Costs and benefits of land fragmentation: Evidence from Rwanda

Daniel Ayalew Ali, Klaus Deininger, Loraine Ronchi

World Bank, Washington DC
dali1@worldbank.org, kdeininger@worldbank.org, lronchi@worldbank.org

We thank the UK Department for International Development and the Knowledge for Change Program for funding support. We thank Rodney Dyer, Sion McGeever, and Cyriaque Harelimana for initiating and supporting data collection as well as the National Land Centre (now Rwanda National Resource Authority) in Kigali, especially Emmanuel Nkurunziza, Didier Sagashya and Thierry Ngoga for their overall support. We would like to thank the editor (Andrew Foster) and anonymous reviewers for their helpful comments and suggestions. We are also very grateful to Siobhan Murray for an excellent Arc GIS related support. The views presented in this paper are those of the authors and do not represent those of the World Bank or its member countries.

Costs and benefits of land fragmentation: Evidence from Rwanda

Abstract: Panel data from Rwanda allow us to explore costs and benefit from land fragmentation in a non-mechanized setting using two methodological improvements, namely (i) a terrain-adjusted measure of travel time/cost required to visit all parcels to measure fragmentation; and (ii) instrumental variable (IV) approaches that use measures for inherited/allocated parcels and past displacement as instruments. Results suggest that fragmentation as measured by travel cost negatively affect yield, intensity of labor use, and technical efficiency while reducing yield variability. With some 7% increase in yields, the size of the estimated impact of potential consolidation remains modest, suggesting that in an un-mechanized setting such as the one studied here, the costs of programs to reduce fragmentation may outweigh the benefits.

JEL Classification: O13, Q12, Q15, Q16

Keywords: land fragmentation, productive efficiency, travel cost, yield, crop shocks, Rwanda

1. Introduction

Concern about land fragmentation as a constraint to agricultural productivity has long featured prominently in the policy debate. Although it is acknowledged that, by reducing susceptibility to localized shocks, spatial diversification of field locations can bring risk diversification benefits, having a cultivator's land distributed in many small fragments, often located at large distances from each other, can tend to increase production cost as more time is needed to move between fragments and set up equipment or as use of machinery will be entirely impossible. Equal inheritance practices in settings with high population growth or land reforms that aimed to award multiple parcels of different type are often believed to imply that households are endowed with 'inefficiently' large numbers of parcels. Illiquid or non-existent land markets and high coordination costs may then cause a collective action problem that makes it difficult for individual land owners to consolidate. Concern about involuntary fragmentation and barriers to market-driven consolidation, in particular high coordination costs and insufficiently liquid markets have thus been key to motivate public consolidation efforts in many parts of the developed world. Often, these were motivated by a desire to 'modernize' agriculture and open the way to large-scale mechanized cultivation rather than evidence on impacts of fragmentation or the results of past consolidation programs.

To assess when and to what extent land fragmentation should concern policy makers, this paper explores impacts of this phenomenon on productivity and management of production risk. The conceptual basis is simple. Having parcels in a diverse set of locations can reduce exposure to risk and allows exploiting micro-variation in climate or soil quality. These diversification benefits need to be weighed against fragmentation-induced costs, in particular more time needed to move between parcels. Net impacts of fragmentation will then vary across settings with external factors such as the opportunity cost of labor, mechanization potential, levels of production risk, and access to other forms of insurance, being key parameters. Benefits from a fragmented holding structure may well outweigh associated costs in settings where mechanization is not an option, most work is performed by family members, or wages are low. At the same time, where labor costs and payoffs from mechanization are high, having holdings with a large number of irregularly shaped parcels that are too small to allow mechanized cultivation is likely to be a clear impediment to higher productivity.

We use parcel level panel data from Rwanda, a country where land is scarce and holdings highly fragmented, to illustrate the underlying factors and contribute to the literature in a number of ways. First, in an un-mechanized setting time spent traveling between parcels is arguably a major channel through which fragmentation will affect outcomes. We thus rely on GPS information on location and altitude of each of the parcels operated by a household together with a digital elevation model (DEM) to calculate friction-adjusted travel time between the homestead and all of the household's parcels. Using this variable as our key measure of fragmentation is justified on conceptual ground and supported by results suggesting that this variable is indeed more robust and has higher explanatory power than fragmentation measures, in particular number of parcels and the Simpson index, that have traditionally been used in the literature.

Second, while it is recognized that analysis of impacts of fragmentation on productivity has to account for the fact that, producers can choose number, size, and location of parcels subject to the constraints imposed by technology and factor markets, few studies addressed the resulting endogeneity. We use exogenous variation in inheritance or allocation of land at parcel- and displacement at household-level to construct instruments that allow addressing potential endogeneity more convincingly. Results suggest that fragmentation, measured by GIS/DEM-based travel time/cost, reduces yields and intensity of labor use but has risk-reduction benefits in terms of reducing the variance of yield at parcel level. The same result obtains for technical efficiency when we draw in a third round of data and use a two-step approach where conventional input use is instrumented with lagged values.

Quantifying the magnitudes involved in Rwanda suggests that, in this setting where mechanization is not (yet) an issue, fragmentation reduces yield and labor intensity, plausibly due to increased time required to move between parcels. Hypothetical amalgamation all of a household's parcels into one would, based on our results, increase yield by some 7%. If increased variability of production outcomes due to consolidation is negligible, capitalizing this value provides an upper limit for the cost of a land consolidation program that can, in the setting at hand, be justified economically. Programs with costs in excess of this value require additional assumptions in terms of future benefit streams (e.g., as a result of mechanization) to be economically justified.

The paper is structured as follows. Section two reviews the literature and lays out the conceptual framework and measurement issues. Section three describes the data, provides descriptive statistics at holding and parcel level, and explores the extent to which expansion of holdings through land market transactions may underlie some of the high levels of fragmentation observed in the data. Section four presents the econometric approach as well as results regarding risk diversification benefits from land fragmentation, yield functions, and estimates of technical efficiency using a frontier production function. Section five concludes by drawing out implications for policy both in Rwanda and beyond.

2. Measurement, conceptual framework, and econometric approach

The literature on impacts of fragmentation has often assumed these to be identical across settings and, in part a consequence, used crude proxies to measure this phenomenon. To provide the conceptual framework for our study we first discuss how exogenous factors may affect impacts of fragmentation. Noting that in our setting time spent moving between parcels is likely to be a key factor, we then describe a more precise way of capturing this variable and its application to our data.

2.1 Context and measurement

While fragmentation may be an issue in many contexts, it is near the top of the agricultural policy agenda in Rwanda, a country with not only the highest population density in Africa (384 inhabitants/km^2 overall and 526/km^2 of agricultural land) and high fragmentation.[1] Negative impacts from population growth-induced fragmentation are widely viewed as having led to environmental challenges (Republic of Rwanda 2004) or conflict (Andre and Platteau 1998). To stem fragmentation, the law prohibits fragmentation of land below a size of 1 ha,[2] though more than 90% of parcels fall below this size (Ali and Deininger 2015).

Fragmented land holding structures may be exogenously imposed by inheritance rules under high population growth (Baker and Miceli 2005; Platteau and Baland 2001) or the modalities of plot individualization in the transition from centrally planned economies (Tanaka 2007). Alternatively, they may be chosen by producers who trade off benefits against costs of fragmentation. A key benefit is the ability to diversify the plot portfolio to include plots with different attributes (cropland, pastures, orchards, vineyards) so as to reduce exposure to price risk or variability of production due to flood, drought, or pests (McCloskey 1975) or ease bottlenecks and smooth out seasonal variability in labor demands (Fenoaltea 1976). Costs accrue because fragmented plot holdings may preclude use of machinery and, depending on the

[1] As 85% of the population have agriculture as their main income source, effective land use and land-related investment are critical for poverty reduction in Rwanda (Republic of Rwanda 2009). In 2008, the average household had only 0.72 ha of land, below the threshold of 0.75 ha estimated to be required to meet a family's subsistence needs with available technology.

[2] The 2005 Organic Land Law (OLL) prohibited sub-division of agricultural land parcels less than one ha and required approval for any subdivision of parcels between 1 and 5 ha. The prohibition of sub-division is upheld in the 2013 OLL revision.

location of plots, may require spending more time to move between plots (Bentley 1987). Ascertaining the impact of fragmentation will thus require analysis of costs as well as benefits.

The simplest empirical measure for fragmentation is the number of parcels. The Simpson index $SI = 1 - \sum a_i^2 / (\sum a_i)^2$ where a is the area of parcel i=1,..,n, links this to variation across parcels in size.[3] To account for the ability to grow a larger number of crops as a potential benefit of fragmentation, we complement this with the Shannon crop diversity index, defined as $SH = -\sum(P_i * lnP_i)$ where P_i is the share of area covered by a specific crop. This index which combines both richness of type (number of crops) and relative abundance (land allocation among crops), equals zero if there is only one crop (i.e., no diversity) and increases with the number of cultivated crops as well as with more even shares by different crops, reaching its maximum when crops are cultivated in equal shares (i.e., P_i=1/J) where J is the maximum possible number of crops cultivated. Although neglected in most relevant studies, variability in parcel altitude provides an additional dimension that may be relevant for risk-reduction and diversification benefits from fragmentation.

As they do not incorporate information on relative plot location, the above measures are at best a noisy proxy for the cost of having to travel between plots. In particular, households with a given number and size distribution of plots will have the same Simpson index irrespectively of whether these plots are located next to each other and the homestead or kilometers away and separated by rivers and mountain ridges. While some surveys include questions regarding travel time between homestead and plots to fill this gap, these are subjective and may suffer from measurement error. Recording parcel-level GPS coordinates for all of a household's parcels in the cell of residence[4] allows us to go beyond this and compute the mean Euclidian distance between homestead and parcel, the length of the path to be traversed to visit all parcels, and the time to walk these in light of terrain characteristics. Use of the latter is justified because, in Rwanda's hilly terrain, a straight line will often not be the most effective way of getting from location A to location B.

To implement this, we use the publicly available Aster Global DEM to derive a 30 by 30 meter slope grid and apply appropriate friction parameters to compute the least cost walking distance between any two points.[5] Figure 1, which displays the DEM together with the location of the homestead and parcels for one of our sample households using 4 parcels (located at an altitude between about 1,500 and 2,100 m), one next to the homestead, illustrates this graphically. It is easily verified that the (red) straight line or Euclidian distance differs significantly from the (blue) least cost path between any pair of parcels. From the pairwise least cost paths between all parcels we choose the minimum path to visit all of a household's parcels starting and ending at the homestead as the measure for our regression (see figure 2 for an illustration).

---

[3] The index varies between 0 and 1 with zero for a completely consolidated operation.

[4] Expanding beyond the cell would have been too costly.

[5] The Aster model is available at http://asterweb.jpl.nasa.gov/. We use Tobler's (1993) hiking function—estimated using Imhof's (1950) empirical data—that models for walking on footpath by taking into account the slope of the terrain. It is an exponential function of slope (S), defined as W = 6*exp{-3.5*abs(S + 0.05)}, that takes an average speed of 5 km/h on a flat terrain.

2.2 Conceptual issues and existing evidence

Studies that include the number of plots cultivated or the Simpson index of land fragmentation that ranges from zero to one, with higher values corresponding to higher fragmentation but fail to account for risk-reduction benefits, often find negative effects of fragmentation. In 5 Chinese provinces (Jilin, Shandong, Jiangxi, Sichuan, Guangdong), a production function for a sample of some 1,000 farmers with 5 key crops (maize, early and late rice, wheat, and tubers) in 1994 suggests a significant negative impact: exogenous addition of one plot is estimated to reduce output by 2 to 10 percentage points depending on the crop (Wan and Cheng 2001).[6] In Bangladesh, for a small sample of 298 farm households with an average of 4.4 parcels per holding in 2000, an estimate from a frontier production function points towards a negative productivity- and efficiency-effects from plot fragmentation (Rahman and Rahman 2008). Estimating a yield function on a small panel from North Vietnam for 200 farmers on very fragmented holdings (mean and median plot numbers of 6.84 and 6 and Simpson index of 0.59 and 0.68, respectively) in 2000/1 similarly suggests that number of plots has a statistically negative effect on yield and increased family labor use and other input expenses (Hung *et al.* 2007). On the other hand, a small sample (227 households) from Hubei province in China finds that the number of plots had no significant effect on crop production (as measured by a production function). Moreover, a program to reduce land fragmentation did not achieve its main goal, possibly because with relatively abundant labor, the scope to use any labor savings in other tasks was limited and low levels of technology implied significant fragmentation-related benefits in terms of risk-reduction (Wu *et al.* 2005).[7]

One way of accounting for potential benefits from fragmentation more explicitly is to explore impacts on variability of output in a panel setting. Kawasaki (2010) applied this approach together with a cost stochastic frontier function on a large panel of Japanese rice farmers (18,000 farms per year from 1995-2006 rice production cost statistics) who cultivate a mean of 4 parcels with a Simpson index of 0.8 together with a cost function. Results show that parcel number and Simpson index increase cost inefficiency and average cost but reduce variance of output. In Japan, high wages and well-functioning insurance markets implied that costs associated with fragmentation outweighed the benefits but this may well be different in other settings (Kawasaki 2010). In Bulgaria, fragmentation (measured by the Simpson index) was found to reduce farm profitability but to boost richness of species as measured by a biodiversity index (Di Falco *et al.* 2010).

If fragmentation increases cost of production due to the need to move between plots and/or the homestead, it may be desirable to measure these costs more explicitly. Indeed, studies going in this direction suggest

[6] The magnitude of these effects is used to argue that efforts to eliminate fragmentation, either by improving functioning of factor markets or by programs aimed specifically at land consolidation, would be worthwhile (Chen *et al.* 2009).
[7] This mirrors evidence from Albania where fragmentation does not seem to negatively affect output and crop abandonment due to ill-functioning land market institutions, in particular the registry seems to be a more potent factor in reducing land use efficiency (Deininger *et al.* 2012).

not only that distances are often highly relevant but also that they are often very weakly correlated with more traditional measures of fragmentation. A cost function for 331 rice farmers on highly fragmented holdings (more than 7 plots on average with a Simpson index of 0.73 and a mean homestead-parcel distance of 16 min.) in three villages of Jiangxi Province in China suggests that cost of production is little affected by the Simpson index, increases with mean homestead-plot distance, and falls somewhat with farm size (Tan *et al.* 2008).[8] A stochastic frontier production function using these data suggests that mean plot-homestead distance reduces technical efficiency whereas number of plots (and mean plot size) both are associated with higher levels of technical efficiency (Tan *et al.* 2008).

The relative importance of fragmentation-induced costs and benefits will depend on wages, the scope for mechanization, and availability of other options to reduce or insure against risk. If labor costs are high and mechanization widely practiced, the costs imposed by small or highly fragmented plots can easily outweigh potential benefits. On the other hand, if options to diversify risk via insurance are limited and labor cost low, benefits from fragmentation may outweigh associated costs (Sengupta 2006). This explains why returns of programs to consolidate holdings, often together with land use planning, in developed countries are reported to have been high (Simons 1987). It also suggests that impacts may change with factor price ratios in the course of economic development. For example, in India it is argued that traditionally productivity losses associated with land fragmentation have been modest and benefits substantial (Heston and Kumar 1983) but this may change as cost of labor relative to machinery increases (Foster and Rosenzweig 2011).

Study of fragmentation impacts in Africa, though not always fully conclusive, highlights the cost-benefit trade-off and yields qualitatively similar conclusions. Data for a small sample (150 and 80 households per region) from three regions in Ghana suggest little, if any, correlation between the level of fragmentation and mean plot-homestead distance.[9] Yield equations provide little evidence of land fragmentation having adversely affecting productivity (Blarel *et al.* 1992). Although Rwanda is characterized by high fragmentation (6 parcels with a median Simpson index of 0.66 per household), there is no significant relationship between levels of fragmentation and yields, with and without efforts to control for endogeneity. While pointing towards risk reduction as a potential reason for persistence of fragmentation, links between fragmentation and variability of output remain tentative at best (Blarel *et al.* 1992).

[8] The Simpson index was found to increases labor cost but reduces costs of fertilizer, seed, and oxen or tractors costs.

[9] Two regions (Analoga and Wassa) are highly fragmented (median number of 5 parcels and Simpson index of 0.77 and 0.66, respectively). On the other hand, fragmentation is very low in the third region (Ejura), with a median parcel number of one and a mean Simpson index of 0.23. However, the correlation between average parcel distance and fragmentation in consistently low and average distance between homestead and parcels is greater in the latter than the former (3.5 km vs. 1.8 km).

3. Data and descriptive evidence

To highlight how our parcel-level panel from 3,600 small Rwandan producers allows us to address some of the issues raised in the literature, we describe these data and compare self-reported travel time between parcels to our GIS/DEM-based measure. Evidence of systematic biases provides a justification for using the more objective measure in relevant regressions.

3.1 Data sources and household/parcel characteristics

We use a nationally representative panel survey of 3,600 households in 300 randomly selected villages conducted in 2010/11 and 2011/12 (see figure 3 for location of sample cells). Detailed information on crop production was collected at parcel level and GPS readings were taken for the homestead and all parcels in the cell of a household's residence.[10] Key descriptive statistics are illustrated in table 1, overall (col. 1) and for households in terciles of the parcel distribution (cols. 2-4) where stars indicate significance of differences between the 1st and the 2nd and 3rd terciles.

With a mean holding size of 0.81 ha, 0.61 ha of which is cultivated, the average sample household has almost 5 parcels on which some 4.5 crops are grown, a figure that increases to 8.3 parcels and 5.6 crops for the top tercile (see appendix figure 1 for the distribution of number of crops cultivated). Some 54% of parcels were acquired via inheritance, more than a third via market transactions (22% through purchase and some 15% via rental arrangements), with about 5% allocated by government. About 8.2% of parcels are in wetland, 5.6% have access to irrigation, and 36% and 16% are reported to be of high or medium soil quality. About 50% are cultivated with grains, 26% with tubers, 23% with trees, and the remainder of about 1% with vegetables. Some 10% of parcels benefit from extension, mean output is US$ 552/ha, with slight variation in terms of the share of tubers vs. grains. The average parcel has been with the current owner for some 18 years. Female headed households have smaller landholdings and fewer parcels than male headed ones. Households with more parcels are also slightly larger, with 2.3, 1.7, and 0.98 members below 14, between 14 and 35, and between 35 and 60, respectively and have marginally higher levels of education (64% and 9.6% having completed primary and secondary education).

Mean parcel area is 0.19 ha and 28% of households had experienced a crop shock in the reference period. The mean self-reported distance between homestead and parcel is 0.17 hours, from 0.13 hours in the bottom to 0.21 hours in the top tercile, slightly higher than the computed minimum cost path of 0.11 hours (0.08 to 0.14). Total travel time from the homestead to visit all parcels and then return home is 0.82 hours overall, from 0.38 hours for the bottom tercile to 1.37 hours for the top one, an increase that is markedly higher in

[10] The 3-round panel survey was designed to assessing the impact of Rwanda's program of land tenure regularization (LTR) and was managed by the World Bank with support from the UK Department for International Development (DFID) and International Growth Centre (IGC).

relative terms than the increase in cultivated (0.53 to 0.74 ha) or owned area (0.59 to 1.13 ha). Simpson and Shannon indexes peak at 0.7 and 1.4 (appendix figures 2 and 3). The association between holding size and number of parcels is weak, i.e., a desire to expand cultivated area does not seem to be the primary motivation for increasing the number of cultivated parcels (see appendix fig. 4).

3.2 Comparing self-reported and computed parcel distances

To provide initial evidence to on the extent to which endogenous adjustment of parcel sizes or numbers by households in response to fragmentation may be a concern, table 2 compares key distances for land rental (panel A) and land purchase (panel B) between participants and non-participants and, more importantly, within the same household between owned and rented or purchased parcels. We report the mean walking time between parcels on the fastest path taking into account slopes from the DEM (cols. 1-2), between parcels and the homestead using the least-cost path based on the DEM (cols. 3-4), and walking time between parcels and the homestead as obtained from households' survey responses (cols. 5-6).

Comparing survey- and GIS-based measures suggests that households' estimates may be biased and that, in setting such as the one studied here, Euclidean distance is a poor measure of travel time. While survey-based estimates of walking distance to owned parcels are reasonably close to those computed using the least-cost path, households seem to systematically overestimate the distance to rented parcels. This is evident from the significant difference in homestead-parcel distances between owners and renters-cum owners (rows 1 and 2, cols. 5-6), contrary to figures using least cost paths computed based on GIS/DEM information. Also, while rented parcels are more distant from the homestead based on owners' estimates (12 vs. 23 min) and GIS/DEM information (10 vs. 16 min), no such difference exists for the least cost path between rented to owned vs. owned parcels only. Land rental markets thus appear to increase homestead-parcel distances, but have no impact on travel time between parcels. Differences are more pronounced for purchased parcels; while there is no significant difference in distance between pure owners and those who also purchased some parcels, mean travel time to purchased vs. own parcels is larger from the homestead (11 vs. 16 min.) and for the within-parcel travel time (17 vs. 23 min).[11] To adjust for potential endogeneity, we use instrumental variable (IV) estimation technique, as discussed in more detail below.

4. Econometric results

Following a brief discussion of the instruments used, we show that IV estimates for yield, labor intensity, and variability of yield produce results that are more consistent than un-instrumented ones, that our cost distance measure routinely outperforms variables used to capture fragmentation in the literature, and that

[11] Interestingly, use of survey responses would lead one to underestimate this difference, suggesting that ideally the impacts of fragmentation should be studied using of GIS/DEM-based information.

consistent results are obtained for a two-step approach to measure technical efficiency. Fragmentation is estimated to significantly reduce yield, labor intensity, and technical efficiency at a rate that is decreasing with the level of fragmentation. Yet, with hypothetical amalgamation of all of a household's parcels estimated to result in a 7% increase in value of output per ha, large-scale interventions to consolidate holdings may be economically justified only if there is an assumption of additional benefits.

4.1 Fragmentation and crop shocks

Before considering impacts of fragmentation on productivity or farm efficiency, we assess if it helps Rwanda's smallholders manage the effects of crop shocks at the household level. A measure for this is obtained from responses to a binary parcel-level question on whether or not production on this parcel had been affected by shocks in the last season. Aggregating this to the household level using parcel area as a weight allows us to obtain a measure of the share of a household's land that was affected by shocks in any given season.[12] Regressing this variable on indicators of fragmentation (number of parcels/crops, Simpson or Shannon index), farm characteristics such as the coefficient of variation in parcels' altitude, Herfindahl index of diversity in soil quality and slope,[13] and household attributes provides a naïve reduced-form estimate of the extent to which any of these factors may help to reduce the likelihood of the farming household, on average, being affected by a crop shock.

Results from this reduced form regression are given in table 3. Key independent variables are the Simpson land fragmentation index (col. 1), the number of parcels (col. 2), the Shannon crop diversity measure (col. 3), and the number of crops (col. 4). Point estimates for the coefficients on all of these are negative, suggesting land fragmentation reduces the likelihood of crop shocks, on average, at the household level. At the same time, magnitude and significance levels are larger and significant at conventional levels for proper fragmentation measures compared to indicators of crop diversity with the number of crops grown being small and insignificant. Regarding other variables, the coefficient of variation for parcel altitude remains insignificant suggesting that differences in elevation may affect output more through crop choice rather than risk diversification. While the lack of a measure of the severity of the shocks or losses associated with it cautions against over-interpretation and implies that the relationships should be taken as indicative only, the magnitude of fragmentation-related reduction of the level effects of crop shocks seems substantial; according to the estimates, adding an additional parcel would reduce the likelihood of shocks as much as

[12] The question was specifically phrased as: "While in the field or at the time of harvest, was there any crop damage/loss?" and it was asked at the parcel level. While we acknowledge that information on the intensity of crop shocks, rather than just a zero-one variable as used here, might have been useful, aggregated to the household level in the way described provides the most appropriate way of testing the hypothesis that land fragmentation can help farmers to manage risks.

[13] The Herfindahl index (fractionalization index for soil type and topography) that captures the probability that two randomly selected parcels from a given household will not belong to the same soil type or topography is defined as $H = \sum_{i=1}^{n} p_i(1 - p_i)$ where i indexes groups, n is the total number of groups (soil type or topography) in each household and p_i is the share of group i in the household (i.e., the area share of a given soil type or topography). It ranges between zero and one, moving from a homogenous to a more diversified land type.

an addition of more than 3 adults to the household. Larger households are less likely to be affected by crop shocks, possibly because they devote more attention to crop management. Female headed households are no more likely to be affected by crop shocks than male-headed ones.

4.2 Yield regression

We explore effects of land fragmentation on agricultural productivity by estimating a reduced form yield equation to measure the direct impacts of land fragmentation on the value of crop output per unit of cultivated land at the parcel level. Assuming a Cobb-Douglas production function, the yield equation is:

$$\ln\left(\frac{Y_{iht}}{L_{iht}}\right) = \beta_0 + \beta \ln L_{iht} + \alpha \ln A_{ht} + \theta' P_{iht} + \delta' H_{ht} + \phi' F_{ht} + V_t + \epsilon_{iht}, \qquad (1)$$

where Y_{iht} denotes total value of crop output on parcel *i* by household *h* at time *t*; L_{iht} is cultivated parcel area and A_{ht} is total operated area at household level, both in hectares; P_{it} is a vector of parcel characteristics including subjective land quality measures (soil type, topography, irrigation), crop dummies, and a measure for whether or not parcel-specific crop shocks were experienced; H_{it} is a vector of household level characteristics; F_{it} is a vector of land fragmentation indicators at household level (number of parcels, Simpson fragmentation index, Shannon crop diversity measure, or the minimum cost travel time between the homestead and all parcels); V_t is a time-varying village specific fixed effects; β, α, θ, δ and ϕ parameters to be estimated and ϵ_{iht} is a random error term. Standard errors are clustered at the household level. This specification allows us to examine the relationship between yield on a given parcel size and land fragmentation conditional on the household's total operated area. We also run a similar specification for the intensity of labor use per hectare at the parcel level.

We first estimate a naïve household random effects regression of yield and labor input use per hectare at the parcel level on either number of parcels or the Simpson land fragmentation index controlling only for parcel and total holding size. Results are rather mixed (see appendix table 1): while number of parcels has a significant positive association with yield, the Simpson index is negatively and significantly associated with the intensity of labor use per hectare. Adding parcel and household level characteristics including the Shannon crop diversity and DEM/GIS based cost distance measures (appendix tables 2 and 3) slightly changes the results as the positive and significant coefficient on number of parcels in the yield function becomes statistically insignificant while the negative relationship between the Simpson index and yield becomes statistically significant, suggesting that fragmentation lowers the value of crop output per hectare at parcel level. While the negative association between land fragmentation and labor use remains unchanged, the positive statistically significant coefficient of the Shannon crop diversity measure suggests that having multiple crops increases intensity of labor use and yield for a given parcel and holding size. Interestingly, GIS/DEM-based travel cost measures are insignificant throughout.

If households' land market participation decisions affect our measures of fragmentation (Deininger *et al.* 2016), these specifications will suffer from endogeneity. For example, if more productive farmers were more likely to acquire land via markets and acquisition of adjacent parcels is difficult, fragmentation would be correlated with farmers' unobserved ability. To address this, we instrument for number of parcels, Simpson and Shannon index, and the total cost distance and its square. Instrumental variables that are highly correlated with fragmentation but orthogonal to productivity are the number and size of parcels inherited or received via allocation by government; total cost distance and its square computed using only parcels that had been inherited or received via government allocation; and an indicator variable that takes the value of one if the household head had been displaced in the past and zero otherwise. Descriptive statistics of the instruments are in the bottom panel of table 1. Past displacement, especially during the 1994 Genocide, is a good instrument as the victims of such displacement were often allocated more consolidated pieces of land in settlement schemes (Kondylis 2008). Instruments that exploit exogenous variation in the modality of land acquisition have been shown to be useful in the literature (Deininger *et al.* 2016; Foster and Rosenzweig 2011; Lai *et al.* 2015).[14] Statistically, our choice of instruments is supported by the fact that in regressions of yield, labor intensity, yield variance, and a two-step estimation of technical efficiency the Hansen J statistic overidentification test does not allow us to reject the validity of the instruments overall.

Estimation results for the yield function are presented in table 4 with number of parcels (cols. 1-2) or the Simpson index (cols. 3-4) as key fragmentation indicators. We complement basic specifications in cols. 1 and 3 with the GIS/DEM-based distance measure and the Shannon index.[15] Although significant if introduced on their own, coefficients on standard measures of fragmentation (parcel number or Simpson index) routinely used in the literature lose significance once GIS/DEM-based measures of fragmentation are added (cols. 2 & 4). The only fragmentation-related variable that remains significant is the cost distance which, over the range of our data, is estimated to affect yield negatively at a decreasing rate.[16] A rough estimate of the net effect of fragmentation, obtained by replacing the actual with twice the average household-parcel cost distance for all of a household's parcels, suggests that amalgamation of all of a household's plots into one would result in a 7% yield increase.

Coefficients on other variables are in line with expectations. A large negative relationship between parcel size and yield (-0.62) as found elsewhere (Ali and Deininger 2015) is partially compensated by a positive

[14] These studies applied similar approaches in different contexts: Deininger *et al.* (2016) examined only inherited paddy and wheat plots in India to investigate the productivity effects of land fragmentation; Foster and Rosenzweig (2011) used inherited land as an instrument for operated land in their study of the efficiency of small-scale farming in India; and Lai *et al.* (2015) exploited previous long-term land assignment as an instrument for land fragmentation on machinery use and crop production in China.

[15] The fact that these results are substantively quite different from those obtained from estimating a simple random effects regression (see appendix table 2) highlights the importance of accounting for endogeneity.

[16] According to our estimates, cost-distance would have a positive effect for households needing more than 5 h to visit all their parcels which is well above what is observed in our data.

albeit much smaller (0.096) coefficient at holding level. Yield on parcels located at greater distance to the homestead is somewhat lower with one additional hour of walking time estimated to reduce yield by 10 percentage points. Parcel-level crop shocks are estimated to reduce yield by 21 to 28 percentage points. Good soil quality is estimated to increase parcel-level yield by 9-10 points, similar to irrigation access. Reliance on extension messages is estimated to increase gross value of crop output per hectare by about 17%,[17] and possibly due to experience, each additional year of land possession is estimated to be associated to a 0.5 percentage point yield increment.

If, as the above suggests, fragmentation reduces output because more time is spent moving between parcels than working on them, one should be able to observe a negative effect of fragmentation on labor intensity. Results from an IV regression where yield is replaced with labor input per hectare in table 5 support this notion. Labor intensity decreases significantly and at a decreasing rate in fragmentation as measured by our GIS/DEM-based cost distance the introduction of which results in loss of significance of the Simpson index and other right hand side variables. Labor intensity decreases in parcel size, distance to the homestead, and the share cultivated with tree crops or grains but increases in use of extension services and (marginally) length of possession and having experienced crop shocks.

While the above clearly point towards a negative effect of fragmentation on mean levels of yield, its negative association with farm-level crop shocks as discussed above (table 3) suggests it may still reduce risk. To address this formally, we use the square of the predicted error term from the mean yield function (see table 4 cols. 2 and 4) to estimate an IV variance function with explanatory variables identical to those used above. Results as reported in table 6 are consistent with the hypothesis that a GIS/DEM-based measure appropriately captures fragmentation. Although part of the effect seems to be captured by the Shannon crop diversity measure, these results also suggest that fragmentation reduces variability of output.

4.3 Frontier production function

Although the findings reported above support the notion that fragmentation reduces output while allowing farmers to diversify risk, a weakness of the yield function approach is its failure to control for conventional inputs. If there was a systematic link between levels of fragmentation and farmers' application of inputs (e.g., if farmers with fragmented holdings used family labor more intensively), yield regression would not provide a proper assessment of the impact of fragmentation on productivity. We thus exploit the panel nature of our data[18] to complement yield regressions with estimation of a deterministic frontier production function

[17] As unobservable attributes such as farmers' ability may affect whether or not extension advice is sought in the first place, this should not be interpreted in a causal sense.
[18] A stochastic production frontier function would have been more appealing to simultaneously estimate the production function and inefficiency parameters, but we decided against it due to the difficulty of addressing potential endogeneity of measured inputs in this framework.

to assess how land fragmentation affects efficiency of production. Using a Cobb-Douglas functional form, the parcel fixed effects production function can be expressed as:

$$\ln\left(\frac{Y_{iht}}{L_{iht}}\right) = \alpha_{ih} + \beta' \ln X_{iht} + \gamma' D_{iht} + \theta' P_{it} + T + v_{iht}, \tag{2}$$

where Y_{iht} denotes the total value of crop output on parcel i by household h at time t; L_{iht} is cultivated parcel area in hectares; X_{it} is a vector of traditional inputs including land, labor, chemical fertilizer, manure and pesticides; D_{iht} is a vector of indicator variables to account for zero values of non-labor variable inputs (Battese 1997);[19] P_{iht} is a vector of parcel characteristics such as soil type, location, topography, tenure, length of possession, water and extension service access, crop choice, seed type, incidence of crop-specific shocks;[20] T is a time dummy; α_{ih} is parcel specific fixed effects; v_{iht} is a random error to account for statistical noise distributed as $N(0, \sigma_v^2)$ and β, γ, and θ are parameters to be estimated. In this context, our estimate of technical inefficiency is defined relative to the most efficient farming household for each parcel as $u_{ih} = \alpha - \alpha_{ih}$ where $\alpha = \max(\alpha_{ih})$ and α_{ih} is the estimated fixed effects of parcel i operated by household h.

The relationship between land fragmentation and the estimated technical efficiency can thus be estimated by IV using the same identifying instruments as the yield function as follows:

$$u_{ih} = \phi' F_h + \beta \ln L_{ih} + \alpha \ln A_h + \delta' Z_h + \xi_{ih}, \tag{3}$$

where L_{ih} is average cultivated parcel area in hectares; A_h is average holding size in hectares at the household level; Z_h is a vector of indicators of managerial ability (gender, age and education of the household head, access to extension) and incidence of crop shocks; F_h is set of indices measuring different aspects of land fragmentation as in the yield function; ϕ, β, α and δ are vectors of parameters to be estimated; and ξ_{ih} is the unobserved random error term.

Although a parcel fixed effects (within) estimator of (2) effectively deals with potential correlation between input choices and unobserved parcel specific heterogeneity, it may still suffer from endogeneity bias due to correlation between unobserved time variant productivity shocks and input uses. To address this, we make use of a third round of panel data collected in early 2015. While failure to collect GPS coordinates makes the third round data unsuitable for estimation of the impact of fragmentation,[21] results from yield and input intensity functions for the three rounds in appendix tables 4 and 5 are similar to those reported earlier (table 4 and 5).

[19] Following the procedure proposed by Battese (1995), variable inputs with zero values are transformed into logarithm form as $\ln\{\mathrm{Max}(X_{it}, D_{it})\}$ where D_{it} is a vector of dummies taking a value of one for zero observations of non-labor inputs.

[20] For empirical examples on the choice of parcel characteristics and environmental conditions that are directly incorporated in the non-stochastic component of the production frontier, see Coelli *et al.* (1999) and Sherlund *et al.* (2002).

[21] A key shortcoming of the third round data is that GPS coordinates of parcels acquired since the second round were not captured. As this would bias the cost distance measure of land fragmentation, the third round data was excluded from the yield and intensity of input use function analysis reported in the previous section.

We thus use lagged input decisions as instruments for current input use and then extract parcel fixed effects to assess the relationship between land fragmentation and technical efficiency.

Coefficients on all conventional inputs except pesticides in the first-step estimations are positive and highly significant and parcel size is inversely related with yield (table 7). Second step estimates of technical inefficiency (table 8) are supportive of our earlier results in a number of ways. While standard measures of land fragmentation including number of parcels and the Simpson as well as Shannon indices are all insignificant, our GIS/DEM-based measure of travel time between the homestead and a household's parcels significantly increases technical inefficiency at a decreasing rate. In addition, both parcel and holding size and education (in some specifications) are estimated to reduce technical inefficiency while crop shocks tend to increase it and extension has no discernible effect.

5. Conclusion and implications

A nationally representative survey from Rwanda, a country characterized by high population density and a large number (4.8) of small (less than 0.2 ha each) parcels per household allows us to explore the impact of land fragmentation in an un-mechanized developing country setting. Three main findings emerge. First, a GIS/DEM-based measure of the friction-adjusted time needed to travel between all of a household's parcels consistently outperforms alternative measures of land fragmentation, suggesting that, in the setting at hand, the main channel for fragmentation to affect crop output is by increasing the amount of time spent traveling between parcels rather than working on them, an interpretation that is supported by labor intensity functions. Second, using IV techniques to adjust for potential endogeneity of fragmentation results in estimates which are more consistent than those obtained from simple OLS regressions and suggesting in particular that, in our setting, fragmentation adversely affects yield, labor intensity, and production efficiency while at the same time reducing yield variability and exposure to parcel specific crop shocks. Amalgamation of all of a household's parcels is estimated to increase yield by 7%. Programs to consolidate holdings the cost of which is in excess of the net present value of this benefit can be justified economically only if additional assumptions on future benefit streams (e.g., as a result of mechanization) are made. While this suggests that, as long as the scope for mechanization is limited, the size of fragmentation-induced yield reductions is likely to remain modest, the literature shows that, once mechanization becomes viable, the size of fragmentation-induced reductions in productivity could increase sharply (Foster and Rosenzweig 2011). Carefully exploring the relevant mechanisms and associated impacts for relatively land-abundant African settings is an area of great relevance for future research.

Table 1: Household level descriptive statistics

Variables	Total	By tercile of parcel distribution 1st	2nd	t-stat	3rd	t-stat
Holding characteristics						
Number of parcels	4.787	2.302	4.416	***	8.316	***
Simpson land fragmentation index	0.523	0.337	0.584	***	0.711	***
Shannon crop diversity index	1.137	0.983	1.159	***	1.315	***
Distance to homestead in hours (self-reported)	0.166	0.125	0.179	***	0.207	***
Minimum cost path to homestead in hours	0.105	0.075	0.112	***	0.137	***
Total cost distance in hours	0.815	0.376	0.834	***	1.367	***
Soil type diversity index	0.194	0.100	0.207	***	0.303	***
Parcel topography diversity index	0.172	0.094	0.183	***	0.265	***
Range of altitude covered in meters	986-2839	986-2799	1158-2833		1150-2839	
Total holding in hectares	0.805	0.588	0.756	***	1.129	***
Total cropped area hectares	0.606	0.528	0.565		0.742	***
Household has parcel outside the cell	0.403	0.235	0.450	***	0.580	***
Share of land located outside the cell	0.074	0.054	0.088	***	0.087	
Area share of rented-in parcels	0.150	0.150	0.173	**	0.129	***
Share of inherited land	0.538	0.548	0.552		0.514	***
Share of purchased land	0.224	0.175	0.198	**	0.311	***
Share of government allocated land	0.048	0.077	0.038	***	0.020	***
Share of rented-in land	0.150	0.150	0.173	**	0.129	***
Household-level land use						
Number of crops grown	4.507	3.602	4.534	***	5.653	***
Value of crop output per hectare (US$/ha)	551.6	588.6	526.8		524.8	
Share of land cultivated with extension input	0.104	0.108	0.101		0.100	
Share of land with crop shock	0.278	0.329	0.283	***	0.208	***
Share of total cultivated area with grains	0.501	0.521	0.496	***	0.479	*
Share of cultivated area with tubers	0.262	0.246	0.276	***	0.271	
Share of cultivated area with tree crops	0.225	0.223	0.217		0.236	**
Share of cultivated area with vegetables	0.012	0.011	0.011		0.013	
Parcel characteristics						
Parcel area in hectares	0.19	0.27	0.15	***	0.15	
Average number of years parcel possessed	18.393	19.565	17.063	***	18.016	**
Share of land located in wet land	0.082	0.068	0.080	*	0.102	***
Share of land with access to irrigation	0.056	0.054	0.058		0.056	
Share of high quality land	0.362	0.369	0.354		0.361	
Share of medium quality land	0.155	0.135	0.165	***	0.172	
Share of poor quality land	0.483	0.495	0.481		0.468	
Household characteristics						
Female headed household	0.290	0.368	0.264	***	0.211	***
Head completed only primary school	0.576	0.522	0.584	***	0.637	***
Head completed secondary education	0.066	0.045	0.064	***	0.096	***
Age of household head	46.794	48.179	45.473	***	46.134	
Members aged 14 and less	2.136	1.985	2.140	***	2.328	***
Adult members between 15 and 35	1.605	1.485	1.634	***	1.735	***
Adult members between 35 and 60	0.854	0.785	0.807		0.984	***
Members aged 60 and above	0.244	0.272	0.231	***	0.218	
Instrumental variables						
Number of inherited land	2.217	1.142	2.163	***	3.654	***
Number of government allocated land	0.167	0.180	0.185		0.134	***
Total area of inherited land in ha	0.343	0.216	0.338	***	0.510	***
Total area of government allocated land in ha	0.083	0.131	0.066	***	0.036	**
Total cost distance of inherited parcels in hours	0.441	0.201	0.444	***	0.747	***
Head was displaced	0.476	0.460	0.477		0.498	
Number of observations	6625	2720	1801		2104	

Source: Own computation from 2010/11 Rwanda LTR Household Survey

Note: Stars indicated significance levels for t-tests of the equality of means for each of the variables between subsequent terciles (*significant at 10%; **significant at 5%; *** significant at 1%).

Table 2: Comparison of distance measures for different types of parcels in minutes

	Least cost path between parcels		Least cost path from home to [ ] parcels:		Own estimate of from home to [ ] parcels:	
	Mean	Sig	Mean	Sig	Mean	Sig
Panel A: By rental market participation						
Between households						
Participants (N = 1219, 1964, 1698)	17.23		11.608		15.865	***
Non-participants (N=1792, 1816, 1822)	18.14		10.416		11.88	
Within household (only participants)[a]						
Owner-cum-tenants						
All rented (N = 970, 1290, 1548)	20.76		16.303	***	22.498	***
Owned	19.09		10.086		12.265	
Owner-cum-fixed renter						
Fixed rented (N =700, 905, 1112)	21.6		17.397	***	24.573	***
Owned	18.88		9.732		12.821	
Owner-cum-free of charge renter						
Free (N = 383, 546, 652)	18.2		13.371	**	18.47	***
Owned	18.64		10.389		11.749	
Panel B: By sales market participation						
Between households						
Participants (N=1496, 1799, 1816)	19.26		11.397	**	15.089	***
Non-participants	18.14		8.607		8.614	
Within household (only participants)[a]						
Purchased (N=878, 1315, 1469)	22.68	**	16.852	***	15.477	**
Other owned	17.26		11.128		17.605	

Source: Own computation from 2010/11 Rwanda LTR Household Survey.

N is the number of households per column in the same order; Sig refer to significance of difference between owned and the category under concern (*significant at 10%; **significant at 5%; *** significant at 1%). An empty "Sig" column implies difference is not significant at 10% or less.

[a] The least cost path for traded in (e.g., rented or purchased) refers to average distance from these parcels to owned parcels whereas for owned parcels it refers to average distance between owned parcels.

Table 3: Naïve household-level regression of having experienced crop shocks on and fragmentation and crop diversity measures

	Fragmentation		Crop diversity	
	Simpson	**No. of parcels**	**Shannon**	**No. of crops**
Fragmentation-related measures				
Simpson land fragmentation index	-0.044**			
	(2.294)			
Log of number of parcels		-0.025***		
		(3.171)		
Shannon crop diversity measure			-0.021**	
			(2.129)	
Log of number crops				-0.016
				(1.536)
Coefficient of variation of parcel altitude	0.170	0.245	0.124	0.128
	(0.906)	(1.284)	(0.667)	(0.684)
Average distance parcel- homestead (min)	-0.000	0.000	-0.000	-0.000
	(0.286)	(0.043)	(0.597)	(0.518)
Parcel characteristics				
Log of total cropped area in hectares	-0.005	-0.002	-0.003	-0.003
	(1.509)	(0.683)	(0.924)	(0.728)
Soil type diversity index	0.017	0.018	0.010	0.010
	(0.833)	(0.910)	(0.532)	(0.503)
Parcel topography diversity index	0.003	0.003	-0.004	-0.005
	(0.159)	(0.155)	(0.214)	(0.248)
Household characteristics				
Age of household head	0.004**	0.004**	0.004**	0.004**
	(2.460)	(2.419)	(2.422)	(2.428)
Age of household head squared	-0.000**	-0.000**	-0.000**	-0.000**
	(2.202)	(2.155)	(2.092)	(2.114)
Female headed household	0.012	0.012	0.013	0.013
	(1.229)	(1.225)	(1.385)	(1.381)
Head completed only primary school	0.007	0.008	0.007	0.007
	(0.871)	(0.963)	(0.793)	(0.788)
Head completed secondary education	0.006	0.008	0.005	0.004
	(0.408)	(0.487)	(0.288)	(0.285)
Members aged 14 and less	-0.003	-0.003	-0.003	-0.003
	(1.110)	(0.988)	(1.090)	(1.096)
Adult members 15 -35	-0.007**	-0.007**	-0.008**	-0.008**
	(2.307)	(2.161)	(2.444)	(2.427)
Adult members 35 - 60	-0.019**	-0.018**	-0.019**	-0.019**
	(2.576)	(2.481)	(2.545)	(2.556)
Members aged 60 and above	-0.019	-0.018	-0.019*	-0.019
	(1.584)	(1.586)	(1.665)	(1.636)
Constant	0.118***	0.131***	0.126***	0.126***
	(2.721)	(2.996)	(2.873)	(2.790)
Number of observations	6,708	6,719	6,719	6,719
R-squared	0.006	0.007	0.006	0.006

Note: Dependent variable is the share of a household's area having been subject to crop shocks, aggregated from parcel-level information as explained in the text.

Absolute value of t-statics in parenthesis: *** significant at 1%; ** significant at 5%;* significant at 10%. Time varying village specific fixed effects are included, but not reported.

Table 4: Parcel level yield function – IV estimation

	(1)	(2)	(3)	(4)
Number of parcels	-0.028***	-0.043		
	(2.897)	(1.149)		
Simpson land fragmentation index			-0.493**	0.455
			(2.561)	(0.865)
Shannon crop diversity measure		0.888		-0.519
		(0.971)		(0.691)
Total cost distance in hours		-0.521**		-0.594**
		(1.974)		(2.208)
Total cost distance squared in hours		0.096*		0.115**
		(1.850)		(2.144)
Household has parcel outside the cell	0.090***	0.076**	0.092***	0.032
	(3.645)	(2.230)	(3.588)	(0.952)
Ln parcel area in hectares	-0.618***	-0.648***	-0.607***	-0.618***
	(65.352)	(23.687)	(80.600)	(39.583)
Ln holding size in hectares	0.052***	0.082***	0.018	0.096***
	(2.826)	(3.133)	(1.517)	(2.977)
Distance from homestead to parcel in hours	-0.130***	-0.113***	-0.128***	-0.103***
	(5.281)	(3.972)	(5.171)	(3.572)
Rented in parcel	0.001	-0.002	0.009	-0.021
	(0.038)	(0.088)	(0.369)	(0.722)
Number of years parcel possessed	0.005***	0.004***	0.005***	0.005***
	(6.096)	(3.291)	(5.733)	(3.862)
Parcel located in wet land	0.038	0.011	0.039	0.046
	(1.323)	(0.278)	(1.365)	(1.235)
Irrigated land	0.086**	0.139**	0.087**	0.079
	(2.111)	(2.274)	(2.172)	(1.344)
Good soil quality	0.098***	0.088***	0.094***	0.098***
	(4.794)	(3.939)	(4.595)	(4.033)
Medium soil quality	0.034	0.032	0.033	0.032
	(1.348)	(1.187)	(1.323)	(1.164)
Use knowledge from extension services	0.171***	0.172***	0.175***	0.156***
	(5.516)	(5.078)	(5.653)	(4.427)
Crop shock	-0.235***	-0.276***	-0.240***	-0.207***
	(10.701)	(5.774)	(10.764)	(4.507)
Share of parcel with grains	-0.094	0.039	-0.098	-0.145
	(1.020)	(0.255)	(1.073)	(1.050)
Share of parcel with tubers	-0.040	0.033	-0.041	-0.055
	(0.432)	(0.300)	(0.446)	(0.509)
Share of parcel with tree crops	0.255***	0.338***	0.246***	0.234**
	(2.687)	(2.818)	(2.590)	(2.080)
Number of observations	19,671	19,671	19,671	19,671
R-squared	0.453	0.408	0.455	0.382
Hansen J statistic (overidentification test)	6.835	3.224	8.390	3.117
Chi-sq P-val	0.336	0.358	0.211	0.374

Notes: Dependent variable is log of output value per ha. Village-time fixed effects and household level characteristics (sex, age and education level of the head the household, and member composition of the household) are included and standard errors are clustered at the household level. Instruments for the potentially endogenous variables of number of parcels or the Simpson land fragmentation index, total cost distance and its squared in hours and the Shannon crop diversity measure include: number of inherited parcels, number of government allocated parcels, size of inherited land in hectares, size of government allocated land in hectares, total cost distance and its squared term computed using only inherited and government allocated parcels, and whether head was displaced in the past. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

Table 5: Parcel level intensity of labor use function – IV estimation

	(1)	(2)	(3)	(4)
Number of parcels	-0.030***	-0.087**		
	(3.487)	(1.992)		
Simpson land fragmentation index			-0.584***	-0.765
			(3.343)	(1.514)
Shannon crop diversity measure		1.914*		0.766
		(1.810)		(1.149)
Total cost distance in hours		-0.509*		-0.615***
		(1.957)		(2.826)
Total cost distance squared in hours		0.090*		0.116***
		(1.779)		(2.578)
Household has parcel outside the cell	0.043**	0.051	0.048**	0.032
	(2.016)	(1.323)	(2.182)	(1.048)
Ln parcel area in hectares	-0.652***	-0.707***	-0.640***	-0.665***
	(77.250)	(22.204)	(97.709)	(50.083)
Ln holding size in hectares	-0.028*	0.006	-0.065***	-0.039
	(1.678)	(0.195)	(6.115)	(1.336)
Distance from homestead to parcel in hours	-0.101***	-0.094***	-0.098***	-0.076***
	(4.600)	(2.769)	(4.462)	(2.895)
Rented in parcel	0.022	0.028	0.030	0.026
	(1.107)	(0.995)	(1.536)	(1.060)
Number of years parcel possessed	0.003***	0.002	0.003***	0.002*
	(4.848)	(1.186)	(4.420)	(1.649)
Parcel located in wet land	0.002	-0.051	0.003	-0.022
	(0.081)	(1.259)	(0.123)	(0.709)
Irrigated land	-0.011	0.088	-0.010	0.039
	(0.323)	(1.289)	(0.311)	(0.789)
Good soil quality	0.042**	0.026	0.037**	0.025
	(2.361)	(1.016)	(2.108)	(1.177)
Medium soil quality	-0.008	-0.010	-0.009	-0.011
	(0.373)	(0.358)	(0.418)	(0.473)
Use knowledge from extension services	0.103***	0.112***	0.107***	0.107***
	(3.863)	(3.171)	(4.095)	(3.762)
Crop shock	0.117***	0.029	0.110***	0.071*
	(6.471)	(0.557)	(5.991)	(1.782)
Share of parcel with grains	-0.384***	-0.115	-0.389***	-0.269**
	(6.368)	(0.717)	(6.480)	(2.487)
Share of parcel with tubers	-0.100*	0.037	-0.101*	-0.032
	(1.653)	(0.359)	(1.676)	(0.413)
Share of parcel with tree crops	-0.307***	-0.145	-0.316***	-0.242***
	(4.746)	(1.238)	(4.940)	(2.895)
Number of observations	19,671	19,671	19,671	19,671
R-squared	0.586	0.377	0.589	0.548
Hansen J statistic (overidentification test)	11.256	1.763	13.630	8.541
Chi-sq P-val	0.081	0.623	0.034	0.036

Notes: Dependent variable is log of total labor days per ha spent on land preparation and planting, field management and input application and harvesting. Village-time fixed effects and household level characteristics (sex, age and education level of the head the household, and member composition of the household) are included and standard errors are clustered at the household level. Instruments for the potentially endogenous variables of number of parcels or the Simpson land fragmentation index, total cost distance and its squared in hours and the Shannon crop diversity measure include: number of inherited parcels, number of government allocated parcels, size of inherited land in hectares, size of government allocated land in hectares, total cost distance and its squared term computed using only inherited and government allocated parcels, and whether head was displaced in the past. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

Table 6: Parcel level estimates of yield variance parameters – IV estimation

	(1)	(2)	(3)	(4)
Number of parcels	0.088*	0.034		
	(1.711)	(1.589)		
Simpson land fragmentation index			0.770	0.481
			(1.176)	(1.455)
Shannon crop diversity measure	-1.515		-0.406	
	(1.215)		(0.389)	
Total cost distance in hours	-0.405	-0.585**	-0.518	-0.600**
	(1.208)	(2.065)	(1.477)	(2.336)
Total cost distance squared in hours	0.119*	0.147***	0.157**	0.172***
	(1.862)	(2.602)	(2.199)	(2.956)
Household has parcel outside the cell	-0.083	-0.042	-0.086*	-0.073*
	(1.402)	(0.985)	(1.752)	(1.755)
Ln parcel area in hectares	0.026	-0.015	0.031	0.024
	(0.640)	(0.879)	(1.281)	(1.475)
Ln holding size in hectares	0.049	0.061*	0.053	0.042
	(1.248)	(1.765)	(1.230)	(1.248)
Distance from homestead to parcel in hours	0.018	0.010	-0.007	-0.005
	(0.358)	(0.233)	(0.145)	(0.114)
Rented in parcel	0.062	0.058	0.028	0.030
	(1.322)	(1.312)	(0.618)	(0.652)
Number of years parcel possessed	0.003	0.001	0.003	0.002
	(1.367)	(0.832)	(1.412)	(1.521)
Parcel located in wet land	0.129*	0.086	0.074	0.063
	(1.901)	(1.508)	(1.145)	(1.091)
Irrigated land	-0.069	0.004	-0.012	0.008
	(0.783)	(0.074)	(0.146)	(0.127)
Good soil quality	0.009	-0.002	0.042	0.036
	(0.248)	(0.062)	(1.145)	(1.070)
Medium soil quality	-0.016	-0.021	0.033	0.031
	(0.361)	(0.519)	(0.751)	(0.719)
Use knowledge from extension services	0.094	0.111*	0.067	0.073
	(1.493)	(1.944)	(1.109)	(1.246)
Crop shock	0.212***	0.149***	0.237***	0.216***
	(3.333)	(4.017)	(3.662)	(5.482)
Share of parcel with grains	-1.208***	-0.976***	-1.119***	-1.059***
	(4.255)	(4.957)	(4.464)	(5.439)
Share of parcel with tubers	-0.833***	-0.717***	-0.848***	-0.817***
	(3.663)	(3.635)	(3.970)	(4.191)
Share of parcel with tree crops	-0.597**	-0.469**	-0.593***	-0.562***
	(2.517)	(2.308)	(2.724)	(2.781)
Number of observations	19,671	19,671	19,671	19,671
R-squared	-0.006	0.055	0.090	0.091
Hansen J statistic (overidentification test)	3.629	5.835	5.420	5.333
Chi-sq P-val	0.304	0.212	0.144	0.255

Notes: Dependent variable is variance of yield as explained in more detail in the text. Village-time fixed effects and household level characteristics (sex, age and education level of the head the household, and member composition of the household) are included and standard errors are clustered at the household level. Instruments for the potentially endogenous variables of number of parcels or the Simpson land fragmentation index, total cost distance and its squared in hours and the Shannon crop diversity measure include: number of inherited parcels, number of government allocated parcels, size of inherited land in hectares, size of government allocated land in hectares, total cost distance and its squared term computed using only inherited and government allocated parcels, and whether head was displaced in the past. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

Table 7: Instrumented (IV) parcel fixed effects estimates of production function

	(1)	(2)
Ln parcel area in hectares	-0.900***	-0.902***
	(58.069)	(58.423)
Ln of labor use in days	0.201***	0.198***
	(6.355)	(6.196)
Ln of chemical fertilizer use in kg	0.430***	0.418***
	(5.083)	(4.951)
Ln of pesticide use in USD	-0.182	-0.173
	(1.577)	(1.504)
Ln of manure use in kg	0.115***	0.110***
	(4.408)	(4.251)
Dummy chemical fertilizer use	0.592***	0.560***
	(2.851)	(2.680)
Dummy pesticide use	-0.206	-0.189
	(1.069)	(0.960)
Dummy manure use	0.646***	0.611***
	(4.518)	(4.265)
Use knowledge from extension services	0.003	0.016
	(0.061)	(0.322)
Crop shock	-0.278***	-0.268***
	(9.144)	(8.863)
Household has parcel outside the cell	-0.094**	-0.090**
	(2.279)	(2.186)
Time dummy	0.184***	0.196***
	(6.898)	(7.343)
Share of parcel with grains		-0.108
		(0.767)
Share of parcel with tubers		0.061
		(0.431)
Share of parcel with tree crops		0.236
		(1.593)
Constant	1.352***	1.357***
	(4.118)	(4.254)
Number of observations	15,258	15,258
R-squared	0.418	0.422

Notes: Dependent variable is value of crop output per ha at parcel level for three survey rounds as explained in the text. Parcel level fixed effects are included throughout. Endogenous inputs (labor, chemical fertilizer and pesticide use) are instrumented using their lagged values. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

Table 8: Results from two-step IV estimation of technical inefficiency

	(1)	(2)	(3)	(4)
Number of parcels	0.072		0.067	
	(1.584)		(1.520)	
Simpson land fragmentation index		-0.030		-0.071
		(0.073)		(0.170)
Shannon crop diversity measure	-1.683	0.105	-1.556	0.165
	(1.492)	(0.221)	(1.411)	(0.349)
Total cost distance in hours	0.124***	0.104***	0.122***	0.103***
	(2.869)	(2.933)	(2.878)	(2.911)
Total cost distance squared in hours	-0.019***	-0.017***	-0.019***	-0.017***
	(2.983)	(3.142)	(3.055)	(3.181)
Ln parcel area in hectares	-0.115***	-0.153***	-0.115***	-0.151***
	(4.571)	(18.480)	(4.682)	(18.305)
Ln holding size in hectares	-0.082***	-0.058***	-0.083***	-0.061***
	(3.553)	(2.895)	(3.655)	(3.071)
Use knowledge from extension services	-0.055	-0.045	-0.071	-0.062
	(1.186)	(1.155)	(1.563)	(1.569)
Crop shock	0.100**	0.039	0.088*	0.029
	(2.050)	(1.162)	(1.846)	(0.860)
Age of household head	0.002	-0.002	0.002	-0.002
	(0.762)	(1.413)	(0.623)	(1.611)
Female headed household	0.134***	0.147***	0.127***	0.139***
	(4.756)	(6.277)	(4.624)	(5.953)
Head completed only primary school	-0.017	-0.055**	-0.024	-0.060***
	(0.479)	(2.410)	(0.685)	(2.620)
Head completed secondary education	-0.026	-0.127***	-0.032	-0.128***
	(0.328)	(2.927)	(0.414)	(2.958)
Number of observations	9,850	9,850	9,850	9,850
Centered R-squared	-0.267	0.070	-0.224	0.065
Hansen J statistic (overidentification test)	2.697	7.089	2.789	6.647
Chi-sq P-val	0.441	0.069	0.425	0.084

Notes: The dependent variables are estimates of technical inefficiency using fixed effects' estimators for which the estimated parameters of the production function, that are consistent irrespective of whether inputs are correlated with household level unobservables or not, are reported in appendix table 3. In this setting, the technical inefficiency estimator is defined relative to the most efficient farming household as: $u_i = \alpha - \alpha_i$ where $\alpha = \max(\alpha_i)$ and α_i is the estimated fixed effects of household *i*. Columns 1-2 and 3-4 are based on estimates of technical efficiency from columns 1 and 2, respectively, of table A8.
Village specific fixed effects are included, but not reported. Instruments for the endogenous variables of number of parcels or the Simpson land fragmentation index, total cost distance and its squared in hours include: number of inherited parcels, number of government allocated parcels, size of inherited land in hectares, size of government allocated land in hectares, total cost distance and its squared computed using only inherited and government allocated parcels, and whether head was displaced in the past. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

Appendix table 1: Naive parcel level yield and intensity of labor use functions: Household random effects (RE) estimates

	Ln value of crop output per ha		Ln number of labor days per ha	
Number of parcels	0.011***		-0.005	
	(2.696)		(1.264)	
Simpson land fragmentation index		-0.071		-0.222***
		(1.437)		(4.948)
Ln parcel area in hectares	-0.586***	-0.591***	-0.631***	-0.632***
	(82.894)	(86.056)	(105.855)	(108.122)
Ln holding size in hectares	0.031***	0.048***	-0.038***	-0.038***
	(2.618)	(4.552)	(3.585)	(4.001)
Constant	4.590***	4.654***	4.462***	4.538***
	(17.889)	(18.175)	(32.826)	(32.018)
Number of observations	20,310	20,310	20,310	20,310
R-squared	0.361	0.361	0.501	0.501

Note: Village-time fixed effects are included and standard errors are clustered at the household level. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

Appendix table 2: RE estimates of yield function controlling for parcel and household level characteristics

	(1)	(2)	(3)	(4)
Number of parcels	0.005	-0.001		
	(1.226)	(0.255)		
Simpson land fragmentation index			-0.133***	-0.292***
			(2.632)	(5.368)
Shannon crop diversity measure		0.297***		0.320***
		(11.742)		(12.537)
Total cost distance in hours		-0.035		0.004
		(1.268)		(0.142)
Total cost distance squared in hours		0.007		0.002
		(1.252)		(0.347)
Household has parcel outside the cell	0.036	0.037*	0.050**	0.056**
	(1.627)	(1.683)	(2.272)	(2.543)
Ln parcel area in hectares	-0.598***	-0.600***	-0.601***	-0.602***
	(81.295)	(81.426)	(84.195)	(83.721)
Ln holding size in hectares	0.012	0.003	0.021*	0.001
	(1.022)	(0.285)	(1.855)	(0.066)
Distance from homestead to parcel in hours	-0.146***	-0.149***	-0.142***	-0.147***
	(5.961)	(6.119)	(5.833)	(6.048)
Rented in parcel	0.017	0.019	0.017	0.020
	(0.751)	(0.825)	(0.721)	(0.880)
Number of years parcel possessed	0.006***	0.006***	0.006***	0.006***
	(7.266)	(7.171)	(7.105)	(6.941)
Parcel located in wet land	0.046*	0.041	0.045*	0.040
	(1.674)	(1.501)	(1.662)	(1.484)
Irrigated land	0.089**	0.099***	0.089**	0.099***
	(2.319)	(2.578)	(2.318)	(2.594)
Good soil quality	0.068***	0.068***	0.068***	0.067***
	(3.533)	(3.514)	(3.508)	(3.487)
Medium soil quality	0.019	0.019	0.019	0.019
	(0.798)	(0.779)	(0.811)	(0.803)
Use knowledge from extension services	0.142***	0.144***	0.143***	0.147***
	(4.638)	(4.725)	(4.669)	(4.816)
Crop shock	-0.206***	-0.216***	-0.208***	-0.220***
	(9.570)	(10.077)	(9.675)	(10.277)
Share of parcel with grains	-0.015	0.012	-0.013	0.016
	(0.165)	(0.135)	(0.140)	(0.174)
Share of parcel with tubers	0.018	0.030	0.020	0.033
	(0.197)	(0.340)	(0.220)	(0.372)
Share of parcel with tree crops	0.316***	0.328***	0.319***	0.331***
	(3.430)	(3.573)	(3.454)	(3.597)
Constant	4.208***	3.879***	4.250***	3.934***
	(13.802)	(12.488)	(14.052)	(12.851)
Number of observations	19,755	19,755	19,755	19,755
R-squared	0.376	0.380	0.376	0.379

Note: Village-time fixed effects and household level characteristics (sex, age and education level of the head the household, and member composition of the household) are included and standard errors are clustered at the household level. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

Appendix table 3: RE estimates of intensity of labor use function controlling for parcel and household level characteristics

	(1)	(2)	(3)	(4)
Number of parcels	-0.006*	-0.014***		
	(1.647)	(3.420)		
Simpson land fragmentation index			-0.239***	-0.434***
			(5.248)	(9.025)
Shannon crop diversity measure		0.385***		0.409***
		(17.135)		(18.132)
Total cost distance in hours		-0.032		-0.001
		(1.236)		(0.035)
Total cost distance squared in hours		0.005		0.001
		(1.058)		(0.264)
Household has parcel outside the cell	0.026	0.027	0.036*	0.041**
	(1.264)	(1.380)	(1.775)	(2.122)
Ln parcel area in hectares	-0.631***	-0.635***	-0.632***	-0.634***
	(101.900)	(102.384)	(103.996)	(104.174)
Ln holding size in hectares	-0.060***	-0.073***	-0.063***	-0.087***
	(5.456)	(6.616)	(6.243)	(8.443)
Distance from homestead to parcel in hours	-0.117***	-0.122***	-0.114***	-0.119***
	(5.692)	(5.873)	(5.562)	(5.769)
Rented in parcel	0.031	0.033*	0.032*	0.036*
	(1.626)	(1.749)	(1.670)	(1.907)
Number of years parcel possessed	0.005***	0.005***	0.005***	0.005***
	(7.644)	(7.490)	(7.460)	(7.181)
Parcel located in wet land	-0.018	-0.023	-0.018	-0.022
	(0.835)	(1.052)	(0.833)	(1.044)
Irrigated land	0.017	0.028	0.017	0.028
	(0.549)	(0.884)	(0.537)	(0.876)
Good soil quality	0.019	0.019	0.018	0.017
	(1.181)	(1.162)	(1.112)	(1.066)
Medium soil quality	-0.000	-0.001	-0.000	-0.002
	(0.004)	(0.067)	(0.006)	(0.075)
Use knowledge from extension services	0.087***	0.090***	0.088***	0.093***
	(3.408)	(3.599)	(3.467)	(3.740)
Crop shock	0.116***	0.105***	0.114***	0.101***
	(6.831)	(6.285)	(6.711)	(6.059)
Share of parcel with grains	-0.366***	-0.333***	-0.365***	-0.330***
	(6.211)	(5.690)	(6.197)	(5.642)
Share of parcel with tubers	-0.096	-0.080	-0.095	-0.077
	(1.632)	(1.370)	(1.606)	(1.321)
Share of parcel with tree crops	-0.301***	-0.286***	-0.300***	-0.285***
	(4.776)	(4.569)	(4.765)	(4.554)
Constant	4.376***	3.959***	4.435***	4.037***
	(23.694)	(21.481)	(23.535)	(21.608)
Number of observations	19,755	19,755	19,755	19,755
R-squared	0.511	0.519	0.511	0.518

Note: Village-time fixed effects and household level characteristics (sex, age and education level of the head the household, and member composition of the household) are included and standard errors are clustered at the household level. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

Appendix table 4: Parcel level yield function – IV estimation (three round data)

	(1)	(2)	(3)	(4)
Number of parcels	-0.032***	-0.060		
	(4.124)	(1.193)		
Simpson land fragmentation index			-0.624***	0.572
			(3.842)	(1.138)
Shannon crop diversity measure		0.924		-0.988*
		(0.793)		(1.658)
Total cost distance in hours		-0.445**		-0.456**
		(2.052)		(2.145)
Total cost distance squared in hours		0.084*		0.090**
		(1.960)		(2.099)
Household has parcel outside the cell	0.094***	0.086***	0.099***	0.044
	(4.238)	(2.700)	(4.297)	(1.396)
Ln parcel area in hectares	-0.660***	-0.688***	-0.649***	-0.650***
	(89.663)	(23.574)	(102.778)	(60.123)
Ln holding size in hectares	0.084***	0.107***	0.047***	0.122***
	(5.604)	(4.971)	(4.532)	(4.198)
Distance from homestead to parcel in hours	-0.111***	-0.104***	-0.106***	-0.085***
	(5.049)	(3.744)	(4.771)	(3.191)
Rented in parcel	0.015	0.008	0.026	-0.001
	(0.703)	(0.318)	(1.201)	(0.021)
Number of years parcel possessed	0.005***	0.004***	0.005***	0.005***
	(6.648)	(3.825)	(6.358)	(5.058)
Parcel located in wet land	0.009	-0.022	0.012	0.038
	(0.385)	(0.476)	(0.476)	(1.124)
Irrigated land	0.100***	0.141**	0.097***	0.074
	(2.882)	(2.518)	(2.871)	(1.565)
Good soil quality	0.085***	0.077***	0.078***	0.087***
	(4.752)	(3.941)	(4.385)	(3.988)
Medium soil quality	0.029	0.027	0.027	0.031
	(1.316)	(1.141)	(1.228)	(1.248)
Use knowledge from extension services	0.190***	0.194***	0.194***	0.186***
	(7.018)	(6.621)	(7.143)	(5.964)
Crop shock	-0.260***	-0.302***	-0.266***	-0.206***
	(14.344)	(5.220)	(14.448)	(5.410)
Share of parcel with grains	-0.113	0.042	-0.115	-0.258*
	(1.313)	(0.209)	(1.338)	(1.946)
Share of parcel with tubers	-0.015	0.070	-0.014	-0.087
	(0.169)	(0.541)	(0.158)	(0.822)
Share of parcel with tree crops	0.281***	0.366***	0.275***	0.210**
	(3.166)	(2.797)	(3.114)	(1.984)
Number of observations	28,435	28,396	28,435	28,396
R-squared	0.479	0.435	0.480	0.361
Hansen J statistic (overidentification test)	11.016	6.403	10.607	3.906
Chi-sq P-val	0.026	0.094	0.031	0.272

Note: Village-time fixed effects and household level characteristics (sex, age and education level of the head the household, and member composition of the household) are included and standard errors are clustered at the household level. Dummies are included in (2) and (4) to control for those households who transacted in or out land during the third round given the travel cost indicators for this round are from the second round due to absence of geographic coordinates for the newly acquired parcels. Instruments for the potentially endogenous variables of number of parcels or the Simpson land fragmentation index, total cost distance and its squared in hours and the Shannon crop diversity measure include: number of inherited parcels, number of government allocated parcels, size of inherited land in hectares, size of government allocated land in hectares, total cost distance and its squared term computed using only inherited and government allocated parcels, and whether head was displaced in the past. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

Appendix table 5: Parcel level intensity labor use function – IV estimation (three round data)

	(1)	(2)	(3)	(4)
Number of parcels	-0.021***	-0.102*		
	(3.209)	(1.655)		
Simpson land fragmentation index			-0.442***	-0.203
			(3.131)	(0.592)
Shannon crop diversity measure		2.164		0.043
		(1.544)		(0.099)
Total cost distance in hours		-0.499**		-0.479***
		(2.049)		(2.925)
Total cost distance squared in hours		0.091*		0.094***
		(1.900)		(2.673)
Household has parcel outside the cell	0.031*	0.038	0.035*	0.010
	(1.721)	(0.970)	(1.887)	(0.472)
Ln parcel area in hectares	-0.709***	-0.765***	-0.702***	-0.713***
	(113.006)	(21.556)	(129.379)	(81.812)
Ln holding size in hectares	0.010	0.041	-0.014	0.018
	(0.812)	(1.440)	(1.601)	(0.904)
Distance from homestead to parcel in hours	-0.060***	-0.068**	-0.056***	-0.040**
	(3.374)	(2.076)	(3.111)	(2.041)
Rented in parcel	0.028	0.021	0.036**	0.022
	(1.603)	(0.776)	(2.057)	(1.178)
Number of years parcel possessed	0.004***	0.002*	0.004***	0.003***
	(6.462)	(1.797)	(6.240)	(4.171)
Parcel located in wet land	-0.005	-0.074	-0.003	-0.006
	(0.263)	(1.432)	(0.156)	(0.253)
Irrigated land	0.025	0.112*	0.023	0.034
	(0.888)	(1.663)	(0.824)	(0.990)
Good soil quality	0.035**	0.024	0.030**	0.027*
	(2.402)	(1.014)	(2.079)	(1.645)
Medium soil quality	0.004	0.002	0.004	0.005
	(0.247)	(0.079)	(0.199)	(0.259)
Use knowledge from extension services	0.127***	0.135***	0.130***	0.131***
	(5.733)	(4.259)	(5.936)	(5.728)
Crop shock	0.092***	-0.009	0.088***	0.087***
	(6.500)	(0.140)	(6.133)	(3.175)
Share of parcel with grains	-0.299***	0.051	-0.301***	-0.282***
	(5.769)	(0.221)	(5.824)	(3.222)
Share of parcel with tubers	-0.066	0.118	-0.066	-0.053
	(1.277)	(0.878)	(1.281)	(0.809)
Share of parcel with tree crops	-0.332***	-0.145	-0.337***	-0.321***
	(6.016)	(1.051)	(6.129)	(4.869)
Number of observations	28,435	28,396	28,435	28,396
R-squared	0.636	0.357	0.637	0.617
Hansen J statistic (overidentification test)	15.068	2.959	15.287	13.952
Chi-sq P-val	0.005	0.398	0.004	0.003

Note: Village-time fixed effects and household level characteristics (sex, age and education level of the head the household, and member composition of the household) are included and standard errors are clustered at the household level. Dummies are included in (2) and (4) to control for those households who transacted in or out land during the third round given the travel cost indicators for this round are from the second round due to absence of geographic coordinates for the newly acquired parcels. Instruments for the potentially endogenous variables of number of parcels or the Simpson land fragmentation index, total cost distance and its squared in hours and the Shannon crop diversity measure include: number of inherited parcels, number of government allocated parcels, size of inherited land in hectares, size of government allocated land in hectares, total cost distance and its squared term computed using only inherited and government allocated parcels, and whether head was displaced in the past. Absolute value of t-statistics in parenthesis: *** significant at 1%; ** significant at 5%; * significant at 1%.

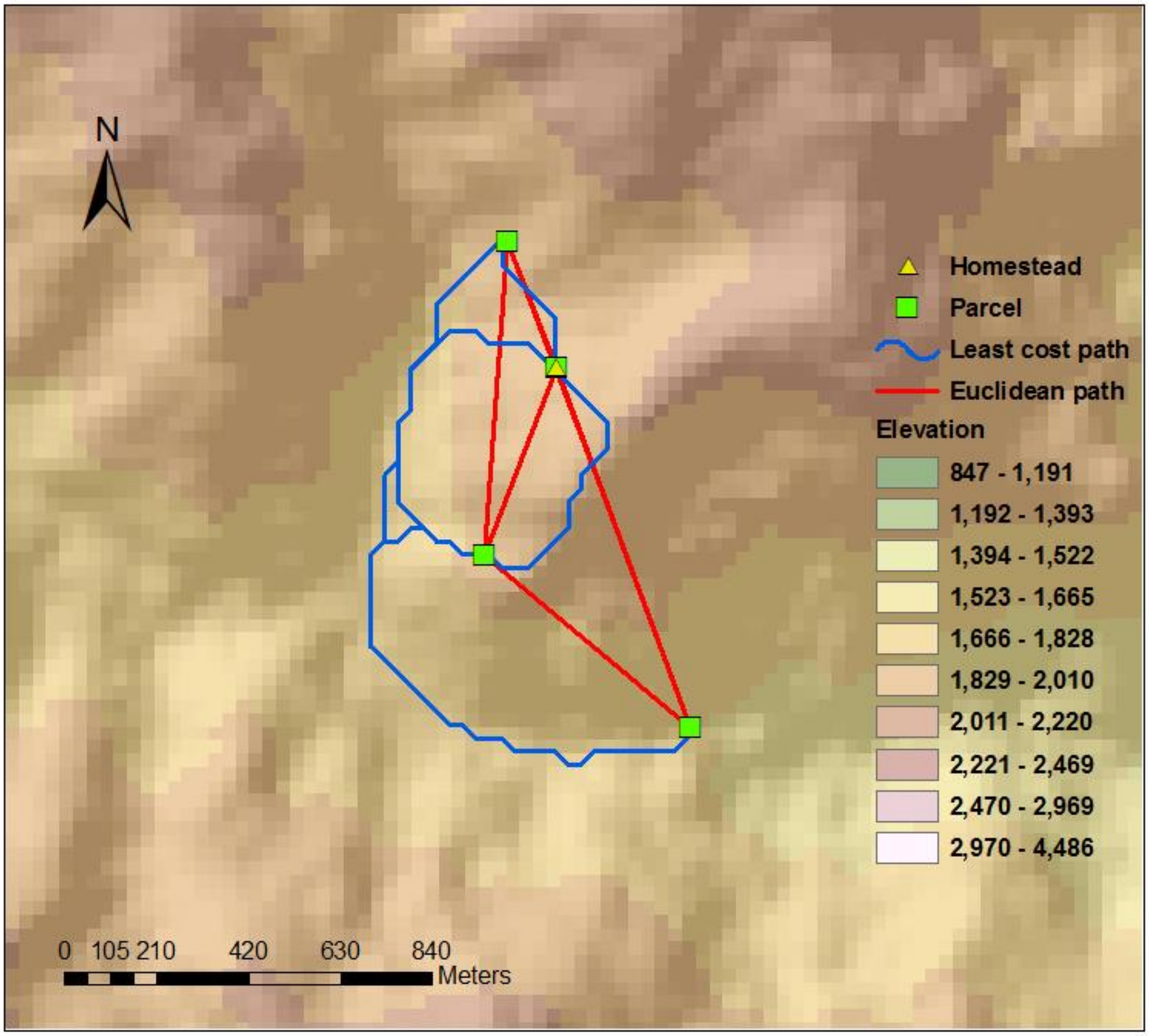


Figure 1: Pair-wise least cost path including the homestead parcel

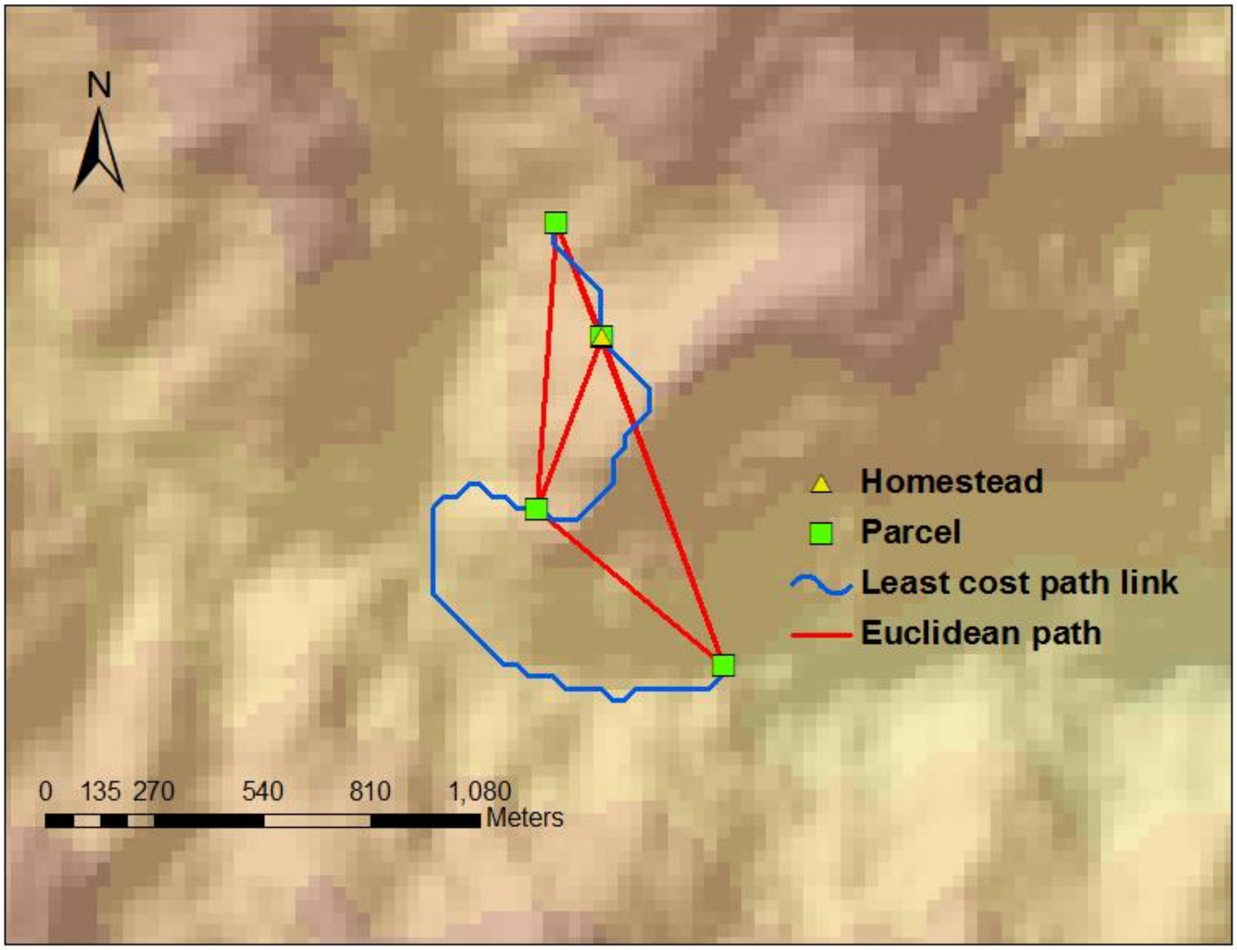


Figure 2: Least cost path link starting and ending at the homestead parcel

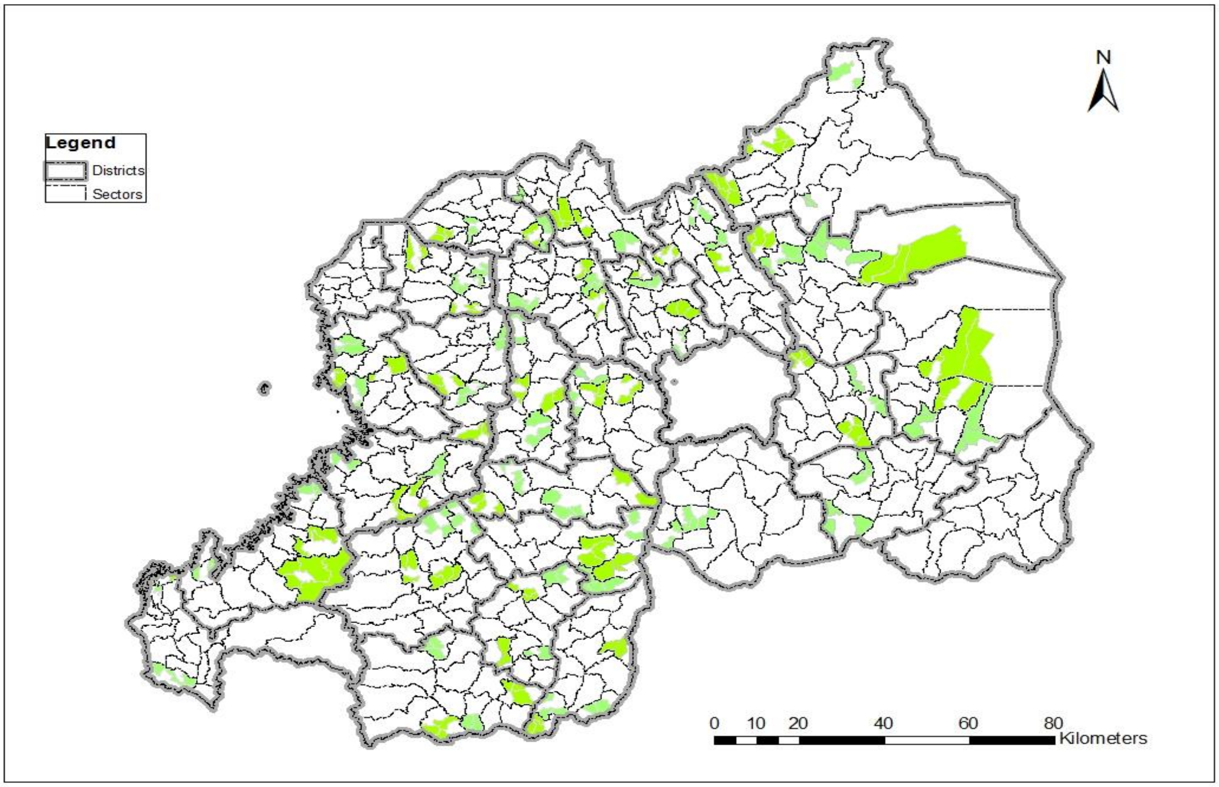
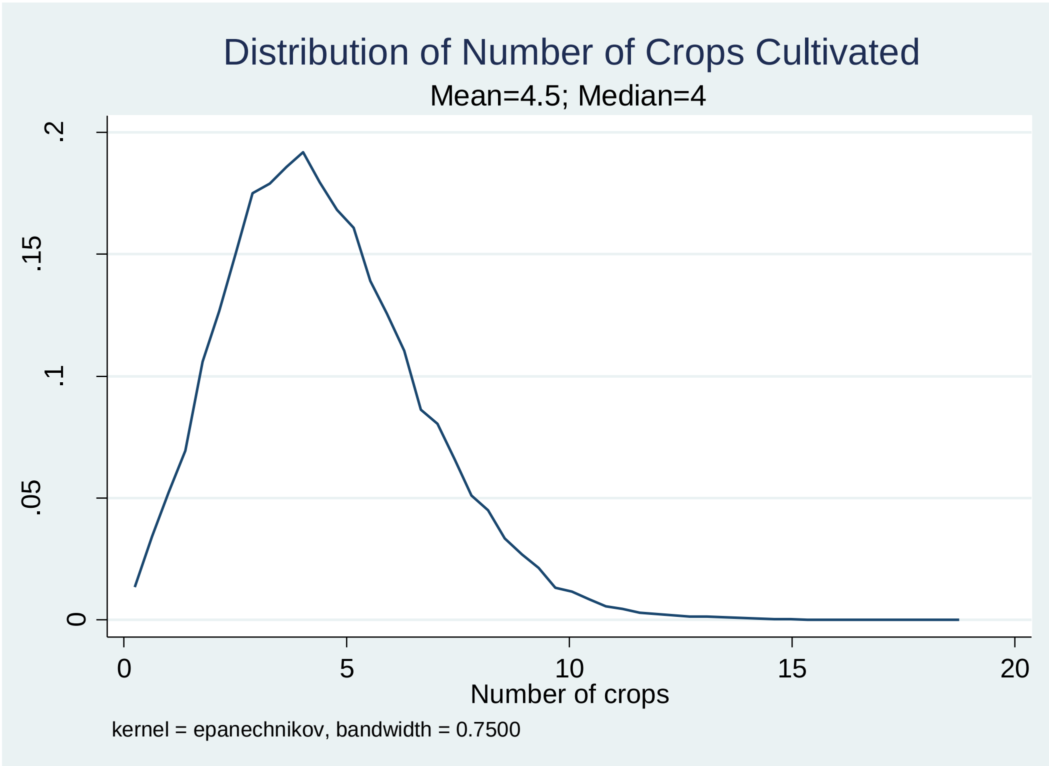
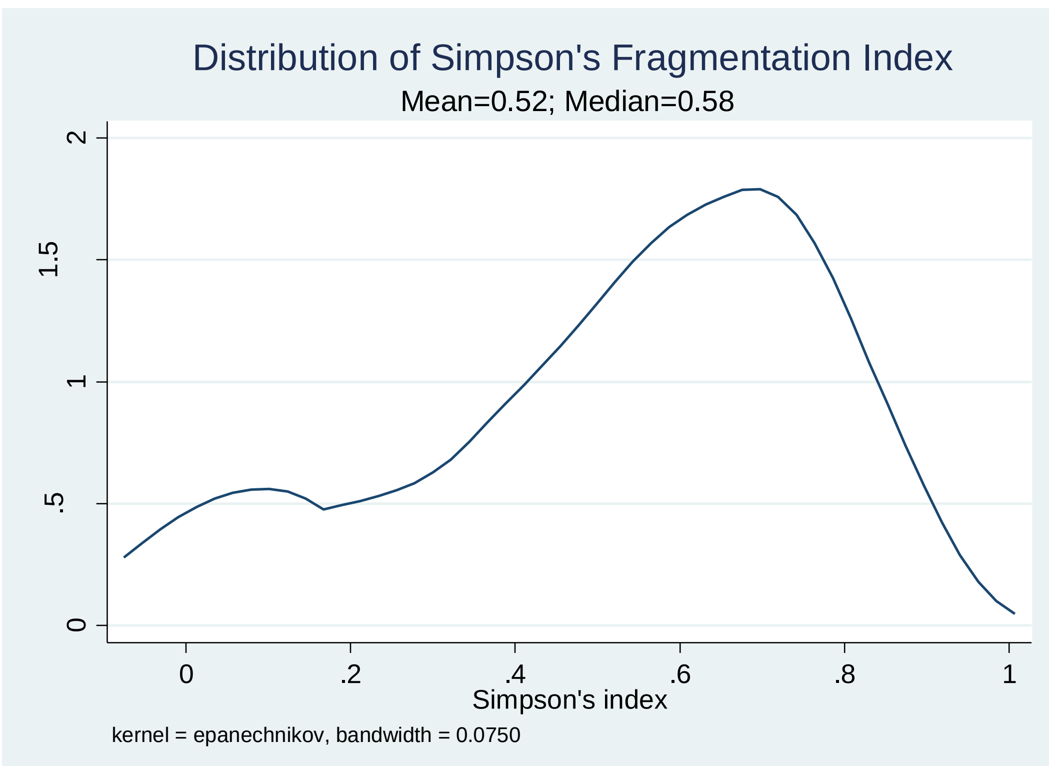


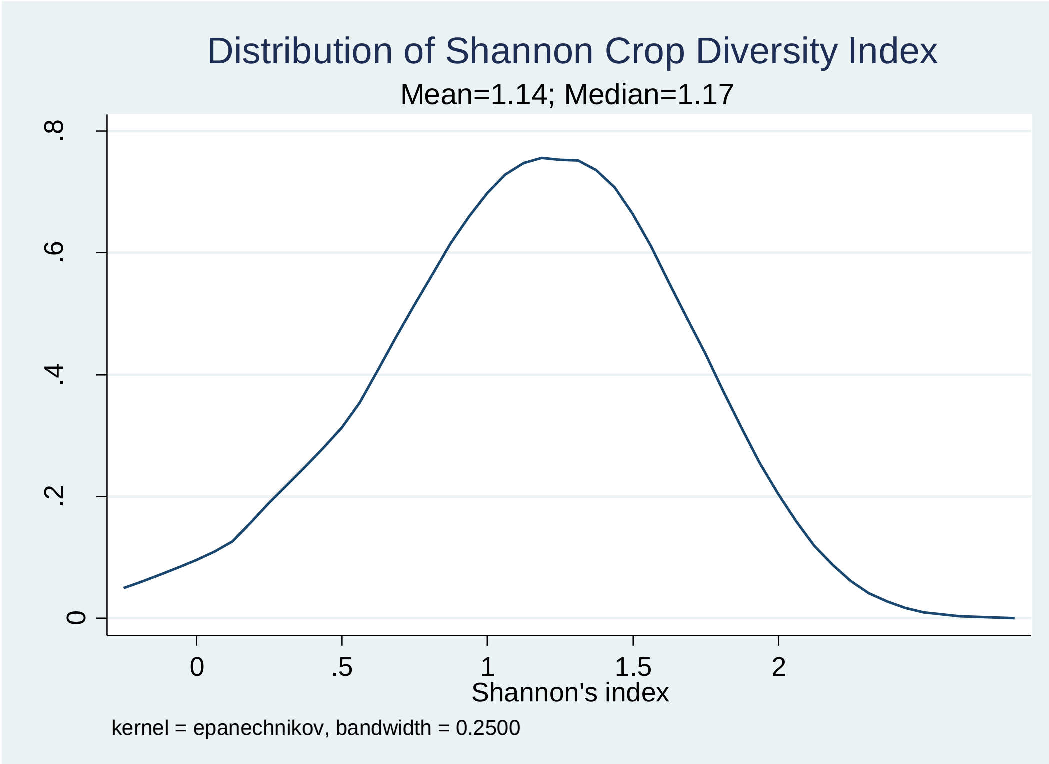
Figure 3: Map of sampled cells (300 villages, 3,600 households)



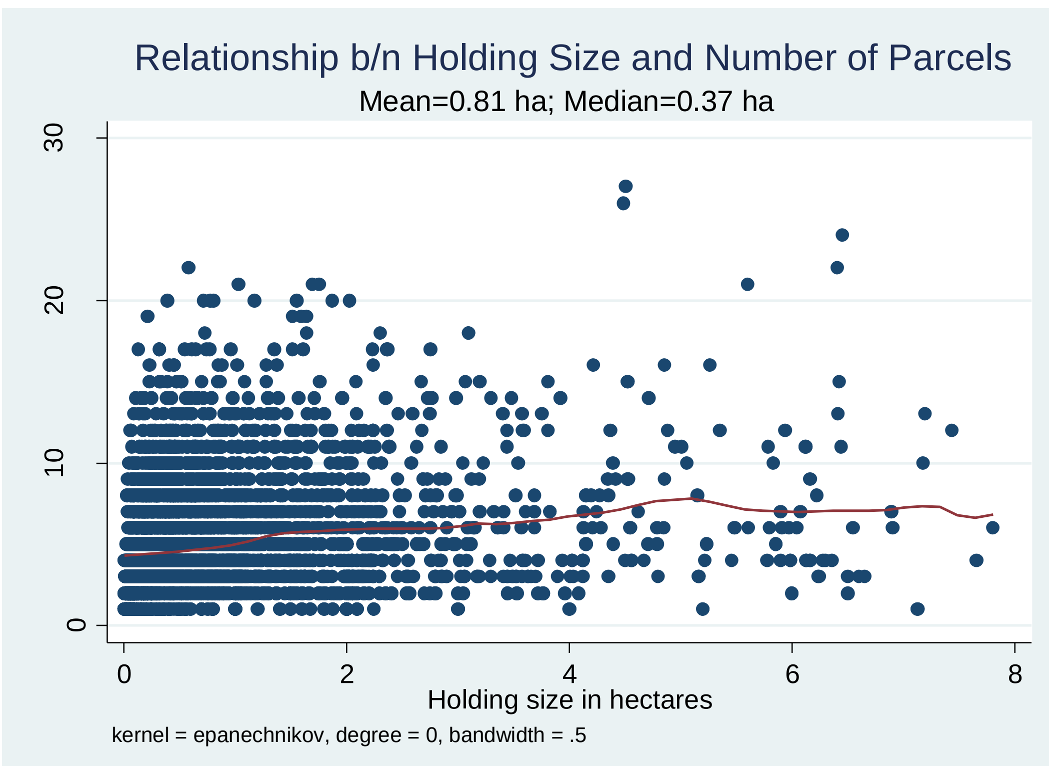
Appendix figure 1: Distribution of number of crops cultivated



Appendix figure 2: Distribution of Simpson's fragmentation index



Appendix figure 3: Distribution of Shannon crop diversity index



Appendix figure 4: Relationship of holding size and number of parcels

References

Ali, D. A. and K. Deininger. 2015. "Is There a Farm Size-Productivity Relationship in African Agriculture? Evidence from Rwanda." *Land Economics*, ***91***(2), 317-343.

Andre, C. and J. P. Platteau. 1998. "Land Relations under Unbearable Stress: Rwanda Caught in the Malthusian Trap." *Journal of Economic Behavior and Organization*, ***34***(1), 1-47.

Baker, M. and T. J. Miceli. 2005. "Land Inheritance Rules: Theory and Cross-Cultural Analysis." *Journal of Economic Behavior and Organization*, ***56***(1), 77-102.

Battese, G. E. 1997. "A Note on the Estimation of Cobb-Douglas Production Functions When Some Explanatory Variables Have Zero Values." *Journal of Agricultural Economics*, ***48***(2), 250-252.

Bentley, J. 1987. "Economic and Ecological Approaches to Land Fragmentation: In Defense of a Much Maligned Phenomenon." *Annual Review of Anthropology*, ***16***(1), 31-67.

Blarel, B., P. Hazell, F. Place and J. Quiggin. 1992. "The Economics of Farm Fragmentation: Evidence from Ghana and Rwanda." *World Bank Economic Review*, ***6***(2), 233-254.

Chen, Z., W. E. Huffman and S. Rozelle. 2009. "Farm Technology and Technical Efficiency: Evidence from Four Regions in China." *China Economic Review*, ***20***(2), 153-161.

Coelli, T., S. Perelman and E. Romano. 1999. "Accounting for Environmental Influences in Stochastic Frontier Models: With Application to International Airlines." *Journal of Productivity Analysis*, ***11***(3), 251-273.

Deininger, K., G. Carletto and S. Savastano. 2012. "Land Fragmentation, Corpland Abandonment, and Land Market Operation in Albania." *World Development*, ***40***(10), 2108-2122.

Deininger, K., D. Monchuk, H. K. Nagarajan and S. K. Singh. 2016. "Does Land Fragmentation Increase the Cost of Cultivation? Evidence from India." *Journal of Development Studies*, ***53***(11), 82-98.

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.

Fenoaltea, S. 1976. "Risk, Transaction Costs, and the Organization of Medieval Agriculture." *Explorations in Economic History*, ***13***(2), 129-151.

Foster, A. and M. R. Rosenzweig. 2011. "Are Indian Farms Too Small? Mechanization, Agency Cost, and Farm Efficiency ", Brown University, Providence, RI.

Heston, A. and D. Kumar. 1983. "The Persistence of Land Fragmentation in Peasant Agriculture: An Analysis of South Asian Cases." *Explorations in Economic History*, ***20***(2), 199-220.

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***(2), 195-211.

Imhof, E. 1950. *Gelande Und Karte.* Eugen Rentsch Verlag, Zurich.

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.

Kondylis, F. 2008. "Agricultural Outputs and Conflict Displacement: Evidence from a Policy Intervention in Rwanda." *Economic Development and Cultural Change*, ***57***(1), 31-66.

Lai, W., B. Roa and Y. Liu. 2015. "Estimating the Effect of Land Fragmentation on Mchinery Use and Crop Production." *Selected paper prepared for presentation at the 2015 agricultural and applied economics association and western agricultural economics association annual meeting*, San Francisco, CA, July 26-28.

McCloskey, D. N. 1975. *The Persistence of English Common Fields*. Princeton University Press, Princeton, N.J.

Platteau, J. P. and J.-M. Baland. 2001. "Impartial Inheritance Versus Equal Division: A Comparative Perspective Centered on Europe and Sub-Saharan Africa." in A. de Janvry, et al. (eds.), *Access to Land Rural Poverty and Public Action*, Oxford University Press, Oxford.

Rahman, S. and M. Rahman. 2008. "Impat of Land Fragmentation and Resource Ownership on Productivity and Efficiency: The Case of Rice Producers in Bangladesh." *Land Use Policy*, ***26***(1), 95-103.

Republic of Rwanda. 2004. "National Land Policy." Kigali.

Republic of Rwanda. 2009. "Strategic Plan for the Transformation of Agriculture in Rwanda Phase Ii (Psta Ii). Final Report." Ministry of Agriculture and Animal Resources, Kigali.

Sengupta, N. 2006. "Fragmented Landholding, Productivity, and Resilience Management." *Environment and Development Economics*, ***11***(4), 507-532.

Sherlund, S. M., C. B. Barrett and A. A. Adesina. 2002. "Smallholder Technical Efficiency Controlling for Environmental Production Conditions." *Journal of Development Economics*, ***69***(1), 85-101.

Simons, S. 1987. "Land Fragmentation and Consolidation: A Theoretical Model of Land Configuration with an Empirical Analysis of Fragmentation in Thailand." *Ph.D.thesis, University of Maryland, College Park*.

Tan, S. H., N. Heerink, G. Kruseman and F. T. Qu. 2008. "Do Fragmented Landholdings Have Higher Production Costs? Evidence from Rice Farmers in Northeastern Jiangxi Province, Pr China." *China Economic Review*, ***19***(3), 347-358.

Tanaka, T. 2007. "Resource Allocation with Spatial Externalities: Experiments on Land Consolidation." *B.E.Journals in Economic Analysis and Policy: Topics in Economic Analysis and Policy*, ***7***(1), 1-31.

Tobler, W. 1993. "Three Presentations on Geographical Analysis and Modeling: Non-Isotropic Geographic Modeling Speculations on the Geometry of Geography Global Spatial Analysis." *Technical Report, 93*, National center for geographic information and analysis.

Wan, G. H. and E. Cheng. 2001. "Effects of Land Fragmentation and Returns to Scale in the Chinese Farming Sector." *Applied Economics*, ***33***(2), 183-194.

Wu, Z., M. Liu and J. Davis. 2005. "Land Consolidation and Productivity in Chinese Household Crop Production." *China Economic Review*, ***16***(1), 28-49.