiar specifications on inefficiency and heteroskedasticity in this study, we decided to maintain the independence assumption while admitting that this strict assumption may generate bias in our estimates for efficiency scores and marginal inefficiency effects.

distributed for any plot on any farm, and we assume $E(e_{ij}) = 0$, $\mathrm{Var}(e_{ij}) = \sigma_e^2$ and $\mathrm{Cov}(\theta_{ij}, e_{ij}) = 0$ for any i and j.

With such a decomposition, we take 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

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

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

$$\begin{aligned}\mathrm{Var}(y_i) &= \mathrm{Var}\left[\frac{1}{A_i}\sum_j^J [(\theta_{ij} + e_{ij}) * A_{ij}]\right] \\ &= \frac{1}{A_i^2}\mathrm{Var}\left[\sum_j^J (\theta_{ij} * A_{ij})\right] + \frac{1}{A_i^2}\sum_j^J \sigma_e^2 A_{ij}^2 \\ &\equiv \sigma_{\theta i}^2 + (1 - \mathrm{SI}) * \sigma_e^2\end{aligned} \tag{8}$$

The second term on the right-hand side of equation (8), $(1 - \mathrm{SI}) * \sigma_e^2$, shows that variance on the whole farm is reduced by spreading out the common stochastic effects σ_e^2 across the plots. The first term, $\sigma_{\theta i}^2$, represents the aggregation of stochastic effects that are specific to each plot, and its effect on yield variability is generally unknown unless the distribution (or at least the variance) of each θ_{ij} is given. We expect $\sigma_{\theta i}^2$ to relate to soil heterogeneity for reasons argued in Hung, MacAulay and Marsh (2007). Moreover, if farmers can match the growing conditions on all plots with the proper crop portfolio (Bellon and Taylor, 1993; Hung, 2006), we expect $\sigma_{\theta i}^2$ to be negatively associated with crop diversification, which is a common strategy for risk reduction.

In this way, yield variance is shown to vary among farms and to depend on several farm-specific factors, justifying our concerns of heteroskedasticity. Specifically, the variance of the common error term v_i has its own explanatory variables; that is

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

where $\boldsymbol{h_i}$ includes a constant, the SI, and variables for plot heterogeneity and crop diversification. Further, some inputs such as labour and fertiliser (Hadri, Guermat and Whittaker, 2003) may affect either or both variance terms. To avoid potential bias in the model 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 $\boldsymbol{\varphi}$:

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

If heteroskedasticity 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 homoskedastic case.

To summarise, the conceptual framework of this study incorporates both the efficiency and risk effects of land fragmentation into an integrated model and allows estimating them simultaneously. Previous studies (e.g. Blarel *et al*., 1992; Goland, 1993; Kawasaki, 2010) directly regressed output variance on certain fragmentation index regardless of the probable efficiency effects. Therefore, our model builds upon more solid conceptual and statistical grounds and can generate more reliable results.

4. Context and data

To empirically assess the impact of land fragmentation, this study examines Tanzanian agriculture, which accommodated 75 per cent of the national population and accounted for 45 per cent of the GDP in 2008. Although a vast area of cropland is available for intensive cultivation, small-sized farming has been the predominant form of production with scarce use of inputs and low levels of productivity overall. In 2008, 37 per cent of the rural population, i.e. more than one-quarter of the national population, lived below the poverty line. Efficient and stable food production carries great significance for Tanzania's millions of impoverished rural citizens as well as its national economy.

There is one interesting historical fact about Tanzanian agriculture. Right after its independence in 1961, Tanzania adopted a communist approach by promoting collective land cultivation and shared labour for its agricultural production. The government relocated about 75 per cent of the population from scattered homesteads and smallholdings to communal villages of 2,000–4,000 residents (Dondeyne *et al.,* 2003; Maoulidi, 2004). Despite governmental efforts, Tanzanian farmers showed strong preferences for individually allocated and cultivated farmland (USAID, 2011). The following administration in the 1980s quickly abandoned the communist approach and installed a new legal framework that supported private property rights and individualised control of farming. The law acknowledged personal rights to land and encouraged productive and sustainable use of land. Since then, farmers have the right to buy, sell, lease and mortgage their plots and to decide on matters such as crop choices and land use. Interestingly, most farmers have chosen to keep multiple plots on their farms rather than consolidate all their land holdings through sales or transfers. In 2008, each rural family owned or cultivated an average of 2.5 plots. The changes in Tanzania's land tenure system in the past several decades may have highlighted the role of scattered land holdings in agricultural production.

The empirical analysis uses the Tanzania Living Standards Measurement Study (LSMS) 2008–2009 data collected by the World Bank. This survey adopted a stratified, multi-stage cluster design to obtain a nationally representative sample. Enumerators interviewed rural family members regarding family socioeconomics and agricultural activities. Information on location, ownership, soil conditions, crop varieties, input use and harvest was collected for each cultivated plot. This study focuses on plots that were grown either partially or fully with annual crops in the long rainy season (March–May), because the production of annual crops differs significantly from that of perennial crops and trees. Our sample contains 1,503 farms with 2,756 plots. Maize is the predominant crop both in terms of occurrence and planting area; other popular annual crops include beans, groundnuts, paddy rice and sorghum.

5. Empirical model

5.1. Deterministic frontier function

The dependent variable of the frontier function is an implicit output quantity derived by dividing the gross returns (in Tanzanian shillings) of all annual crops on each farm by the corresponding composite output price. The composite output price is the sum of individual crop prices weighted by crop value shares. Since farms have different crop portfolios, the composite output price varies among farms and depends on the crops grown and their quantities. Most studies on land fragmentation have adopted a similar method to aggregate multiple outputs in spite of the potential aggregation bias.[3] In the Tanzanian case, almost 70 per cent of the farms produce more than one type of crop. Focusing on one crop (such as maize) will not only significantly reduce the sample size but also overlook the concurrence between land fragmentation and crop diversification and the associated implications (to be discussed later). Table 1 presents the definitions and descriptions of all the explanatory variables, which are also discussed below.

Explanatory variables of the frontier function include land (in acres) and three types of labour (in person-days), i.e. labour on land preparation/planting, weeding and harvesting. Other inputs such as fertiliser, irrigation and pesticide are rare in Tanzania and are thus excluded. Few farmers have access to draft animals or farm machinery. Instead, the most common farm implement is the hand hoe with all the households in the sample having at least one. Hence, this model includes the number of hand hoes as well as a dummy variable for the use of any draft animal or machinery. To account for weather effects, explanatory variables also include the average temperature and average precipitation during the wettest quarter in the long rainy season of 2008–2009.

5.2. Explanatory variables of inefficiency

The mean inefficiency function in equation (5) contains explanatory variables related to land fragmentation, household characteristics, other productive activities and growing conditions for crops. Specifically, we use the SI, farm

3 One recent exception is Kawasaki (2010), which focuses exclusively on rice farms in Japan.

Table 1. Variable definitions

Label	Definition and description
labour1	Total person-days of both family labour and hired labour spent on land preparation and planting of annual crops on each farm
labour2	Total person-days of both family labour and hired labour spent on weeding of annual crops on each farm
labour3	Total person-days of both family labour and hired labour spent on harvest of annual crops on each farm
labour	Sum of all three types of labour; i.e. labour = labour1 + labour2 + labour3
land	Area (in acres) of annual crops
precipitation	Precipitation (in millimetres) of wettest quarter, from monthly climatology
temperature	Average temperature (in Celsius) of the wettest quarter, from monthly climatology, multiplied by 10
hoe	Number of hand hoes for farming in each household
machinery dummy	=1 if using any farm machine (e.g. tractor, thresher) or draft animals; =0 otherwise
perennial	Ratio of land area for perennial crops/fruit trees to area for annual crops
age	Average age (in years) of family workers working on family farm
education	Average education (in years) of family workers working on family farm
male labour	Male labour as a proportion of the sum of male labour and female labour
children	The number of children younger than five divided by family labour
hired labour	Hired labour as a proportion of the sum of hired labour and family labour
nutrient1	=1 if nutrient availability reports 'No constraint'; =0 otherwise
nutrient2	=1 if nutrient availability reports 'Moderate constraint'; =0 otherwise
oxygen1	=1 if oxygen availability reports 'No constraint'; =0 otherwise
oxygen2	=1 if oxygen availability reports 'Moderate constraint'; =0 otherwise
workability1	=1 if workability reports 'No constraint'; =0 otherwise
workability2	=1 if workability reports 'Moderate constraint'; =0 otherwise
SI	The SI to measure land fragmentation, as calculated in equation (1)
distance1	Average distance (in kilometres) of all plots to farm owner's home, weighted by plot area
distance2	Average distance (in kilometres) of all plots to the most nearby road, weighted by plot area
distance3	Average distance (in kilometres) of all plots to the closest market, weighted by plot area
plot area	Area (in acres) of each plot
plot heterogeneity	The number of different land profiles (including soil type, erosion type and steepness of slope), divided by the number of plots on each farm
crop diversification	The number of crop varieties on each farm
crop_SI	The SI to measure crop diversification, similar to SI in equation (1) but replacing the area for each plot (A_j) with area for each crop variety

size, an interaction term between the two and three distance-related variables (i.e. the average distance from farm to home, to the closest road and to the closest market, weighted by plot area) to characterise land fragmentation. Household characteristics include the average age (in years) and average education (in years) of family workers. Two variables, *male labour* and *hired labour*, are constructed to account for the potential efficiency differences between genders and between family labour and hired labour. Activities other than growing annual crops may affect efficiency as well. We include the variables *children* and *perennial* to represent the relative intensity. Finally, we consider three factors most relevant to crop growth, i.e. nutrient availability, oxygen availability to roots and workability for field management, to represent the growing conditions on each farm.[4]

Table 2 provides the summary statistics for the key variables in this study. It shows that most farms in the sample have a small size with a mean of 4.96 acres, and 95 per cent of them is smaller than 15 acres. The SI reports a mean value of 0.25 and a median of 0.20. Finally, about three-quarters of the plots are located within 3 km from either home or the nearest road.

5.3. Explanatory variables for heteroskedasticity

The two variance terms in equations (9) and (10) include explanatory variables related to plot heterogeneity, crop diversification and fragmentation. Since geo-referenced data of growing conditions are available only at the farm level, we use farmers' self-reported information on soil type, erosion type and steepness of slope to construct a plot heterogeneity measure, which is calculated as the number of soil profiles divided by the number of plots. Crop diversification is measured as the number of crops on each farm.

Furthermore, researchers have long assumed certain inputs, such as labour and fertiliser, to affect production risk in addition to output (Hurley, 2010). For example, Antle and Crissman (1990) found labour to be risk reducing, while Villano and Fleming (2006) argued that labour increases output variability. In stochastic production frontier models, the variance of either or both the one-sided and two-sided errors may be affected by producers' input use (Schmidt, 1986; Hadri, 1999; Hadri, Guermat and Whittaker, 2003). Hadri, Guermat and Whittaker (2003) reported that expenditure on labour and machinery increases the variability in efficiency, whereas land area and fertiliser cost have the opposite effect. This study considers labour for its possible effects on production risk.

4 Using geo-referenced homestead location data, the LSMS survey has imported soil and terrain data from the Harmonized World Soil Database at a resolution of 0.083 degree (approximately 10-km grids). The area in each grid is large enough to cover a typical farm, but not fine enough to generate plot-level information. To include categorical variables 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'.

Table 2. Summary statistics of key variables

	Unit	Mean	SD	Min.	First quartile	Median	Third quartile	Max.
labour1	person-day	66.23	72.34	1	22	44	84	645
labour2	person-day	62.07	72.16	1	21	41	76	703
labour3	person-day	50.35	77.06	1	11	25	59	1,282
precipitation	millimetre	580.90	190.03	231	419	542	706	1,284
temperature	degree	233.34	27.08	157	214	232	255	282
hoe	(count)	3.17	1.83	1	2	3	4	17
perennial	(ratio)	0.05	0.24	0.00	0.00	0.00	0.00	5.00
age	year	36.50	13.58	0.00	27.00	32.67	42.00	97.00
education	year	4.74	2.67	0.00	3.00	5.00	7.00	12.00
male labour	(ratio)	0.45	0.25	0.00	0.33	0.50	0.57	1.00
children	(ratio)	0.37	0.46	0.00	0.00	0.25	0.50	3.00
hired labour	(ratio)	0.09	0.17	0.00	0.00	0.00	0.10	1.00
land	acre	4.96	11.88	0.01	1.25	2.50	5.25	337.50
SI	–	0.25	0.26	0.00	0.00	0.20	0.49	0.88
distance1	kilometre	3.11	6.44	0.00	0.00	1.5	3.00	90
distance2	kilometre	1.90	3.01	0.00	0.06	1.00	2.5	42
distance3	kilometre	7.76	9.03	0.00	2.00	5.00	10.00	90.00
plot area	acre	1.74	4.60	0.0025	0.50	1.00	2.00	150.00
plot heterogeneity	(ratio)	0.88	0.22	0.20	0.75	1.00	1.00	1.00
crop diversification	(count)	2.18	1.11	1	1	2	3	7
crop_SI	–	0.36	0.27	0.00	0.00	0.45	0.57	0.83

6. Estimation and results

This study uses the Stata® package *sf_model* developed by Kumbhakar, Wang and Horncastle (2015) to estimate a stochastic production frontier, assuming a truncated normal distribution for the inefficiency term. This model allows for exogenous explanatory variables for inefficiency and heteroskedasticity in both variance terms. Before proceeding to model estimates, we first address two problems in the empirical production literature.

The first problem is the endogeneity associated with inputs in the production function. For non-experimental data like the LSMS data, endogeneity may exist because the observed use of inputs, especially labour, is not predetermined but is chosen by producers in some optimal fashion, such as profit (or returns) maximising or cost minimising. Failure to control for the unobservable factors, such as risk preferences or expectations, will generate inconsistent frontier estimates and efficiency scores (e.g. Marschak and Andrews, 1944; Mundlak, 1994).

Nevertheless, we have sufficient reasons to question the extent to which the above arguments for endogeneity apply to the Tanzanian context, and we believe endogeneity should not be a primary concern for this study. Among many others, Collinson (1983) and Makeham and Malcolm (1986) distinguished between two basic types of farm-operating objectives: profit-maximisation on market-oriented farms and household sustenance on subsistence-oriented farms. Most Tanzanian farmers belong to the second type and their primary farming objective is to produce enough food and fibre for household consumption. It is very likely that Tanzanian farmers use available resources (land, labour etc.) to maximum capacity instead of premediating some optimal plans. In this sense, use of inputs can be taken as 'fixed' or 'independent' so that the usual exogeneity assumption can be adopted in this context.[5]

The second problem is the theoretical consistency (e.g. monotonicity and quasi-concavity) of the estimated production frontier. Although the functional form of the production (or cost, distance etc.) frontier may seem tangential to the main interest of studies with stochastic frontier applications, it is necessary to check those regularity conditions to ensure the estimated function behaves well and satisfies the underpinning theoretical assumptions. Supplementary Material Section 3 (in supplementary data at *ERAE* online) provides a thorough discussion of this issue and the possible implications on the estimates.

6.1. Model selection and hypothesis tests

This study follows Kumbhakar and Lovell's procedure (2003) to determine the variance structure. Specifically, we start with the model *HUV* where

5 As pointed out by one reviewer, unobserved characteristics of farmers and farm households may affect their input capacity and agricultural production. Although the mean inefficiency function includes a few variables related to farmers' characteristics, it still may not completely eliminate the concerns over endogeneity, thus leading to biased estimates.

labour appears in both variance terms, while *SI, plot heterogeneity* and *crop diversification* appear in the variance of the two-sided error term. Then we estimate the model *HU* where heteroskedasticity appears only in the one-sided error u and model *HV* where heteroskedasticity appears only in the two-sided error v. Finally, model *HO* is estimated with homoskedasticity in both error terms (Table 3). Using likelihood-ratio tests, we find the model *HV* to be the statistically preferred model with four explanatory variables: *SI*, *labour*, *plot heterogeneity* and *crop diversification* (Table 3, Parts 3 and 4). Discussions below are based on model *HV* unless otherwise noted.

Table 3, Part 4 shows that the coefficient estimate of the SI in the variance function for v (equation (9)) is −0.568 and statistically significant at 1 per cent level. The negative sign suggests that the higher the SI (i.e. more fragmented farm), the lower the output variability, a relation that is consistent with the conceptual framework. The same relation is also reported for *crop diversification* with an estimate of −0.098. *Plot heterogeneity* reports a positive, although statistically insignificant, coefficient estimate. Lastly, *labour* is reported to be risk increasing with a statistically significant coefficient estimate, a finding consistent with those in many studies. As a tentative explanation, labour is likely to become more heterogeneous in quality and productivity as its quantity increases, thus causing more variations in output.

Regarding the determinants of inefficiency (Table 3, Part 2), coefficient estimates for *perennial* and *male labour* have the expected signs and both are statistically significant at the 1 per cent level. For example, the more effort is required for perennial crops, the less efficient is the production of annual crops. In comparison, *age*, *education* and *children* result in statistically insignificant estimates, suggesting these household characteristics are unlikely to affect technical efficiency in this sample. Curiously, the coefficient estimate for *hired labour* is −3.79 with a *p*-value of 0.017. The negative sign in the inefficiency function indicates that proportionally, more hired labour is linked to higher technical efficiency. This finding contradicts the common belief that hired workers are less efficient than family workers since the former lack farm-specific experience and are difficult to supervise (Feder 1985; Binswanger and Rosenzweig, 1986). This counter-intuitive estimate for *hired labour* may be attributed to bias caused by omitted variables such as the age and sex of hired workers. In Tanzania's rural labour market, men are more likely to work outside than women, such that hired labour has a relatively higher ratio of male workers than family labour does. If we believe men are more productive in farming than women but fail to control for the gender differential, hired labour will seem to be more efficient than family labour, *ceteris paribus*. Unfortunately, the LSMS data do not contain the necessary information for a further exploration. For variables associated with soil conditions, the only counter-intuitive estimates are reported for nutrient availability. The two associated dummy variables, *nutrient1* and *nutrient2*, report positive estimates that are significant only at the 10 per cent level. This confusing finding may arise from measurement errors in this factor since the data are collected on the pixel level instead of on the farm level.

Table 3. Model estimates and model comparisons (N = 1,503)

Part 1: Production frontier function (equation (3)):

Models[a]	HUV	HU	HO	HV	HV_crop
labour1 × *labour1*	−0.0114	−0.00301	−0.00748	−0.0166	−0.0156
	(0.0305)	(0.0278)	(0.0268)	(0.0286)	(0.0286)
labour1 × *labour2*	−0.0456	−0.0282	−0.0624	−0.0497	−0.0487
	(0.0531)	(0.0499)	(0.0505)	(0.0517)	(0.0517)
labour1 × *labour3*	0.0715	0.0384	0.0537	0.0733*	0.0729*
	(0.0451)	(0.0425)	(0.0425)	(0.0429)	(0.0428)
labour1 × *land*	−0.0226	−0.0412	−0.0110	−0.0171	−0.0150
	(0.0430)	(0.0432)	(0.0417)	(0.0417)	(0.0416)
labour1 × *precipitation*	0.00411	−0.0205	0.0370	0.00834	0.0105
	(0.131)	(0.127)	(0.129)	(0.131)	(0.131)
labour1 × *temperature*	0.601*	0.656**	0.584*	0.621**	0.636**
	(0.317)	(0.319)	(0.320)	(0.315)	(0.315)
labour1 × *hoe*	−0.0587	−0.0345	−0.0631	−0.0600	−0.0589
	(0.0765)	(0.0875)	(0.0768)	(0.0763)	(0.0762)
labour1 × *dummy*	−0.0264	0.0408	0.0486	−0.0370	−0.0425
	(0.111)	(0.134)	(0.105)	(0.108)	(0.108)
labour2 × *labour2*	0.0125	−0.00320	0.0166	0.0110	0.0119
	(0.0377)	(0.0422)	(0.0380)	(0.0375)	(0.0374)
labour2 × *labour3*	0.0317	0.0370	0.0432	0.0301	0.0286
	(0.0397)	(0.0383)	(0.0395)	(0.0391)	(0.0391)
labour2 × *land*	−0.0268	−0.0145	−0.0299	−0.0246	−0.0281
	(0.0419)	(0.0465)	(0.0420)	(0.0413)	(0.0413)
labour2 × *precipitation*	−0.0151	−0.0302	0.00782	0.000177	0.0174
	(0.145)	(0.150)	(0.140)	(0.140)	(0.140)

labour2 × temperature	0.0531	−0.0137	0.0718	0.0655	0.0307
	(0.314)	(0.349)	(0.321)	(0.314)	(0.315)
labour2 × hoe	0.0342	−0.00682	0.000740	0.0338	0.0363
	(0.0802)	(0.0926)	(0.0811)	(0.0798)	(0.0797)
labour2 × dummy	0.182	0.146	0.215*	0.191	0.200
	(0.135)	(0.170)	(0.128)	(0.133)	(0.133)
labour3 × labour3	−0.0924***	−0.0797***	−0.102***	−0.0983***	−0.0985***
	(0.0252)	(0.0227)	(0.0218)	(0.0220)	(0.0219)
labour3 × land	0.0329	0.0382	0.0734**	0.0349	0.0356
	(0.0350)	(0.0368)	(0.0349)	(0.0348)	(0.0348)
labour3 × precipitation	0.0975	0.0787	0.0840	0.0942	0.0892
	(0.103)	(0.101)	(0.102)	(0.102)	(0.102)
labour3 × temperature	−0.0441	0.0273	−0.0377	−0.0598	−0.0511
	(0.254)	(0.235)	(0.255)	(0.251)	(0.251)
labour3 × hoe	0.0905	0.110	0.0784	0.0882	0.0870
	(0.0619)	(0.0681)	(0.0610)	(0.0610)	(0.0610)
labour3 × dummy	−0.0187	−0.0297	−0.0826	−0.0150	−0.0146
	(0.110)	(0.122)	(0.107)	(0.110)	(0.110)
land × land	0.0448**	0.0378***	0.0198	0.0428**	0.0415**
	(0.0182)	(0.0138)	(0.0184)	(0.0176)	(0.0176)
land × precipitation	0.0699	0.107	0.0139	0.0514	0.0423
	(0.118)	(0.118)	(0.115)	(0.113)	(0.112)
land × temperature	−0.479*	−0.567**	−0.523*	−0.492*	−0.494*
	(0.262)	(0.288)	(0.269)	(0.261)	(0.261)
land × hoe	−0.0640	−0.0434	0.0294	−0.0563	−0.0555
	(0.0673)	(0.0633)	(0.0631)	(0.0644)	(0.0641)

(*continued*)

Table 3. (*continued*)

Part 1: Production frontier function (equation (3)):

Models[a]	HUV	HU	HO	HV	HV_crop
land × *dummy*	0.0239	0.00419	0.0631	0.0191	0.0169
	(0.0969)	(0.111)	(0.0958)	(0.0971)	(0.0969)
precipitation × *precipitation*	0.327	0.231	0.289	0.337	0.346
	(0.238)	(0.251)	(0.249)	(0.237)	(0.238)
precipitation × *temperature*	−0.614	−0.226	−0.415	−0.617	−0.643
	(0.749)	(0.695)	(0.774)	(0.747)	(0.750)
precipitation × *hoe*	0.217	0.250	0.292*	0.227	0.217
	(0.172)	(0.195)	(0.175)	(0.172)	(0.172)
precipitation × *dummy*	0.154	0.0331	0.0970	0.172	0.179
	(0.339)	(0.383)	(0.328)	(0.335)	(0.335)
temperature × *temperature*	1.355	2.229	1.753	1.201	1.161
	(1.520)	(1.431)	(1.498)	(1.496)	(1.497)
temperature × *hoe*	−0.645	−0.536	−0.487	−0.646	−0.650
	(0.427)	(0.477)	(0.440)	(0.427)	(0.427)
temperature × *dummy*	0.716	0.518	0.690	0.701	0.652
	(0.972)	(1.254)	(0.927)	(0.970)	(0.970)
hoe × *hoe*	0.120	0.0962	0.0952	0.119	0.114
	(0.0814)	(0.0988)	(0.0804)	(0.0814)	(0.0814)
hoe × *dummy*	0.0347	0.0275	−0.0298	0.0437	0.0383
	(0.168)	(0.178)	(0.164)	(0.169)	(0.169)
labour1	−3.153*	−3.347*	−3.223*	−3.252*	−3.360*
	(1.890)	(1.840)	(1.865)	(1.883)	(1.882)
labour2	−0.298	0.244	−0.525	−0.440	−0.361
	(2.024)	(2.063)	(2.018)	(2.004)	(2.004)

labour3	−0.0282	−0.283	0.0947	0.108	0.0970
	(1.542)	(1.378)	(1.521)	(1.513)	(1.511)
land	2.738*	2.956*	2.713*	2.890*	2.960*
	(1.552)	(1.641)	(1.595)	(1.542)	(1.537)
precipitation	−1.351	−2.100	−2.101	−1.531	−1.525
	(5.070)	(5.024)	(5.296)	(5.086)	(5.103)
temperature	−12.93	−24.96	−18.61	−11.32	−10.70
	(17.36)	(16.12)	(17.06)	(17.01)	(16.99)
hoe	1.970	1.162	0.750	1.924	2.004
	(2.544)	(2.774)	(2.600)	(2.548)	(2.551)
dummy	−5.004	−3.238	−4.695	−5.046	−4.829
	(5.773)	(7.676)	(5.566)	(5.784)	(5.783)
constant	51.87	86.78*	70.97	48.34	46.71
	(53.28)	(49.33)	(53.08)	(52.56)	(52.49)

Part 2: Mean inefficiency function (equation (5)):

Models	HUV	HU	HO	HV	HV_crop
perennial	0.520**	0.913**	0.583***	0.482***	0.467***
	(0.206)	(0.382)	(0.145)	(0.134)	(0.129)
age	0.00458	0.0216*	0.00814**	0.00338	0.00284
	(0.00551)	(0.0127)	(0.00387)	(0.00298)	(0.00269)
education	0.000462	0.0101	0.00110	−0.00443	−0.00539
	(0.0226)	(0.0456)	(0.0204)	(0.0163)	(0.0148)
male labour	−0.461*	−0.897	−0.478***	−0.386***	−0.365***
	(0.273)	(0.586)	(0.174)	(0.150)	(0.137)
children	−0.103	0.00299	−0.101	−0.0642	−0.0542
	(0.137)	(0.235)	(0.106)	(0.0937)	(0.0847)

(*continued*)

Table 3. (*continued*)

Part 2: Mean inefficiency function (equation (5)):

Models	HUV	HU	HO	HV	HV_crop
hired labour	−4.818	−6.934*	−1.832***	−3.793**	−3.382***
	(3.850)	(3.679)	(0.539)	(1.587)	(1.289)
nutrient1	0.335	0.530	0.170	0.306*	0.266
	(0.230)	(0.497)	(0.194)	(0.178)	(0.175)
nutrient2	0.348	0.560	0.321*	0.321*	0.286*
	(0.222)	(0.499)	(0.177)	(0.170)	(0.168)
oxygen1	−0.477	−1.505*	−0.775**	−0.376*	−0.355*
	(0.397)	(0.777)	(0.326)	(0.197)	(0.187)
oxygen2	−0.668	−2.040*	−0.976***	−0.499*	−0.468**
	(0.638)	(1.081)	(0.370)	(0.258)	(0.236)
workability1	−0.309	−0.862*	−0.383**	−0.266**	−0.250**
	(0.237)	(0.518)	(0.175)	(0.113)	(0.0998)
workability2	−0.244	−0.903*	−0.329*	−0.194*	−0.176*
	(0.228)	(0.515)	(0.181)	(0.113)	(0.106)
land	0.0155	0.0119	−0.665***	0.0126	0.0110
	(0.0143)	(0.0773)	(0.156)	(0.00910)	(0.00857)
*land * SI*	−0.000241	0.0154	0.0541	0.000449	0.00138
	(0.0111)	(0.0896)	(0.204)	(0.0101)	(0.00996)
SI	−0.546	−2.196**	−0.542*	−0.466***	−0.201
	(0.364)	(0.923)	(0.284)	(0.167)	(0.229)
crop_SI	–	–	–	–	−0.337
	–	–	–	–	(0.247)
distance1	−0.0718	−0.106	−0.0311	−0.0636**	−0.0601**
	(0.0484)	(0.0724)	(0.0199)	(0.0294)	(0.0274)

distance2	0.0354	0.0598	0.0266	0.0286	0.0274
	(0.0311)	(0.0571)	(0.0247)	(0.0219)	(0.0209)
distance3	−0.0136	−0.0316	−0.0194**	−0.00995	−0.00939
	(0.0156)	(0.0217)	(0.00978)	(0.00856)	(0.00798)
constant	1.288**	1.650**	3.058***	1.166***	1.173***
	(0.514)	(0.816)	(0.481)	(0.353)	(0.330)

Part 3: One-sided error variance function (equation (10)):

Models	HUV	HU	HO	HV	HV_crop
labour	0.00120	0.000778**	–	–	–
	(0.000878)	(0.000393)	–	–	–
constant	−2.525	−0.0305	−0.701***	−3.342**	−3.689**
	(2.378)	(0.467)	(0.235)	(1.576)	(1.747)

Part 4: Two-sided error variance function (equation (9)):

Models	HUV	HU	HO	HV	HV_crop
SI	−0.616***	–	–	−0.568***	−0.564***
	(0.213)	–	–	(0.172)	(0.171)
labour	0.000396*	–	–	0.000440**	0.000440**
	(0.000231)	–	–	(0.000208)	(0.000207)
plot heterogeneity	0.0340	–	–	0.0446	0.0476
	(0.211)	–	–	(0.201)	(0.201)
crop diversification	−0.0888	–	–	−0.0980*	−0.101**
	(0.0585)	–	–	(0.0511)	(0.0508)

(*continued*)

Table 3. (*continued*)

Part 4: Two-sided error variance function (equation (9)):

Models	HUV	HU	HO	HV	HV_crop
constant	0.148	−0.439***	−0.422***	0.162	0.169
	(0.244)	(0.0573)	(0.0619)	(0.228)	(0.228)
Log Likelihood	−1,989.97	−2,009.46	−2,003.59	−1,990.20	−1,989.37

Note: [a]In the HUV model labour appears in both variance terms, while SI, plot heterogeneity and crop diversification appear in the variance of the two-sided error term. In the model, HU heteroskedasticity appears only in the one-sided error u and in model HV heteroskedasticity appears only in the two-sided error v. Model HO is estimated with homoskedasticity in both error terms.
Standard errors in parentheses; ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.
All explanatory variables (except dummy) in the frontier model are in the logarithmic form.

6.2. Land fragmentation and technical efficiency

Our primary interest lies in the variables related to land fragmentation. The coefficient estimate for the SI in the mean inefficiency function is −0.466 with a p-value of 0.005. This can be interpreted as follows: the more fragmented the farm is, the more efficient the production. This relationship contradicts traditional wisdom, although it is robust to various model specifications. Other dimensions of land fragmentation (i.e. farm area, its interaction term with the SI and the three distance variables) are jointly significant according to a log-likelihood test (LR = 19.07, $p = 0.0019$). This finding confirms the need to use more than one measure to characterise land fragmentation.

The intriguing positive relationship between land fragmentation and technical efficiency reported in this study demands re-examining the arguments surrounding fragmentation and the underlying assumptions. In a general sense, Simons (1987) summarised that disadvantages of land fragmentation include the need for higher physical inputs due to increased labour and travel time among plots, the inability to use certain farming equipment and greater difficulty with pest control and supervision. However, Bentley (1987) emphasised that fragmentation would become a constraint to productivity only if it impedes the ability to use machinery in areas with decreasing agricultural population. Further, Fenoaltea (1976) argued that fragmentation could even benefit productivity by smoothing out seasonal labour use over a portfolio of plots with different attributes and in various locations. At low levels of mechanisation, it is possible that the benefits outweigh the costs. Recently, Di Falco *et al*. (2010) showed that land fragmentation could benefit farm productivity and profitability by fostering crop diversification. The reason being that polyculture, compared to monoculture, can facilitate complementary resource use and thus improve productivity (Cardinale *et al*., 2007).

The ambivalence in the conceptual assessments of land fragmentation is in line with the mixed empirical evidence. Most studies so far use the number of plots as the metric for land fragmentation. Some (e.g. Wan and Cheng, 2001) find a negative efficiency impact, while others (e.g. Tan *et al*., 2008) report statistically insignificant results. Applying stochastic frontier models to the 2005 Albanian LSMS data, Deininger, Savastano and Carletto (2012) discovered that the number of plots has a statistically significant, positive effect on output. Including both a SI and the number of plots for Chinese data, Chen, Huffman and Rozelle (2009) found that technical efficiency increases when the number of plots falls in the first quartile and starts to decrease when the number falls in the highest quartiles.

In Tanzania, each farm cultivates on average less than three plots when all crops are considered, and less than two plots when only annual crops are considered. The average farm size in this sample is about five acres (i.e. two hectares) and the median is only 2.5 acres (Table 2). The low level of fragmentation and small farm size are likely to have limited the negative impact of fragmentation. For the potential gains from fragmentation, the LSMS data do not allow an investigation of the labour-smoothing hypothesis. As a tentative exploration of the indirect effect of fragmentation through biodiversity, we construct a SI for crop varieties, *crop_SI*, as a rough metric of biodiversity. Specifically, *crop_SI* is calculated using the cultivated area for each crop in the same fashion as in equation (1); the higher the value the higher the biodiversity. The correlation coefficient between *SI* and *crop_SI* is as high as 0.88. Further, *crop_SI* is included in the mean inefficiency function as an additional explanatory variable. The results (Table 3, Part 2, model *HV Crop*) show that the coefficient estimate for *SI* remains negative but becomes statistically insignificant, and the same is true for *crop_SI*; the two estimates are jointly significant and negative with a *p*-value of 0.004. This finding may suggest that land fragmentation can benefit technical efficiency by way of biodiversity, although the causal relationship cannot readily be established. Crop diversification may result from other causes, such as risk management. We need to examine the extent to which land fragmentation has caused crop diversification before assigning the efficiency effects.

6.3. Efficiency scores and marginal efficiency effects

After discussing the coefficient estimates, we further derive farm-specific efficiency scores and the marginal efficiency effects of explanatory variables in equation (5). The *JLMS* estimator proposed by Jondrow *et al.* (1982) and the *BC* estimator by Battese and Coelli (1988) indicate very similar efficiency scores for this sample (Table 4). Consider, for example, the *BC* estimator. The average efficiency for the sample is 0.73, implying that these farms produce on average 73 per cent of the output predicted by the frontier function. It also shows the large variations between the most and least efficient farms, with a standard deviation of 0.18. This confirms the observation that agricultural production in Tanzania is characterised by low productivity and a tremendous variation across its many agro-ecological zones (USAID, 2011).

Table 5 reports the summary statistics for marginal inefficiency effects, which are calculated from the parameter estimates in the mean inefficiency function (equation (5)) and vary from observation to observation. Increasing *male labour* by 0.10 will on average decrease/increase the inefficiency/efficiency by 2.5 percentage points (= 0.10 × 0.25). Upgrading the workability of land from 'Severe Constraints' (the base category) to 'No or Slight Constraints' will increase efficiency by about 17 percentage points. The

Table 4. Summary statistics of efficiency score estimates ($N = 1{,}501$)

	Mean	SD	Min.	First quartile	Median	Third quartile	Max.
JLMS estimates	0.72	0.18	0.05	0.58	0.72	0.89	0.99
BC estimates	0.73	0.18	0.06	0.59	0.73	0.89	0.99

Note: The *JLMS* estimates are derived using the estimator proposed by Jondrow *et al*. (1982); the *BC* estimates are derived using the estimator proposed by Battese and Coelli (1988). Two observations are dropped out by the software during the calculation.

Table 5. Marginal inefficiency effects (N = 1,501)

	Mean	SD	Min.	First quartile	Median	Third quartile	Max.
perennial	0.31	0.18	0.00	0.13	0.38	0.47	0.48
age	0.00	0.00	0.00	0.00	0.00	0.00	0.00
education	−0.00	0.00	−0.00	−0.00	−0.00	−0.00	−0.00
male labour	−0.25	0.14	−0.39	−0.38	−0.31	−0.11	−0.00
children	−0.04	0.02	−0.06	−0.06	−0.05	−0.02	−0.00
hired labour	−2.42	1.41	−3.79	−3.72	−3.01	−1.04	−0.01
nutrient1	0.20	0.11	0.00	0.08	0.24	0.30	0.31
nutrient2	0.20	0.12	0.00	0.09	0.25	0.31	0.32
oxygen1	−0.24	0.14	−0.38	−0.37	−0.30	−0.10	−0.00
oxygen2	−0.32	0.19	−0.50	−0.49	−0.40	−0.14	−0.00
workability1	−0.17	0.10	−0.27	−0.26	−0.21	−0.07	−0.00
workability2	−0.12	0.07	−0.19	−0.19	−0.15	−0.05	−0.00
land	0.01	0.00	0.00	0.00	0.01	0.01	0.01
SI	−0.30	0.17	−0.47	−0.46	−0.37	−0.13	−0.00
distance1	−0.04	0.02	−0.06	−0.06	−0.05	−0.02	−0.00
distance2	0.02	0.01	0.00	0.01	0.02	0.03	0.03
distance3	−0.01	0.00	−0.01	−0.01	−0.01	−0.00	−0.00

estimated marginal inefficiency effects of the SI range from 0.00 to −0.47, suggesting widely varying effects from farm to farm. We use a thought experiment of land consolidation to give two rough interpretations of those farm-specific marginal effects. At the sample mean of −0.30, consolidating all the multi-plot farms into single-plot farms in our sample (the average SI = 0.25) will increase the average inefficiency by about 0.075 (i.e. 0.30 × 0.25), causing the average efficiency score to drop from 0.73 to 0.66. Alternatively, and aggregating from the marginal changes on each farm, the same consolidation experiment would reduce the average efficiency score from 0.73 to 0.58.[6]

For robustness, we explore alternative model specifications, such as using aggregated labour input instead of three disaggregated labour inputs, using alternative weather variables and using an alternative measure of crop diversification, and find no substantial changes to our major conclusions.[7]

7. Summary and concluding remarks

To summarise, this study develops a framework incorporating land fragmentation into both technical efficiency and production risk, to quantitatively

6 Usually, marginal effects in nonlinear models should not be interpreted over such a large range in the associated independent variable, so neither interpretation provided here is accurate. However, a marginal change in the SI is difficult to interpret practically, and the most meaningful counterfactual for comparison would be the case where all farms are consolidated. So our goal here is to give a rough interpretation of the marginal effects instead of an accurate estimate of productivity changes. We thank one reviewer for pointing out this issue.

7 The detailed results are available from the author upon request.

evaluate its impact on agricultural production. We hypothesise that land fragmentation can mitigate risk by diversifying production among separate land plots with different agro-ecological conditions. Applying a stochastic production frontier model to the Tanzania LSMS data, this study finds robust evidence to support the risk-reduction hypothesis. Meanwhile, land fragmentation is found to be positively associated with technical efficiency, thus benefiting farmers in terms of both efficiency and risk management (i.e. 'double dividends'). The positive relationship between fragmentation and technical efficiency contradicts conventional wisdom. Further explorations suggest that the low level of fragmentation in Tanzanian agriculture may have limited its negative impact on technical efficiency, while crop diversification spurred by fragmentation may have increased technical efficiency, leaving the net effect to be positive. Compared to previous studies, this study emphasises and confirms the need to fully characterise land fragmentation and its various effects.

Given the improved analytical framework and robust results, this study generates useful insights for countries such as Tanzania, where smallholding and traditional agriculture practices still prevail. Our analysis suggests that the small plot size and rare use of machinery can minimise the potential gains from land consolidation beyond the associated transaction costs, while land fragmentation may provide the desired benefits for farmers as a partial insurance against risk. Nevertheless, the findings in this study should not be interpreted as a conclusive recommendation on land fragmentation, whose roles depend crucially on the specific contexts. The vast differences in agro-ecological conditions, socio-economic constraints and farming traditions warn against any hasty generalisation of land fragmentation and once-and-for-all consolidation efforts.

Last but not least, future research on land fragmentation can improve on the work in this study in a few ways. Due to data availability, this study aggregates multiple crops by value into an implicit output index without comparing the associated production costs. Obtaining better data and estimating a profit frontier model could avoid potential aggregation bias and allow explicit consideration of the effects of fragmentation on labour costs as well as allocative efficiency. Better data can also allow researchers to relax some of the restrictive assumptions in this study, such as input exogeneity and independence between the two error terms. These assumptions may be violated in certain situations and bias the results and conclusions of this study. Future research should also develop a more general analytical framework that incorporates efficiency, production risk and risk preferences, and allow decision makers to make trade-offs between risk reduction and efficiency losses.[8] This effort will be meaningful not only to the immediate topic of land fragmentation but also to the broad field of efficiency analysis.

8 Kumbhakar (2002) developed an analytical model that allows the specification and estimation of risk preferences, production risk and technical efficiency. The empirical applications of his model have been very limited, probably due to the complicated estimation procedure.

Supplementary data

Supplementary data are available at *ERAE* online.

Funding

This research has received financial support from the International Science and Technology Practice and Policy (InSTePP) Center at the University of Minnesota and the HarvestChoice project funded by the Bill and Melinda Gates Foundation.

Acknowledgements

I am grateful to Journal Editor Ada Wossink, three anonymous reviewers, Terrance Hurley, Philip Pardey, Marc Bellemare, David Smith and Jaya Jha for their helpful comments.

References

Aigner, D., Lovell, C. A. K. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6: 21–37.

Antle, J. M. and Crissman, C. C. (1990). Risk, efficiency, and the adoption of modern crop varieties: evidence from the Philippines. *Economic Development and Cultural Change* 38: 517–537.

Battese, G. and Coelli, T. (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 Taylor, J. E. (1993). Folk soil taxonomy and the partial adoption of new seed varieties. *Economic Development and Cultural Change* 41(4): 763–786.

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.

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

Blarel, B., Hazell, P., Place, F. and Quiggin, J. (1992). The economics of farm fragmentation: evidence from Ghana and Rwanda. *World Bank Economics Review* 6: 233–254.

Cardinale, B. J., Wright, J. P., Cadotte, M. W., Carroll, I. T., Hector, A., Srivastava, D. S., Loreau, M. and Weis, J. J. (2007). Impacts of plant diversity on biomass production increase through time due to complementary resource use: a meta-analysis. *Proceedings of the National Academy of Sciences, USA* 104: 18123–18128.

Caudill, S., Ford, J. and Gropper, D. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroskedasticity. *Journal of Business Economics and Statistics* 13: 105–111.

Chen, Z., Huffman, W. E. and Rozelle, S. (2009). Farm technology and technical efficiency: evidence from four regions in China. *China Economic Review* 20: 153–161.

Collinson, M. (1983). *Farm Management in Peasant Agriculture*. Boulder, Colorado: Westview Press.

Deininger, K., Savastano, S. and Carletto, C. (2012). Land fragmentation, cropland abandonment, and land market operation in Albania. *World Development* 40: 2108–2122.

Di Falco, S., Penov, I., Aleksiev, A. and van Rensburg, T. M. (2010). Agrobiodiversity, farm profits and land fragmentation: evidence from Bulgaria. *Land Use Policy* 27: 763–771.

Dondeyne, S., Vanthournout, E., Wembah-Rashid, J. A. R. and Deckers, J. A. (2003). Changing land tenure regimes in a matrilineal village of South Eastern Tanzania. *Journal of Social Development in Africa* 18: 7–32.

Fan, S. and Chan-Kang, C. (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: 225–235.

Feder, G. (1985). Relation between farm size and farm productivity: the role of family labour, supervision and credit constraints. *Journal of Development Economics* 18: 297–313.

Fenoaltea, S. (1976). Risk, transaction costs, and the organization of medieval agriculture. *Explorations in Economic History* 13: 129–152.

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.

Hadri, K. (1999). Estimation of a doubly heteroskedasticity stochastic frontier cost function. *Journal of Business Economics and Statistics* 17(3): 359–363.

Hadri, K., Guermat, C. and Whittaker, J. (2003). Estimating farm efficiency in the presence of double heteroskedasticity using panel data. *Journal of Applied Economics* 6: 255–268.

Heston, A. and Kumar, D. (1983). The persistence of land fragmentation in peasant agriculture: an analysis of South Asian cases. *Explorations in Economic History* 20: 199–220.

Huang, C. J. and Liu, J. T. (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 Vietnam. Unpublished Doctoral Thesis, University of Sydney.

Hung, P. V., MacAulay, T. G. and Marsh, S. P. (2007). The economics of land fragmentation in the North of Vietnam. *Australian Journal of Agricultural and Resource Economics* 51: 195–211.

Hurley, T. M. (2010). A review of agricultural production risk in the developing world. HarvestChoice Working Paper. https://harvestchoice.org/publications/review-agricultural-production-risk-developing-world.

Jabarin, A. S. and Epplin, F. M. (1994). Impact of land fragmentation on cost of production wheat in the rain-fed region of Northern Jordan. *Agricultural Economics* 11: 191–196.

Jondrow, J., Lovell, C., Materov, I. and Schmidt, P. (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 Burton, S. P. (1982). Land fragmentation, a fundamental rural spatial problem. *Progress in Human Geography* 6: 475–494.

Kumbhakar, S. C. (2002). Specification and estimation of production risk, risk preferences and technical efficiency. *American Journal of Agricultural Economics* 84: 8–22.

Kumbhakar, S. C., Ghosh, S. and McGuckin, J. T. (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 Lovell, K. C. A. (2003). *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Kumbhakar, S. C., Wang, H. J. and Horncastle, A. P. (2015). *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*. New York: Cambridge University Press.

Liu, E. M. (2013). Time to change what to sow: risk preferences and technology adoption decisions of cotton farmers in China. *The Review of Economics and Statistics* 4: 1386–1403.

Maoulidi, S. (2004). *Critical analysis of land laws - a study*. Dar es Salaam: Land Rights Research Institute (Haki Ardhi). www.hakiardhi.org/index.php?option=com_docman&task=doc_download&gid=72&Itemid=80.

Markeham, J. P. and Malcolm, L. R. (1986). *The Economics of Tropical Farm Management*. New York: Cambridge University Press.

Marschak, J. and Andrews, W. (1944). Random simultaneous equations and the theory of production. *Econometrica: Journal of the Econometric Society* 12: 143–205.

McCloskey, D. N. (1976). English open fields as behavior towards risk. In: P. J. Uselding (ed.), *Research in Economic History: An Annual Compilation*. Connecticut: Jai Press, 17–63.

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

Meeusen, W. and van den Broeck, J. (1977). Technical efficiency and dimension of the firm: some results on the use of frontier production functions. *Empirical Economics* 2: 109–122.

Mundlak, Y. (1994). On the (in)efficiency of empirical dual production functions and related issues. University of Chicago, Unpublished paper.

Nguyen, T., Cheng, E. J. and Findley, C. (1996). Land fragmentation and farm productivities in China in the 1990s. *China Economic Review* 7: 169–180.

Niroula, G. S. and Thapa, G. B. (2005). Impacts and causes of land fragmentation, and lessons learned from land consolidation in South Asia. *Land Use Policy* 22: 358–372.

Schmidt, P. (1986). Frontier production functions. *Econometric Reviews* 4: 289–328.

Simons, S. (1987). Land fragmentation and consolidation: a theoretical model of land configuration with an empirical analysis of fragmentation in Thailand. Ph.D. Dissertation. University of Maryland, College Park.

Soini, E. (2005). Land use change patterns and livelihood dynamics on the slopes of Mt. Kilimanjaro, Tanzania. *Agricultural Systems* 85: 306–323.

Tan, S., Heerink, N., Kruseman, G. and Qu, F. (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.

United States Agency for International Development (USAID) (2011). USAID Land Tenure Tanzania Profile. http://usaidlandtenure.net/tanzania.

Villano, R. and Fleming, E. (2006). Technical inefficiency and production risk in rice farming: evidence from Central Luzon Philippines. *Asian Economic Journal* 20: 29–46.

Wan, G. H. and Cheng, E. (2001). Effects of land fragmentation and returns to scale in the Chinese farming sector. *Applied Economics* 33: 183–194.