ECONOMICS OF LAND FRAGMENTATION: EFFECTS ON PRODUCTIVITY, TECHNICAL EFFICIENCY AND CROP DIVERSITY IN TIGRAY, NORTHERN ETHIOPIA

SELEMON ASSEFA NEGASH

NORWEGIAN UNIVERSITY OF LIFE SCIENCES
SCHOOL OF ECONOMICS AND BUSINESS
MASTER THESIS, 30 CREDITS, 2013

ECONOMICS OF LAND FRAGMENTATION:

Effects on Productivity, Technical Efficiency and Crop Diversity in Tigray, Northern Ethiopia

A Thesis Submitted in Partial Fulfillments of the Requirement for the Degree of Master of Science in Economics

Selemon Assefa Negash

THE SCHOOL OF ECONOMICS AND BUSINESS

NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Ås 2013

Declaration

I, Selemon Assefa Negash, declare that this is my own original work, and applications of all other materials are acknowledged. This thesis has not been submitted to any other University than UMB for any type of academic degree.

Signature ______________________________

Date and Place ______________________________

SELEMON ASSEFA NEGASH

Acknowledgement

No such project can be attained without the support, inspiration and encouragement of grate persons. First of all, I am very grateful to Professor Stein T. Holden, my supervisor, for his fascinating guidance, thought-provoking suggestions, encouragements and patience during my thesis writing. I had such a wonderful privilege of working with you; and I hope this is only the beginning.

I am also greatly thankful to all my family and friends who have been encouraging me during the frustrating and difficult time I have been through following my late father's death. I thank you all, again.

Finally, my especial thanks and love have to go to my wife Kidist Berie and my daughter Hermon Selemon. Nothing is possible without your extraordinary tolerance, love, support and hope; and nothing is impossible, with you on my side.

Abstract

In this study I investigated the effect of land fragmentation (LF) in Northern Ethiopia on 1) farm productivity, 2) efficiency and 3) crop diversity using stochastic production frontier (SPF) analysis and farm household model (FHHM) with factor market imperfections. The analysis is carried out at plot level mainly in a plot-panel framework using a cross-sectional sample data of 421 households and their corresponding 1918 plots. Along with "plot size" and "farm size", I used three other land fragmentation indicators that are widely used in the literature: "number of operated plots", "distance to plots" and "SI-index" (where larger index means highly fragmented).

Applying different econometric specifications, on 1) the dominant-crop model, 2) two-main-crops model and 3) aggregated-output models, I found no evidence to the conventional claim that land fragmentation could be detrimental to productivity or efficiency; in fact, the results indicate to the opposite. A non-trivial positive and significant association is observed between 1) number of plots and productivity, technical efficiency and crop diversity and 2) between SI-Index and crop diversity. A negative association is observed between farms size and productivity, technical efficiency and crop diversity. These signs of positive implications of fragmentation on productivity and efficiency can be explained by its indirect effect through diversification than by diseconomies of scale.

Signs of negative implication of land fragmentation are observed only in the dominant crop model that ignores diversification and farm integration. However, such analysis is methodologically inconsistent when farm production involves diversification, since higher fragmentation can counterbalance the negative impact through diversification.

Table of Contents

List of Tables

List of Figures

1. INTRODUCTION

Land is the major source of wealth and livelihood in rural Ethiopia and land policy is often regarded as part of welfare and food security policy (Haile et al., 2005). Population growth and the growing number of landless people have been the major challenges to policy makers of the country (Ghebru and Holden, 2008, Haile et al., 2005, Holden and Yohannes, 2002). In the past, policy makers have reacted to these challenges through frequent land reforms that focused mainly on egalitarian land redistribution to protect people and enhance food self-sufficiency, using land distribution as a safety net (Holden and Yohannes, 2002). Consequently, the per capita land holding, particularly, in the highlands of Ethiopia has considerably contracted (Deininger et al., 2006, Haile et al., 2005, Holden and Yohannes, 2002). Whether this contraction resulted in declining productivity per unit area is linked to the *long-standing* inverse-productivity debate (Cotula, 2009).

Recently, however, the Government of Ethiopia (GOE) emphasized on introducing large scale farms in parts of the country not only because it is believed to promote economic growth but also because it is perceived as an apt to structural transformation of the economy.[1] Clearly, the government favors transforming the small scale "traditional" agriculture by opening investment doors to the *emerging* economies to get the most out of their technologies (Cotula, 2009, MoFED, 2010). This favoring of the GOE seems to be in line with the recent argument made by Collier (2008) who brought the debate on scale farming back to stage. Collier (2008) argued for introducing large scale farming in Africa as a way to create development. While small farmers are believed by many as "poor but efficient" (Schultz, 1964), technology is believed to change this relationship. So seems the motive for the GOE to embark on large scale investment on "green infrastructures" with the vision to transform the agricultural sector as well as the fundamental structure of the economy (MoFED, 2010).

But the question remains, what led the GOE to radically switch its policy from emphasizing on small scale agriculture, for nearly two decades, to promoting large scale farming? Are there economies of scale in the agricultural sector in Ethiopia? Are small farmers who are operating on fragmented plots performing poorly? Could this be the reason for opening up the large scale,

[1]The Key Note addressed by the late PM. Meles Zenawi (2011) at the opening of the 6th African Economic Conference on "Green Economy and Structural Transformation in Africa" made it explicit that the *only* way to sustain structural transformation in Africa is through "Green Economy". Online resource, accessed 10.05.2013. http://www.africaneconomicconference.org/2011/updates/Speech%20by%20HE%20Meles%20Zenawi.pdf

nearly, land sales market?[2] Since the strength of land as a safety net is eroding with the rapidly growing population, perhaps the GOE is seeking to provide alternative employment opportunities to the young (landless) population. Yet, the issue of economies of scale must be implicit to this view as well, since large scale farming is favored over entitling the young to land.

With Constant Returns to Scale and proper functioning of the factor markets, land fragmentation, I would argue, should not create any inefficiency. In this study, I investigated the performance of small scale farmers operating on fragmented plots, in the Ethiopian Highlands, specifically in Tigray region. Tigray is a typical region characterized by traditional small scale farming that has persisted for centuries. That makes it conducive for investigating productivity and technical efficiency of small farmers and to assess whether there economies of scale in the sector.

The average land holding in this region is among the lowest in the country, roughly estimated to be around 1 ha per household (Ghebru and Holden, 2008, Holden et al., 2011, Segers et al., 2010). According to the statistics from Tigray Regional Plan and Finance Bureau (2011), more than 80 % of the regional population, relies on cultivating 1.23 million ha of arable land most of which are highly degraded and fragmented. Growing at 2.5% (CSA 2007), the increase in population size in this region further escalates the demand for more land despite the very little scope to increase the supply of land. Land, as a state property, has been periodically redistributed by the government in order to equitably accommodate the landless population (Ghebru and Holden, 2008, Holden et al., 2011). Consequently, farm households possess small and fragmented plots with varying degrees of fertility, size, and distance from homestead (Beyene et al., 2006); and presumably farm productivity might be constrained by the pronounced degree of land fragmentation (Rahman and Rahman, 2009) or possibly has been improved, if land fragmentation can serve as a safety net (Blarel et al., 1992).

So what is land fragmentation? Land fragmentation is defined as the practice of farming a number of spatially isolated small plots of owned or rented land by the same famer (McPherson, 1982). It can be caused by external factors, such as population pressure and land policy, or by farmers' own choice as a rational decision to maximize benefits and/or minimize risks (Blarel et al., 1992). Conventionally, land fragmentation was regarded as a detrimental factor to both farmers and the economy. Owing to this, policies that encourage land consolidation were top policy priorities among policy makers (Bentley, 1987). But recent studies, among others, by

2

Bentley (1987) and Blarel et al. (1992) challenged this view for it can also be beneficial both to the farmers and the economy.

While there are evidences across the world on negative aspects of farm fragmentation, such as constraining technological adoptions and investment on land (Niroula and Thapa, 2005), there are also positive aspects of it in terms of reducing price and production risks, smoothing labor supply, diversifying crop production, and matching soil types with the necessary food crops (Blarel et al., 1992, Di Falco et al., 2010, Hung et al., 2007).

In Tigray, where the rural people depends mainly on *rain fed* agriculture, farming is a risky venture. Farmers who rely on subsistence farming, like those in Tigray, are likely to be risk-averse. A fundamental behavior of a risk-averse farmer is making decisions that emphasize the objective of minimizing variations in total production (income) than maximizing total production (income). Crop choice and agricultural diversification decisions are the most common and key decisions often made by farmers in Tigray. This indicates that *if* land fragmentation, as claimed by Bentley (1987) and Blarel et al. (1992), can be beneficial in minimizing production and price risks as well as in promoting crop and agricultural diversity, then it is possible to have demand-driven land fragmentation in Tigray, besides to the supply driven fragmentation. Thus, understanding the actual cause and effect of land fragmentation will be of crucial for appropriate policy recommendation.

Two compelling motives are behind this study: local and global. The first motive is owing to the fact that the effect of land fragmentation on productivity and technical efficiency in Tigray is not well investigated. Despite being an important feature of the agrarian system, land fragmentation in Tigray did not get adequate scholarly attention as it should deserve. Mostly, it appeared only as an auxiliary to some other main objectives; even then, it is only one aspect of it, namely distance, which appeared commonly; see for instance, Holden et al. (2009) Holden et al. (2011) and Pender and Gebremedhin (2006).

While the 1990’s land reform and regional land policies emphasized on plot diversification to protect people and enhance food self-sufficiency using land distribution as a safety net, some studies revealed that land fragmentation, defined in terms of distance to plot, has constrained traditional soil fertility management practices (Corbeels et al., 2000), farming operations (Beyene et al., 2006) and investment on land (Holden et al., 2009). However, no study, to the best of my knowledge, attempted to investigate its effect on productivity and efficiency in the Ethiopian highlands. Given the claim, elsewhere, that land fragmentation, if induced by egalitarian land redistribution, may improve food security and equity among farm households (Bentley, 1987, Blarel et al., 1992), its effect on productivity and efficiency in Tigray begs for thorough investigation; thus, the intention to contribute towards this gap.

The second motive emanates from the fact that the discourse on the economics of land fragmentation is yet to resolve; and hence, my attempt toward contributing to the ongoing discourse. Several studies that are conducted across the world show mixed result vis-à-vis the effect of land fragmentation on productivity. Some found unambiguously negative effect of land fragmentation on production and recommended for land consolidation, for example, in Bangladesh (Rahman and Rahman, 2009), Pakistan (Parikh and Shah, 1994), Jordan (Jabarin and Epplin, 1994) and Rwanda (Bizimana et al., 2004). Others found ambiguous or insignificant effect of fragmentation on production and efficiency and demanded policy makers to be cautious in promoting land consolidation programs such as in Bulgaria (Di Falco et al., 2010), Vietnam (Hung et al., 2007), as well as Ghana and Rwanda (Blarel et al., 1992). In China alone the result is mixed. Nguyen et al., (1996) found land fragmentation to have significant economic cost while Tan et al., (2008) found ambiguous results.

These mixed results might have emerged from variations in socio-cultural, political, economic, and environmental features of the study areas; as well as, I would argue, due to methodological reasons.[3] For instance, for the Chinese case, Nguyen et al., (1996) derived their conclusion relying only on one indicator of land fragmentation while Tan et al., (2008) used three indicators and found mixed results for each indicator. Moreover, some of the earlier works concentrate mainly on cross-sectional analyses using simple linear regression models that have obvious limitations as to account for unobserved plot and household heterogeneity.[4]

[3] Numerous factors can be listed within each features: a) Socio-cultural characteristics: including demography and resource endowment; b) Economic characteristics: including factor and product market functioning; access to credit, infrastructure and insurance; c) Political factors: including land policy and other policy priorities that reduce pressure on land; d) Environmental features: including agro-ecological settings, uncertainties and agricultural systems; all are likely to explain the mixed results.

[4] Not least recent studies reach conclusion using single (few) econometric model(s); posing a question about the robustness of the results. Thus, I intended to launch my analysis to rely on more than one indicator and several econometric specifications as a test for robustness of results. Detailed explanation on the type of indicators I used and methodological approach I followed in this study is provided in Chapter 5 and 6, respectively.

Moreover, the debate on the relationship between landholding size and productivity has not yet been resolved (Niroula and Thapa, 2005). Despite the conventional view of economies of scale, Schultz (1964) found inverse relationship between farm size and productivity in developing countries. Yet, others argued that this inverse relationship could be due to unobserved land quality and labor market imperfections (Benjamin, 1995, Bhalla, 1988, Bhalla and Roy, 1988, Udry, 1996) or due to bias in plot/farm size measurement (Holden et al., 2001) than due to diseconomies of scale (Niroula and Thapa, 2005). In fact, recently, Collier (2008) argued for large scale commercial agriculture for their production efficiency, cost advantage and facilitating innovation. This study also attempts to contribute toward this debate.

The overall purpose of the study is, however, to investigate the current situation of land fragmentation in Tigray, Northern Ethiopia, and to examine its relationship to and effect on productivity, technical efficiency and crop diversity. To this end, comparative static analysis, based on the farm household model, is used to examine the relationships between land fragmentation and productivity. Empirical analysis is also carried out based on the theoretical framework of the farm household model to investigate its effect. The following three research problems (R1, R2, and R3) are then set to be empirically tested:

R1. Whether land fragmentation reduces farm productivity.

R2. Whether land fragmentation could be correlated with soil quality variation that partly is unobservable and that can affect crop choice.

R3. Whether land fragmentation causes technical inefficiency.

The analysis relies on plot level primary data of cross-sectional structure. The data was collected during summer 2010 which constitutes plot level observations of 421 farm households operating on 1918 plots. Econometric analysis is carried out using farm household model (FHH), to analyze plot productivity and on-farm crop diversity, and using the stochastic frontier model (SF) to analyze technical efficiency. Different econometric specifications are used in order to check robustness of the results.

The rest of the paper is organized as follows. Chapter 2 offers background information on land fragmentation, general back ground as well as specific to the study area. It discusses fundamental concepts of land fragmentation and provides a brief explanation of the evolution of land fragmentation in Tigray. Chapter 3 reviews and summarizes the findings of various empirical researches that have been conducted across the world, early works as well as recent findings. Chapter 4 presents the theoretical approaches that are used to investigate the economic implications of land fragmentation. Chapter 5 briefly introduces the study area, sampling procedure and data type. It also offers detailed explanation of the variables used in the different model analyses along with their expected signs. Chapter 6 demonstrates the methodological approaches used to address the three research problems. Chapter 7 presents and discusses the findings with respect to each of the research problems. Chapter 8 concludes.

The Appendix part offers additional result summary tables that complement my arguments in Chapter 7 while addressing the main research objective. The tables in Appendices B, C and D are intended mainly to illustrate robustness of the results corresponding to each of the research problems.

2. BACK GROUND

2.1. Land Fragmentation: Definition and Overview

2.2. Definition

Land fragmentation is a common feature of agrarian economy. In earlier literatures, it is defined as the practice of farming a number of spatially separated plots of owned or rented land by the same farmer (McPherson, 1982); and as a type of land ownership pattern where "a single farm consists of numerous discrete parcels, often scattered over a wide area" (King and Burton, 1982). In this study, however, the first definition is used with slight modification where land fragmentation is perceived as the practice of farming distinct plots (not necessarily spatial separated parcels) of owned and/or rented land by the same farmer. The reason for this modification is that farmers, as a common practice in rural Ethiopia, may temporarily divide a homogenous unit of land into two or more plots on which multiple cropping is practiced in one season, and perhaps mono-cropping in the next season. It is based on this definition and realization of plots that this study analyses the economics of land fragmentation.

In earlier studies, land fragmentation was often considered as an obstacle for agricultural development (Bentley, 1987, Di Falco et al., 2010, Hung et al., 2007, Jabarin and Epplin, 1994). As a result, several countries have been encouraging land consolidation policies including Kenya, Tanzania and Rwanda (Blarel et al., 1992), China (Nguyen et al., 1996), Vietnam (Hung et al., 2007), and Bulgaria (Di Falco et al., 2010). However, some literature questioned this view and found that land fragmentation can be an adaptive strategy and under certain circumstances can have beneficial effects (Blarel et al., 1992, Tan et al., 2008). For instance, it can facilitate risk management through crop diversification (Di Falco et al., 2010) and operating on scattered plots to reduce the risk of total loss from flood, drought, fire and other perils; it may enable households spread their own labor over the seasons, and, if induced by egalitarian land redistribution, it may improve food security and equity among farm households (Bentley, 1987, Blarel et al., 1992).

On the other hand, land fragmentation may have some obvious problems. It can cause difficulty in management and supervision of fragmented plots, although how significant remains an empirical issue to solve. It can also lead to increased travelling time between fields which may, in turn, induce loss of working hours as well as higher transport cost of inputs and outputs.

Negative externalities are likely to co-exist with fragmentation in terms of reducing the scope for irrigation, soil and water conservation (SWC) structures and adoption of other agricultural technologies. Furthermore, fragmentation may have greater potential for disputes between neighbors (Blarel et al., 1992).

2.2.1. Causes of Land Fragmentation

In line with the pros and cons of land fragmentation, two conventional explanations appeared in the literature with regard to the emergence and persistence of land fragmentation, namely the *demand-side* explanation and the *supply-side* explanation (Bentley, 1987, Blarel et al., 1992). While the demand-side explanation views land fragmentation as a choice variable for farmers, the supply-side explanation treats it as an exogenous imposition on farmers resulting from, for instance, population pressure, inheritance, and land scarcity. In what follows a brief overview of these explanations is presented.

The *demand-side* explanation presumes free choice and views fragmentation as a rational response of farmers. In this case, the private benefits of land fragmentation might exceed its private cost (Blarel et al., 1992, Di Falco et al., 2010). For instance, it is argued that in the presence of heterogeneous land quality, farmers may prefer to operate on scattered plots as a safeguard. Where there is higher likelihood of production risk (such as localized hailstorm, thunderstorm, or flood) and lower risk spreading mechanism (such as insurance and credit), farmers may rationally operate on spatially dispersed plots to spread risk and reduce the variance of total output. Similarly in the absence of labor market, land fragmentation can facilitate smoothing labor supply across the scattered plots particularly during peak seasons (Blarel et al., 1992).

Likewise, in cases of commodity market failures, farmers may overcome the risk of household consumption by adopting several products on fragmented lands. Furthermore, whenever diseconomies of scale exist for a given crop production with respect to individual parcels, farmers may split the parcels into fragmented plots to optimize production by harvesting diversified crops. Such practices are very common in Northern Ethiopia, particularly in Tigray. Yet, whether these seemingly *rational* choices of farmers are supported by empirical evidence will be discussed in Chapter 3 and will be part of the analysis in this study.

The *supply-side* explanation focuses on rather involuntary land fragmentation imposed on farmers by various exogenous factors. The fact that land is scarce in most agrarian countries makes land fragmentation inescapable as long as there is population pressure and a corresponding egalitarian land distribution policy. With a growing population pressure not only the already arable lands but also communal properties might be turned into fragmented holdings. Similarly, land scarcity may lead to fragmented holdings as farmers in quest of additional land will tend to accept any available parcel of land within reasonable distance of hours (Tan et al., 2006). Where labor is chip, as is the case in most African and Asian countries, crop production is mainly carried out by hand cultivation and animal traction which is suited to small scale and self-sufficient production. Under such circumstances fragmentation is likely to increase. Legal rights and customs for partible inheritance and farmers desire to provide each of several heirs with land of similar quality is another possible explanation for land fragmentation (*Ibid*). Moreover, the presence of imperfect land market, laws that restrict land transaction, and missing credit markets as well as limited off-farm employment opportunities are also among exogenous factors of land fragmentation (Blarel et al., 1992, Tan et al., 2006).

For the Tigray case, land fragmentation might be explained mainly by population pressure and the land allocation process. Several authors argued that the egalitarian principle of land allocation with an attempt to accommodate the landless fairly, by emphasizing on plot diversification based on plot quality and location, has led to fragmented holdings (Beyene et al., 2006, Holden and Yohannes, 2002, Segers et al., 2010). A recent land proclamation issued by the Tigray Regional State (TLP 2007)[5] indicated that each men and women, above 18, who reside in rural Tigray, is entitled to a share of the village land, to at least 0.25 ha per household, on a usufruct basis. Moreover, the proclamation constitutes legal provision of partible inheritance that might play a role in further promoting land fragmentation in the region. Nonetheless, such arguments need to be supported empirically which is beyond the scope of the study; yet, it is recognized as a potential for future investigation.

2.2.2. Measuring Land Fragmentation

Since there is no standard measurement of land fragmentation (Bentley, 1987, Hung et al., 2007), it is difficult to determine whether a given farm household is 'very fragmented' or 'less

[5] TLR (2007). A regulation to determine the Administration and Using of Rural Land. Tigray Regional State, Ethiopia.

fragmented' as compared to others. However, several authors make use of some potential indicators to measure land fragmentation. Among the potential indicators that appeared in the literature include: farm size, number of plots, (average) plot size, plot shape, (average) plot distance, spatial distribution, the size distribution of the fields, and/or the Simpson Index (SI). The most common ones are, however, number of plots, the Simpson Index, average plot distance, and average plot size. This study will make use of the later common indicators, except average plot size. Detailed explanation on each of the indicators employed in this study is provided in chapter 5.

2.3. Overview of Land Fragmentation in Tigray

Provided with a brief background on the definition, causes and implication of land fragmentation, one may wonder on how land fragmentation came into existence in the study area. While still a thorough empirical investigation is needed, in this session I attempt to provide a brief overview of land fragmentation processes and consolidation practices in Tigray based on reviewing speculations of some studies.

2.3.1. The Evolution of Land Fragmentation

Prior to 1975, it is argued, land tenure system in Ethiopia was characterized by great inequality, insecurity, eviction, and underutilization in which a vast majority of land was concentrated in the hands of few but powerful absentee landlords (Deininger et al., 2003). Only persons who could trace their paternal and maternal ancestry to the village founders held *resti* rights on the village farmland (Segers et al., 2010). Although, land was less fragmented during this period, it greatly lacks fairness in its distribution which ultimately brought the regime to its end.

Following the overthrow of the imperial regime, land reform was initiated by the Marxist government (the *Derg*) when land was nationalized and declared as "public property" under proclamation 31, 1975. The government then provided user rights to the cultivators with highly restrictive transferability rights of any kind except inheritance to immediate family members (Deininger et al., 2006, Haile et al., 2005, Segers et al., 2010). During its reign, the *Derg* prohibited not only the transfer of land rights but also the hiring of labor putting the landless households in a more challenging situation (*Ibid*). Since the non-farm sector was underdeveloped and the scope for expansion of cultivated land was limited, the government had to redistribute land from the relatively land-rich to the relatively land-poor households (Holden and Yohannes, 2002). As a result, the Derg continued to frequently redistribute land in order to maintain its egalitarian principle of land allocation with each household receiving land proportional to family size (*Ibid*). Needless to say, this frequent land redistribution coupled with the (exceptional) land policy provision of partible inheritance and prohibition of land renting[6] must have played significant role toward land fragmentation in the country.

Following the collapse of the Derg regime, the current government overtook power in 1991 and withheld the major policies of the tenure system with few notable exceptions. Land remained public property and farmers were still granted lifetime user rights; however, regional governments were set autonomous over the land that resides within the region. Land rentals, as well as hiring of labors, have been officially permitted, and the frequency of land redistribution was to be reduced (Deininger et al., 2006, Haile et al., 2005, Holden and Yohannes, 2002, Holden et al., 2011, Segers et al., 2010). While the latter two points can be regarded as a move toward reversing land fragmentation, the ever growing population pressure and the existing law of partible inheritance are still decisive in furthering land fragmentation in the region in particular and in the country in general.

In Tigray, the regional government had already declared an end to administrative land redistribution in the first half of the 1990s (Deininger et al., 2003, Segers et al., 2010). During its last redistribution, all arable land in each village was initially classified based on local knowledge into three classes: first on the basis of fertility, as *Regwid* (fertile), *Machelay* (intermediate) and *Reqiq* (less fertile) (Segers et al., 2010) and, next on the basis of vicinity to homestead as *Gedena* (adjoining the homestead), *Dehri-bet* (near the homestead), and *Wofri* (the farthest plot from the homestead) (Beyene et al., 2006). Then, each land from each class was randomly assigned to farmers using lottery system, except in the first case where Gedena were raffled among the owners of the houses (Beyene et al., 2006, Segers et al., 2010). In this way land of different quality and size was *fairly* distributed among farm households taking family size into consideration. As a result, large households received either more scattered plots, or relatively larger plots, compared to their counterparts.

According Segers et al. (2010) and Haile et al. (2005) each adult farmer who was residing in the villages of Tigray during the redistribution was, in principle, entitled to a certain size of land, at

[6] Some authors argued that land renting practices can consolidate land into the hands of households who rented in the land, eg., Tan et al. (2006).

least 0.5*ha* However, practically it was difficult to maintain this principle due to shortage of arable land and population pressure. Thus, some unfortunate farmers received below the minimum size. Moreover, with time, new born babies and landless people (such as immigrant farmers from other provinces) emerged, bringing new challenges to the local administers. To deal with these challenges, the local administrators had only two opportunities; i.e., lands of people who died without heirs and few converted communal lands. These lands, although small in magnitude, were allocated to the most senior landless and land-poor farmers (Segers et al., 2010). Most youngsters, however, acquired land through inheritance from parents or grandparents, often involving division of land among several heirs. Consequently, not only the average land holding did persistently fall, but also the number of operational plots increased simultaneously over time (*Ibid*).

Clearly, this indicates that the land allocation process forced by the population pressure could be the leading factor of aggravating land fragmentation in the region. In fact, similar experience was reported in China by Tan et al. (2006) who found implementation of such an egalitarian principle of land distribution to be the major factor of land fragmentation in China.

2.3.2. Farm Dynamics and Operational Holdings in Tigray

Comparing average land holding in Tigray over time, Beyene et al. (2006) argued that average land holding shrunk from 1.7 ha (varying from 0.37 to 3.5 ha) in 1964 down to 1.2 ha (ranging from 0.5 to 2.0 ha) in 1994, with over 60% of the households owning less than 1.0 ha. Furthermore, in 2001, the average holding is reported to be 0.5 ha, ranging from 0.1 to 1 ha, with 2.46 plots per household, on average (Deininger et al., 2003, Segers et al., 2010).

Table 1 Trends in per capita land holding in Tigray

Year	**Land Per Capita (ha)**			**Average no. of plots per HH**	**Regional land Gini Coefficient**	**Sources**
	Mean	Max	Min			
1964	1.7	3.5	0.37			Beyene et al. (2006)
1994	1.2	2	0.5			Beyene et al. (2006)
2001	0.5	1	0.1	2.46		(Deininger et al., 2003)
2005	0.25*				0.45	Segers et al. (2010)
2009	0.2*					Holden et al. (2009)

*The authors estimated average land holding per HH. Based on the data I have in this study, I assumed average family size to be around 5 to convert their estimation to per capita land holding.

In 2005, the regional land Gini-coefficient was estimated ~0.45 (Segers et al., 2010). Recently, following the 2007 regional Land Administration and Use Proclamation (LAUR 2007)[3], Holden et al. (2009) and Holden et al. (2011) estimated the average land holding per household to be around 1ha. They have also reported improvements in land rental market mainly due to land certification. Summary of the trends observed in per capita land holding is presented in Table 1.

2.3.3. Traditional Land Management Practices and Land Fragmentation

Traditionally farmers in Tigray identify three types of plots based on plot quality and location from homestead, namely: *Gedena*, *Dehri-Bet*, and *Wofri* (Beyene et al., 2006). *Gedena* possess good quality and is located near the vicinity of the homestead and embodies significant cultural and social values; *Dehri-Bet* is located further from the homestead; and *Wofri* is the furthest of all. The cultural significance and location of the plot determines farm management practices and transactions of the land which, in turn, determines the productivity and technical efficiency of land. Beyene et al. (2006) argued that farmers tend to operate lands of the highest cultural significance – the *Gedena* – on their own and with better management and closer inspections instead of renting them out.

As the distance to plots increases, farmers seem to attach less symbolic value and put lesser effort on their management. Corbeels et al. (2000) argued with land fragmentation, particularly distance, farmers have increasingly *abandoned* traditional soil fertility management practices (such as fallowing, manuring, terracing, and using crop residues). This can led to poor soil fertility and, perhaps, to decline in productivity. Moreover, the more distant plot from the homestead, the higher the likelihood to rent it out (Beyene et al., 2006). Elsewhere, significant productivity difference is reported between own operated and rented out plots (Kassie and Holden, 2007).

One of my objectives is to test the validity of these arguments, i.e., whether a decline in productivity can be observed due to fragmentation (distance), whether directly or through poor soil fertility status or participation in rental market.

2.3.4. Land Consolidation Practices in Tigray

Some argued that proper functioning of land market can serve as a means of land consolidation (Nguyen et al., 1996, Niroula and Thapa, 2005, Tan et al., 2006). Therefore, one may expect the existence of land rental market in Tigray to play a role in consolidating land. However, Beyene et al. (2006) claimed that physical factors such as plot proximity and land quality are less significant criterion during land transaction. Instead, farmers consider trustworthiness, evaluated based on social relationship, as the most decisive factor, which implies that plots are not moving to plots but rather to trustworthy tenants. On the other hand, Ghebru and Holden (2012) argued that land rental market in Tigray is mainly characterized by "'Reverse-Share-Tenancy' where landlords are poor in non-land resources […] while tenants can be best described as non-land asset rich landowners", which implies that plots are moving mainly to non-land assets, not necessarily to plots.

Jointly, these arguments indicate that the land rental market may not necessarily consolidate land. Renting practice may consolidate land 1) in terms of increasing the size of total operational land holdings, if land moves from the relatively land-poor to the relatively land-rich; or, in the 'Reverse-Share-Tenancy' case, if the relatively land-rich landlord rent-out substantial size of her/his land so that the size of land operated by the tenant is greater than the size of the landlords land endowment; 2) in terms of scaling up plot size, if the tenant is lucky enough to rent-in plots adjacent to his own plot; or 3) in terms of reducing distance to plot, if the physical proximity of plots is closer to the tenants' residence than the landlords'. Yet, we have the limitation to explain the degree of fragmentation if both total operational land holding (farm size) and number of plots increase in the same direction.

Despite this argument, however, there have been actual land consolidation practices by local farmers soon after the redistribution. Haile et al. (2005) and Segers et al. (2010) reported that farmers who have formal user rights on plots that are far from their own home but near other farmer's home have exchanged plots. Alternatively, exchanging parties did obtain plot(s) adjacent to field(s) for which they already have the use rights, thereby reducing the fragmentation of their landholding. Moreover, upon the death of a parent or grandparent, there exists a legal restriction concerning the division of landholdings among heirs to prevent land from further fragmentation. Landholding of a deceased person can be divided among heirs legally, only if no plots smaller than 0.25ha are created (Segers et al., 2010). Likewise, the practice of "joint ownership" of inherited land, where heirs decide to jointly own and manage their ancestral land, is another important land consolidation practice in the region.

3. LITERATURE REVIEW

A number of studies have examined land fragmentation in different countries and regions covering a wide range of topics including: explanations of land fragmentation, cost estimates of fragmentation, effect of land fragmentation on agricultural production, and methods used to test the effects of land fragmentation on agricultural production. In most cases, there are plenty of unresolved issues often reflected by contrary findings. In what follows a brief review of literature on each of these points will be presented. Taking no side for the moment, exemplary arguments from both sides will be presented.

3.1. Deriving Forces of Land Fragmentation

Several studies examined causes of land fragmentation in different parts of the world. For instance, empirical analysis on factors of land fragmentation in China by Tan et al. (2006) reported the egalitarian land distribution policy of China, which aimed at equitable "asset distribution", to be the major deriving force of land fragmentation. Cultivation of high value added crops, particularly in the suburban area, was also found to associate with higher land fragmentation. Furthermore, grain quota, off-farm employment opportunities, and proper functioning of land rental markets were found to associate with lower land fragmentation. Similarly, in South Asia Niroula and Thapa (2005) found higher dependency on agriculture as the main cause of land fragmentation. Other structural problems including the law of inheritance of paternal property, lack of progressive tax on inherited land, heterogeneous land quality, and an underdeveloped land market were identified as significant determinants of land fragmentation.

3.2. Land Fragmentation, Productivity and Technical Efficiency

Studies on the link between land fragmentation, farm productivity and efficiency are mixed and inconclusive. For example, in Bangladesh Wadud (2003) and Rahman and Rahman (2009) highlighted that land fragmentation has a significant detrimental effect both on productivity and efficiency. Similarly, Parikh and Shah (1994) reported that land fragmentation reduces efficiency in rice production in Pakistan. In Rwanda, too, Bizimana et al. (2004) found greater economic inefficiency in highly fragmented plots. In contrast, Blarel et al. (1992) concluded that land fragmentation in Rwanda and Ghana "does not seem to have any adverse impact on the productivity of land." Although, Di Falco et al. (2010) (Bulgaria) and Hung et al. (2007)

(Vietnam) reported a reduction in productivity due to land fragmentation, the net effect is inconclusive due to the positive impact revealed through enhancing agro-biodiversity and labor efficiency, respectively. Such contradiction is not limited to cross country results, rather it is also observed on studies within a given country. In China, for instance, while Wan and Cheng (2001) and Chen et al. (2009) reported significant detrimental effect of land fragmentation on productivity and efficiency, Wu et al. (2005) found no significant effect. Similar trends are also observed with respect to the link between land fragmentation and production cost, for instance, Jabarin and Epplin (1994), Nguyen et al. (1996), and Tan et al. (2008). These mixed findings can be explained by variations in methodological approaches, unobserved plot heterogeneity, or variations in market, agro-ecological, demographic and cultural conditions.

3.3. Land Fragmentation and Production Cost

Land fragmentation as reflected in average plot size was reported to have significant economic cost in the production of Wheat in Jordan (Jabarin and Epplin, 1994) and in the production of Maize, Rice and Wheat (major grain crops) in China (Nguyen et al., 1996). In contrast, captured by three indicators, the effect of land fragmentation on the cost of rice production in China was reported to be inconclusive for two reasons (Tan et al., 2008): First, although farm size and distance to plots were found to associate with production cost, number of plots and plot size distribution were not. Second, labor efficiency was significantly and positively correlated to fragmentation, specifically to the number of plots a household operated on. It is argued that this significant positive correlation can "counterbalance the negative impact of fragmentation" and hence the net effect is ambiguous (Tan et al., 2008). Despite such contradictions, however, most authors (Bentley, 1987, Blarel et al., 1992, Nguyen et al., 1996, Niroula and Thapa, 2005, Tan et al., 2008, Wan and Cheng, 2001) agreed that distance to plot have an economic cost in terms of time wasted travelling from plot to plot, with few exceptions (such as Ilbery (1984)). Moreover, the view that land fragmentation deters adoption of agricultural innovations is more or less undisputable (Bentley, 1987, Niroula and Thapa, 2005, Rahman and Rahman, 2009).

3.4. Land Fragmentation and Labor Productivity and Efficiency

With respect to impact on labor productivity and efficiency, authors such as Bentley (1987), Blarel et al. (1992), and Hung et al. (2007) found similar result as Tan et al. (2008) that land fragmentation facilitates labor smoothing activities, thereby elevating efficiency of labor utilization. Contrarily, Jia and Petrick (2011) argued that "land fragmentation indeed leads to lower agricultural labor productivity" in China. Similarly, Corbeels et al. (2000) argued that land fragmentation in Tigray made soil fertility management increasingly labor intensive which led farmers to "abandon soil fertility management practices" that resulted in poor soil fertility status.

3.5. Land Fragmentation and Risk Management

With regard to the link between fragmentation and risk, there is a general consensus, supported by ample evidence, that in areas with higher production and market risks, farming fragmented plots can be a viable mitigation measure. Blarel et al. (1992) found that risk, measured by the variance of total farm income per hectare, declines linearly as the Simpson Index (a measure of land fragmentation) increases. They then concluded that "fragmentation increases the diversity of agro-climatic conditions available to farmers [that] leads to more diversified cropping pattern". Likewise, Wu et al. (2005) found high share of land devoted to grain production in China as "a risk aversion strategy to ensure household self-sufficiency". Consequently, Wu et al. (2005) cautioned policy makers' effort toward consolidating land. Consistent to this, Sikor et al. (2009) argued that land fragmentation in Albania "is not a rigid constraint on production but resulted from producers' strategic risk spreading". As a result, Sikor et al. (2009) emphasized on "the need to support desirable adaptations initiated by local communities" instead of advocating for centrally initiated land consolidation programs. Moreover, it is even argued that "high-risk areas should be left fragmented" (Bentley, 1987).

3.6. Land Fragmentation and Integrated Farm Productivity

The scope of most economic evaluation of land fragmentation is narrowly focused in the sense that efficiency analysis is often limited to one or few aspects of production. Clearly, such approaches are likely to underestimate (or overestimate) farm efficiency particularly in areas that are characterized by mixed farming; i.e., simultaneous production of crops, vegetable, and fruits as well as animal husbandry. To this end, Rosset (2000) demonstrated evidently that small farms, often characterized by farming fragmented plots, are more productive, more efficient, and able to contribute more to economic development than large farms if efficiency measurement involves total output than a single crop; namely, the output of all crops on a designated plots – including various grains, fruits, vegetables, fodder, animal production, and so on. Similarly, investigating the link between agro-biodiversity and land fragmentation at micro-scale, Di Falco et al. (2010)

concluded that farm fragmentation is positively correlated with the number of crops (farm-biodiversity) which, in turn, is positively correlated with farm profitability. On the other hand, del Corral et al. (2011) identified that profit in the Spanish dairy farms increases in the range between 9.4 and 14% owing to land consolidation program.

3.7. Review of Policy Recommendations

In their book titled "Modern Macroeconomics: its origins, development and current state", Snowdon and Vane (2005:7) wrote "economists tend to disagree more over theoretical issues, empirical evidences, and the choice of policy instruments than they do over the ultimate objectives of policy." Quite similar trend is observed in the literature on economics of land fragmentation. Despite the significant difference observed in their empirical findings, most authors tend to converge into almost tantamount policy prescriptions toward the same goal.

Three dominant prescriptions are proposed by authors, regardless of their positive, negative or ambiguous evidences. The first policy prescription emphasizes on creating off-form and non-farm employment opportunities. The rationale is that doing so will reduce pressure on land and retard further land fragmentation. Examples include: Blarel et al. (1992), Nguyen et al. (1996), Niroula and Thapa (2005), Rahman and Rahman (2009), Tan et al. (2006) and Wu et al. (2005).

The second prescription emphasizes on promoting rural markets particularly land, labor, food, and credit markets (Examples: *Ibid*; Tan et al. (2008); Di Falco et al. (2010)). Holden et al. (2012) demonstrated that promoting the non-land market alone could ensure production efficiency with out the need for land market. They argued that the standard neoclassical household model (Singh et al., 1986) can give efficient outcomes even without land market, given the non-land factor markets function appropriately. Thus, promoting not only the land market but also the non-land factor market alone can enhance productivity. Similarly, it is argued that the availability of such markets can enhance "the ability of farmers to adjust optimally the extent of fragmentation (or consolidation) of their holdings over time" (Blarel et al., 1992).

The third dominant prescription is based on the belief that factor markets can correct the side-effects of land fragmentation; and that it demands limited government intervention; Examples include: Nguyen et al. (1996) and Rahman and Rahman (2009), to whom land fragmentation is costly and detrimental to productivity, Niroula and Thapa (2005), to whom the implication of land consolidation is not clear, and Blarel et al. (1992) and Di Falco et al. (2010), to whom land fragmentation is beneficial for risk management and crop diversity (Blarel et al., 1992, Di Falco et al., 2010). All ended up recommending less (modest) government intervention. The main rationales are the existence of mixed results elsewhere, the uncertainty on the overall effects of land fragmentation and the doubt that farmers may not perceive land fragmentation as a problem (Niroula and Thapa, 2005). Other recommendations include adoption of new technologies, expansion of extension system and infrastructure development (See Table 2 for summary).

Table 2 Summary of Some of Land Fragmentation Literatures

Author(s)	Study Area	Effect on	Method	Indica-tors	Result	Conclusion	Recommendation
Blarel et al., (1992)	Ghana Rwanda	Production	Pooled OLS	SI, N D, A	↑ Yield ↑ Cost **Inconclusive**	LC is unlikely to increase productivity significantly	Focus on reducing root causes of LF. Promote land & non-land factor markets.
Di Falco et al., (2009)	Bulgaria	Agro-biodiversity; Farm Profit	2SLS (Village fixed effect)	N, D, A	↓ Profit ↑ Agro-biodiversity (↑ Profit) **Inconclusive**	Policy measure of LC must carefully maintain the net effect of LF between Agro-biodiversity and profit.	Instead of LC improve functioning of land, labor, credit and food markets, and access to improved technology and off farm employment.
Hung et al., (2007)	Vietnam	Labor Efficiency; Land Productivity	Standard FHHM (Frontier Regression)	N, N*labor	↓ Production ↑ Labor use **Inconclusive**	Real benefits to FHHs from LC may not be apparent until the real opp.cost of farm labor begins to rise.	Consolidate by creating new off-farm jobs and movement of agricultural labor force to other sectors of the economy.
Jabrin and Epplin (1994)	Jordan	Production Cost; Efficiency	GLS	A	↑ Cost ↓ Production Efficiency	LF is indeed an impediment to efficient wheat production.	Consolidate by encouraging land market.
Nguyen et al., (1996)	China	Production cost; Productivity	Production Function	A	↑ Cost ↑ Productivity **Inconclusive**	Outcome could be expensive in terms of output forgone.	LC with less gov't intervention; improve land market, grain market and access to credit
Tan et al., (2008)	China	Cost Efficiency	FHHM	A, D SI,	↑ Cost ↓ Cost **Inconclusive**	The net impact on total production cost is not significant.	LC can stimulate technological adoption, but also can reduce agricultural employment and increase the rural labor surplus.
Rahman and Rahman (2008)	Bangla desh	Production Efficiency	SPFA (MLE)	N	↓ Productivity and efficiency	Productivity and efficiency are adversely affected by land fragmentation in Bangladesh.	Address the structural causes underlying the process of LF: law of inheritance and political economy of the agrarian sector.
Parkikh and Shah (1994)	Pakistan	Production Efficiency	SPFA (MLE)	N	Negative Relationship (no causality identified)	LF can be result of technical inefficiency rather than a cause of it.	Increased education and availability of credit along with land consolidation would improve efficiency.

SI= Simpson Index; N= Number of Plots; D=Average plot distance; A=Average Plot size; FHHM= Farm Household Model (Production Approach); SPFA= Stochastic Production Frontier Analysis approach; MLE=Maximum Likelihood Estimate; LF= Land Fragmentation; LC= Land Consolidation; HH= Household; TC=Transaction Cost;

4. THEORETICAL FRAMEWORK

Various theoretical frameworks appear in land fragmentation literatures in analyzing its economic implications. Stochastic production frontier models and farm household models are the common ones. As already argued in the introduction chapter, the observed mixed result in the literature of land fragmentation could be attributed, somehow, to methodological approaches. Thus far, authors tend not to include research methodology as explanatory factor. Following the methodological approach of Holden et al. (2001), more than one methodology is applied in this study in order to check the validity and robustness of the results. Thus, follows a brief explanation of analytical frameworks relevant to addressing the three research questions.

4.1. Farm Household Model

Most of the empirical works on economic analysis of land fragmentation tend to use the standard production function with various model specifications and functional forms. However, since such analyses are dealing with small farmers who operate on fragmented plots, often in an imperfect factor market settings, production function can be appropriate analyzed if it is built upon farm household models that take account of other farm decisions than mere production decision. In the absence of perfect market production and consumption decisions are intricate (Sadoulet and De Janvry, 1995). Some authors, like Hung et al. (2007), built their model on the standard farm household model that assumes farm households' production and consumption decisions to be separable. Since this assumption is mostly unrealistic, particularly with respect to the study area (Tigray), where factor markets are imperfect (Ghebru and Holden, 2008, Holden et al., 2011), productivity analysis has to be carried out based on farm household model with market imperfections as suggested by Sadoulet and De Janvry (1995) and Udry (1996). Similar approaches were used by Tan et al. (2006) and Holden et al. (2001) although with different objectives. The former's emphasis was on uncovering the effect of land fragmentation on cost efficiency (as opposed to production efficiency) in rural China and the latter's approach was a general analysis of land productivity with market imperfection in rural Ethiopia. This study is built on the approach by Holden et al. (2001) with land fragmentation indicators being the main variables of interest.

4.1.1. Productivity With-in a Farm Household Model Approach

The theory of farm household deals with the decision making process of rural households that attempt to attain their goals and aspirations using their limited resources (such as land, labor, and oxen) and choosing among alternative productive activities (such as mono-cropping, multiple cropping, on-farm or off-farm activities) (Singh et al., 1986). The model accounts for interdependence of household production and consumption decisions. The following model illustration is due to Sadoulet and De Janvry (1995) and Holden et al. (2001).

Suppose a household (i) with a given level of agricultural ability (α_i) and endowed with a fixed amount of land $(\bar{A}_i)$, labor $(\bar{L}_i)$, and capital $(\bar{K}_i)$, where capital includes oxen (traction power), and other farm inputs $(\bar{X}_i)$ has a utility maximization problem defined by:

$$Max\ U_i = u(c_i;\zeta) = u(c_i^p, c_i^m, c_i^l;\zeta) \tag{4.1}$$

where c_i is vector of consumption goods of household (i) which is a combination of goods that are produced $\left(c_i^p\right)$, market purchased (c_i^m) and home time (leisure) (c_i^l); and ζ is a set of household characteristics.

Assuming perfect market and perfect information and that household (i) operates on P numbers of plots, plot level output $\left(y_{ip}\right)$ can be defined as a function of the household specific agricultural ability (α_i), operating plot size (a_{ip}), labor (l_{ip}), capital (k_{ip}), and a vector of other inputs employed (x_{ip}) (such as seed, fertilizer, manure, herbicide, and pesticide):

$$y_{ip} = \alpha_i q\left(a_{ip}, l_{ip},\ x_{ip}, k_{ip};\varsigma\right) \tag{4.2}$$

where $\sum_p a_{ip} = A_i$, A_i is the total operating area of household i; $\sum_p l_{ip} = L_i$, L_i is the total labor used in the production by household i; $\sum_p k_{ip} = K$, K_i is the total operating capital of household i; $\sum_p x_{ip} = X_i$, X_i is the total inputs used by household i and ς is the farm (plot) characteristics which is assumed to include the degree of land fragmentation along with land quality, soil depth, soil type, slope, etc… as factors that affect farm productivity;

Assuming production decision is made at household level, the farm level aggregated production function will have the form:

$$Y_i = \alpha_i Q(A_i, L_i,\ X_i, K_i;\varsigma) \tag{4.3}$$

which is subject to the following resource constraints:

$$\begin{aligned} A_i &= \bar{A}_i + A_i' \\ L_i &= \bar{L}_i + L_i' \\ K_i &= \bar{K}_i + K_i' \\ X_i &= \bar{X}_i + X_i' \end{aligned} \tag{4.4}$$

where A_i' is the net land rented-in or rented-out, negative if rented-out; L_i' is the net labor hired-in or hired-out, negative if hired-out; K_i' is the net capital rented-in or rented-out, negative if rented-out; and X_i' the net other inputs bought or sold in the market, negative if sold. Alternatively, we can define the superscript $' \in \{s, *, b\}$, so that s denotes households (net) selling the factor, $*$ denotes self-sufficient households, and b denotes households (net) buying the factor. All other notations are as previously defined.

The utility function is, therefore, subject to the following budget constraint:

$$\sum p_j c_i^j = p_q \alpha_i Q(A_i, L_i, X_i, K_i; \varsigma) - p_x X_i' - p_l L_i' - p_a A_i' - p_k K_i' \tag{4.5}$$

where p_j is vector of prices of consumption goods that are produced (p_p), market purchased (p_m) and leisure p_l; p_q, is vector of output prices; p_x is vector of input prices; p_a, p_l, and p_k are prices of land, labor, and capital, respectively; $Q(.)$ is the production function as previously defined.

4.1.2. Productivity, the Farm Household Model and Market Imperfections

Equation (4.5) assumes perfect market and perfect information to mean that households make decisions solely based on market prices and that production and consumption decisions are separable.

However, farm households are often located in an environment highly characterized by a number of market failures for some of its products and factors. With market failures and institutional constraints households may encounter highest transaction cost and imperfect information that may cause non-separability of production and consumption decisions (Holden et al., 2001). Under such circumstances, farm households' decisions rely on shadow prices or price bands that can be markedly different from market prices (Sadoulet and De Janvry, 1995). This can be demonstrated mathematically as follows.

Let the institutional constraints that govern participation in markets, access to credit, insurance and infrastructures are provided by:

$$\boldsymbol{z} \leq \boldsymbol{z_{max}}(\zeta, \varsigma, \tau), \tag{4.6}$$

where $\boldsymbol{z}$ is a vector of choice variables $c_i^j, A_i, L_i, X_i,$ and K_i; and $\boldsymbol{z_{max}}$ is the vector of maximum value for the choice variables determined by household characteristics (ζ), farm/plot characteristics (ς), and institutional characteristics (τ).

The Lagrangian associated with this constrained maximization problem can be formulated as:

$$\mathcal{L} = u(c_i; \zeta) - \lambda\left[p_p c_i^p + p_m c_i^m + p_l c_i^l - p_q \alpha_i Q(A_i, L_i, X_i, K_i; \varsigma) + p_x X_i' + p_l L_i' + p_a A_i' + p_k K_i' \right] - \mathcal{u}_z(z_{max} - z)$$

Where λ is the Lagrange multiplier of the budget constraint and $\mathcal{u}_z$ is the multiplier related to the labor, land, capital, and other material inputs. This yields the following first order conditions for:

Consumption goods: $\frac{\partial \mathcal{L}}{\partial c_i^j} = \frac{\partial u(.)}{\partial c_i^j} = \lambda p_j;$ where $j \in \{p, m, l\}$ (4.7A)

Land factor: $\frac{\partial \mathcal{L}}{\partial A} = p_q \alpha_i \frac{\partial Q(.)}{\partial A} = p_a + \frac{\mathcal{u}_A}{\lambda} = p_a + \lambda_a = p_a'$ (4.7B)

Labor Factor: $\frac{\partial \mathcal{L}}{\partial L} = p_q \alpha_i \frac{\partial Q(.)}{\partial L} = p_l + \frac{\mathcal{u}_L}{\lambda} = p_l + \lambda_l = p_l'$ (4.7C)

Capital: $\frac{\partial \mathcal{L}}{\partial K} = p_q \alpha_i \frac{\partial Q(.)}{\partial K} = p_k + \frac{\mathcal{u}_k}{\lambda} = p_k + \lambda_k = p_k'$ (4.7D)

Agricultural inputs: $\frac{\partial \mathcal{L}}{\partial X} = p_q \alpha_i \frac{\partial Q(.)}{\partial X} = p_x + \frac{\mathcal{u}_X}{\lambda} = p_x + \lambda_x = p_x'$ (4.7E)

Where λ_zs are the endogenous markups on the price of institutionally constrained factors of production; p_z's are the respective shadow prices of the factors of production; the superscript $' \in \{s, *, b\}$, where s denotes household selling the factor, $*$ denotes self-sufficient household, and b denotes household buying the factor; attributed to the associated transaction costs we have that $p_z^s < p_z^* < p_z^b$. These prices vary not only among households but also between villages (Sadoulet and De Janvry, 1995). Thus, it is assumed that, in the absence of fully functioning factor markets, factor prices may be defined by village (v) and household (ζ) characteristics:

$$p' = p(v, \zeta) \tag{4.8}$$

where p' is a vector of shadow prices and v is a vector of village specific characteristics.

Since factor markets in Tigray area are assumed to be incomplete (Ghebru and Holden, 2008, Holden et al., 2011), price bands in which some households selling factors, some being self-sufficient, and others buying them, are considered as a more realistic representation of the factor markets Holden et al. (2001). Assuming product market is *complete*, the budget constraint function will then have the form:

$$\sum p_j c_i^j = p_q \alpha_i Q(A_i, L_i, X_i, K_i; \varsigma) - p_x' X_i' - p_l' L_i' - p_a' A_i' - p_k' K_i' \tag{4.9}$$

Nesting the resource constraint equations in (4.4) to (4.9) gives:

$$\sum p_j c_i^j = p_q \alpha_i Q(\bar{A}_i + A_i', \bar{L}_i + L_i', \bar{K}_i + K_i', X_i; \varsigma) - p_x' X_i' - p_l' L_i' - p_a' A_i' - p_k' K_i' \tag{4.10}$$

The Lagrangian associated with the constrained maximization problem can be formulated as:

$$\mathcal{L} = u(c_i; \zeta) - \lambda\left[\sum p_j c_i^j - p_q \alpha_i Q(\bar{A}_i + A_i', \bar{L}_i + L_i', \bar{K}_i + K_i', \bar{X}_i + X_i'; \varsigma) + p_x' X_i' + p_l' L_i' + p_a' A_i' + p_k' K_i'\right] \tag{4.11}$$

and the FOCs for the factors of production can be summarized as (the FOCs for consumption goods, by assumption, are the same as (4.7A)):

$$\frac{\partial \mathcal{L}}{\partial z'} = p_q \alpha_i \frac{\partial Q(.)}{\partial z'} = p_z' \; ; where\, ' \in \{s, *, b\} \; and \; z \in \{A_i, L_i, X_i, K_i\} \tag{4.12}$$

[7]

As previously illustrated $p_z^s < p_z^* < p_z^b$. The implication of this is that marginal productivity of land for each household depends on whether the household is a net seller, self-sufficient or net-buyer of the various factors of production (Holden et al., 2001). Consequently, solving the FOCs gives a set of reduced-form equations for the vector of optimal choice variables $\boldsymbol{z}$, $z \in \{c_i^j, A_i, L_i, X_i, K_i\}$

$$\boldsymbol{z} = f(\bar{\boldsymbol{z}}, p_q, p_j, p_a', p_l^i, p_k', p_x', \zeta, \varsigma, v) \tag{4.13}$$

where $\bar{\boldsymbol{z}}$ is the set of household resource endowments, $\{\bar{A}_i, \bar{L}_i, \bar{K}_i, \bar{X}_i\}$; p_q is output price; p_j is vector of price of consumption goods; p_z's are shadow prices of land, labor, capital and other inputs, respectively; ζ, ς *and* v are household, plot and village specific characteristics, respectively.

[7] Extended form of the FOCs is provided in Appendix A.

Equation (4.13) implies that the optimal choice of factor input in the production process is determined not only by its own shadow price but also by the shadow prices of other factor inputs (along with market prices of production and consumption goods). Dependence on the various shadow prices, in turn, implies that producers are facing different price ratios in the different factor markets: $p_z^s < p_z^* < p_z^b$. If, as the law of demand forecasts, a higher price causes lower demand for resources, this would affect productivity at plot level. Holden et al. (2001) suggested an approach to test this hypothesis empirically: if land productivity at plot level is significantly determined by household factor endowments $(\bar{A}_i, \bar{L}_i, \bar{K}_i, \bar{X}_i)$, then it is a sign of factor market imperfection and significant transaction cost.

Based on the above theory, productivity per unit area can be defined as a function of household choice variables, household resource endowments, prices, and farm and village characteristics. Combining equations (4.3), (4.8) and (4.13) yields such a reduced form of farm production:

$$\boldsymbol{Y} = \boldsymbol{f}\left(\alpha_i, \boldsymbol{Q}\left(\boldsymbol{z}\left(\bar{\boldsymbol{z}}, p_q, p_j, p_z'; \zeta, \varsigma, v\right); \zeta\right)\right) \tag{4.14A}$$

$$\boldsymbol{Y} = g(\bar{\boldsymbol{z}}, \alpha_i, \zeta, \varsigma, v) \tag{4.14B}$$

where $\boldsymbol{Y}$ is a vector of outputs; $\boldsymbol{z} \in \{A_i, L_i, X_i, K_i\}$ are the actual factor inputs employed in the production process**;** $\bar{\boldsymbol{z}} = \{\bar{A}_i, \bar{L}_i, \bar{K}_i, \bar{X}_i\}$ is the set of household resource endowments**;** α_i is the household specific agricultural ability; ζ, ς and v are household, farm/plot and village specific characteristics, respectively.

Needless to say, if farm level production is determined by the variables in (4.14), then plot level production is also determined by the same variables. Similarly, plot level productivity is determined by the same variables. The later will be used as a dependent variable in the forthcoming analyses of effect of land fragmentation on productivity where land fragmentation is considered as component of farm characteristic (ς).

Implicit in (4.14) is the assumption that optimal decisions about production are made at household (farm) level as opposed to plot level. This also implies that plots do compete over resources and that the decisions are endogenous. This assumption will have fundamental implication in empirical analysis as it also implies violation of the assumption of independent distributions, which is often a common requirement in econometric models.

4.1.3. Functional Form specifications Within the Farm Household Model

The common functional form that dominates in the farm household production analysis is the Cobb-Douglas (full translog) production function and the semi-translog production function. Hung et al. (2007) argued that problem of multicollinearity occurs when using the full translog functional form. This may suggest that the semi-translog function can be a feasible alternative. While I rely mainly on the semi-trasnlog form, three functional forms are employed in this study as a method of testing robustness of the results. Detailed explanations on model and functional form specification are provided in Chapter 6.

4.2. Cereal Crop diversity, the Farm Household Model and Market Imperfections

Van Dusen (2000) and Benin et al. (2004) argued that farmer's decision about which and the extent of crop diversity and varieties to grow can be understood in the context of the theory of the farm household. This would then suggest that the conceptual framework of farm household model that developed so far can be used to analyze the on-farm crop diversity as well. In fact, practically, on-farm crop diversity is determined by farm physical characteristics, household characteristics, local market conditions and other similar factors that determine productivity. Thus, Equation (4.13) can provide the base for econometric estimation to examine the factors affecting crop diversity where the optimal choice is an outcome of choices made in constrained optimization problem rather than an explicit choice. Consequently, crop diversity, assumed to be measured by diversity richness (D_i) which is defined as the total number of crops produced at household level, can be expressed in the same conceptual form as the reduced form of farm productivity in (4.14):

$$\boldsymbol{D_i} = g(\bar{\boldsymbol{z}}, \alpha_i, \zeta, \varsigma, v) \quad (4.15)$$

Similar expression is used by Benin et al. (2004), Van Dusen (2000) and Van Dusen and Taylor (2005) with Benin et al. (2004) added exogenous income into the model which is the sum of remittances, food aid, gifts and pension. Unfortunately, data on such income is missing, although it is an important type of income in the study area. Nonetheless, its omission may not be costly since Benin et al. (2004) found it to have insignificant effect on crop diversity, in Tigray.

4.3. The Stochastic Production Frontier Model

Stochastic production frontier model is widely used method of measuring technical efficiency in the farming sector. The model assumes the existence of technical inefficiency of the different farm households involved in production, such that for specific factor inputs the level of production is less than what would be the case if the farm households were fully technically efficient (Coelli, 1995).

The model assumes competition among producers with some objective such as avoiding waste, by maximizing output from a given set of inputs or by minimizing input in producing a given level of output (Kumbhakar and Lovell, 2003). This notion of productive efficiency is commonly referred to as technical efficiency as opposed to economic efficiency in which the objective of producers is to maximize profit or income or minimize per unit cost of production (*Ibid*).

Given the standard production function, technical efficiency analysis using the stochastic frontier model involves two stages: The first stage estimates the standard production function while the second stage estimates technical efficiency. In the first stage the standard production function (output per unit area) can be estimated using the set production factors such as land, labor, fertilizer and other inputs. The residuals from the first stage estimation are used in the second stage to estimate technical efficiency. The residuals are assumed to consist of two components, one to account for pure random effects and the other to account for technical inefficiency (TIE). Technical efficiency, in the second stage, is estimated using the second component of the error term. Once technical efficiency is computed, identifying determinants of this efficiency is straight forward. Detailed estimation procedure is presented in chapter 6.

5. SAMPLING, DATA AND VARIABLE DEFINITION

5.1. The Study Area

The study is conducted in one of Ethiopia's highly degraded and fragmented region, Tigray. Agriculture, in this region, is one of the significant economic activities on which more than 80% of the people rely for their livelihood (TRPFB, 2011). Of the total land, about 65% is under cultivation, farmed by smallholders, within mixed crop-livestock systems (*Ibid*). In the last few decades, farming in Tigray is characterized by a sharp decline in land holding and prevalence of fragmented holdings (See Table 1). The agricultural system in this region is characterized by heavy reliance on rainfall which is mostly mono-modal, extremely variable, erratic and strongly influenced by topography.

Cereal crops provide the major means of livelihood in the mixed farming system. The main cereal crops are Teff (*Eragrostis teff*) (an endemic cereal crop), barley (*Hordeum vulgare*), and wheat (*Triticum vulg*). Along with cereal crops farmers also produce pulse crops which are also important part of crop rotation; the main pulse crops include: horse beans (*Vicia faba*), field peas (*Pisum sativum*), chickpeas (*Cicer arietinum*), lentils (*Lens culinaris*) and flax (*Linum usitatissimum*). Livestock raring particularly cattle, sheep and goats are also major component of the farming system. Crops and livestock are highly integrated where crop residues provide major share of livestock feed and livestock provide important source of manure and draught power.

Influenced by topography and underlying geological formation, soil characteristics in the region is highly variable. More than 50% of the soils are shallow, very low in organic matter, and extremely deficient in both total nitrogen and available Phosphorus, but moderately sufficient in potassium (Beyene et al., 2006). Farming practices in the region are characterized by negative soil nutrient balances, thus reducing the fertility of the soils, eventually leading to lower yields of the major crops and declining land productivity (Corbeels et al., 2000, Tamene and Vlek, 2008).

The heavy reliance on farm land with less appropriate management has facilitated the intensity of tillage erosion which contributes to sheet and rill erosion. Both sheet and rill erosion along with gully formation are considered as the most important degradation processes in the region (Nyssen et al., 2004, Tamene and Vlek, 2008). It seems that anthropogenic factor was dominant both in facilitating and curbing erosion processes in the region. Recently, through targeted interventions and mass mobilizations, the rural society shows progress in controlling and reversing degradation process (Nyssen et al., 2004, Nyssen et al., 2007, Nyssen et al., 2008). For example, Nyssen et al. (2007) found that estimated soil loss due to sheet and rill erosion in 2006 was only 68 % of its rate in 1975. Yet, the annual rate of soil loss is higher than the rate of soil formation (Tamene and Vlek, 2008).

5.2. Sampling and Data

The School of Economics and Resource Management (IØR) at the Norwegian University of Life Science (UMB) is running a NORAD-funded (NOMA) collaborative MSc-program in Development and Natural Resource Economics together with four African Universities – of which Mekelle University (MU) is amongst. In collaboration with MU, IØR carried out a survey in summer 2010 as part of a follow up survey to 500 households since 2006. The recent cross-sectional data that was collected at plot level in 2010 is used for the purpose of this study. The data covers around 450 households and over 2000 plots. It contains a wide range of variables including: output data (type and amount), input data (land, labor, seed, fertilizers, draft power and herbicide and pesticide usage) household characteristics (education, age and sex of the household head, household size and labor force …), plot characteristics (ownership, distance, size, soil quality, slope, depth...), data on market participation (credit, labor, land rental markets …) as well as on soil erosion, land degradation and conservation activities. Owing to missing important data, particularly data on the dependent variable, some households are dropped in making use of the data for this study; specifically, 421 households and their corresponding 1918 plots are used. The exclusion criteria is in such a way that it does not follow systematic trend and that if there is a missing output observation of a single plot, the household is excluded entirely whether she/he has observed outputs for other plots or not. Of the numerous plot level observations, only relevant observations are extracted for the purpose of this study as illustrated in Table 3.

Table 3 Variable Definitions and Expected Signs

Variable			Descriptive Statistics			Expected sign in analyzing		
List	**Type**	**Definition**	**Ob**	**mean**	**SD**	**PR**	**DIV**	**TE**
		Dependent Variables						
netput_v	cont	Net output value (birr/ha)	1859	4119	135	DV	DV	
netput_w	cont	wheat equivalent net out put (Q/ha)	1888	8.58	0.25	DV	DV	
cropdiv	count	Crop diversity (no. of crop cultivated at household level)	1918	3.51	0.03			DV
		Factors of Production						
plotsize	cont	Plot size planted, ha	1905	0.24	0.01	+		+
fam.labr	Cont	Farm labor engaged per ha	1902	137.7	4.11	+		+
hirdlabor	Cont	Hired Labor per ha	1894	4.99	0.49	+		+
ddap	dum	Dummy for dap fertilizer use, 1 if used	1891	5.68	0.25	+		+
durea	dum	Dummy for urea fertilizer use, 1 if used	1886	5.73	0.25	+		+
dmanure	dum	Dummy for manure use, 1 if used	1843	232.9	24. 8	+		+
dherbpest	dum	Dummy for herbicide & pesticide use, 1 if used	1918	0.1	0.01	+/-		+/-
oxen	count	Number of Oxen owned by the household per ha	1899	2.36	0.05	+	+/-	+
		Plot Characteristics						
distance	Cont	Distance from homestead to plot, minutes	1871	22.22	0.77	-	+/-	-
no.of.plot	count	number of household operated plots	1918	5.61	0.06	+/-	+	+/-
si-index	index	SI Index	1918	0.72	0	+/-	+	+/-
sfarmsz	cont	Total farm size less operated plot size; ha	1905	1.04	0.03	+/-	+	+/-
landqlty	catg	Land quality rank: 1=Poor, 2=Medium, 3=Good	1883	1.83	0.02	+	+/-	+
slope	catg	1= flat 2= moderate, 3= steep, 4= Very Steep	1883	1.34	0.02	-	+	-
soiltype	catg	1= Loam, 2= Clay, 3=Sand, 4=Silt	1879	2.55	0.03	+/-	+/-	+/-
erosion	dum	1=sever erosion, 4= less erosion	1913	3.2	0.03	+	+/-	+
soildepth	catg	1=deep, 2=medium, 3=shallow	1879	2.22	0.02	-/+	+/-	-/+
		Household Characteristics						
hh-size	count	Household size	1918	5.58	0.06	+/-	+	+/-
hh-edu	Cont	Education level of household head, years	1918	1.58	0.03	+/-	+/-	+/-
hh-age	Cont	Age of household head, years	1904	57.02	0.34	+/-	+/-	+/-
hh-sex	dum	Sex of household head; 1 if male, 0 otherwise	1889	0.81	0.01	+	+	+
fem.labor	Cont	Household female labor force per ha	1902	15.85	1.36	+	+	+
malelabor	Cont	Household male labor force per ha	1902	15.18	0.74	+	+	+
adulteqvt	cont	Household adult-equivalent labor force per ha	1902	46.2	2.99	+	+	+
		Other Control Variables						
village	dum *	Village Dummies	1918	10.75	0.13			
avgmedpr	cont*	Average Median Price	1911	6.49	0.04			
plotrent	dum	Dummy for participation in rental market; 1 if participating	1876	0.35	0.01	+/-	+/-	+/-
totalvisit	count	Number of plot visits for farm operations	1918	42.6	0.98	+		+
solevisit	count	Number of sole supervisions	1538	24.18	0.68	+		+
lvaiable		Log of variable						

Legends- 2nd row: Ob = Number of Observations; PR= Productivity; DIV=Crop Diversity;TE= Technical Efficiency

5.3. Defining Variables and Expected Signs

5.3.1. Defining the Dependent Variable

Analysis of land productivity is often considered as an indicator of production efficiency. The justification behind this approach is the belief that in subsistence farming there is no competition between farmers with variable land holding size. However, it is argued that since subsistence farmers produce not only for direct consumption but also for selling their products in order to purchase other goods and service from the market, their competence can be better evaluated if cost of production is also considered in the analysis (Niroula and Thapa, 2005). Hence, net financial gain per unit of land should be the criterion for evaluation of land use efficiency (Holden et al., 2001, Niroula and Thapa, 2005, Wattanutchariya and Jitsanguan, 1992).

Owing to this argument, net output value, the income earned from less cost incurred on each unit land is used as a dependent variable in this study for two reasons. First, as opposed to the use of quantity of output as a dependent variable, this approach allows a consistent inter-plot comparison and efficiency evaluation where farming involves production of diversified crops. It is also a consistent way of aggregating farm output at household level for inter-household comparison and efficiency evaluations, although it may be sensitive to price variations.

To handle price sensitivity, some authors tend to test robustness of results by running sensitivity analysis using major crop(s) only. However, application of such technique in a plot level analysis should be carried out with caution as it might lead to bias attributed to the likely loss of a significant number of observations.

Second, since the data has limited information about cost of production (for example, with respect to cost of labor, draft power, fertilizer and pesticide) 'net output value', which subtracts value of seed (purchased or own seed) from gross output value, is used instead of net-profit. Some authors treated 'amount of seed' or 'value of seed' as explanatory variable. However, the problem with such method is that it is susceptible to endogeneity bias since decision on quantity of seeds presupposes decision on seed selection which is also an endogenous feature of the dependent variable. Therefore, besides computing 'net output value' in this way, the method has the advantage of minimizing potential endogeneity problem.

Net output value is computed as follows:

$$Netput\_v = \text{Output Value} - \text{Cost of Seed}$$

$$Netput\_v = p_i(q_i - s_i)$$

where $Netput\_v$ is expressed in local currency (Birr) per unit area

q_i is output in kg of the i^{th} commodity harvested in a given plot

p_i is the annual median price per kg of commodity i

s_i is seed in kg of commodity i

For the purpose of converting the values of output and seed cost, price data is used from the Central Statistical Agency of Ethiopia where for all types of commodities the annual median retail price has been gathered from markets in seven main towns in Tigray during the 2010/2011 production year. For each crop type the annual median price is used. Strong assumption is made that for grain crops seasonal variation in price is not significant and hence seed value and output value are estimated at the median price. However, caution is needed with this assumption as the region was suffering from higher inflation during the specified production year. For vegetables and fruits, prices in the seed and output markets are usually not the same, as a result the respective prices are used.

Alternative to the 'net output value' approach, some authors use the "wheat-equivalent yield" approach to aggregate total farm output (measured in Kg/ha) and use as a dependent variable, for instance (Wu et al., 2005). This alternative is also used in this study to test robustness of the result. While price information is still necessary to convert all outputs to wheat-equivalent yields, it involves, however, use of price ratio. The value can be considered as transformed output. Wheat-equivalent yield is computed as:

$$Netput\_w = {}^{p_i}\!/_{p_w}\,(q_i - s_i)$$

Where $Netput\_w$ is Wheat-Equivalent Net Output in kg per unit area; p_w is the annual median price of wheat.

5.3.2. Land Fragmentation Indicators

The main explanatory variables on which this study is emphasizing are those indicators that can potentially capture land fragmentation. As explained in Chapter 3, there is no standard unit of measuring land fragmentation. It is difficult, especially, to compare two farmers in terms of degree of land fragmentation. Hence, the common consensus is to use potential indicators that can capture features of land fragmentations.

In the absence of standard measurement of land fragmentation several indicators have been used in the literature. The most commonly used and arguably potential indicators that appeared in the literature include: the number of plots, average plot size, distance to plot, and the Simpson Index (Monchuk et al., 2010, Tan et al., 2006). Some studies involved use of only one of these indictors while others used all of them. However, both extremes are open to critique for the following two reasons:

First, using only one indicator may not sufficiently capture the effect of land fragmentation since different aspects of land fragmentation (indicators) may show different results. For instance, recall the mixed result reported from China about economic cost of land fragmentation. Nguyen et al. (1996) reported significant economic cost of land fragmentation in China using average plot size as the only indicator of land fragmentation. On the other hand, Tan et al. (2008) uses three indicators and found mixed results. Specifically speaking, farm size and plot distance had significant effects on production cost while the number of plots and plot size distribution (captured by the SI) had no effect. Consequently, the effect of land fragmentation on production cost remains inconclusive. Similarly, using only one indicator, namely the number of plots, Rahman and Rahman (2009) found significant negative effect of fragmentation on productivity and efficiency. Nevertheless, it is argued that such method of quantifying land fragmentation is flawed as it ignores distance (Bentley, 1987).

Second, use of all of the indicators may lead to multicollinearity problem due to the direct relationship between plot number, average plot size and farm size. Taking these two points into account, the following three indicators are selected for the purpose of this study, namely: the number of plots (N), distance to plot (D) and the Simpson Index (SI). While the data on all of the indicators is available, average plot size is excluded to avoid potential collinearity problem.

In what follows a brief explanation of these indicators and their expected signs is presented.

Distance to plot (distance) – is the average distance from the homestead to the plots which is assumed to capture the spatial distribution of plots within a farm. Although inter-plot distance can have potential effect on productivity[8], there is a gap in the data to handle this distance where only about 10% of the sample plots are documented with GPS coordinates. Hence, the analysis is limited to the average homestead-plot distance. The distance variable is determined in minutes of walk as opposed to in kilometers for this is a more realistic representation of distance given the steep and hilly terrain features of the study area. However, such definition of distance might also be susceptible to measurement error bias with respect to two reasons: First, the estimated minutes of walk between homesteads to plots by the respondents are solely on the basis of *subjective* guesstimates. Hence, there might be possible bias compared to actual measurement, for instance, using timer. Second, where there is nearly accurate estimates, there might still be bias attributed to age, sex, and health status of the respondent. In other words, the same distance might take different time to walk for different persons. This suggests that some correction measures or assumptions are needed in dealing with distance as one of the main explanatory variable. Once the measures or assumptions are taken, distance to plot is expected to have a negative relationship with productivity due to loss of time and its effect on management operations, such as supervision and manure application, if labor is scarce.

The maximum distance to plot estimated by respondents is 210 minutes of walk with an average of 22.28 minutes. As illustrated in Figure 1C, the distribution is highly skewed to the left for both male and female households.

Number of plots (no.of.plots) – the number of plots is determined following the definition of plot provided in Chapter 2. Recall that it is a common practice in Tigray to divide parcels into temporary plots for a given season. Thus, this variable is comprised of not only the spatially separated parcels but also the various plots within a parcel. This definition of plot is consistent to the demand-side explanation of land fragmentation. Moreover, with respect to methodological analysis, this definition of plot can be consistently estimated with cross-sectional data setting, for the number of plots per household may vary over years with cross-sectional time-series data.

[8] Not only distance between homesteads and plots, but also inter-plot distance can have potential effect on productivity. If, for example, a farmer has two small plots situated at a closer distance to each other, s/he may not necessarily need to travel twice while it is possible to undertake farm operations in both plots simultaneously in just a single trip. Obviously, econometric models should take this into account. However, due to the limited data available, the effect of inter-plot distance on production is not included in this study. But it is recognized as a potential research gap for the future.

In other words, with this definition of plot, although more realistic, analysis of cross-sectional time-series data with panel data models would have been highly sensitive to attrition bias. This is because of the fact that, besides the likelihood of missing sample households over time, the number of plots per observed household may (endogenously) vary over years. Such attrition bias, however, can be minimized using cross-sectional data of a single year within which plot-panel analyses can be carried out using the standard panel data models to account for unobserved heterogeneity. Nonetheless, cautious treatment of endogeneity is required.[9]

The number of plots operated per household ranges between 1 and 18 with an average of 5.8. The distribution, as illustrated in Figure 1B, is skewed to the left for both male and female households.

The expected sign for this variable is ambiguous. On the one hand, a large number of plots may enable farmers benefit from variation in soil quality and local agro-climatic conditions as well as by smoothing their labor supply over the seasons. Moreover, at a given average plot size, acquiring more number of plots, for example through renting plots in, can be expected to have positive scale effects. On the other hand, for a given farm size, more number of plots implies that the farm is more fragmented that could make farming operation more difficult and could impede various technological adoptions. If the first effect exceeds the later, the overall impact of number of plots on productivity and efficiency will be positive; else it will be negative, and hence ambiguous. In fact these two results are also reflected commonly in the literature; see for instance Nguyen et al. (1996), Rahman and Rahman (2009) and Tan et al. (2010).

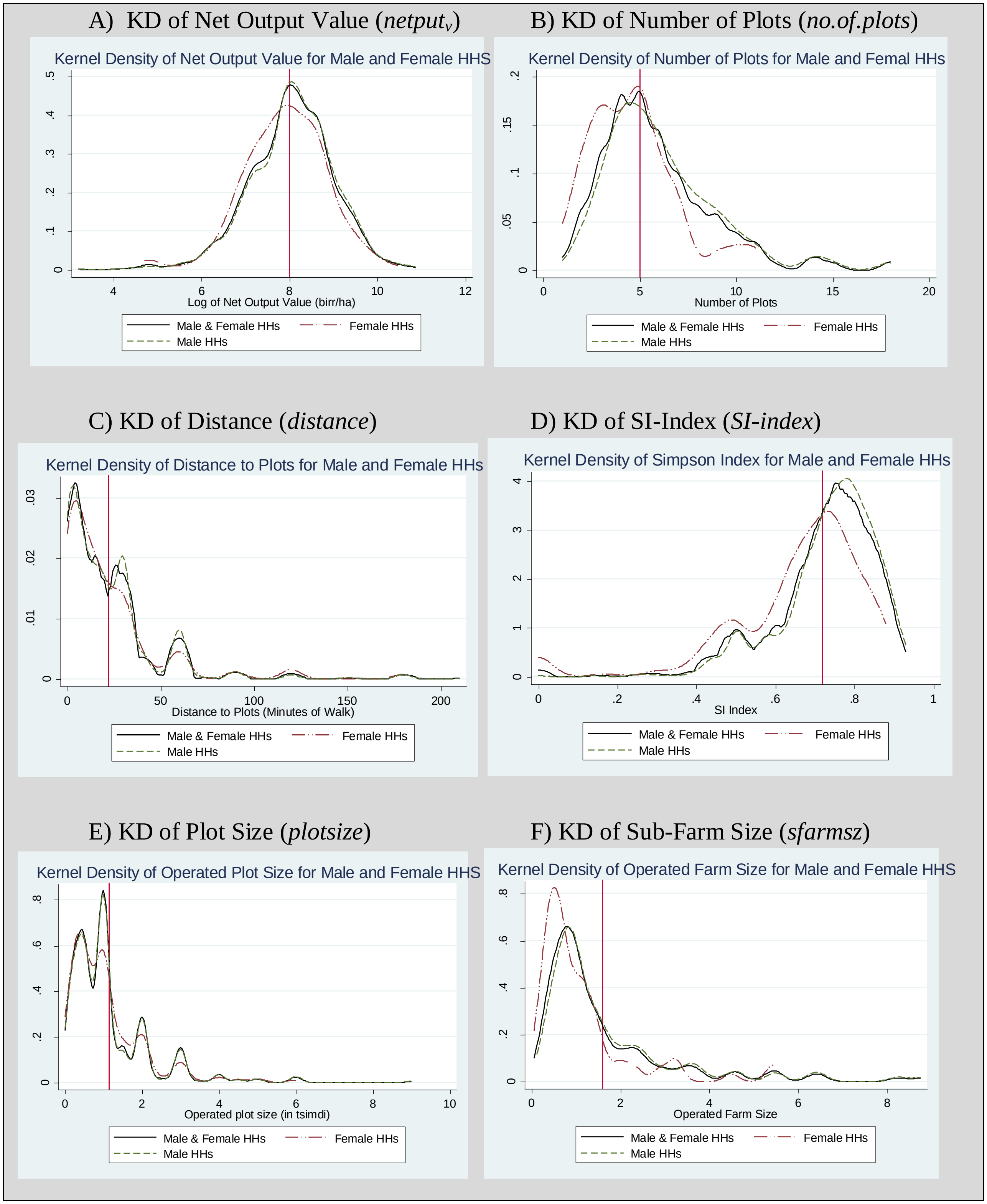
The Simpson index (SI-index) **-** is determined by the number of plots, average plot size, and the plot size distribution. The index does not capture farm size and distance to the plots. Mathematically, the index is defined as:

$$SI = 1 - \sum_{i=1}^{n} a_i^2 / (\sum_{i=1}^{n} a_i)^2$$

where n is the number of plots and a_i is the area of each plot. SI is the index between zero and one, with a higher value of SI indicating a larger degree of land fragmentation.

[9] Recall that our definition of plots includes the subjective demand-side plot fragmentations. This implies that decision to operate on more (or less) number of plots is partly endogenous to the household's farm decision and partly exogenously determined by external factors, such as population pressure, land policy and inheritance customs. Previous studies have already demonstrated that crop selection decision is endogenous and affects production (productivity) (Blarel et al. p244). So is assumed here with respect to the partial decision of households on how many plots to operate for a given production season.

Figure 1 Kernel Distribution (KD) of the dependent variable and land fragmentation indicators for male- and female-headed households



The expected sign for SI-index is also ambiguous. Higher index value (larger degree of fragmentation) can be positively associated with productivity through facilitating labor smoothing and risk management. It may also reduce productivity through constraining technological adoption and making management operations difficult.

The descriptive statistics on SI index shows that land fragmentation is highly pronounced in the study area. While the index ranges between 0 and 0.93, the average index value is 0.73 which is closer to 1 – the extreme index value of land fragmentation. Moreover, the distribution, as illustrated in Figure 1D is highly skewed to the right indicating that most of the observed households operate on highly fragmented plots.

5.3.3. Other Plot Characteristics

Operated plot size and sub-farm size: In analyzing the effect of land fragmentation on productivity, distinction is made between operated plot size and sub-farm size since both have important economic implications.[10] The variable sub-farm size is the difference between the total operated farm size and the operated plot size.

Earlier economic literatures suggested that large farms have cost advantage over small farms due to economies of scale; as a result small farms were believed to operate at subsistence (Ellis, 1993). But this conventional view was questioned by Schultz (1964) who argued that farm families in developing countries are "efficient, but poor" and thus "there is comparatively little significant inefficiency in the allocation of the factors of production in traditional agriculture." Nevertheless, with access to new technologies the "traditional agriculture" could change.

Recently, Deininger and Byerlee (2011: 28-30) identified three critical factors determining whether there are economies of scale in farming: access to credit and insurance, lumpy inputs (such as machinery and skills) and the nonagricultural wage rate. First, since land is ideal collateral and cost of borrowing is a declining function of owned farm size, larger farm size will have a cost advantage due to economies of scale. Second, farms that rely on their own machinery could produce economies of scale and increase the optimum operational farm size. Third, large

[10] In addition to operated plot size and farm size, owned plot and farm size are also recognized as potential for economic implication if there is significant difference between operated and owned plot size and farm size. A significant difference would suggest that significant size of land is left fallowed (or uncultivated for some reason), which, in turn, implies that there is significant endogenous decision on plot selection and operated farm size. This would cause selection bias and affect result of econometric models. However, the data does not support for significant difference between operated and owned plot size as well as farm size, for there are only few plots left fallowed. Hence, plot selection decision is treated synonymous to crop selection decision and is assumed to be explained by the factors included in the model.

scale farming can be profitable if wage rates go above a certain threshold level. For most of the 20th century, farm size and wage rate in the US moved closely in the same direction.

Earlier studies argued that in areas where labor is abundant, there could be an inverse relationship between productivity and operated land size, at the cost of labor inefficiency. However, Bhalla (1988) and Bhalla and Roy (1988) argued that once the econometric model allows to control for unobserved land heterogeneity and labor market, which was missing in earlier works, the inverse productivity relationship has to vanish. In other words, the inverse productivity hypothesis can be explained only by unobservable land quality and labor market imperfection, if the theory of economies of scale is to be consistent. Therefore, after controlling for land quality, plot size is expected to be positively and significantly correlated with productivity, if labor market is perfect and productivity of labor is efficient. A negative relationship would then manifest unobservable land qualities (Udry, 1996), labor market imperfection (Bhalla and Roy, 1988) or measurement bias (Holden et al., 2001).

Productivity of a given plot can be adversely affected if the total operated farm size is large given limited factors of productions. The forth coming models will include a variable that subtracted operated plot size from the total operated farm size to capture effect of factor competition (or coordination) between the different plots on productivity.

Land quality – is a dummy variable to control productivity differentials attributed to land heterogeneity. Farmers were allowed to rate their plots as “poor”, “medium” and “good” quality: 657 plots were rated as poor quality, 789 plots were rated as medium quality, and 437 plots were rated as good quality. A positive relationship is expected between land quality and productivity.

Slope, soil type and soil depth: In addition to rating land quality in general, farmers were also allowed to rate their plots in terms of slope, soil type and soil depth. As a result, 1459 plots were regarded as *gentle or flat land*, 299 plots as *moderately* inclined plots, and 116 plots as *steep* while 9 plots as *very steep*. In terms of soil type, 422 plots were categorized as *loamy* soil, 540 plots as *clay* soil, 381 plots as *sandy* soil, and 536 plots as *silt* soils. Similarly, in terms of depth, the numbers of plots classified as deep, medium and shallow were 535, 499, and 845, respectively. Including these variables in the model is assumed to capture productivity/efficiency variations attributed to plot quality. A positive relationship is expected if plots tend to be gentle.

The expected sign of soil type and soil depth is ambiguous depending on the type of crop cultivated.

5.3.4. Farm Inputs and Management Operations

Labor data is available in two forms: the total farm household labor used during the entire farm operation and labor hired from the labor market. In labor intensive farming system, the effect of farm labor on productivity is tremendous. In both cases, positive and significant relationship with productivity is expected.

Fertilizer: Some plots have used synthetic fertilizer (urea and/or dap), some have used organic fertilizer (manure) while others did not use at all. Of the total 1918 sample plots, the number of plots that used Urea, DAP, and manure fertilizers are 797, 784, and 536, respectively[11]. Given technically appropriate fertilizer application, the use of fertilizer, synthetic or organic, is normally expected to associate with productivity positively. However, it should also be noted that there may be a selection bias related to plots selected for fertilizer use which can make it impossible to get reliable estimates of the marginal effect of fertilizer on land productivity; hence dummy variables are used. Moreover, since the data on manure fertilizers posses mixed units of measurements and seemingly biased guesstimates, dummy variables are used.

Herbicide and Pesticide: this is a dummy variable used to capture the effect of herbicide and pesticide on productivity. Of the total 1918 plots, 225 plots used herbicides and/or pesticides. In areas where labor is scarce, herbicides can substitute labor devoted to weeding, inspection, and land preparations; hence a positive and significant effect on productivity is expected if labor is scarce. Unless an outbreak occurs, the common pesticide in the study areas is rodenticide. A positive relationship is expected if rodent or other pest infestation is a problem. In both cases a positive relationship is expected given herbicides and pesticides are appropriately applied.

Degree of management operations: there are two potential indicators of the degree of management in the data: the total number of plot visits during whole farm operations and the number of sole supervisions. The first captures the intensity of management operations, such as land preparation, weeding, etc. The second captures the degree of monitoring and supervision. Closer supervision and management is expected to associate with productivity positively.

[11] Note that there are overlaps here. Some plots actually used all or two of these fertilizers. The number of plots that did not use any of these fertilizers at all is 721 plots.

5.3.5. Household Characteristics

Age of household head (*hh-age*) – the effect of age on productivity is indeterminate depending on whether older farmers are more experienced and rich in farming knowledge or slower to accept new technology and relatively less energetic than their young counterparts. The average age of heads of households is around 56 years.

Sex of household head (*hh-sex*) – for female headed household productivity is likely to be lower compared to their male counterparts due to the problem they face with respect to male labor and oxen which are usually insufficient. Most of the time female headed households rent out their land to overcome this problem. However, studies also show that the rented out plots are relatively inefficient due to "lower enforcement ability" and "tenure insecurity" by female households (Bezabih and Holden 2006; Holden and Bezabih 2008). Therefore, whether it is a rented-out plot or own-operated plot, productivity is expected to be lower for female household heads. Of the total 1918 plots, 385 plots belong to female headed households.

Education of household head (*hh-edu*) – the effect of education on productivity can be positive or negative. It can be positive when a higher level of education leads to a better assessment of the importance and complexities of production decisions which, in turn, results in a better management of farming practice. It can be negative when farmers with higher education tend to pay more attention to the non-farm work. The years of education of household head ranges between 0 and 7 years of education with the average being 1.7 years.

Household size per ha – a large household size usually implies more laborers that can be devoted to management operations or more children and older persons that can be devoted to supervisions. In this case, a positive relationship is expected to be observed with productivity. On the other hand, age proportion and working ability of household members may affect this relationship. If a significant proportion of the household members are infants and very old or disabled persons, then it implies that more adult labor force, particularly female labor force, may be required devoted to childrearing, caring for elders and disabled persons. In this case, a negative relationship is will be observed. Hence, the expected sign of household size and its effect on productivity is ambiguous. The average household size in the sample is 5.6 ranging between 1 and 11 heads.

Household adult-equivalent labor force per ha: Alternative to household size is the household adult-equivalent labor force which takes into account only to those who can contribute to farm labor, including child labor. A positive and significant relationship with productivity is expected if the labor market operates less efficiently. Yet, age proportion of the members in the labor force matters. Noting that two child labors are assumed to be equivalent to one adult labor, the effect of adult equivalent labor force on productivity can be weak if it mainly comprised of child labor.

Adult labor force per ha: can be used as an alternative variable to adult-equivalent labor force. However, owing to the existence of a fairly strict gender division of labor in the study area, the household adult labor force is appropriately divided by gender between: *household male labor force per ha* and *household female labor force per ha*. Such cultural division of labor may reduce the substitutability of male and female labor. Consequently, the shortage of one type of labor may cause inefficiency if labor market does not function well. Moreover, one would need to control for off-farm activity to get the exact effect of male and female labor forces. In our case, data on off-farm and non-farm activity is not complete as result we are unable to control for it. But it is expected that if the labor market functions well, there will be no significant effect of male or female labor force on productivity. And if there is significant opportunity for off-farm and non-farm activities, a negative relationship between productivity and these labor forces can be observed. The average household adult, male and female labor forces per hectare in the sample are 29.1, 13.7 and 15.3, respectively.

Oxen Ownership: If a farm household owns oxen, it can prepare land more timely and carefully and is expected to be more efficient. If the owning household participates in oxen-rental market, a trade-off between production efficiency and benefits gained from the rental activities can be observed. If markets for draft power work efficiently, productivity differences attributed to oxen ownership will not be observed, since those who can afford to rent-in oxen can be equally efficient as owners of the oxen. So, the expected sign is positive but it can be insignificant unless it is controlled for market participation and the level of household income. The average number of owned oxen is 2.35 ranging between 0 and 9 oxen.

5.3.6. Other Control Variables

Village or wereda (district) dummies – are alternative dummy variables to capture the effect of differences in price, agro-climatic variations, ecological diversities, topographic differences, and differences in local land, labor, credit and food markets. The data consists of 17 villages and 11 districts. In the forthcoming models, where it is relevant, the *village* dummy will be used to clustering errors while alternatively *district* will be used to show district specific effects. More details on the techniques will be provided later in the Chapter 6.

Participation in Rental Market: is a dummy variable to control for the effect of participation in land rental market on productivity and efficiency. Economic theory suggests that *Marshallian inefficiency* is likely to be observed in share-tenancy arrangements causing lower input use and lower net output value on plots that are sharecropped (Kassie and Holden, 2007, Stiglitz, 1974). Since share-cropping is a common arrangement of land rental market in the study area, a negative sign is expected if Marshallian inefficiency is observed. However, the result might also be affected by selection bias which is an endogenous decision of renting in or out of selective plots (Kassie and Holden, 2007). So, care should be given in interpreting the results. Yet, like all the other endogenous variables in the model, we do not have suitable instrument(s) for this variable, however, all the possible factors that could affect renting in or out of plots are assumed to be incorporated in the model. Of the total 1918 plot observations, 785 plots are involved in land rental markets.

5.4. Summary of Variables

Dependent Variable:

- net output value ($netput_v$), net wheat equivalent yield gain ($netput_w$), crop diversity

Plot-varying explanatory variables

- distance, soil type, slope, soil depth, land quality, total farm labor, hired labor, dap, urea, manure, operated plot size, operated farm size, plot renting, herbicide and pesticide usage, total visit, sole visit

Plot-invariant explanatory variables

- HH size, HH head’s age, HH head’s sex, HH head’s education, SI index, number of plots, male labor force, female labor force, adult-equivalent labor force, owned oxen

6. ESTIMATION METHODS

In the first chapter, three research questions were explicitly stated to which this study is intended to address. So far we are equipped with the relevant background information, theoretical models, empirical reviews, as well as explanations of the data, variables and their expected signs. In what follows the estimation procedures that are followed to address each of the research questions will be discussed.

6.1. Productivity Analysis Using the Farm Household Model

Productivity analysis using the FHH model is carried out based on equation (4.14) which is provided in Chapter 4:

$$\boldsymbol{Y} = \boldsymbol{f}\left(\alpha_i, \boldsymbol{Q}\left(\boldsymbol{z}\left(\bar{\boldsymbol{z}}, p_q, p_j, p_z'; \zeta, \varsigma, v\right); \zeta\right)\right) \tag{4.14A}$$

$$\boldsymbol{Y} = g(\bar{\boldsymbol{z}}, \alpha_i, \zeta, \varsigma, v) \tag{4.14B}$$

Where $\boldsymbol{Y}$ is a vector of net output values as defined in Chapter 5.3.1; $\boldsymbol{z} \in \{A_i, L_i, X_i, K_i\}$ are the actual factor inputs employed in the production process**;** $\bar{\boldsymbol{z}} = \{\bar{A}_i, \bar{L}_i, \bar{K}_i, \bar{X}_i\}$ is the set of household resource endowments**;** α_i is the household specific agricultural ability which is unobservable; ζ, ς and v are household, plot and village specific characteristics, respectively.

Three functional forms are employed in analyzing the production function in (4.14):

i. The *full translog* form

$$lnY_{ip} = ln\boldsymbol{X}_{ip}^{'}\beta + \alpha_i + u_{ip} \tag{6.1}$$

ii. The *semi translog* form with *some* of the independent variables being in linear form

$$lnY_{ip} = ln\boldsymbol{X}_{ip}^{'}\beta_1 + \boldsymbol{X}_{ip}^{'}\beta_2 + \alpha_i + u_{ip} \tag{6.2}$$

iii. The *Log-Lin* form with *all* of the independent variables being in linear form

$$lnY_{ip} = \boldsymbol{X}_{ip}^{'}\beta + \alpha_i + u_{ip} \tag{6.3}$$

where lnY_{ip} is log of the dependent variable, for household *i* operating on plot *p*, defined in two ways – either in terms of net output value (netput_v) expressed in birr/ha or wheat-equivalent net output (netput$_w$) expressed in quintal/ha; $\boldsymbol{X}_{ip}$ is a vector of exogenous variables that are observed for household i on plot p, the vector includes both plot variant as well as invariant variables and dummy variables; α_i are the unobserved household specific effects; u_{ip} are the idiosyncratic errors.

Three additional models are also identified based on crop output considered in the analysis: (*a*) the aggregated output model where all outputs are converted in to monetary value or wheat-equivalent yield as discussed in Chapter 5; (*b*) the dominant crop (*teff*) model where analysis is carried out only using net output (value) of the dominant crop; and (*c*) the two-main crop (*teff-wheat*) model where analysis is carried out based on net output value and wheat equivalent yield of the two crops.

The intention is to evaluate the results whether they are robust to changing functional forms, econometric specifications and output types. The variables of interest being the land fragmentation indicators, various estimation methods are employed within the context of panel data model.

The first estimation method employed is the Pooled OLS model; where for the *full translog* form, for example, can be demonstrated as:

$$lnY_{ip} = ln\boldsymbol{X}_{ip}^{'}\beta + v_{ip}; \text{where } v_{ip} = \alpha_i + u_{ip} \qquad (6.4)$$

The model assumes that all exogenous variables in the model are uncorrelated with the unobservable household specific characteristics as well as the idiosyncratic errors $\left(E\left[\boldsymbol{X}_{ip}\alpha_i\right] = E\left[\boldsymbol{X}_{ip}u_{ip}\right] = 0\right)$. This method will not give consistent estimates if this assumption failed to hold. While strong assumption to maintain practically, the model is used as a reference and evaluate the alternative estimators and properties of the residuals.

The second alternative model considered is the Random Effects (RE) model

$$lnY_{ip} = ln\boldsymbol{X}_{ip}^{'}\beta + \alpha_i + u_{ip} \qquad (6.5)$$

which requires the assumptions of strict exogeneity and orthogonality to obtain consistent estimates of $\boldsymbol{\beta}$. Strict exogeneity is the assumption that the idiosyncratic error term should not correlate with any of the regressors, observed or unobserved, in all plots: $\left(E\left[u_{ip}|\alpha_i, X_{i1}, \dots X_{iP}\right] = 0\right)$. Orthogonality is the assumption that the unobserved heterogeneity, or random household specific effect, should not correlate with any of the observed regressors: $(E[\alpha_i|X_{i1},\dots,X_{iP}] = E[\alpha_i] = 0)$. RE estimators are consistent *only* if these two assumptions hold. In the standard random effect analysis a further assumption of homoscedasticity and no serial correlation is required for statistical inference.

Further, if normal distribution of the error terms can be assumed, more efficient random effects can be estimated by the maximum likelihood (ML) method. However, since plot level activities, for a given household, cannot be entirely independent and uncorrelated, the *iid* and normality assumptions may not hold. As a result the Random Effects Feasible Generalized Least Square (RE FGLS) estimator is proposed alternative to the standard and ML random effect estimators, since it is efficient under the first two assumptions regardless of heteroskedasticity and auto-correlations. As for the first two assumptions of RE model, the Hausman test can be used.

The third alternative is the Fixed Effects (FE) model

$$lnY_{ip} = ln\boldsymbol{X}'_{ip}\beta + \alpha_i + u_{ip} = ln\boldsymbol{X}'_{ip}\beta + v_{ip} \tag{6.6}$$

which allows the observed regressors to correlate with household specific fixed effects, i.e., it does not require the orthogonality assumption. Under the strict exogeneity assumption the fixed effects within estimator $\left[\ln(Y_{ip} - \bar{Y}_i) = \ln(\boldsymbol{X}_{ip} - \bar{\boldsymbol{X}}_i)'\beta + (u_{ip} - \bar{u}_i)\right]$ can provide consistent estimates of $\boldsymbol{\beta}$ for plot-varying regressors, despite allowing weak endogeneity of regressors. For a short panel the household specific fixed effects may not be consistently estimated. For the standard within estimator a further assumption of homoscedasticity and no serial correlation is required for statistical inference and efficiency of estimates. Where the last two assumptions do not hold, a robust variance matrix can be computed by clustering standard errors at the village level.

Practically, the RE model has an advantage over the FE model *if* its assumptions hold, for it provides more efficient estimates including for the household specific effects and the plot invariant regressors. But, the orthogonality assumption of RE is less likely to hold practically and thus FE estimator may be preferred at a cost of losing estimates of plot-invariant regressors. But since two of the three variables of our primary interest are plot invariant, namely the number of plots and the SI index, unfortunately, the fixed effect model will not give us the most required information. Thus, while still testing the two models is relevant using the Hausman test, other alternative models are also proposed in case the test rejected the orthogonality assumption of the RE model.

The Hausman-Taylor (HT) model is the fourth alternative model employed to address the first objective of this study. Unlike the two extremes (the RE model which assumes exogeneity of *all* regressors and the random household specific effect or the FE model which allows endogeneity of *all* the regressors and the household specific fixed effect), the HT model allows *some* of the regressors to be correlated with the household specific effect (Baltagi et al., 2002). The HT estimators are based upon an instrumental variable estimator which uses both the individual means of the strictly exogenous variables as instruments. The choice of strictly exogenous regressors is, therefore, of crucial importance that needs to be tested. Testing can be carried out using Hausman test that compares performance of the FE and HT estimators.

6.2. Analysis of Crop Diversity

Analysis of crop diversity is carried out using diversity *richness* as a dependent variable which is the total number of crops produced at household level. The variable is plot invariant and has the feature of count variables where its relationship with the exogenous variables may not be appropriately handled with linear regression models. Since OLS regression can predict negative or non-integer values, distinct methodology for limited dependent variables is generally recommended to deal with such data (Wooldridge, 2010). To this end, the Poisson and the negative binomial regression models within panel data framework are nominated to this analysis.

Derived from the reduced farm household model of crop diversity in (4.15) the general structure of the regression equation for on-farm crop diversity analysis can be expressed in a plot-panel form as:

$$D_i = \boldsymbol{X}'_{ip}\beta + \alpha_i + u_{ip} \tag{6.7}$$

where D_i represents the diversity richness of the i^{th} household; $\boldsymbol{X}_{ip}$ is a vector of exogenous variables that accounts for household, farm and village characteristics. Included in farm characteristics are the various plot characteristics that determine crop selection including the three land fragmentation indicators of our prime interest; α_i are the unobserved heterogeneity, and u_{ip} are the idiosyncratic errors.

The vector $\boldsymbol{X}_{ip}$ consists of plot variant and invariant explanatory variables, summarized as follows along with their expected signs (also summarized in Table 3):

Plot-variant variables: Distance (+/-), soil type (+/-), slope (+), soil depth (+/-), land quality (+/-), operated farm size (+), plot renting (+/-)

Plot-invariant variables: household size (+), age (+/-), sex (+), education (+/-), SI index (+), number of plots (+), owned oxen (+/-), male labor force (+), female labor force (+), adult-equivalent labor force (+), operated farm size (+)

Using the basic Poisson regression model, diversity richness (D_i), the non-negative dependent (count) variable, is assumed to have a Poisson distribution that is completely determined by the conditional mean $\mu(\boldsymbol{X})$ having a density function of the form:

$$f(D_i|\boldsymbol{X}) = \frac{e^{-\mu(X)}[\mu(\boldsymbol{X})]^D}{D!} \qquad (6.8)$$

where $\mu(\boldsymbol{X}) > 0$ and $\mu(\boldsymbol{X}) \equiv E\left(D_i|\boldsymbol{X_{ip}}\right)$; D_i is the diversity richness for each household and $\boldsymbol{x_{ip}}$ is a vector of observed exogenous variables including dummy variables.

The Poisson distributional assumption imposes restrictions on the conditional moments of D_i. Typical of such restriction is the equality of the conditional mean and variance of D_i:

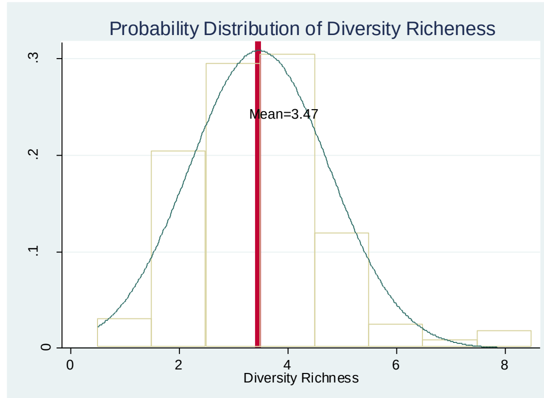
$$var\left(D_i|\boldsymbol{X_{ip}}\right) \equiv E\left(D_i|\boldsymbol{X_{ip}}\right) \qquad (6.9)$$

implying that the usual assumption of homoscedasticity is not appropriate for Poisson data. However, Wooldridge (2010) suggested that this restriction has often violated in applications. As a result, weaker assumptions are proposed to allow for any positive constant variance-mean ratio: $var\left(D_i|\boldsymbol{X_{ip}}\right) \equiv \sigma^2 E\left(D_i|\boldsymbol{X_{ip}}\right)$. Empirically, σ^2>1 (over-dispersion) is relevant when the variance is greater than the mean; and $\sigma^2 < 1$ (under-dispersion) is relevant when the variance is less than the mean (*Ibid*).

In our case, the sample data is characterized by (under-dispersion), which is not common in empirical works (Wooldridge, 2010). Being under-dispersion (or over- dispersion) implies, however, that there is an extra heterogeneity across the Poisson mean which, in turn, may result in less efficient results. Consequently, two alternatives methods are considered to deal with it. The first method uses the Poison-normal model which keeps the above weaker assumption that allows under-dispersion and makes further assumption that the under-dispersion is normally distributed. Where the under-dispersion is not normally distributed, assuming it may be gamma distributed, the alternative negative binomial model is used which is also analogous to Poisson gamma model. In short, the Negative binomial model is used in case the restriction on equality of mean and variance is not necessary. The following graph illustrates the distribution of D around its mean (3.47) with variance of (1.7).

Figure 2 Distribution of Diversity Richness:

The values of Diversity Richness ranges between 1 and 8 with mean 3.47 and standard deviation of 1.3. While under-dispersed, the distribution seems normal, which is the rationale behind using the Poisson normal model.



I used common mean function in applications, the exponential (the log-linear) parametric model:

$$E\left(D_i|\boldsymbol{X_{ip}},\alpha\right) = \exp\left(\boldsymbol{X}'_{ip}\boldsymbol{\beta}\right) = m(\boldsymbol{X_{ip}},\boldsymbol{\beta}) \text{ ; Poisson Pooled Model} \qquad (6.10A)$$

$$E\left(D_i|\boldsymbol{X_{ip}},\alpha_i\right) = \alpha_i m(\boldsymbol{X_{ip}},\boldsymbol{\beta}); \quad \text{Random Effects (RE) Poisson Model} \qquad (6.10B)$$

Estimation of the parameters under the Poisson normal model is carried out using the Conditional Maximum Likelihood Estimators (CMLE) which is generally fully efficient for random samples under the assumptions of Poisson distribution (Wooldridge 2010).

Once the parametric models of 6.10 are estimated using the Pooled Poisson and the RE Poisson regressions in STATA, the partial effect of a continuous variable x_j is interpreted as:

$$\frac{\partial E(D_i|\boldsymbol{X_{ip}})}{\partial x_j} = \exp\left(\boldsymbol{X}'_{ip}\boldsymbol{\beta}\right)\beta_j \qquad (6.11A)$$

which also implies that the partial effects of x_j on $E(D_i|\boldsymbol{X_{ip}})$ depend on $\boldsymbol{X}$; β_j is interpreted as:

$$\beta_j = \frac{\partial E(D_i|\boldsymbol{X_{ip}})}{\partial x_j} . \frac{1}{E(D|\boldsymbol{x})} = \frac{\partial \log[E(D_i|\boldsymbol{X_{ip}})]}{\partial x_j} \qquad (6.11B)$$

implying that $100\hat{\beta}_j$ is the semielasticity of $E(D_i|\boldsymbol{X_{ip}})$ w.r.t. x_j.

Furthermore, since using the parametric model in 6.10 is analogous to using $\log(D_i)$ as a dependent variable in linear regression model (Wooldridge 2010), a log-linear regression model of this sort is also used as a starting point of analyzing the diversity richness. The model is adjusted for heteroskedasticity and autocorrelations using the robust variance-covariance matrix by clustering error terms at the village level. For the limited dependent variable model bootstrapping method is applied by resampling households using 200 replications.

In order to compare the Poisson estimates with the linear model estimates, average partial effect (APE) of the Poisson estimate is used which is roughly provided by $\overline{D}_i\hat{\beta}_j$.[12]

6.3. Efficiency Analysis Using the Stochastic Frontier Model

This model can be applied both with cross-sectional as well as panel data sets. As Coelli (1995) suggested in the presence of substantial measurement errors and weather-related random disturbances in analyzing plot level data, stochastic frontier production model can be applied to estimate the technical efficiency scores for different plots. Following this suggestion, I applied cross-sectional as well as panel data frontier models on the data to analyze effect of land fragmentation on technical efficiency of farm producers.[13]

6.3.1. Cross-Sectional Stochastic Frontier Models

Recall, from chapter 4, that the frontier model involves two stage analyses. In the first stage, estimation of the production function is carried out by assuming the production function to be explained by the standard Cobb-Douglas and/or semi-translog functional forms. I used these functional forms since they are the common forms in stochastic frontier models (Greene, 1990). It is argued that functional specification has a noticeable though rather a small impact on estimated efficiency (Wu et al., 2005).

Consider a stochastic production function model over a cross-sectional observation expressed as:

$$y_i = f(x_i, \beta)e^{\varepsilon_i} \tag{6.12A}$$

$$lny_i = \alpha + X_i'\beta + \varepsilon_i \tag{6.12B}$$

Where y_i is the scalar output of plot $i, i = 1, \dots, n$, x_i is a vector of N inputs used in plot i, $f(x_i, \beta)$ is the production frontier, β is a row vector of an unknown parameter to be estimated, α is the intercept, and ε_i is a stochastic error term consisting of two independent elements:

$$\varepsilon_i = v_i - u_i \tag{6.13}$$

[12] I used the *margin* command in STATA to compute the APEs of all x_js and evaluating the results.

[13] The main reference for the econometric specifications in this session is based on Kumbhakar and Lovell (2003)

where v_i is the symmetrical component that accounts for random variation in output of the i^{th} plot due to factors outside the farmer's control, such as weather and disease. v_i is conventionally assumed to be independently and identically distributed (*iid*) as $N(0, \sigma_v^2)$. (NB: This assumption will be maintained throughout the paper). u_i is the non-negative technical inefficiency component of the error term. It is non-negative because theoretically no farmer can exceed the ideal *frontier* of perfect efficiency (zero inefficiency).

Four common distributional assumptions emerge in the literature w.r.t. u_i: half normal, exponential, Truncated Normal, and gamma distributions. Depending on the distributional assumptions of u_i, different models emerge (Kumbhakar and Lovell, 2003).[14]

Assuming the half normal distribution of u_i and an independent distribution of v_i and u_i w.r.t. the regressors (x_i) and to each other, the frontier of each plot is given by combining (6.12) and (6.13) as:

$$Y_i = f(x_i, \beta) e^{v_i - u_i} \tag{6.14}$$

With a variance and standard deviation of ε is given by

$$\sigma^2 = \sigma_u^2 + \sigma_v^2 \qquad \text{and} \tag{6.15}$$

$$\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2} \tag{6.16}$$

The ratio of the standard errors being

$$\lambda = {}^{\sigma_u}/_{\sigma_v} \tag{6.17}$$

The measure of efficiency at individual plot level can be obtained from the error terms ($\varepsilon_i = v_i - u_i$) as a function of the expected value of u_i conditional on ε_i:

$$E[u_i|\varepsilon_i] = \left(\frac{\sigma_u \, \sigma_v}{\sigma}\right) \left[\frac{\phi(\varepsilon_i\lambda/\sigma)}{1-\Phi(\varepsilon_i\lambda/\sigma)} - \frac{\varepsilon_i\lambda}{\sigma}\right] \tag{6.18}$$

where $\Phi(.)$ and $\phi(.)$ are the standard normal cumulative distribution and density functions evaluated at $(\varepsilon_i\lambda/\sigma)$.

[14] Kumbhakar and Lovell (2000) argued that the first assumption $[v_i \sim iid\, N(0, \sigma_v^2)]$ is conventional; the second assumption $[u_i \sim iid\, N^+(0, \sigma_u^2)]$ is based on the plausible proposition that the mean value of technical inefficiency is zero, with increasing values of technical inefficiency becoming increasingly less likely. Thus, modeling using the other assumptions can be used alternatively to check robustness of the results to changing assumptions.

Once point estimates of u_i are obtained, in the second stage estimates of the Technical Efficiency (TE) scores for each plot can be computed as:

$$TE_i = exp[E\{u_i|\varepsilon_i\}], = exp\{-\hat{u}_i\} \qquad (6.19)$$

where $\hat{u}_i$ is $E(u_i|\varepsilon_i)$.[15]

6.3.1.1. Determinants of Technical Efficiency

Once consistent Technical Efficiency Score is computed for each plot, one can identify the determinant of technical efficiency by setting the efficiency score as a function of household, farm and village characteristics:

$$TE_i = z_i\gamma + \eta_i \qquad (6.20)$$

where z_i is a $(1xk)$ vector of household, farm and village characteristics, including land fragmentation, that may influence technical efficiency of plots and that does not contain any variable in x_i, γ is a $(kx1)$ vector of unknown parameter to be estimated, and η_i is a random disturbance term assumed to be *iid* as $N(0, \sigma_\eta^2)$.

6.3.1.2. Model Specifications and Distributional Assumptions

All the parameters involved in the model ($\alpha, \beta, \gamma, \sigma_v, \sigma_u$ and TE) are estimated using the standard maximum likelihood Estimators (MLE). Given the specified assumptions of distribution, the MLE is generally the most efficient estimation procedure in the class of estimators that use information on the distribution of the endogenous variable conditional on exogenous variables (Wooldridge, 2010). As a result the MLE is widely used for its consistency and efficiency (Kumbhakar and Lovell, 2003); hence the reason for using it in this study.

Kumbhakar and Lovell (2003:73) argued that for a cross-sectional data setting "the estimates of technical efficiency in (6.19) are less consistent because the variation associated with the distribution of $(u_i|\varepsilon_i)$ is independent of *i*." In order to overcome this limitation of cross-sectional data other distributional assumptions on u_i are suggested including the exponential, truncated-normal and gamma distributions. I used the half-normal and the exponential distributional

[15] Kumbhakar and Lovell, 2000 argued that the "estimates of technical efficiency are inconsistent because the variation associated with the distribution of $(u_i|\varepsilon_i)$ is independent of i. unfortunately this appears to be the best that can be achieved with cross-sectional data" (pp78). This might be one reason for the mixed findings in the literatures of land fragmentation efficiency analysis with the Stochastic Frontier Model.

assumptions. Since the cross-sectional estimates can generally perform poorly in the presence of plot specific unobservable effects, I cannot rely totally on this approach to evaluate the effect of land fragmentation on technical efficiency; thus, the reason to employ the panel frontier model.

6.3.2. Efficiency Analysis Using Panel Data Stochastic Frontier Models

An important advantage of panel data model is its ability to control for plot specific unobservable effects which can be correlated with other explanatory variables in the model, given assumptions about the unobservable effect hold.

In general, the principle of panel data model is the same as that of the cross-sectional model. Careful specification of models, functional forms and distributional assumptions are still required except that some of the distributional assumptions of u_{ip} can be relaxed in the panel data model.

The general form of panel data frontier model is given by:

$$y_{ip} = f(x_{ip}, \beta)e^{\varepsilon_{ip}} \tag{6.21A}$$

$$lny_{ip} = \alpha_i + X_i'\beta_{i1} + X_{ip}'\beta_{i2} + v_{ip} - u_{ip} \tag{6.21B}$$

where y_{ip} is the net output value gained from plot $p, p = 1, \ldots, P$ by household $i, i = 1, \ldots, N$; x_i are vectors of household-level (plot-invariant) regressors; x_{ip} are vectors of household-plot-level regressors, α_i are household-specific effects (for pooled OLS $\alpha_i = \alpha$, which is an intercept common to all producers); $\varepsilon_{ip} = v_{ip} - u_{ip}$ is the stochastic error term consisting of two independent elements; v_{ip} is the idiosyncratic error which is always assumed to be *iid* $N(0, \sigma_v^2)$; u_{ip} is error component consisting of plot level non-negative technical inefficiency and unobserved heterogeneity. When the model is specified in log form, u_{ip} is interpreted as the percentage deviation of observed performance from the ideal frontier (Greene, 2005).

Distributional and independent assumptions of u_{ip}, though often flexible in panel data model, are the same as in cross-sectional frontier model. They vary between half normal $N^+(0, \sigma_u^2)$, exponential, gamma and truncated normal distributions. I used all assumptions except the gamma distribution in analyzing robustness of the results with the half normal distribution being the central assumption. Moreover, w.r.t. assumptions on technical efficiency, I used both plot variant and plot invariant technical efficiency models to further test robustness of my results.

The *plot variant* panel model allows technical inefficiency and unobserved heterogeneity components of u_i to vary with plots while the *plot invariant* panel model does not. Econometric models usually assume the later (Kumbhakar and Lovell, 2003). I analyzed both models with Fixed Effect and Random Effect Models.

6.3.2.1. Plot Invariant Fixed Effects Model

The simplest panel data model is the fixed effect model which allows the unobserved heterogeneity to correlate with the explanatory variable. In order to adopt this model to frontier analysis, a minor modification on the form of u_i is required such that $u_i \geq 0$. While v_{ip} is still assumed to be *iid* $N(0, \sigma_v^2)$, distributional assumption is not needed for u_i and it is allowed to correlate with the plot variant regressors X_{ip} and the error component v_{ip}. Since it is treated as a fixed effect, it will be part of the household specific intercept to be estimated along with the other parameters. The model can be estimated by applying MLE on:

$$lny_{ip} = \alpha_i' + X_{ip}'\beta_{i2} + v_{ip} \tag{6.22}$$

where $\alpha_i' = \alpha_i + u_i$ are household specific intercepts. (6.23)

By applying within estimator on (6.22) and employing normalization we get:

$$\hat{\alpha}_i = \max_i\{\hat{\alpha}_i'\} \text{ , and} \tag{6.24}$$

$$\hat{u}_i = \hat{\alpha}_i - \hat{\alpha}_i' \tag{6.25}$$

which ensures that $\hat{u}_i > 0$ and producer specific technical efficiency to be

$$\widehat{TE}_i = exp\{-\hat{u}_i\} \tag{6.26}$$

Determinants of efficiency are then identified by regressing[16]

$$\widehat{TE}_i = z_{ip}\gamma + \eta_{ip} \tag{6.27}$$

6.3.2.2. Plot Invariant Random Effects Model

Unlike the fixed effect model, the random effect model does not allow u_i to correlate with either the exogenous variables or v_{ip}. However, analogous to the fixed effect model, v_i is still assumed *iid* $N(0, \sigma_v^2)$ and u_i is still required to be non-negative with no requirement for distributional

[16] Note that STATA result provides technical inefficiency.

assumption, although it is required to be randomly distribute with constant mean and variance. The virtue of this randomness assumption is that it allows plot invariant regressors to be part of the model (Kumbhakar and Lovell, 2003).

Mathematical formulation of the model can be illustrated by rewriting equation (6.21B) which is the general form of panel data model:

$$lny_{ip} = \alpha_i + X_i'\beta_{i1} + X_{ip}'\beta_{i2} + v_{ip} - u_{ip} \qquad (6.21B)$$

Recall that u_{ip} is assumed to be plot invariant; rewriting this equation as

$$lny_{ip} = [\alpha_i - E(u_i)] + X_i'\beta_{i1} + X_{ip}'\beta_{i2} + v_{ip} - [u_i - E(u_i)] \qquad (6.28A)$$

$$= \alpha_i^* + X_i'\beta_{i1} + X_{ip}'\beta_{i2} + v_{ip} - u_i^* \qquad (6.28B)$$

where $\alpha_i^* = \alpha_i - E(u_i)$ and $u_i^* = u_i - E(u_i)$

Equation (6.28B) shows that plot invariant regressors can be included in the model. By applying FGLS estimator on (6. 28B), which is often appropriate for large N, we can estimate $\hat{\alpha}_i^*, \hat{\beta}_{i1}$ and $\hat{\beta}_{i2}$; and from the residuals $\hat{u}_i^*$ can be estimated as

$$\hat{u}_i^* = E[lny_{ip} - \hat{\alpha}_i^* - X_i'\hat{\beta}_{i1} - X_{ip}'\hat{\beta}_{i2}] \qquad (6.29)$$

By means of normalization of equation (6.29), estimates of $\hat{u}_i$ can then be obtained as

$$\hat{u}_i = \max_i\{\hat{u}_i^*\} - \hat{u}_i^* \qquad (6.30)$$

As usual, technical efficiency is then estimated as

$$\widehat{TE}_i = exp\{-\hat{u}_i\} \qquad (6.26)$$

and the determinants of efficiency are obtained by regressing

$$\widehat{TE}_i = z_{ip}\gamma + \eta_{ip} \qquad (6.27)$$

6.3.2.3. Plot Variant Fixed Effects and Random Effects Models

The assumption that technical efficiency is constant and does not vary along with plots is a strong one given the fact that plots may vary considerably in size and quality, and that factors of production are deployed accordingly. As the number of plots increase, it is unlikely that one would find this assumption to be tenable. As a result, one would require making this assumption flexible in order to allow technical efficiency to vary with plots.

Equation (6.21B) can be slightly modified to accommodate for this flexible assumption as

$$lny_{ip} = \alpha_p + X_i'\beta_{i1} + X_{ip}'\beta_{i2} + v_{ip} - u_{ip} \tag{6.30}$$

So that α_p (as apposed to α_i which is the random household specific effect) entered the model to represent the common intercept of farmers who operate on the p^{th} plot (where p =1,…P) (Cornwell et al., 1990, Kumbhakar and Lovell, 2003:108). Rewriting equation (6.30) gives the structural form of plot variant panel model:

$$lny_{ip} = \alpha_{ip} + X_i'\beta_{i1} + X_{ip}'\beta_{i2} + v_{ip} \tag{6.31}$$

$$\alpha_{ip} = \alpha_p - u_{ip} \tag{6.32}$$

where u_{ip} is plot varying technical inefficiency component of the error term, and α_{ip} is the intercept for producer i operating on plot p.

The standard two stage method of frontier analysis is no more technically feasible for a $N \times P$ panel to obtain estimates of all N*P intercepts (α_{ip}), slope parameters and variances (Kumbhakar and Lovell, 2003). Instead, Cornwell et al. (1990) suggested an approach to deal with by specifying

$$\alpha_{ip} = \varphi_{i1} + \varphi_{i2}p + \varphi_{i3}p^2 \tag{6.33}$$

The quadratic specification of α_{ip} minimizes the number of intercept parameters to be estimated to N*3 and allows technical efficiency to vary with plots and in a different manner for each producer.

If $\varphi_{i2} = \varphi_{i3} = 0$ for all i, then the model collapses to plot-invariant technical efficiency model.

If $\varphi_{i2} = \varphi_2\ \forall i$ and $\varphi_{i3} = \varphi_3\ \forall i$, then the model becomes fixed effect model with household specific intercepts φ_{i1} and a quadratic term in plots common to all households given by $(\varphi_i p + \varphi_i p^2)$ (Kumbhakar and Lovell, 2003). This implies that either technical efficiency is household specific that varies with plots in the same manner for all households or it is household specific and plot invariant with the quadratic term capturing effect of technical change.

Estimation can be carried out in several ways. A random and fixed effect approaches are suggested by Cornwell et al. (1990). In order to show how, first equations (6.31) and (6.33) will be nested to provide

$$lny_{ip} = X_i'\beta_{i1} + X_{ip}'\beta_{i2} + w_{ip}'\partial_i + v_{ip} \tag{6.34}$$

where $w_{ip}' = [1, p, p^2]$ and $\partial_i' = [\varphi_{i1}, \varphi_{i2}, \varphi_{i3}]$

Two strategies are proposed to estimate plot varying technical efficiency. First, by ignoring u_{ip} from (6.31) and consequently $w_{ip}'\partial_i$ from (6.34), we estimate $\beta_{nj}s$. Then, ∂_i can be estimated by regressing the residuals $(lny_{ip} - X_i'\hat{\beta}_{i1} - X_{ip}'\hat{\beta}_{i2})$ for household i on w_{ip}'. The fitted values from this regression provide $\hat{\alpha}_{ip}$, which is a consistent estimator of α_{ip} in (6.33) as $P \rightarrow \infty$. Then, estimate of α_p can be provided by:

$$\hat{\alpha}_p = \max_i\{\hat{\alpha}_{ip}\}; \tag{6.35}$$

and from (33) it follows that:

$$\hat{u}_{ip} = \hat{\alpha}_p - \hat{\alpha}_{ip} \tag{6.36}$$

Needless to say, once $\hat{u}_{ip}$ is estimated, plot varying technical efficiency can be estimated by

$$\widehat{TE}_{ip} = exp\{-\hat{u}_{ip}\} \tag{6.26}$$

and the determinants of plot varying technical efficiency will be obtained by regressing

$$\widehat{TE}_{ip} = z_{ip}\gamma + \eta_{ip} \tag{6.27}$$

All models are estimated using the integrated statistical software STATA 2012.[17]

[17] For the panel data stochastic frontier model, I also used the sfpanel (Stochastic frontier models for panel data) package, version 1.3.2., designed by Belotti, Federico, et al. "Stochastic frontier analysis using Stata." (2012).

7. RESULTS AND DISCUSSION

7.1. The Effect of Land Fragmentation on Productivity

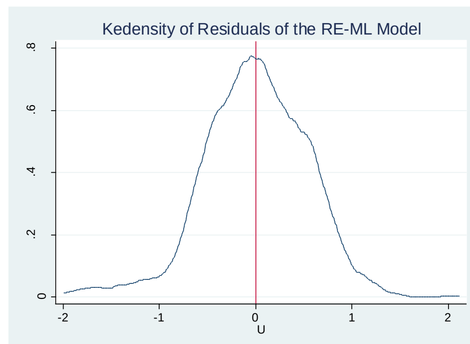
In analyzing this effect, seven panel data models repeated over three functional forms, three types of dependent variables and two dimensions of the dependent variable were used. In this session, the estimated results of the panel data models using the log-semi-log functional form are presented for the *aggregated* crops model, *two*-main crops model and *dominant* crop model using *net output value* ($log(netput_v)$) as a dependent variable. These results, I believe, will sufficiently illustrate my argument and address the research objective; nonetheless, as a means of checking robustness of the results, particularly, for the estimated results of the land fragmentation indicators, I also summarized the results obtained using the Lin-Lin and Log-Log functional forms and using wheat equivalent net output ($log(netput_w)$) as a dependent variable in Appendix B. Tables 4 – 6 summarize estimated result of each panel data model along with the standard errors and the level of significance, determined at the 0.05 level.

I started my analysis with the pooled model on the *aggregated* output model where heteroskedasticity was detected using the White's general test although the Breusch-Pagan test failed to detect it. However, since the Breusch-Pagan test relies on strict normality assumption, which did not hold, I relied on White's test for its flexibility to distributional form. The Skewness/Kurtosis, Shapiro-Francia, and Shapiro-Wilk tests revealed that normality did not hold for the error terms. In Chapter 5, it was already indicated that for plot-panel data the distributions can hardly be *iid* normal which is confirmed by the above tests. Furthermore, the Ramsey RESET test gives evidence for model miss-specification.

To use as a starting point of my analysis, I applied clustering of error terms at village level to correct for the observed heteroskedasticity and autocorrelation. Although, the pooled estimator, after correcting for heteroskedasticity and autocorrelation, provides estimates that are the same in sign and level of significance for the variables of interest as expected, the estimates cannot be unbiased or inconsistent if the model, as revealed by the Ramsey RESET test, is miss-specified. Thus, I proceeded to the next model.

The RE estimates, as already discussed in Chapter 6, are consistent only if the strict exogeneity and orthogonality assumptions simultaneously hold, in which case it will be efficient compared to the FE estimator that does not rely on the orthogonality assumption. Using the standard Hausman-Wu test, I found no evidence for the violation of orthogonality which implies that the RE is the appropriate estimator. However, this test result is valid only if the assumption of homoscedasticity and no serial correlation holds. But the assumption did not hold as tested by both the standard Breusch and Pagan Lagrangian multiplier test for random effects and another test for unbalanced panels.[18] Thus, I cannot rely on it totally. Instead, using the robust standard error, clustered at village level, I estimated the random effects by the RE-GLS regression.

Figure 3 Kernel Density of Residuals



Assuming normality of the error terms, I also estimated the random effects by the maximum likelihood (ML) method to get efficient estimates. The assumption of normality was reasonable as indicated by the kernel density of the error terms after the Random-effects ML regression which is presented in Figure 3, although tests of normality do not support it. Yet, I estimated using a bootstrap technique with 200 replications to get robust standard errors. The estimated results of random effects by the GLS and ML regressions are presented in column 2 and 4 of Table 4, while the fixed-effects within estimator is presented in column 3, after corrected for heteroskedasticity and autocorrelation.

In all models, where relevant, errors are clustered at the *village* level to compute robust standard errors. Interpretation of the robustness of the errors should be taken cautiously as the numbers of clustering groups (villages) are less than the minimum threshold amount required to do clustering.[19] Where clustering is not relevant, the bootstrap method with 200 replications is used as an alternative to compute robust standard errors. The type of techniques used to compute the robust standard errors in each model is indicated on the last rows of the result summary tables.

[18] I used the *xttest1* command in Stata, proposed by W. Sosa-Escudero and A. K. Bera, to test for heteroskedasticity in unbalanced error component models. The command is an extension of the standard *xttest0* command that computes seven specification tests for balanced error component models.

[19] It is argued that if the number of clusters is small, the critical values will substantially increase relative to those computed from the standard Normal (*t* with large d.f.). As a rule of thumb, Nichols and Schaffer (2007) in a report titled "*Clustered standard errors in Stata*" suggest that the data should have at least 20 balanced clusters or 50 reasonably balanced clusters. Rogers' seminal work (Stata Tech.Bull., 1993) suggested that no cluster should contain more than five per cent of the data. The number of reasonably balanced village clusters in the models is 18, each consisting of, on average, 5.5% of the observation. This indicates that the clustering fails to meet both requirements, and hence its result should be taken cautiously.

So far, the illustrated models deal with the assumption of strict exogeneity. The estimates will not be consistent if this assumption violates. There are sufficient reasons that cast doubt over the likely violation of this assumption. The decisions to use fertilizers, manure, herbicide and pesticide; to hire labor or to participate in land market are more likely to be endogenous. Consequently, the Hausman-Taylor (HT) model is used to deal with such endogeneity where variables including fertilizers (dap and urea), manure, herbicides and pesticides, plot renting, hired labor, sub-farm size, and plot size are treated as endogenous in the two stage model. Plot size is considered partially as endogenous to account for the demand side fragmentation; and sub-farm size to account for fallowing practices, renting in and out of lands. Crop diversity, recognizing as a potential endogenous variable that affects productivity, is dropped in all models since all the variables that are assumed to determine crop diversity are already in the model.

The estimated result for the HT model along with its robust standard errors is presented in column 5 Table 4. Using the standard Hausman-Wu test, as suggested by Baltagi et al. (2003), the HT model is found to be appropriate; or in technical terms, the choice of strictly exogenous variables is found to be appropriate. However, while the HT can be preferred, due to the observed heteroskedasticity in the model, I cannot have full confidence in the test to rely totally on the HT model; thus, the need for additional econometric specifications.

Some plots, as reported in the survey data, have experienced unexpected production loss due to flood and other events. As a result the net output for such plots is found to be negative. Of the total 1918 observed plots, 195 plots are of this type. These plots are automatically dropped from the analysis given the dependent variable is used in its logarithmic form. If the loss of the observations is systematic, the standard regression method will not give consistent estimates. Perhaps, this could be the reason for the frequent detection of violation of normality in the previous models. To deal with this problem, considering the dependent variable as a latent variable – that is only observed if it has positive value – the random effects Tobit (RE-T) model is used treating the observation as if it is left censored at zero. The estimated result is presented in column 6 along with its bootstrapped standard errors computed from 200 replications.

Finally, the feasible GLS regression is used to allow estimation in the presence of within panel auto-correlation and across panel correlation and heteroskedasticity. Comparison of results estimated by allowing *only* for across panel correlation and heteroskedasticity (GLS_1) against results estimated by allowing for *both* within and across panel correlations and heteroskedasticity

(GLS_2) by Hausman test reveals significant difference between the two models. This indicates the potential for inconsistency in estimators if the within panel autocorrelation is not recognized and treated properly, and hence the rationale to include the GLS_2 model as an alternative.

Table 4 Determinants of productivity for the *aggregated* output model

log(net output value) (br/ha)	**POOLED**	**RE**	**FE**	**RE-MLE**	**HT**	**RE-TOBIT**	**RE-GLS**
hh-edu	-0.002	-0.019		-0.017	-0.007	-0.023	-0.026
hh-age	0.001	0.000		0.000	-0.002	0.003	0.001
hh-sex	0.275***	0.287*		0.285***	0.254*	0.436***	0.254***
lfem.labor	-0.102	-0.124		-0.122	0.152	-0.121	-0.105*
lmalelabor	-0.013	-0.023		-0.023	0.054	-0.093	-0.044
ladulteqvt	-0.063	0.017		0.008	0.219	0.076	-0.078
ldist	0.029	0.008	-0.004	0.010	0.006	0.037	0.024
lno.of.plot	0.498**	0.565*		0.561**	0.425*	0.420*	0.515***
si-index	-0.475	-0.344		-0.365	-0.311	-0.058	-0.439
lsfarmsz	-0.108	-0.119	-0.091	-0.119*	-0.002	-0.092	-0.097**
lplotsize	-0.419***	-0.331	0.255	-0.344***	0.382	-0.340*	-0.477***
landqlty	0.071	0.103*	0.129*	0.099*	0.066*	0.091*	0.098***
slope	-0.005	-0.038	-0.074**	-0.033	-0.013	-0.011	-0.057*
soiltype	-0.008	-0.001	0.011	-0.003	0.017	-0.027	-0.013
soildepth	-0.053	-0.030	-0.002	-0.033	-0.066	-0.098**	-0.031
erosion	0.080**	0.060*	0.026	0.063*	0.018	0.049	0.069***
lfam.labr	0.299***	0.327***	0.341***	0.325***	0.331***	0.353***	0.286***
hirdlabor	0.004*	0.005**	0.004*	0.004*	0.004	0.002	0.005**
ddap	0.114	0.118	0.147	0.115	0.232	0.031	0.058
durea	-0.061	-0.067	-0.097	-0.064	-0.139	0.095	-0.036
dmanure	0.088	0.052	0.046	0.054	0.081	0.176**	0.049
dherbpest	-0.034	-0.060	-0.133	-0.056	-0.175	-0.055	-0.011
loxen.own	0.003	0.022	0.455	0.020	0.177	0.020	0.013
plotrent	-0.188**	-0.208*	-0.531***	-0.203**	-0.239	-0.178*	-0.180***
ltotvist	0.085	0.068	0.006	0.071	0.048	0.012	0.119***
solevisit	0.002*	0.003*	0.005*	0.003*	0.003*	0.003*	0.002*
_cons	4.923***	4.652***	5.585***	4.674***	-1.490	4.522***	4.954***
N	966.000	966.000	966.000	966.000	1008.000	966.000	966.000
SE	Clustered	Clustered	Clustered	Bootstrap	Bootstrap	Bootstrap	Bootstrap

Legends: In the first row: POOLS = the Pooled OLS estimates; RE= the Random Effects model; MLE= Random Effects Estimated by Maximum Likelihood Estimator; HT= the Hausman-Taylor Model; Tobit= the random effects tobit model; GLS= the Generalized Least Square Estimator for Panel Data. All Models used "Net output value (birr/ha)" as a dependent variable and the log-semi-log functional form ($\ln Y_{ip} = \ln X'_{ip}\beta_1 + X'_{ip}\beta_2 + \alpha_i + u_{ip}$).
In models with bootstrapped standard errors, each bootstrapping is replicated on 200 re-sampled observations; Clustering is done at village level; *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

It is based on these alternative approaches that the main findings are summarized in Table 4. Similar approach is followed in all the subsequent models and functional forms, including for the results summarized in Tables 5 and Table 6, and in Appendix B.

Table 4 presents estimation results of the Pooled, RE, FE, RE-ML, HT, RE-T and GLS models, respectively, using the log-semi-log functional form with $log(netput_v)$ being the dependent variable. Table 5 presents similar result but only for the dominant crop (henceforth, the t*eff* model) and Table 6 presents for the main two crops model (henceforth, the *teff-wheat* model). All statistical tests are performed at the 0.05 level.

The effect of numbers of operated plots on productivity is found to be positive and significant across all models except, of course, for the FE model which does not provide estimates for plot invariant regressors[20] (Table 4, 5 and 6). The result is consistent with the findings reported by Tan et al. (2010), for instance. The result is also robust to changing functional forms and model specifications; See Table B.1., in Appendix B, for summary of the result of robustness test with different model specifications. The following discussion is based on these tables.

For the aggregated output model, the effect of number of plots on productivity is clearly positive and significant with all models and functional forms; i.e. it is fully robust in all model types. For the two main crops model, the estimated result is, more or less, robust, with some deviations observed with the Lin-Lin model using net-output value as a dependent variable. However, for the dominant crop model, while the predicted values are is still positive in all models, the significance (at the 0.05 level) disappeared with some model changes: in terms of functional form, less robust result is observed with the Lin-Lin model, and in terms of econometric specification, less robust result is observed with the HT model. Particularly, the estimated result of the HT model for the dominant crop model offers interesting interpretation. Since, the HT model controls for some of the potential endogenous variables, the predicted significance of number of plots for the dominant crop became less robust, while it does not affect the results in the teff-wheat and the aggregate output models.

This may indicate that had the study relied only on dominant crop productivity analysis, as some authors already did, the result would have been less robust; implying that statistical inference

[20] In fact, the FE with in estimator tends to predict unexpected signs to some variables such as land quality, erosion and hired labor. Since, the Hausman test also favored RE over FE, the FE is least preferred next to the Pooled estimator.

based on dominant crop analysis might be biased as it does not represent the farming system fully and can yield less robust results for the following reason.

Operating on more than one plot allows farmers to diversify production and thereby, manage agricultural risk relatively efficiently. That crop diversification is positively associated with the number of plots operated is investigated and evidenced in this study, which is presented in session 7.2; and that farm productivity (profitability) is positively associated with the number of crops (on-farm diversity) is investigated and confirmed by Di Falco et al. (2010). Thus, the positive and significant effect of number of plots on productivity is reasonable and expected particularly for analysis with two or more crop production. Besides its effect through diversity and risk management, number of plots can further affect productivity positively through its role on smoothing farm labor supply. This is also reasonable principally when the effect of farm labor on productivity is strong, positive and significant across all models, functional forms and forms of the dependent variable (Table 4, 5 and 6).

The SI index is found to have negative but insignificant effects in all models and functional forms when productivity analysis involves at least two crop outputs (tables 4 and 6). This result is consistent with the finding reported by Wu et al. (2005) where the dependent variable was aggregated output (of cereals and vegetables). On the other hand, for the teff model, I found significant negative relationship between land fragmentation (the SI index) and productivity (Table 5) which is also consistent with findings of studies on productivity of main crop such as by Rahman and Rahman (2009) on rice productivity and by Jabarin and Epplin (1994) on wheat productivity. This seemingly contradictory effect of SI-index is robust to changing functional forms and model specifications (See Table B.3, in Appendix B).

Yet, interestingly, the combined result reinforces the previous argument that evaluating land fragmentation dealing only with dominant crop productivity is methodologically inconsistent when farm production involves diversification,[21] since the SI index can counterbalance the negative impact of fragmentation on management and technology adoption mainly by improving risk management through diversification. The finding presented in session 7.2 is also consistent with the argument that higher SI index encourages farmers to diversify thereby counterbalance its detrimental effect. This is the most likely reason why a negative but insignificant effect is observed in the non-single crop productivity analysis.

[21] Supporting evidence to this argument is also illustrated in Appendix B, Table B.1.

Table 5 Determinants of productivity for the *Dominant* Crop (*Teff*) model

log(Net Output Value) (br/ha)	**POOL**	**RE**	**FE**	**HT**	**RE-TOBIT**	**RE-GLS**
hh-edu	-0.029	-0.043		-0.041	-0.043	-0.042**
hh-age	0.001	-0.000		-0.006	-0.000	0.001
hh-sex	0.174	0.170		0.208	0.170	0.209***
lfem.lab.force	0.006	-0.008		0.196	-0.008	0.046
lmal.lab.force	0.154	0.135		0.213	0.135	0.070
ladult.eqvt	-0.405	-0.380		-0.474	-0.379*	-0.37***
ldistance	-0.027	-0.013	-0.019	-0.014	-0.014	-0.008
lno.of.plots	0.844*	0.946**		0.788	0.945***	0.810***
si-index	-2.410*	-2.621**		-2.867	-2.619**	-2.29***
lsfarmsz	-0.225	-0.231	-0.805	-0.080	-0.231*	-0.24***
lplotsize	-0.359	-0.405	-0.123	-0.129	-0.403*	-0.4***
landqlty	0.133	0.119	-0.008	0.061	0.120	0.118***
slope	0.096	0.109	0.052	0.098	0.109	0.122***
soiltype	-0.036	-0.039	-0.068	-0.042	-0.038	-0.026
soildepth	0.024	0.016	0.009	0.006	0.016	0.004
erosion	0.101	0.074	-0.044	0.007	0.075	0.111***
ltot.farm.lob	0.350***	0.394***	0.644*	0.470	0.393**	0.338***
Hird.labor	0.001	0.001	-0.001	0.001	0.002	0.002
ddap	-0.080	-0.112	-0.489	-0.456	-0.110	0.026
durea	0.253	0.282	0.682	0.584	0.280	0.176***
dmanure	0.167	0.141	0.246*	0.140	0.141	0.120**
dherbpest	-0.250**	-0.236*	-0.311	-0.242	-0.237	-0.31***
Loxen.own	0.027	-0.008	0.052	0.066	-0.007	0.030
plotrent	-0.067	-0.077	-0.206	-0.052	-0.077	-0.009
ltotvist	0.014	-0.003	-0.332	-0.099	-0.002	0.028
solevisit	-0.001	-0.000	-0.003	-0.000	-0.000	-0.000
_cons	6.240***	6.251***	5.991**	1.049	6.247***	6.022***
N	280.000	280.000	280.000	283.000	280.000	280.000
SE	Clustered	Clustered	Clustered	Bootstrap	Bootstrap	Bootstrap

Legends: In the first row: POOLS = the Pooled OLS estimates; RE= the Random Effects model; MLE= Random Effects Estimated by Maximum Likelihood Estimator; HT= the Hausman-Taylor Model; Tobit= the random effects tobit model; GLS= the Generalized Least Square Estimator for Panel Data. All Models used "Net output value (birr/ha)" as a dependent variable and the log-semi-log functional form ($\ln Y_{ip} = \ln X'_{ip}\beta_1 + X'_{ip}\beta_2 + \alpha_i + u_{ip}$). Bootstrapping is replicated on 200 re-sampled observations; Clustering is done at village level; *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

The relationship between plot distance and productivity is not clear. The result is not robust to changes in functional forms and forms of the output variable. First, it is found to have statistically insignificant effect on productivity for the log-semi-log function for all of the models and forms of the dependent variable, but its sign is mixed. It's negative for the teff and teff-wheat models (Table 5 and 6), but positive though very small in magnitude in the aggregated output model (Table 4). Using the log-linear model it is found to have positive and significant effect for the aggregated output model, positive but insignificant for the teff-wheat model; and negative and insignificant for the teff model. See the summarized results of distance estimated by all models in table B.2., in Appendix B. This inconsistency could be attributed to measurement error which was already indicated in Chapter 5 where the limitations and the apparent measurement errors with the distance data were discussed.[22]

Contrary to what was expected, plot size is found to associate negatively with productivity in all regressions except in the FE and HT models for the aggregated output (Table 4). Though not robust the effect is significant at the 0.05 level for the non-dominant crop models. Similar results were reported in earlier studies such as by Blarel et al. (1992) in Ghana and Rwanda and by Holden et al. (2001) in Ethiopia even after controlling for a wide range of land quality variables that was assumed to explain the inverse productivity relationship.

Several reasons can explain this relationship except diseconomies of scale which I have little evidence to concluded so. First, it could be due to the labor intensive production system that predominates in the area where land poor households are often labor endowed. Thus, small plots can be more efficient at the expense of labor inefficiency which is, as Bhalla and Roy (1988) argued, a typical sign of labor market imperfection. This justification makes sense particularly with the result in Table 5 that predicts the inverse productivity relationship even for the dominant crop model. Second, it could be due to the tendency that small plots are allocated to vegetable production where their net output value per unit area tends to be larger than that of cereals. Third, as Udry (1996) suggested it could be due to unobservable land qualities.[23] Fourth, as Holden et al. (2001) hypothesized, it could be due to the prevalence of small variation in farm size within the farm community in Ethiopia owing to the relatively egalitarian land distribution system. Finally, since the result is not robust (see Table B4, in Appendix B) to all models and functional forms, particularly in its significance level, the inverse-relationship can be justified by

[22] To deal with this bias, I tried using standardized distance variable computed as: $(d_{ip} - \bar{d}_i)/\bar{d}$, where d_{ip} is distance of each plot, $\bar{d}_i$ is average distance at household level, and $\bar{d}$ is average distance for all observations. The result was basically the same, inconclusive.

[23] A separate analysis, its result of which is not presented here, seems to support this hypothesis. Introducing an interaction term, between operated plot size and land quality, in to the models provides significant positive relationship with productivity; although the positive relationship that was observed, as expected, between land quality and productivity turned negative after introducing the interaction term.

the possible bias in plot size measurement where plot size was estimated based on local method during land distribution. The HT estimate, specifically, in Table 4 seems to support this justification since plot size is predicted to positively associate with productivity after treating it as an endogenous variable. Even more supportive of this reasoning is the opposite predictive signs that are observed by the HT model for the partial outputs in tables 5 and 6. Empirically, Holden et al. (2001) argued that plot sizes in the area are biased downwards particularly on larger plots with difficult terrains. This suggests the need for cross-checking robustness of the result with plots measured by the standard measurements techniques, for example, using GPS coordinates or remotely sensed images; thus, a potential research gap for further study.

Operated sub-farm size, which subtracted plot size from the total operated farm size, is found to have negative, though insignificant, effect on plot level productivity. The result is contrary to what economic theories of scale suggest – large farms necessitate less coordination. However, as Bardhan (1973) suggested the observed negative relationship is likely to be the result of an inverse relationship between farm size and other inputs due to production uncertainties and market imperfections than the result of scale diseconomies. Thus, the result may reflect inter-plot factor competition given the limited factors endowments by households in an environment with imperfect factor market. Table B5, in Appendix B, summarizes estimated results of this variable using different econometric specifications.

As expected, male headed operated plots are found to be significantly more productive than female headed households. This is consistent with the results described by Bezabih and Holden (2006), Holden and Bezabih (2008) and Ghebru and Holden (2008). A possible explanation stems from the traditional gender-based division of farm labor. Traditionally, farming in the area is highly characterized by gender-based division of labor where women are mostly in charge of domestic activities and men are in charge of land management. Only men can till the land with oxen; as a result, women possessed limited ability to till the land (Holden et al., 2011). A common empirical finding in the highlands of Ethiopia reveals that female headed households usually face problems with respect to male labor and oxen which are often insufficient (Bezabih and Holden, 2006, Ghebru and Holden, 2008, Holden et al., 2011). Consequently, female headed households are usually characterized by renting out plots. Their participation in land rental market is claimed to increase significantly due to land certification (Holden et al., 2011). In this study, 26% of the plots operated by female households have involved in land rental market.

Table 6 Determinants of productivity for the Two *Dominant* Crops (*Teff* and *Wheat*) Model

log(net output value) (br/ha)	**POOL**	**RE**	**FE**	**HT**	**RE-TOBIT**	**RE-GLS**
hh-edu	0.019	0.008		0.036	0.009	0.001
hh-age	0.005	0.003		0.001	0.004	0.004**
hh-sex	0.295*	0.288**		0.291	0.290**	0.305***
lfem.lab.force	-0.045	-0.051	0.066	-0.010	-0.048	0.003
lmal.lab.force	-0.030	-0.008		0.046	-0.011	-0.054
ladulteqvt	-0.134	-0.164		-0.217	-0.159	-0.170**
ldistance	-0.022	-0.033	-0.019	-0.002	-0.032	0.009
lnum.of.plots	0.527*	0.72***		0.7**	0.68***	0.57***
SI-Index	-0.858	-1.214		-0.899	-1.178	-0.733*
lofarmsz	-0.176	-0.159	-0.722	-0.168	-0.158*	-0.16***
lplotsize	-0.492**	-0.485*		-0.469	-0.49***	-0.46***
landqlty	0.097	0.096*	0.090	0.081	0.097	0.083***
slope	-0.004	-0.036	-0.132	-0.062	-0.024	0.013
soiltype	-0.026	-0.033	-0.054	-0.029	-0.031	-0.020
soildepth	-0.013	0.000	0.013	-0.033	-0.002	-0.026
erosion	0.098	0.094	0.061	0.047	0.097*	0.077***
ltot.fam.lobr	0.34***	0.37***	0.45**	0.35**	0.36**	0.34***
hird.lobor	0.003	0.003*	0.001	0.001	0.003	0.004*
ddap	0.083	0.064	0.051	0.132	0.066	0.018
durea	-0.004	-0.000	-0.053	-0.127	0.004	0.048
dmanure	0.077	0.111	0.175	0.194	0.103	0.057
dherbpest	-0.119	-0.157	-0.267	-0.131	-0.143	-0.079
loxen.own	-0.057	-0.061	0.209	-0.034	-0.058	-0.001
plotrent	-0.214*	-0.259*	-0.506***	-0.281	-0.245**	-0.176***
ltot.vist	0.025	0.028	-0.058	0.014	0.030	0.098***
solevisit	0.002	0.001	0.003	0.002	0.001	0.001
_cons	5.006***	5.138***	5.418***	-0.890	5.125***	4.735***
N	511.000	511.000	511.000	529.000	511.000	511.000
SE	Clustered	Clustered	Clustered	Bootstrap	Bootstrap	Bootstrap

Legends: In the first row: POOLS = the Pooled OLS estimates; RE= the Random Effects model; MLE= Random Effects Estimated by Maximum Likelihood Estimator; HT= the Hausman-Taylor Model; Tobit= the random effects tobit model; GLS= the Generalized Least Square Estimator for Panel Data. All Models used "Net output value (birr/ha)" as a dependent variable and the log-semi-log functional form ($\ln Y_{ip} = \ln X'_{ip}\beta_1 + X'_{ip}\beta_2 + \alpha_i + u_{ip}$). Bootstrapping is replicated on 200 re-sampled observations; Clustering is done at village level; *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

However, the relationship between the dummy variable for participation in land rental market and productivity is found to be negative and significant for the teff and the aggregate output models (Table 4 and 6) but insignificant for the teff and wheat model (Table 5). The negative sign is consistent in all models even after treating the variable as endogenous variable in the two stage HT model, though its significance, at the 0.05 level, disappeared (Tables 4 and 6).

Possible reason that explains this relationship is the existence of higher transaction cost and Marshallian inefficiency in the study area. The rental market in the study area is dominated by sharecropping arrangements (Holden et al., 2011). Investigating the efficiency of sharecropping arrangements in the Ethiopian highlands, Deininger et al. (2006) provides empirical support to the hypothesis of Marshallian inefficiency. They argued that the inefficiency is attributed mainly to market imperfection, suggesting that the hypothesis of Marshallian inefficiency is reasonable if there are market imperfections. Holden et al. (2001), on the other hand, argued that market imperfection arises from higher transaction cost and information asymmetries. But the existence of substantial transaction costs and information asymmetries in the land rental market in Tigray is already documented by Ghebru and Holden (2008). Therefore, if there are substantial TCs and information asymmetries then it may cause market imperfection which, in turn, may cause Marshallian inefficiency.

It is also argued that the existence of significant transaction cost causes the land rental market to operate less efficiently and undermines its potential to contribute in poverty reduction (Holden et al., 2012). Thus, there is ample evidence to hypothesize that higher transaction cost and Marshallian inefficiency can explain the observed negative relationship between plot renting and productivity. However, it should be note that variation in productivity and efficiency may be observed depending on whether the plot is rented out or rented in as well as whether the rented plot is own operated, operated by blood-relatives or non-relatives. Detailed analysis of this sort is beyond the scope of the study; however, Bezabih and Holden (2006), Holden et al. (2011) and Kassie and Holden (2007) have already addressed it.

Land quality, in the aggregated output model, is predicted to have positive and significant effect on productivity using all models except the pooled regression. For the teff and teff-wheat models, although positive, it has insignificant effect on productivity, at the 0.05 level. Other sets of land quality indicators are predicted to have insignificant effect on productivity

All household characteristics, other than sex of the household head, are found to have generally insignificant effect on or relationship with productivity. Most models predict negative association between household labor force (male, female and adult equivalent) per hectare and productivity, although not significant. This prediction, however, is more likely to reflect the general correlation between productivity and the household labor force than any causality. Perhaps, other omitted factors such as household members' participation in off- and non-farm activities, could explain this association. Some members of the household may be, for instance, students or off-farm workers who have little contribution in farm activity.

Among household factor endowments, oxen ownership per hectare has less significant effect. One reason could be due to the smallest variation in households' oxen endowment where the average holding is slightly above a pair of oxen, 2.35, with more than 95% having at least single ox. Given, the smaller average land holding, oxen ownership more than a pair may not result in significant productivity difference.

7.2. The Effect of Land Fragmentation on Crop Diversity

Table 7 and 8 present estimation results of the determinants of diversity richness using the translog and linear functional forms, respectively. Estimation is carried out based on Pooled OLS (with the dependent variable in log-form), Pooled Poisson, Random Effect Poisson, Pooled Negative Binomial and Generalized Negative Binomial models.[24] In all models, except for the Pooled OLS model, robust standard errors are computed by the bootstrapping technique on resampled observations with 200 replications. For the Pooled OLS model, robust standard errors are generated by clustering errors at village level.

The average number of crops harvested by households is 3.47 ranging between one and eight. Only 3% of the household samples cultivated one crop,[25] while 17% of them cultivated at least 5 crops. The fraction of households that cultivated 2, 3 and 4 crops are 21%, 30% and 31.3% respectively. The dominant crop cultivated by the largest share of households is *Teff* which comprises of 31% of the total plots, followed by wheat (21.5%) and barley (10%).

Generally speaking, the models predict strong and positive effect of land fragmentation on crop diversity. Of the three indicators, only distance was found to have no significant effect on diversity despite the claim that location and distance to plots are important factors that determine diversification (Benin et al., 2004). Although the result is robust to changing functional forms,

---

[24] Since the dependent variable is household level observation, as opposed to plot level, there is little room to exploit the potential advantage of panel data models. FE-Poisson model is impossible with such data. Thus I used the RE Poisson to account for unobserved heterogeneity while I rely more on pooled estimates.

[25] This fact alone indicates how misguided it would be to analyze productivity considering only the dominant (single) crops which, I think, was one limitation of earlier works.

the previously (in session 7.1) suspected measurement bias of distance may have its play in this result, too. Interestingly, the SI-index and number of plots are found to have strong positive and significant effect on diversification as expected. The result is consistent with the assertion that operating on fragmented plots may encourage farmers to diversify more (Benin et al., 2004, Blarel et al., 1992, Di Falco and Chavas, 2006, Di Falco et al., 2010). A high level of diversity on more fragmented plots may appear desirable owing to higher expected revenue, lower risk exposure, agro-ecological diversity and heterogeneity in land quality.

All other factors being equal, farm size is associated negatively and significantly with crop diversity. This is consistent with the above explanation that fragmented plots encourage farmers to diversify more. Similar finding was reported by Di Falco et al. (2010) although the opposite was reported by Benin et al. (2004). Among the relevant household characteristics that are included in the model, education and age of the household head, household size and adult equivalent labor force are found to determine crop diversity significantly. While age of the household head affects diversification positively and significantly, education of the household head, contrary to what was expected, is estimated to affect diversification negatively and significantly. The later prediction is even more robust to changing functional form, while the effect of age of the household head is found to be less robust in the sense that the predicted association disappeared in the Lin-Lin model if the normality assumption of the Poisson distribution does not hold, i.e. if the negative binomial model is the appropriate one.

Possible explanations to this prediction could be that with age comes experience and thus household heads who are relatively older may tend to have the skill to manage diversified crops. On the other hand, relatively younger household heads, may lack the skill to manage diversified crops or they may tend towards specialization (Van Dusen and Taylor, 2005). Moreover, older households may be more risk averse and tend to diversify for fear of production loss.

Similarly, more years of education may mean relatively better confidence and less risk averse and hence the tendency to diversify less, if diversification is a copping mechanism than a risky venture in itself. In fact, in a drought-prone rain-fed agricultural systems diversification is claimed to have a major potential role in maintaining yields under adverse shocks, reducing exposure to risk and lowering risk premium (Blarel et al., 1992, Di Falco and Chavas, 2006). Moreover, in developing countries like Ethiopia, where access to primary education is a recent phenomenon, particularly in Tigray, more number of school years is likely to associate with relatively younger ages; thus, as already argued, if the relatively younger lacks skill, then they will tend to diversify less, if diversification is skill demanding.

Table 7 Determinants of on-farm crop diversity: The *Semi-Log* model

Diversity	POOLED	POISSON	POISSON-RE	N.BINOMIAL
hh-age	0.003***	0.003***	0.003***	0.003***
hh-sex	0.032	0.036	0.026	0.036
hh-edu	-0.023**	-0.029***	-0.024***	-0.029***
hh-size	-0.003	-0.006	-0.003	-0.006
lfem.labor	-0.008	-0.016	-0.010	-0.016
lmalelabor	-0.029	-0.028	-0.024	-0.028
ladulteqvt	0.104**	0.132***	0.095***	0.132***
lofarmszh	-0.078***	-0.081***	-0.085***	-0.081***
si-index	0.933***	0.766***	0.728***	0.766***
distance	0.000	0.000	0.000	0.000
no.of.plot	0.054***	0.053***	0.061***	0.053***
soildepth	-0.004	-0.006	-0.005	-0.006
Landqlty	0.012	0.008	0.010	0.008
Soiltype	-0.007	-0.000	-0.004	-0.000
Slope	0.074***	0.073***	0.058***	0.073***
loxen.own	-0.039***	-0.063***	-0.043***	-0.063***
_cons	-0.177	-0.023	0.035	-0.023
N	1446	1446	1446	1446
SE	CLUSTERED	BOOTS	BOOTS	BOOTS
Wald chi2		1438.55	2464.74	1395.01
Prob > chi2	(.)	0.0000	0.0000	0.0000
Log Likelihood	(.)	-2417.73	-2409.24	-2417.75

Legend: POOLED = the Pooled OLS model with the dependent variable in log form, as explained in Chapter 6.2; POISSON = the Pooled Poisson Regression Model; POISSON-RE = the Random Effects Poisson Regression Model; N.BINOMIAL = the Pooled Negative Binomial Regression Model; SE = Robust standard error computed either by bootstrapping technique (*boots*) with 200 replications or clustering the error terms at village level (*clustered*); *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

Furthermore, more education may mean more opportunities to off-farm and non-farm activities and hence limited time and devotion to labor intensive farming activities. Thus, the more educated ones may tend to diversify less, if diversification is labor demanding and the opportunity cost of labor is higher to the relatively more educated households. More education may also mean more tendencies toward specialization. The specialization and opportunity cost arguments are supported by Van Dusen and Taylor (2005), although Benin et al. (2004) stated opposite finding where the young and more educated households were found to diversify more.

The opportunity cost explanation seems to be supported by the observed positive and significant association between household size, adult equivalent labor force and crop diversity, although the result is less robust to changing functional form. Where the household size is significantly associated with diversity, the adult equivalent labor force is not, and the vice versa. Nonetheless, it sufficiently indicates that diversity is associated with some degree of labor endowment. Besides the household size, the significance of adult labor force may indicate the role of children and elder people on diversification, for example, through supervision.

Oxen endowment is found to be another determining factor of diversification, although negatively associated. Contrary to the findings reported by Benin et al. (2004), I found households who own relatively more oxen cultivating significantly less diversified crops. However, this significant relationship is not robust to changing the functional form. Yet, the inverse relationship is maintained in all models and functional forms. Three possible reasons may explain this relationship: First, oxen-rich farm households may be more concerned with crop residues in order to feed their oxen. If so, they are likely to concentrate on cultivating few crops that can offer good amount or quality of fodder than on diversification.

Second, households who own more oxen are often relatively rich. Sometimes land-poor households may own more oxen on which their livelihood is based, through renting their oxen out, though a rare case. If diversification, as already argued, is a copping mechanism, then the relatively (oxen-) poor households are likely to tend to diversify more. Finally, as Benin et al. (2004) argued, oxen-rich households may tend to focus on cultivating crops that are more valuable but demands for intensive ploughing such as teff. This explanation is a likely scenario, although I found little evidence for the teff case where of the total number of plots allocated to teff production only 54% were cultivated by households who own more than a pair of oxen.

Finally, the terrain feature of the cultivated plot was found to determine diversification positively and significantly; implying that households who operate on steeper terrains gravitate toward diversification.

Table 8 Determinants of on-farm crop diversity: The *Lin-Lin* Model

cropdiv	POOLS	POISSON	POISSON-RE	N.BINOM.	G.N.BINOM
hh-age	0.001*	0.001**	0.001*	0.001	0.001
hh-sex	0.034	0.022	0.019	0.036	0.021
hh-edu	-0.032***	-0.035***	-0.028***	-0.037**	-0.035**
hh-size	0.022***	0.022***	0.020***	0.021**	0.022**
fem.labor	0.000	0.000	0.000	0.000	0.000
malelabor	-0.000	0.000	-0.000	0.000	0.000
adulteqvt	0.000	0.000	0.000	-0.000	0.000
ofarmszh	-0.058***	-0.061***	-0.062***	-0.057***	-0.061***
si-index	0.829***	0.740***	0.673***	0.782***	0.725***
distance	-0.000	0.000	-0.000	0.000	0.000
no.of.plot	0.064***	0.064***	0.072***	0.063***	0.064***
soildepth	-0.007	-0.009	-0.008	-0.010	-0.009
landqlty	0.015	0.009	0.012	-0.003	0.009
soiltype	-0.005	-0.001	-0.004	-0.002	-0.001
slope	0.075***	0.070***	0.055***	0.072***	0.070***
oxen.own	-0.000	-0.000	-0.000	-0.000	-0.000
_cons	0.024	0.151*	0.205**	0.157	0.161
N	1784.000	1784.000	1784.000	1784.000	1784.000
SE	CLUSTERED	BOOTS	BOOTS	BOOTS	BOOTS
R^2/Wald/LR-chi2	R^2 =0.495	Wald=1700.21	Wald=3303.03	LR = 416.26	LR = 416.26
Prob > chi2	-	0.0000	0.0000	0.0000	0.0000
Log-likelihood	-	-2948.6	-2936.6082	-2948.6	-2948.6

Legend: POOLED = the Pooled OLS model with the dependent variable in log form, as explained in Chapter 6.2; POISSON = the Pooled Poisson Regression Model; POISSON-RE = the Random Effects Poisson Regression Model; N.BINOM = the Pooled Negative Binomial Regression Model; G.N.BINOM = the Generalized Negative Binomial Regression Model; SE = Robust standard error computed either by bootstrapping technique (*boots*) with 200 replications or clustering the error terms at village level (*clustered*); *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

Moreover, in an attempt to further assess the potential effect of agro-ecological and price variations on diversification, I run additional model that introduced the *wereda* (district) dummy variable to account for *wereda* specific effects and found the above results to be robust. However, after controlling for the *wereda* specific effects, land quality and soil depth turned to have significant effect on diversification. Plots with good land quality and deep soil seem to be favorable to diversification. The result is tabulated and presented in Appendix C.

7.3. The Effect of Land Fragmentation on Technical efficiency

Efficiency analysis is carried out using cross-sectional and panel data frontier models. The results are summarized in Tables 9 and 10. More supplementary results are also provided in Appendix D. The main finding is, basically, in line with the findings in session 7.1. Table 9 summarizes the results estimated by employing 5 cross-sectional and 2 panel data frontier models, using the log-log functional form. The production function is assumed to be determined by factors including: plot size, fertilizers (dap, urea and manure), herbicide and pesticide, total labor used and oxen ownership per ha, as a proxy to amount of draft power used in the production process. The fertilizer variables entered the model as dummies, to make sense of estimating its effect.[26] Household, plot and village characteristics are used to predicted variations in *technical inefficiency*. Column 1 portrays estimation result of the cross-sectional frontier model for aggregated output using net output value as a dependent variable and assuming a normal-half normal (h) distribution of the error components.

Table 9 Stochastic Frontier Models: The *Log-Log* Model

	Cross-Sectional Frontier Model					Panel-Frontier M.	
Prod-Function	**fh1**	**fx1**	**fx2**	**fh3**	**fh4**	**feh**	**bc95**
lplotsize	-0.37***	-0.35***	-0.29**	-0.169	-0.31***	-0.282	-0.44***
dmanure	0.078	0.075	0.105	0.195	0.129	0.069	-0.055
durea	-0.053	-0.034	-0.084	0.209	0.001	-0.127	-0.082
ddap	0.146	0.118	0.169*	-0.062	0.083	0.145	0.216
dherbpest	-0.002	-0.002	-0.002	-0.01**	-0.003	-0.002	-0.004
lfam.labr	0.32***	0.32***	0.29***	0.35***	0.35***	0.38***	0.36***
loxen.own	-0.039	-0.038	-0.003	0.07	-0.034	0.437	-0.099
_cons	6.699***	6.486***	0.412*	0.182	0.299		7.08***

[26] Since fertilizer application rate in the study area is fixed, by officials, for the same type of crop, analyzing its marginal effect will be *misleading*, as each additional observation (of the same crop type) does not add any information in estimating the effect of a unit change of fertilizer use on productivity; Instead, the marginal effect would only reflect the difference in application rate for the different crops, which is exogenous to farmers economic decision. Thus, I introduced the dummy variables so that I can see the difference easily between plots that used fertilizer and that did not. One exception could be manure, as the rate is not fixed; however, since I suspected of measurement error, I preferred to treat it as a dummy along with the others.

$ln(\sigma_v^2)$-	-1.23***	-0.99***	-1.12***	-1.44***	-1.61***	-1.05***	-1.47***
Technical Efficiency							
hh-sex	0.316*	0.516*	0.79***	0.562	0.529*	-2.567*	0.162
lhh-age[a]	0.063	0.071	-0.266	0.489	0.297	0.179	-0.009
lhh-edu[a]	0.178	0.206	0.166	0.368	0.48**	0.493	0.231
ldist	0.068	0.104	0.103	-0.112	-0.043	0.585*	-1.04***
si-index	-0.265	-0.52	-0.754	-7.288*	-2.596	-15.471	-4.933
lno.of.plot	0.894*	1.298*	1.414**	3.13**	1.977*	7.58***	4.144*
lsfarmsz	-0.264*	-0.375*	-0.559**	-0.762**	-0.526*	-4.698*	-1.17
lsolev	0.310***	0.446***	0.343**	0.113	0.203	2.951	0.881
landqlty	0.159	0.241	0.175	0.399*	0.325*	5.11***	0.769
slope	-0.037	-0.061	-0.081	0.282	-0.056	-1.204	-0.121
soiltype	-0.036	-0.064	0.002	-0.036	-0.118	0.868	-0.21
soildepth	-0.099	-0.125	-0.304*	-0.053	-0.111	0.136	-0.458
Plotrent	-0.345*	-0.481*	-0.203	-0.25	-0.606*	-3.971	-0.631
lfem.labor	-0.047	-0.055	0.316	0.657*	0.433*	0.605	-0.347
lmalelabor	0.065	0.078	0.248	0.358	0.381	3.058	-0.7
ladulteqvt	-0.266	-0.333	-0.669*	-1.256*	-0.918*	-0.822	-0.644
_cons	-1.871	-1.537	-0.929	-0.384	-1.549	-24.512	2.1
N	959	959	1001	279	525	959	959
SE	Bootst.	Bootst.	Bootst.	Bootst.	Bootst.	Clustered	Clustered
Wald chi2	140.27	123.75	124.82	83.65	121.80	382.49	227.76
Log-pseudo-likelihood	-1136.42	-1138.52	-1104.0	-270.73	-503.49	-862.182	-941.19

[a] In the panel model the variables entered the model in linear form since the full-log model could not return results.
Legends: fh = frontier model with normal-half-normal distribution; fx = frontier model with normal-exponential distribution; The numerical subscripts: 1 = Aggregated crop model, in net-output-value (birr/ha); 2 = Aggregated crop model, in wheat-equivalent net output (quintal/ha); 3 = Dominant crop model (teff), in net-output value (birr/ha); 4 = Two main crops model (teff and wheat), net-output value (birr/ha). feh = the true fixed effect model with normal-half-normal distribution; bc95 = random effects model with normal-truncated distribution. SE is robust standard error computed either by bootstrapping with 200 replications or clustering the error terms at village level; *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

Column2 portrays results estimated similar to column 1, but changing the distributional assumption to normal-exponential (x). Column 3 portrays result estimated similar to column 2, but using wheat equivalent net output as a dependent variable. Column 4 and 5 portray estimated result similar to column 1 but for the dominant and two crop output models, respectively. The last two columns depict results estimated by the true fixed effects model assuming half normal distribution and random effect model assuming truncated-normal[27] distribution of the error component u_i, respectively.

The results are generally in line with the results presented in session 7.1. In the first stage analysis of the frontier models, all models predict negative and significant association between productivity and plot size. The possible reason that can explain this relationship is already provided in session 7.1. Similarly, strong positive and significant effect of total farm labor is predicted. The other inputs are estimated to have no significant effect on productivity in all models, with two exceptions. For the dominant crop model, herbicide and pesticide usage is predicted to associate negatively and significantly. This is more likely to be a mere association than causality, since farmers in the area usually tend to use herbicide and pesticide on plots that are already (highly) infested. The negative relationship is likely to reflect loss due to infestation than due to the use of chemicals. Of course, the effect of inappropriate application of chemicals cannot be ruled out, but such significant negative effect is less expected.

Technical efficiency is found to have significant association with sex of the household head, sub-farm size, number of plots operated, land quality, plot supervision, and plot renting. All models predict a positive and significant effect of number of plots on technical efficiency, implying that farm households operating with few plots are relatively inefficient. This result is consistent with my finding in analyzing productivity, where with more plots productivity is predicted to increase. The possible reasons are already discussed in session 7.1.

Again, the effect of distance on technical efficiency is not clear. Some models predicted negative relationship while others predicted positive relationship. For the teff and wheat-teff models distance is estimated to increase inefficiency though not significantly. While this relationship is normally expected, the opposite is observed for the aggregated output model, to which I have no better explanation than enhancing my suspicion of measurement bias. More supporting of this suspicion is the result predicted by the panel frontier model, in the last two columns. For the same aggregated crop model using net output value as a dependent variable, the same functional form and the same assumption of plot variant technical inefficiency, but differing in the distributional assumption of u_{ip}, the models predict quite opposite result; with the half-normal

[27] In stochastic frontier literatures the normal-truncated distributional assumption is usually regarded as less efficient compared to the other three distributional assumptions. But I included it here because the other models could not give results.

distribution significant ($p<0.05$) *positive* effect of distance on technical efficiency is estimated while with the truncated-normal distribution significant ($p<0.001$) *negative* effect is predicted. Moreover, changing the functional form, in Table 10, the estimated magnitude resulted almost null. Consequently, the measurement bias for this variable seems inevitable as manifested in all analyses including the productivity and crop diversity analyses. Thus, further investigation is necessary with actual measurements than subjective estimates of distance in the future.

Technical efficiency decreases significantly with the increase in SI-index for the dominant crop model (column 4) while it is insignificant with rest of other models. The insignificant relationship is robust to changing functional form and distributional assumptions for the aggregated crop model, Table 10. The result is similar to the result in session 7.1; thus, the inference will also be the same: analyzing fragmentation effect with respect to dominant crop model, when farming involves diversification, may give misleading, or less robust, result.

In all models, except in the random effects model (column 7), farm size is predicted to increase inefficiency significantly. As previously discussed, this association is more likely to reflect the inter-plot factor competition due to market imperfection than the result of scale diseconomies. In fact, assuming fixed average plot size, the positive association between number of plots and efficiency can be a sign of positive scale effect. Thus, if not due to factor market imperfection or unobserved plot heterogeneity, the predicted inverse size-efficiency relationship is more likely the result of measurement bias. Measurement bias is already suspected in session 7.1 with plot size estimations, suggesting the need for further robustness check,

To this end, further random effects frontier models are employed using log-semi-log functional form and changing the assumption of plot invariant technical efficiency to plot variant. As illustrated in Table 10, the result does not persist; the significant inverse relationship between farm-size and efficiency disappeared. Thus, the less robust result may further strengthen the hypothesis that the inverse relationship may be explained by measurement bias than due to diseconomies of scale.

Male households are found to be relatively more efficient than their female counterparts. The difference is significant at the 0.05 level and, more or less, robust to model changes. Although less robust, land quality and plot supervision are also found to affect efficiency positively while plot renting affects negatively.

Table 10 Stochastic Frontier Models: The *Semi-Log* Model

Prod-Function	xfh1	xfh2	xfh3	xfh4
lplotsize	-0.143**	-0.201***	-0.159**	-0.214***
dmanure	0.134**	0.135***	0.087	0.083*
durea	0.053	0.005	0.050	0.002
ddap	0.048	0.085	0.048	0.084
dherbpest	-0.000	-0.000	-0.000	-0.000
lfam.labr	0.346***	0.319***	0.329***	0.302***
loxen.own	0.043	0.024	0.041	0.025
_cons	6.798***	0.448***	6.881***	0.529***
$ln\sigma^2 = \sigma_u^2 + \sigma_v^2$	-0.133	0.178	-0.064	0.226
Technical Inefficiency				
hh-sex	0,073	0,176**	0,134	0.223**
hh-age	0.000	-0,001	-0,002	-0,002
hh-edu	-0,024	-0,018	-0,024	-0,02
ldist	0.000	0.000	0.000	0.000
si-index	-0,487	-0,603	-0,12	-0,184
lno.of.plot	0.393***	0.394***	0,184	0,216
lsfarmsz	-0,006	-0,026	-0,025	-0,037
lsolev	0,001	0,001	0,003	0,002
landqlty	0.000	0.000	0.000	0.000
slope	0,001	0,001	0,001	0,001
soiltype	0.000	0.000	0.000	0.000
soildepth	-0,001	-0,001	-0,001	-0,001
plotrent	-0,026	-0,022	-0,069	-0,042
lfem.labor	-0,012	0,038	-0,004	0,059
lmalelabor	-0,041	-0,039	-0,004	-0,019
ladulteqvt	0,054	0,006	0,014	-0,031
_cons	-0.988***	-0.790***	-1.126***	-0.948***
N	1075	1075	1075	1075
SE	Bootstrapped	Bootstrapped	Bootstrapped	Bootstrapped
Wald	330.87	413.36	327.12	410.77
Log likelihood	-1239.476	-1240.367	-1240.367	-953.780

Legends: xfh1 = Panel data frontier model with net output value (for all crops) used as a dependent variable, assuming half-normal distribution of u_i and plot **invariant** technical efficiency; xfh2 = Panel data frontier model with wheat-equivalent net output (for all crops) used as a dependent variable, assuming half-normal distribution of u_i and plot **invariant** technical efficiency; xfh3 = Panel data frontier model with net output value (for all crops) used as a dependent variable, assuming half-normal distribution of u_i and plot **variant** technical efficiency; xfh4 = Panel data frontier model with wheat-equivalent net output (for all crops) used as a dependent variable, assuming half-normal distribution of u_i and plot **variant** technical efficiency. All are estimated by the random effects method. SE is robust standard error computed either by bootstrapping with 200 replications; *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

Table 10 summarizes results based on the semi-log functional form for the aggregated output model. The first two columns assumed plot invariant technical efficiency while the later assume plot variant efficiency. Columns 1 and 3 used net output value as a dependent variable while the others used wheat-equivalent net output as dependent variable. The models are used to further check robustness of results.

8. CONCLUSION

Land fragmentation is a common feature of agrarian society. Population pressure and egalitarian land reforms are usually understood as the principal factors of land fragmentations in Ethiopia. While policy makers in the past emphasized on plot diversification to protect people and enhance food self-sufficiency using land distribution as a safety net, the effect of this plot diversification on productivity and efficiency seemed to get little scholarly attention. Yet, the Government of Ethiopia appeared to radically shift its land policy from encouraging the small scale agriculture, for nearly two decades, to promoting large scale commercial farming. Whether this shift has emanated from evaluating the performance of the small scale agriculture is open to investigation, which gives raise to the first motive of this study. The intention toward contributing to the ongoing, yet unresolved, debate on the relationship between land fragmentation and productivity was another motive.

In this study I investigated the effect of land fragmentation (LF) in Northern Ethiopia on 1) farm productivity, 2) efficiency and 3) crop diversity using stochastic production frontier (SPF) analysis and farm household model (FHHM) with factor market imperfections. The analysis is carried out at plot level mainly in a plot-panel framework using a cross-sectional sample data of 421 households and their corresponding 1918 plots. Along with "plot size" and "farm size", I used three other land fragmentation indicators that are widely used in the literature: "number of operated plots", "distance to plots" and "SI-index" (where larger index means highly fragmented).

First, derived from the FHH model, productivity analysis in terms of net output value (and alternatively wheat equivalent net output) was carried out using different panel data specifications. The persistence of heteroskedasticity and non-normality of the error terms under the different models created the need to test the robustness of the results to different econometric model specifications.

I applied Fixed Effects, Random Effects, Hausman-Taylor, Random Effects Tobit and the panel data GLS models on 1) dominant-crop model, 2) two-main-crops model and 3) aggregated-output models as a means to test robustness of results to different model specifications. The following important results were revealed (all statistical tests were performed at the 0.05 level).

Productivity is associated with "number of plots operated" positively and significantly as expected. This association is robust to all model changes. The effect of "distance" is generally predicted to be insignificant. But the direction of its effect is not clear. It is predicted to have negative effect for the dominant crop model, mixed for the two-main crops model and positive for the aggregated output model. Further, it is less robust to changing functional forms triggering suspicion of measurement bias. The effect of "SI-index" was found to be negative and significant in the dominant crop model but insignificant in the two-crops and aggregated-output models; this trend is robust with all models and functional forms. The combined result offers important interpretation in that evaluation of land fragmentation by considering only the dominant crop is methodologically inconsistent when farm production involves diversification, since higher fragmentation can counterbalance the negative impact through diversification.

Moreover, inverse productivity relationship is observed at plot and farm level in the non-dominant crop model, even after controlling for land quality and other plot characteristics. Generally speaking, the result is robust to changing econometric specifications. Given the significant positive effect of number of plots, which may indicate the likelihood of positive scale effects, the inverse relationship is suspected to reflect effect of factor market imperfection (Bhalla and Roy, 1988), unobserved land quality heterogeneity (Udry, 1996), or plot size measurement bias (Holden et al., 2001) than diseconomies of scale (Niroula and Thapa, 2005).

Second, extended from the FHH model, crop diversity analysis in terms of diversity richness was carried out using panel data models for limited dependent variables, namely the Pooled and the Random Effects Poisson models, the Pooled Negative Binomial model and the Generalized Negative Binomial model on two functional forms: Lin-Lin and Lin-Log models. In all models, the result revealed that on-farm crop diversity is positively and significantly associated, as expected, with "number of plots operated" and "SI-index"; and negatively and significantly related to "farm size". These results are robust to all model specifications and to changing functional forms. Important inference from these results can be drawn; land fragmentation can encourage on-farm crop diversification.

Third, technical efficiency analysis was carried out by employing cross-sectional and panel data stochastic frontier analysis using different distributional assumptions as a means of checking robustness of results. Similar to productivity analysis, three models was framed: model for 1) dominant-crop, 2) two-main-crops and 3) aggregated-output. Moreover, further robustness check was performed assuming technical efficiency as 1) plot variant and 2) plot invariant. The main results include:

Farmers operating on *more* plots are relatively efficient from the view point of aggregated output analysis. While positively associated with efficiency, it is not significant in the dominant crop(s) model. Larger SI-index causes significant reduction in efficiency from the dominant crop point of view, but it is estimated to have insignificant reduction effect for the non-dominant crop models. The result is more or less robust. The effect of distance on efficiency is not clear. Though less robust to changing functional form, technical efficiency is predicted to associate inversely with farm size. This is consistent to the conclusion drawn in the previous page on productivity-size relationship.

In general, I found no evidence to the conventional claim that land fragmentation could be detrimental to productivity or efficiency; in fact, the results indicate to the opposite. I found productivity, efficiency and crop diversity to associate positively and significantly with the number of plots operated; and crop diversity to associate positively and significantly with SI-index. Thus, the increase in productivity and efficiency due to fragmentation is probably attributed to its indirect effect through diversification than due to diseconomies of scale. In fact, for a constant average plot size, the positive and significant association between number of operating plots and productivity or technical efficiency may witness the existence of economies of scale; altogether, suggesting that for the same total farm size, several small plots are preferred to one large plot. But this may give rise to the question 'how small do plots get before we get economies of scale?' that invites for further investigation.

Finally, it should be note that these findings are based on the existing data quality some of which may introduce measurement bias in to the models due to the subjective (plot distance) and traditional (plot size) estimation methods involved during data collection.

REFERENCES

BALTAGI, B. H., BRESSON, G. & PIROTTE, A. 2003. Fixed effects, random effects or Hausman–Taylor?: A pretest estimator. *Economics letters,* 79**,** 361-369.

BARDHAN, P. K. 1973. Size, Productivity, and Returns to Scale: An Analysis of Farm-Level Data in Indian Agriculture. *Journal of Political Economy,* 81**,** 1370-1386.

BELOTTI, F., DAIDONE, S., ILARDI, G. & ATELLA, V. 2012. Stochastic frontier analysis using Stata.

BENIN, S., SMALE, M., PENDER, J., GEBREMEDHIN, B. & EHUI, S. 2004. The economic determinants of cereal crop diversity on farms in the Ethiopian highlands. *Agricultural Economics,* 31**,** 197-208.

BENJAMIN, D. 1995. Can unobserved land quality explain the inverse productivity relationship? *Journal of Development Economics,* 46**,** 51-84.

BENTLEY, J. W. 1987. Economic and Ecological Approaches to Land Fragmantation: In Defense of A Much-Maligned Phenomenon. *Annual Review of Anthropology,* 16**,** 31-67.

BEYENE, A., GIBBON, D. & HAILE, M. 2006. Heterogeneity in land resources and diversity in farming practices in Tigray, Ethiopia. *Agricultural Systems,* 88**,** 61-74.

BEZABIH, M. & HOLDEN, S. Tenure insecurity, transaction costs in the land lease market and their implications for gendered productivity differentials. 26th International Conference of the International Association of Agricultural Economists, Brisbane, Australia, 2006.

BHALLA, S. S. 1988. Does land quality matter?: Theory and measurement. *Journal of Development Economics,* 29**,** 45-62.

BHALLA, S. S. & ROY, P. 1988. Mis-Specification in Farm Productivity Analysis: The Role of Land Quality. *Oxford Economic Papers,* 40**,** 55-73.

BIZIMANA, C., NIEUWOUDT, W. L. & FERRER, S. R. D. 2004. Farm size, land fragmentation and economic efficiency in southern Rwanda. *Agrekon,* 43**,** 244-262.

BLAREL, B., HAZELL, P., PLACE, F. & QUIGGIN, J. 1992. The Economics of Farm Fragmentation: Evidence from Ghana and Rwanda. *The World Bank Economic Review,* 6**,** 233-254.

CHEN, Z., HUFFMAN, W. E. & ROZELLE, S. 2009. Farm technology and technical efficiency: Evidence from four regions in China. *China Economic Review,* 20**,** 153-161.

COELLI, T. J. 1995. Recent developments in frontier modelling and efficiency measurement. *Australian Journal of Agricultural and Resource Economics,* 39**,** 219-245.

COLLIER, P. 2008. The Politics of Hunger: How Illusion and Greed Fan the Food Crisis. *Foreign Affairs,* 87**,** 67-79.

CORBEELS, M., SHIFERAW, A. & HAILE, M. 2000. *Farmers' knowledge of soil fertility and local management strategies in Tigray, Ethiopia*, IIED-Drylands Programme.

CORNWELL, C., SCHMIDT, P. & SICKLES, R. C. 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics,* 46**,** 185-200.

COTULA, L. 2009. *Land grab or development opportunity?: agricultural investment and international land deals in Africa*, Iied.

DEININGER, K., AYALEW, D. & ALEMU, T. 2006. Land Rental in Ethiopia: Marshallian Inefficiency or Factor Market Imperfections and Tenure Insecurity as Binding Constraints? *World Bank. Washington DC. Processed*.

DEININGER, K., JIN, S., GEBRE SELASSIE, H. S., ADENEW, B. & NEGA, B. 2003. Tenure Security and Land-Related Investment: Evidence from Ethiopia. *SSRN eLibrary*.

DEININGER, K. W. & BYERLEE, D. 2011. *Rising global interest in farmland: can it yield sustainable and equitable benefits?*, World Bank Publications.

DEL CORRAL, J., PEREZ, J. A. & ROIBAS, D. 2011. The impact of land fragmentation on milk production. *Journal of Dairy Science,* 94**,** 517-525.

DI FALCO, S. & CHAVAS, J.-P. 2006. Crop genetic diversity, farm productivity and the management of environmental risk in rainfed agriculture. *European Review of Agricultural Economics,* 33**,** 289-314.

DI FALCO, S., PENOV, I., ALEKSIEV, A. & VAN RENSBURG, T. M. 2010. Agrobiodiversity, farm profits and land fragmentation: Evidence from Bulgaria. *Land Use Policy,* 27**,** 763-771.

ELLIS, F. 1993. *Peasant economics*, Cambridge University Press.

GHEBRU, H. & HOLDEN, S. 2008. Factor market imperfections and rural land rental markets in Northern Ethiopian Highlands. *The Emergence of Land Markets in Africa: Assessing the Impacts on Poverty, Equity and Efficiency***,** 74-92.

GHEBRU, H. H. & HOLDEN, S. T. 2012. Reverse Share-Tenancy and Marshallian Inefficiency: Bargaining Power of Landowners and the Sharecroppers' Productivity. *CLTS Working Papers.* Centre for Land Tenure Studies, Norwegian University of Life Sciences.

GREENE, W. 2005. Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis,* 23**,** 7-32.

GREENE, W. H. 1990. A gamma-distributed stochastic frontier model. *Journal of econometrics,* 46**,** 141-163.

HAILE, M., WITTEN, W., ABRAHA, K., FISSHA, S., KEBEDE, A., KASSA, G. & REDA, G. 2005. Research Report 2. Land Registration in Tigray, Northern Ethiopia.

HOLDEN, S., SHIFERAW, B. & PENDER, J. 2001. Market Imperfections and Land Productivity in the Ethiopian Highlands. *Journal of Agricultural Economics,* 52**,** 53-70.

HOLDEN, S. & YOHANNES, H. 2002. Land redistribution, tenure insecurity, and intensity of production: A study of farm households in Southern Ethiopia. *Land Economics,* 78**,** 573-590.

HOLDEN, S. T. & BEZABIH, M. 2008. Gender and land productivity on rented land in Ethiopia. *The Emergence of Land Markets in Africa: Impacts on Poverty, Equity and Efficiency***,** 179-196.

HOLDEN, S. T., DEININGER, K. & GHEBRU, H. 2009. Impacts of low-cost land certification on investment and productivity. *American Journal of Agricultural Economics,* 91**,** 359-373.

HOLDEN, S. T., DEININGER, K. & GHEBRU, H. 2011. Tenure Insecurity, Gender, Low-cost Land Certification and Land Rental Market Participation in Ethiopia. *The Journal of Development Studies,* 47**,** 31-47.

HOLDEN, S. T., OTSUKA, K. & PLACE, F. M. 2012. *The Emergence of Land Markets in Africa:" Impacts on Poverty, Equity, and Efficiency"*, Routledge.

HUNG, V., PHAM, MACAULAY, P., T. GORDON & MARSH, S. P. 2007. The economics of land fragmentation in the north of Vietnam*. *Australian Journal of Agricultural and Resource Economics,* 51**,** 195-211.

ILBERY, B. W. 1984. Farm Fragmentation in the Vale of Evesham. *Area,* 16**,** 159-165.

JABARIN, A. S. & EPPLIN, F. M. 1994. Impacts of land fragmentation on the cost of producing wheat in the rain-fed region of northern Jordan. *Agricultural Economics,* 11**,** 191-196.

JIA, L. & PETRICK, M. 2011. How land fragmentation affects off-farm labor supply in China: Evidence from household panel data. *German Association of Agricultural Economists*

(GEWISOLA), 51st Annual Conference. Halle, Germany, September 28-30, 2011 ; http://purl.umn.edu/114522

KASSIE, M. & HOLDEN, S. 2007. Sharecropping efficiency in Ethiopia: threats of eviction and kinship. *Agricultural Economics,* 37**,** 179-188.

KING, R. & BURTON, S. 1982. Land fragmentation: Notes on a fundamental rural spatial problem. *Progress in Human Geography,* 6**,** 475-494.

KUMBHAKAR, S. C. & LOVELL, C. K. 2003. *Stochastic frontier analysis,* Cambridge University Press.

MCPHERSON, M. 1982. Land fragmentation: a selected literature review *Development Discussion Paper,* 141.

MOFED 2010. Growth and Transformation Plan (GTP) 2010/11 - 2014/15. *In:* DEVELOPMENT, T. F. D. R. O. E. M. O. F. A. E. (ed.). Addis Ababa.

MONCHUK, D., DEININGER, K. & NAGARAJAN, H. 2010. Does land fragmentation reduce efficiency: Micro evidence from India. *Paper prepared for presentation at the Agricultural & Applied Economics Association 2010 AAEA,CAES, & WAEA* Joint Annual Meeting, Denver, Colorado, July 25-27, 2010.

NGUYEN, T., CHENG, E. & FINDLAY, C. 1996. Land fragmentation and farm productivity in China in the 1990s. *China Economic Review,* 7**,** 169-180.

NIROULA, G. S. & 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.

NYSSEN, J., HAILE, M., MOEYERSONS, J., POESEN, J. & DECKERS, J. 2004. Environmental policy in Ethiopia: a rejoinder to Keeley and Scoones. *The Journal of Modern African Studies,* 42**,** 137-147.

NYSSEN, J., MUNRO, N., HAILE, M., POESEN, J., DESCHEEMAEKER, K., HAREGEWEYN, N., MOEYERSONS, J., GOVERS, G. & DECKERS, J. 2007. Understanding the environmental changes in Tigray: a photographic record over 30 years. *Tigray Livelihood Papers,* 3.

NYSSEN, J., POESEN, J., MOEYERSONS, J., HAILE, M. & DECKERS, J. 2008. Dynamics of soil erosion rates and controlling factors in the Northern Ethiopian Highlands–towards a sediment budget. *Earth Surface Processes and Landforms,* 33**,** 695-711.

PARIKH, A. & SHAH, K. 1994. MEASUREMENT OF TECHNICAL EFFICIENCY IN THE NORTH-WEST FRONTIER PROVINCE OF PAKISTAN. *Journal of Agricultural Economics,* 45**,** 132-138.

PENDER, J. & GEBREMEDHIN, B. 2006. Land management, crop production, and household income in the highlands of Tigray, Northern Ethiopia: An econometric analysis. *Strategies for sustainable land management in the East African highlands,* 107.

RAHMAN, S. & RAHMAN, M. 2009. Impact of land fragmentation and resource ownership on productivity and efficiency: The case of rice producers in Bangladesh. *Land Use Policy,* 26**,** 95-103.

ROSSET, P. 2000. The Multiple Functions and Benefits of Small Farm Agriculture in the Context of Global Trade Negotiations Development *Development* 43.

SADOULET, E. & DE JANVRY, A. 1995. *Quantitative development policy analysis,* Johns Hopkins University Press Baltimore.

SCHULTZ, T. W. 1964. Transforming traditional agriculture. *Transforming traditional agriculture.*

SEGERS, K., DESSEIN, J., HAGBERG, S., TEKLEBIRHAN, Y., HAILE, M. & DECKERS, J. 2010. Unravelling the dynamics of access to farmland in Tigray, Ethiopia: The ‘emerging land market’ revisited. *Land Use Policy,* 27**,** 1018-1026.

SIKOR, T., MÜLLER, D. & STAHL, J. 2009. Land Fragmentation and Cropland Abandonment in Albania: Implications for the Roles of State and Community in Post-Socialist Land Consolidation. *World Development,* 37**,** 1411-1423.

SINGH, I., SQUIRE, L. & STRAUSS, J. 1986. A Survey of Agricultural Household Models: Recent Findings and Policy Implications. *The World Bank Economic Review,* 1**,** 149-179.

SNOWDON, B. & VANE, H. 2005. *Modern macroeconomics: its origins, development and current state*, Northampton, MA, USA, Edward Elgar Publishing, Inc.

STIGLITZ, J. E. 1974. Incentives and Risk Sharing in Sharecropping. *The Review of Economic Studies,* 41**,** 219-255.

TAMENE, L. & VLEK, P. L. 2008. Soil erosion studies in northern Ethiopia. *Land Use and Soil Resources.* Springer.

TAN, S., HEERINK, N., KRUSEMAN, G. & 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.

TAN, S., HEERINK, N., KUYVENHOVEN, A. & QU, F. 2010. Impact of land fragmentation on rice producers’ technical efficiency in South-East China. *NJAS - Wageningen Journal of Life Sciences,* 57**,** 117-123.

TAN, S., HEERINK, N. & QU, F. 2006. Land fragmentation and its driving forces in China. *Land Use Policy,* 23**,** 272-285.

TRPFB 2011. Five years (2010/11-2014/15) Growth and Transformation Plan. . Tigray Regional Plan and Finance Burea.

UDRY, C. 1996. *Efficiency and market structure: testing for profit maximization in African agriculture*, Department of Economics, Northwestern Univ.

VAN DUSEN, M. & TAYLOR, J. E. 2005. Missing markets and crop diversity: evidence from Mexico. *Environment and Development Economics,* 10**,** 513-531.

VAN DUSEN, M. E. 2000. *In situ conservation of crop genetic resources in the Mexican milpa system.* University of California.

WADUD, M. A. 2003. Technical, Allocative, and Economic Efficiency of Farms in Bangladesh: A Stochastic Frontier and DEA Approach. *The Journal of Developing Areas,* 37**,** 109-126.

WAN, G. H. & CHENG, E. 2001. Effects of land fragmentation and returns to scale in the Chinese farming sector. *Applied Economics,* 33**,** 183-194.

WATTANUTCHARIYA, S. & JITSANGUAN, T. 1992. Increasing the scale of small-farm operations in Thailand. *Extension Bulletins*.

WOOLDRIDGE, J. M. 2010. *Econometric Analysis Cross Section Panel*, MIT press.

WU, Z., LIU, M. & DAVIS, J. 2005. Land consolidation and productivity in Chinese household crop production. *China Economic Review,* 16**,** 28-49.

APPENDIX

A. Expansion of the household model

Eq. (4.12) can be further expanded for each factor and types of household as follows:

$$\frac{\partial \mathcal{L}}{\partial z'} = p_q \alpha_i \frac{\partial Q(.)}{\partial z'} = p'_z \text{ ; } where\,' \in \{s,*,b\} \text{ } and \text{ } z \in \{A_i, L_i, X_i, K_i\} \tag{48}$$

Type	Land	Labor	Capital	Other Inputs
Seller	$\frac{\partial \mathcal{L}}{\partial A^s} = p_q \alpha_i \frac{\partial Q(.)}{\partial A^s} = p_a^s$	$\frac{\partial \mathcal{L}}{\partial L^s} = p_q \alpha_i \frac{\partial Q(.)}{\partial L^s} = p_l^s$	$\frac{\partial \mathcal{L}}{\partial K^s} = p_q \alpha_i \frac{\partial Q(.)}{\partial K^s} = p_k^s$	$\frac{\partial \mathcal{L}}{\partial X^s} = p_q \alpha_i \frac{\partial Q(.)}{\partial X^s} = p_x^s$
Self-suf	$\frac{\partial \mathcal{L}}{\partial A^*} = p_q \alpha_i \frac{\partial Q(.)}{\partial A^*} = p_a^*$	$\frac{\partial \mathcal{L}}{\partial L^*} = p_q \alpha_i \frac{\partial Q(.)}{\partial L^*} = p_l^*$	$\frac{\partial \mathcal{L}}{\partial K^*} = p_q \alpha_i \frac{\partial Q(.)}{\partial K^*} = p_k^*$	$\frac{\partial \mathcal{L}}{\partial X^*} = p_q \alpha_i \frac{\partial Q(.)}{\partial X^*} = p_x^*$
Buyer	$\frac{\partial \mathcal{L}}{\partial A^b} = p_q \alpha_i \frac{\partial Q(.)}{\partial A^b} = p_a^b$	$\frac{\partial \mathcal{L}}{\partial L^b} = p_q \alpha_i \frac{\partial Q(.)}{\partial L^b} = p_l^b$	$\frac{\partial \mathcal{L}}{\partial K^b} = p_q \alpha_i \frac{\partial Q(.)}{\partial K^b} = p_k^b$	$\frac{\partial \mathcal{L}}{\partial X^b} = p_q \alpha_i \frac{\partial Q(.)}{\partial X^b} = p_x^b$

B. Robustness of Results of Productivity analysis

Robustness of the results were tested using: 3 functional forms (Log-Semi-Log, Log-Lin, and Log-Log forms), 3 dependent variables (aggregated output model, dominant crop model, two main crops model) and 2 dimensions of the dependent variable (net output value and net wheat-equivalent output) repeated over the 7 econometric specifications for panel data. It would be tedious to present all the result here, but I summarized only the predicted values of land fragmentation indicators in the following five tables.

Log – Semi-Log Model

$$lnY_{ip} = \beta_o + \beta_{i1} ln\boldsymbol{X}_{ip1} + \beta_{i2}\boldsymbol{X}_{ip2} + \gamma_i D_{ip} + \alpha_i + u_{ip}$$

Log – Log Model

$$lnY_{ip} = \beta_o + \beta_i ln\boldsymbol{X}_{ip} + \gamma_i D_{ip} + \alpha_i + u_{ip}$$

Log – Lin Model

$$lnY_{ip} = \beta_o + \beta_i \boldsymbol{X}_{ip} + \gamma_i D_{ip} + \alpha_i + u_{ip}$$

Where lnY_{ip} is the *log* of net output value (Group 1) or wheat-equivalent net output (Group 2) of aggregated output (A), dominant crop (*teff*) (1) or two-main crops (*teff & wheat*) (2); $\boldsymbol{\beta}_i$ is the vector of unknown parameters to be estimated; $\boldsymbol{X}_{ip}$ is a vector of linear exogenous continuous variables conditioning Y_{ip}; $ln\boldsymbol{X}_{ip}$ is a vector of exogenous continuous variables expressed in *logarithmic* form; α_i the unobserved heterogeneity; D_{ip} is a vector of dummy variables; and u_{ip} is the idiosyncratic error terms.

B.1. Robustness of the effect of *number of plots* on productivity

No.of.plots	**RE**	**MLE**	**HT**	**TOBIT**	**GLS**
Group 1 - Net output Value (birr/ha) – Dependent variable					
SL-A	0.565*	0.561**	0.572**	0.561***	0.515***
SL-1	0.946**	0.945**	0.788	0.945**	0.810***
SL-2	0.713***	0.682**	0.698**	0.682**	0.572***
Lin-A	0.068*	0.065**	0.065**	0.044***	0.073*
Lin-1	0.107	0.106**	0.051	0.106**	0.082***
Lin-2	0.084	0.081**	0.05	0.081**	0.074***
Ln-A	0.5	0.436**	0.542*	0.436**	0.402**
Ln-1	0.909**	0.909**	0.839	0.909**	0.813***
Ln-2	0.73***	0.669**	0.685*	0.669**	0.57***
Group 2 - Net wheat equivalent output (Quintal/ha) – Dependent variable					
SL-A	0.514*	0.512**	0.425*	0.513**	0.558***
SL-1	0.918**	0.918***	0.788	0.861**	0.779***
SL-2	0.702**	0.695**	0.698*	0.658**	0.508***
Lin-A	0.080**	0.078***	0.079**	0.075***	0.060***
Lin-1	0.108	0.108**	0.051	0.096**	0.089***
Lin-2	0.082*	0.082**	0.050*	0.076**	0.073***
Ln-A	0.525*	0.520**	0.433*	0.519**	0.537***
Ln-1	0.886**	0.886**	0.839	0.834**	0.779***
Ln-2	0.687**	0.678**	0.685*	0.643**	0.503***
SE	Clustered	Bootstrapped	Bootstrapped	Bootstrapped	Bootstrapped

Legends: In the first column: SL = the log-semi-log functional form; Ln = the full log functional form; Lin = the log-lin form; A = the Aggregated output model; 1 = the Dominant crop model (*teff*); and 2 = the two main crops model (*teff* and wheat).

In the first row: RE= the Random Effects model; MLE= Random Effects Estimated by Maximum Likelihood Estimator; HT= the Hausman-Taylor Model; Tobit= the random effects tobit model; GLS= the Generalized Least Square Estimator for Panel Data. All Models in Group 1 used "Net output value (birr/ha)" as a dependent variable, while in Group 2 "Net Wheat-Equivalent Output (quintal/ha)" is used as a dependent variable.

In models with bootstrapped standard errors, each bootstrapping is replicated on 200 re-sampled observations; otherwise, the standard error is clustered at village level.*significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

B.2. Robustness of the effect of *distance* on Productivity

Distance	RE	FE	MLE	HT	TOBIT	GLS
Group 1 - Net output Value (birr/ha) – Dependent variable						
SL-A	0.008	-0.004	0.01	0.001	0.01	0.024
SL-1	-0.013	-0.019	-0.014	-0.014	-0.014	-0.008
SL-2	-0.033	-0.019	-0.032	-0.002	-0.032	0.009
Lin-A	0.002**	0.002*	0.002*	0.002*	0.002**	0.002*
Lin-1	0.001	0.002	0.001	0.002	0.001	0.002**
Lin-2	0.000	0.000	0.000	0.000	0.000	0.000
Ln-A	0.034	0.037*	0.036	0.003	0.036	0.037
Ln-1	-0.012	-0.005	-0.013	-0.014	-0.013	-0.004
Ln-2	-0.033	-0.021	-0.032	0.000	-0.032	0.007
Group 2 - Net wheat equivalent output (Quintal/ha) – Dependent variable						
SL-A	0.022	-0.003	0.024	0.006	0.021	0.044**
SL-1	-0.007	-0.024	-0.007	-0.014	-0.004	0.001
SL-2	-0.011	0.001	-0.01	-0.002	-0.015	0.013
Lin-A	0.002***	0.002	0.003**	0.002*	0.002**	0.003***
Lin-1	0.002	0.002	0.002	0.002	0.002	0.002**
Lin-2	0.001	0.000	0.001	0.000	0.000	0.001**
Ln-A	0.023	-0.002	0.025	0.007	0.023	0.050***
Ln-1	-0.009	-0.018	-0.009	-0.014	-0.007	0.002
Ln-2	-0.011	-0.002	-0.01	0.000	-0.015	0.009
SE	Clustered	Clustered	Bootstrap	Bootstrap	Bootstrap	Bootstrap

Legends: In the first column: SL = the log-semi-log functional form; Ln = the full log functional form; Lin = the log-lin form; A = the Aggregated output model; 1 = the Dominant crop model (*teff*); and 2 = the two main crops model (*teff* and wheat).

In the first row: RE= the Random Effects model; MLE= Random Effects Estimated by Maximum Likelihood Estimator; HT= the Hausman-Taylor Model; Tobit= the random effects tobit model; GLS= the Generalized Least Square Estimator for Panel Data. All Models in Group 1 used "Net output value (birr/ha)" as a dependent variable, while in Group 2 "Net Wheat-Equivalent Output (quintal/ha)" is used as a dependent variable.

In models with bootstrapped standard errors, each bootstrapping is replicated on 200 re-sampled observations; otherwise, the standard error is clustered at village level.*significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

B.3. Robustness of effect of *SI-index* on productivity

SI-index	RE	MLE	HT	TOBIT	GLS
Group 1 - Net output Value (birr/ha) – Dependent variable					
SL-A	-0.344	-0.365	-0.438	-0.365	-0.439
SL-1	-2.621**	-2.619*	-2.867	-2.619*	-2.293***
SL-2	-1.214	-1.178	-0.899	-1.178	-0.733*
Lin-A	-0.275	-0.263	-0.263	-0.203	-0.132
Lin-1	-0.808	-0.803	-0.847	-0.803	-0.832***
Lin-2	-0.519	-0.517	-0.382	-0.517	-0.466**
Ln-A	-0.509	-0.207	-0.395	-0.207	-0.396
Ln-1	-2.551**	-2.551*	-2.865	-2.551*	-2.303***
Ln-2	-1.212	-1.171	-0.77	-1.171	-0.779*
Group 2 - Net wheat equivalent output (Quintal/ha) – Dependent variable					
SL-A	-0.07	-0.072	-0.311	-0.05	-0.303
SL-1	-2.290**	-2.294*	-2.867	-2.171*	-1.657***
SL-2	-0.965	-0.953	-0.899	-0.905	-0.18
Lin-A	-0.303	-0.288	-0.213	-0.239	-0.082
Lin-1	-0.816	-0.816	-0.847	-0.707	-0.765***
Lin-2	-0.468	-0.471	-0.382	-0.384	-0.310*
Ln-A	-0.121	-0.118	-0.3	-0.095	-0.302
Ln-1	-2.247**	-2.250*	-2.865	-2.146*	-1.691***
Ln-2	-0.954	-0.938	-0.77	-0.892	-0.286
SE	Clustered	Bootstrap	Bootstrap	Bootstrap	Bootstrap

Legends: In the first column: SL = the log-semi-log functional form; Ln = the full log functional form; Lin = the log-lin form; A = the Aggregated output model; 1 = the Dominant crop model (*teff*); and 2 = the two main crops model (*teff* and wheat).

In the first row: RE= the Random Effects model; MLE= Random Effects Estimated by Maximum Likelihood Estimator; HT= the Hausman-Taylor Model; Tobit= the random effects tobit model; GLS= the Generalized Least Square Estimator for Panel Data. All Models in Group 1 used "Net output value (birr/ha)" as a dependent variable, while in Group 2 "Net Wheat-Equivalent Output (quintal/ha)" is used as a dependent variable.

In models with bootstrapped standard errors, each bootstrapping is replicated on 200 re-sampled observations; otherwise, the standard error is clustered at village level.*significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

B.4. Robustness of effect of *Plot Size* on productivity

Plot Size	**RE**	**FE**	**MLE**	**HT**	**TOBIT**	**GLS**
Group 1 - Net output Value (birr/ha) – Dependent variable						
SL-A	-0.331	0.255	-0.344*	0.167	-0.344**	-0.477***
SL-1	-0.405	-0.123	-0.403	-0.129	-0.403	-0.395***
SL-2	-0.485*		-0.485**	-0.469	-0.485**	-0.464***
Lin-A	-1.333***		-1.357***	-1.357***	-1.433***	-1.160***
Lin-1	-0.675***	-0.492	-0.679**	-0.319	-0.679**	-0.851***
Lin-2	-1.242***	-0.996	-1.254***	-1.239***	-1.254***	-1.269***
Ln-A	-0.463**	-0.321	-0.350**	0.169	-0.350**	-0.423***
Ln-1	-0.415		-0.414*	-0.039	-0.414*	-0.364***
Ln-2	-0.520*		-0.521**	-0.57	-0.521**	-0.503***
Group 2 - Net wheat equivalent output (Quintal/ha) – Dependent variable						
SL-A	-0.299	0.225	-0.313*	0.382	-0.302*	-0.458***
SL-1	-0.340*	-0.26	-0.341	-0.129	-0.336	-0.376***
SL-2	-0.494**		-0.496**	-0.469	-0.481***	-0.492***
Lin-A	-1.418***		-1.436***	-1.269***	-1.343***	-1.442***
Lin-1	-0.719***	-0.512	-0.719**	-0.319	-0.664**	-0.804***
Lin-2	-1.301***	-1.071*	-1.310***	-1.239***	-1.187***	-1.350***
Ln-A	-0.337	0.205	-0.352**	0.369	-0.339**	-0.489***
Ln-1	-0.351*		-0.351	-0.039	-0.345	-0.349***
Ln-2	-0.520**		-0.523***	-0.57	-0.506***	-0.512***
SE	Clustered	Clustered	Bootstrap	Bootstrap	Bootstrap	Bootstrap

Legends: In the first column: SL = the log-semi-log functional form; Ln = the full log functional form; Lin = the log-lin form; A = the Aggregated output model; 1 = the Dominant crop model (*teff*); and 2 = the two main crops model (*teff* and wheat).

In the first row: RE= the Random Effects model; MLE= Random Effects Estimated by Maximum Likelihood Estimator; HT= the Hausman-Taylor Model; Tobit= the random effects tobit model; GLS= the Generalized Least Square Estimator for Panel Data. All Models in Group 1 used "Net output value (birr/ha)" as a dependent variable, while in Group 2 "Net Wheat-Equivalent Output (quintal/ha)" is used as a dependent variable.

In models with bootstrapped standard errors, each bootstrapping is replicated on 200 re-sampled observations; otherwise, the standard error is clustered at village level.*significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

B.5. Robustness of the effect of Operated *Sub-Farm Size* on productivity

S-Farm size	**RE**	**FE**	**MLE**	**HT**	**TOBIT**	**GLS**
Group 1 - Net output Value (birr/ha) – Dependent variable						
SL-A	-0.119	-0.091	-0.119	-0.006	-0.119*	-0.097**
SL-1	-0.231	-0.805	-0.231*	-0.08	-0.231*	-0.240***
SL-2	-0.159	-0.722	-0.158*	-0.168	-0.158*	-0.162***
Lin-A	-0.134	1.076**	-0.128**	-0.128**	-0.078**	-0.093
Lin-1	-0.212*		-0.212**	-0.009	-0.212**	-0.176***
Lin-2	-0.199*		-0.200**	-0.104	-0.200**	-0.197***
Ln-A	-0.092	-0.092	-0.098	0.021	-0.098	-0.056
Ln-1	-0.225	-0.827	-0.225*	-0.051	-0.225*	-0.231***
Ln-2	-0.151	-0.727	-0.149	-0.177	-0.149	-0.167***
Group 2 - Net wheat equivalent output (Quintal/ha) – Dependent variable						
SL-A	-0.168*	-0.039	-0.170**	-0.002	-0.149**	-0.157***
SL-1	-0.237	-0.715	-0.236*	-0.08	-0.229*	-0.281***
SL-2	-0.186	-0.605	-0.186**	-0.168	-0.178**	-0.206***
Lin-A	-0.170*	1.161***	-0.166***	-0.123	-0.152***	-0.141***
Lin-1	-0.221*		-0.221**	-0.009	-0.207**	-0.195***
Lin-2	-0.202*		-0.203***	-0.104	-0.198***	-0.212***
Ln-A	-0.153	-0.041	-0.155**	0.02	-0.135*	-0.139***
Ln-1	-0.228	-0.724	-0.227*	-0.051	-0.221*	-0.273***
Ln-2	-0.178	-0.606	-0.179*	-0.177	-0.170*	-0.208***
SE	Clustered	Clustered	Bootstrap	Bootstrap	Bootstrap	Bootstrap

Legends: In the first column: SL = the log-semi-log functional form; Ln = the full log functional form; Lin = the log-lin form; A = the Aggregated output model; 1 = the Dominant crop model (*teff*); and 2 = the two main crops model (*teff* and wheat).

In the first row: RE= the Random Effects model; MLE= Random Effects Estimated by Maximum Likelihood Estimator; HT= the Hausman-Taylor Model; Tobit= the random effects tobit model; GLS= the Generalized Least Square Estimator for Panel Data. All Models in Group 1 used "Net output value (birr/ha)" as a dependent variable, while in Group 2 "Net Wheat-Equivalent Output (quintal/ha)" is used as a dependent variable.

In models with bootstrapped standard errors, each bootstrapping is replicated on 200 re-sampled observations; otherwise, the standard error is clustered at village level.*significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

C. Further Analysis of Crop Diversity

Controlling for *wereda* specific effects

Diversity Richeness	POOLS	POISSON	POISSON-RE	N.BINOM	G.N.BINOM
hh-age	0.002*	0.002*	0.001	0.002*	0.001
hh-sex	0.032	0.031	0.026	0.031	0.026
hh-edu	-0.027***	-0.028***	-0.027	-0.028***	-0.027
hh-size	-0.011*	-0.013**	-0.011	-0.013**	-0.011
lfem.labor	0.013	0.007	0.011	0.007	0.011
lmalelabor	-0.079***	-0.080***	-0.078	-0.080***	-0.078
ladulteqvt	0.133***	0.157***	0.135	0.157***	0.135
lofarmszh	-0.060**	-0.046*	-0.053	-0.046*	-0.053
si-index	0.985***	0.834***	0.808***	0.834***	0.808***
distance	0.000	0.000	0.000	0.000	0.000
no.of.plot	0.057***	0.055***	0.060***	0.055***	0.060***
soildepth	-0.022*	-0.023**	-0.021	-0.023**	-0.021
landqlty	0.026*	0.023*	0.022	0.023*	0.022
soiltype	-0.01	-0.005	-0.008	-0.005	-0.008
slope	0.050***	0.051***	0.042*	0.051***	0.042
loxen.own	-0.045***	-0.068***	-0.055*	-0.068***	-0.055*
_Iwereda_3	0.026	0.036	0.018	0.036	0.018
_Iwereda_4	-0.02	-0.055*	-0.045	-0.055*	-0.045
_Iwereda_5	-0.039	-0.051*	-0.05	-0.051*	-0.05
_Iwereda_6	0.039	0.032	0.034	0.032	0.034
_Iwereda_7	-0.015	0.008	0.000	0.008	0.000
_Iwereda_8	0.041	0.059*	0.053	0.059*	0.053
_Iwereda_9	0.220***	0.245***	0.228**	0.245***	0.228**
_Iwereda_10	0.272***	0.245***	0.245**	0.245***	0.245**
_Iwereda_11	-0.063*	-0.057*	-0.06	-0.057*	-0.06
_Iwereda_12	-0.131***	-0.158***	-0.150*	-0.158***	-0.150*
_cons	-0.085	0.071	0.104	0.071	14.993
N	1446	1446	1446	1446	1446
SE	CLUSTERED	BOOTS	BOOTS	BOOTS	BOOTS

Legend: POOLED = the Pooled OLS model with the dependent variable in log form, as explained in Chapter 6.2; POISSON = the Pooled Poisson Regression Model; POISSON-RE = the Random Effects Poisson Regression Model; N.BINOM = the Pooled Negative Binomial Regression Model; G.N.BINOM = the Generalized Negative Binomial Regression Model; SE = Robust standard error computed either by bootstrapping technique (*BOOTS*) with 200 replications or clustering the error terms at village level (*CLUSTERED*); *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

D. Further Efficiency Analysis using the Frontier Models:

D1. Cross-Sectional Frontier Model: *Log-Semi-Log* Form

	fh1	fx1	fx2	fh3	fh4
Production Function					
lplotsize	-0.373***	-0.345***	-0.293**	-0.17	-0.307***
dmanure	0.078	0.075	0.105	0.197	0.128
durea	-0.054	-0.034	-0.084	0.211	-0.001
ddap	0.146	0.119	0.171*	-0.064	0.084
dherbpest	-0.002	-0.002	-0.002	-0.013**	-0.003
lfam.labr	0.318***	0.323***	0.292***	0.352***	0.346***
loxen.own	-0.039	-0.038	-0.004	0.07	-0.033
_cons	6.702***	6.488***	0.415*	0.19	0.307
$ln(\sigma_v^2)$	-1.230***	-0.989***	-1.183***	-1.442***	-1.623***
Technical inefficiency					
hh-sex	-0.316*	-0.514*	-0.780***	-0.607	-0.508*
hh-age	-0.002	-0.003	0.003	-0.005	-0.007
hh-edu	-0.086	-0.101	-0.081	-0.139	-0.241**
ldist	-0.068	-0.104	-0.103	0.108	0.043
si-index	0.244	0.485	-0.792	7.416*	2.51
lno.of.plot	-0.887*	-1.284*	-1.393*	-3.177**	-1.960*
lsfarmsz	0.262*	0.372*	0.556**	0.753**	0.519*
lsolev	-0.311***	-0.447***	-0.345**	-0.103	-0.202
landqlty	-0.161	-0.244	-0.178	-0.402*	-0.327*
slope	0.036	0.061	0.081	-0.282	0.049
soiltype	0.036	0.064	-0.004	0.039	0.118
soildepth	0.096	0.122	0.301*	0.049	0.109
plotrent	0.341*	0.476*	0.197	0.245	0.590*
lfem.labor	0.047	0.057	-0.308	-0.680*	-0.436*
lmalelabor	-0.062	-0.072	-0.227	-0.41	-0.378
ladulteqvt	0.26	0.322	0.64	1.325*	0.909*
_cons	1.844	1.534	1.949	-1.2	1.065
N	959	959	1001	279	525
SE	BOOTS	BOOTS	BOOTS	BOOTS	BOOTS

Legends: fh = frontier model with normal-half-normal distribution; fx = frontier model with normal-exponential distribution; The numerical subscripts: 1 = Aggregated crop model, in net-output-value (birr/ha); 2 = Aggregated crop model, in wheat-equivalent net output (quintal/ha); 3 = Dominant crop model (teff), in net-output value (birr/ha); 4 = Two main crops model (teff and wheat), net-output value (birr/ha). SE= Bootstrapped (BOOTS) standard errors with 200 replications; *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.

D2. Cross-Sectional Frontier Model: *Log-Lin* Form

	fh1	**fx1**	**fx2**	**fh3**	**fh4**
Production Function					
plotsize	-1.277***	-1.234***	-1.273***	-0.875***	-1.290***
dmanure	0.148**	0.139**	0.137**	0.248**	0.194***
durea	-0.017	0.005	-0.047	0.270*	0.003
ddap	0.139	0.107	0.142	-0.141	0.062
dherbpest	-0.001	-0.002	-0.001	-0.007	-0.001
fam.labr	0.002***	0.002***	0.002***	0.002*	0.001*
oxen.own	0.001	0.001	0.001	0.003*	0.002
_cons	8.757***	8.495***	2.325***	2.255***	2.502***
$ln(\sigma_v^2)$	-1.176***	-0.963***	-1.144***	- 1.524***	- 1.637***
Technical Inefficiency					
hh-sex	-0.052	-0.114	-0.322	-0.277	-0.063
hh-age	-0.007	-0.01	-0.005	-0.012	-0.014*
hh-edu	-0.066	-0.081	-0.07	-0.01	-0.098
distance	-0.003	-0.005	-0.006	-0.003	-0.001
si-index	-0.536	-0.732	-0.583	0.992	0.678
no.of.plot	-0.041	-0.058	-0.179**	-0.260*	-0.199**
sfarmsz	0.181*	0.263*	0.452***	0.477*	0.444***
solevisit	-0.009**	-0.014**	-0.012**	-0.009	-0.005
landqlty	-0.145	-0.213	-0.139	-0.105	-0.269*
slope	-0.063	-0.084	-0.036	-0.447*	-0.191
soiltype	0.025	0.04	0.049	0.024	0.126
soildepth	0.208*	0.297*	0.410***	0.091	0.125
plotrent	0.219	0.268	0.217	0.153	0.388*
fem.labor	0.003	0.005	-0.002	-0.032**	-0.025**
malelabor	-0.001	-0.001	-0.006	-0.030*	-0.025**
adulteqvt	0.001	0.001	0.004	0.029**	0.020***
_cons	0.867	0.044	-0.26	1.027	0.8
N	1311	1311	1371	378	713
SE	BOOTS	BOOTS	BOOTS	BOOTS	BOOTS

Legends: fh = frontier model with normal-half-normal distribution; fx = frontier model with normal-exponential distribution; The numerical subscripts: 1 = Aggregated crop model, in net-output-value (birr/ha); 2 = Aggregated crop model, in wheat-equivalent net output (quintal/ha); 3 = Dominant crop model (teff), in net-output value (birr/ha); 4 = Two main crops model (teff and wheat), net-output value (birr/ha). SE= Bootstrapped (BOOTS) standard errors with 200 replications; *significant at 5 per cent; ** significant at 1 per cent; ***significant at 0.1 per cent.