CREDIT Research Paper

No. 17/04

Crop Diversification and Technical Efficiency in Afghanistan: Stochastic Frontier Analysis

by

Hayatullah Ahmadzai

Abstract:
Using Stochastic Frontier Analysis (SFA), this paper centered on analysing the impact of Crop Diversification (CD) on farm level technical efficiency in Afghanistan. Data from a household level survey conducted in 2013-2004 by the Central Statistic Organization (CSO) is used in the analysis. The results revealed that adoption of a diversified portfolio of crops by the farmers significantly improves technical efficiency. In addition, access to extension services, farm size, cattle, oxen and tractor ownership by the farm households, and regional variables were other important factors that significantly affect technical efficiency. It is evident from the results that the estimated technical efficiency indices from the preferred truncated normal distribution range from 1.5% to 99.29%, with a sample mean of 71.9%. The basic SFA model was investigated for potential endogeneity in crop diversification. Instrumental Variable (IV) method was employed to correct for endogeneity in crop diversification. The results of the IV estimation reveal that failing to account for endogeneity in the basic model leads to a downward bias which is consistent with attenuation bias (measurement error in CD implies that OLS coefficients are biased towards zero, so one would predict IV coefficients greater in absolute size). The results of crop diversification index showed the presence of a relatively low level of crop diversification. Maximum likelihood estimation of translog stochastic frontier model shows that land, labour, and other purchased inputs (fertilizer, seeds, pesticides usage) have positive impact on farm revenues. The results show an evidence of constant returns-to-scale.

JEL Classification: O12, Q12, O13, Q18, D24

Keywords: Agricultural economics, Afghanistan, applied econometrics, technical efficiency, crop diversificaiton

Centre for Research in Economic Development & International Trade, University of Nottingham

CREDIT Research Paper

No.17/04

Crop Diversification and Technical Efficiency in Afghanistan: Stochastic Frontier Analysis

by

Hayatullah Ahmadzai

Outline

The Author
Hayatullah Ahmadzai is a PhD student in the School of Economics at the University of Nottingham. Email: hayatullah.ahmadzai@nottingham.ac.uk

Acknowledgements

The author is grateful to Professor Oliver Morrissey and Dr Simon Appleton for their useful comments and support. Thanks to the IDB for thier financial support to make this work possible. The usual disclaimer applies.

Research Papers at www.nottingham.ac.uk/economics/credit

I. Introduction

Measuring economic performance of a farm requires an understanding of production decisions and the levels of technical efficiency. Technical Efficiency as a precondition for economic efficiency safeguards the economic viability and sustainability of a farm. Farm productivity can be improved by adopting technology such as introduction of new machinery, chemicals, and improved seed varieties. Alternatively, productivity can be enhanced by changing how factors are combined to improve the efficiency by which inputs are being transformed into output such that higher outputs are produced from the same level of inputs and technology (Coelli, 1995). Production decisions by farmers also affect the level of technical efficiency and overall productivity of a farm, for instance decisions by farmers to shift away from specialization towards adopting a diversified production system.

Empirical research suggests that farmers in developing countries fail to exploit fully the production technology and production resources and often make inefficient decisions. This study attempts to measure farm-specific technical efficiency of crop farmers and identify potential factors determining technical efficiency using Stochastic Frontier Analysis (SFA) techniques. The main interest is to analyse the degree or extent of crop diversification and its impact on the levels of technical efficiency in Afghanistan.

As investment in the farm sector increases production and production efficiency, contributing to economic growth, farmers are likely to switch from subsistence agriculture based on self-sufficiency to profit and income-oriented decision making, henceforth farm output is accordingly more responsive to market trends. This transition from subsistence food production to a commercially oriented system typically involves crop diversification (Minot et al. 2006; Ibrahim et al. 2009; and Nguyen 2014). Hence, the choice and the extent of crop diversification may depend on the degree of commercialization[1] of the farms (i.e. subsistence, semi-commercial, or fully commercial systems).

Since emerging out of conflict and establishing a modern economy in 2001, Afghanistan has undergone drastic economic policy change. The economy was on the verge of collapse due to conflict and political instability, lack of a sound economic policy and inefficiencies of public institutions. However, when international aid agencies began to pledge aid to support the economy, particularly agricultural economy, transition towards a fully market-led system began. Many challenges and uncertainties have resulted from policy changes made over the last fifteen years, all of which have influenced farming practices and production decision making in the country's faltering progress towards a market economy.

Against this backdrop and keeping the importance of recent changes in the afghan agriculture economy in view, it is important to investigate the levels of technical efficiency at the farm level and its determinants in Afghanistan. Identifying determinants of technical efficiency is a major task in efficiency analysis. Therefore, it is also essential to examine how production decisions by farmers, particularly crop diversification strategies as a major

[1] Describes the extent of market participation. In substance farming system, production mainly takes place for the household consumption, in semi-commercial system part of the produce is sold and part of it is consumed by the household, and in fully commercial system most of the production takes place for the market.

factor, affect the level of technical efficiency. To better introduce this study, it is essential to firstly highlight the recent developments and significance of the agriculture sector in Afghan economy.

1.1. Background

Afghanistan current population is 31.63 million with an average annual growth rate of about 2.5%. More than 80 percent of the country's population, and nearly 90 percent of the poor, live in rural areas, and agriculture plays an important role in their livelihoods and income.

According to the World Bank data, Afghan economy experienced a steady growth at an annual rate of 9.4% between 2003 and 2012 (Figure 1). International aid is an important part of the GDP growth. Official development aid and military assistance grew steadily from US$404 million in 2002 to more than US$15.7 billion in 2010. About a third of this aid went into the development and civilian infrastructure. The resulting development outcome is significant, for instance the GDP per capita (in current international $PPP) increased from $896 in 2002 to $1,925 in 2012 (Figure 1).

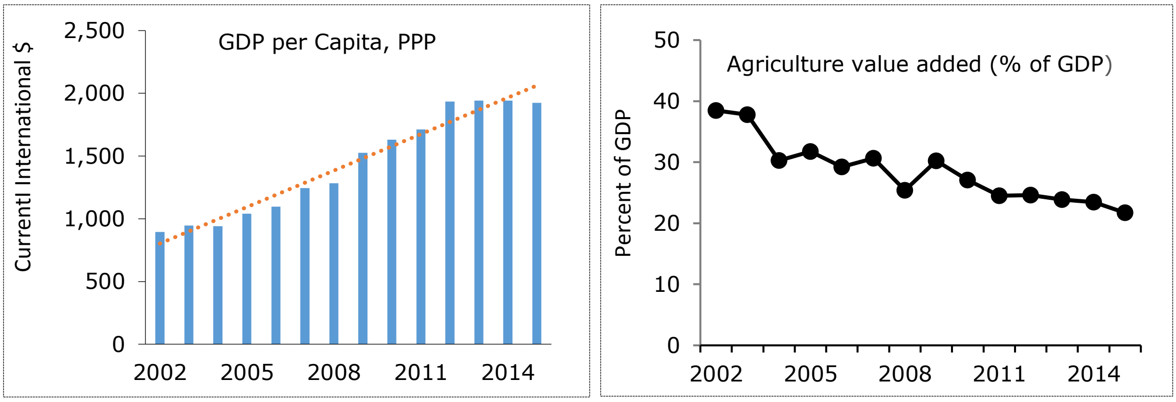

Figure 1: GDP per capita, PPP (current international $) and agriculture growth (Value added as % of National GDP)
Source: World Bank, International Comparison Program Database

Investment in the farming sector substantially increased aiming to improve crop yields and expand market opportunity for Afghan products. Among domestic sources, agriculture is central to the country's economy. The sector is the second largest contributor to GDP growth after the services sector. The share of agriculture in national GDP slightly dropped since 2002 (Figure 1), possibly due to the revitalization of other sectors such as the service and manufacturing industry.
Agriculture continues to be a strategic sector in the economic development of Afghanistan in terms of its potential for contributing to household income, food security, and employment. It accounted for 25% (excluding the opium poppy economy) of the national GDP and employed about 40% of the total workforce in 2014.

Despite its substantial share in the country's economy, the performance of the agricultural sector remains poor and the country is not self-sufficient with almost 40% of the population living below the national poverty line in 2013/14. The performance of agriculture may be attributed to the productivity gap due to factors such as lack of knowledge on the efficient utilization of available inputs and limited use of modern agricultural inputs.

In line with Afghanistan's national agriculture development framework, achieving self-sufficiency in basic food crops and promoting economic growth and development via improvement in agriculture production and productivity remains on the top of the governments strategic goals. In this context, the measurement of existing efficiency in agricultural production becomes important.

Agriculture is dominated by small-scale farm households with an average farm size of 7 Jeribs (about 1.5 hectares)[2]. Wheat occupies the major portion of the agriculture land and as the main staple food crop and a major source of calories plays a critical role in food security. Other important crops include maize, rice, barley, fodder crops, potato, and other high value crops such as vegetables and fruits. Although wheat is important for food security, market-based production requires farmers to diversify production at the farm level to enhance their cash income.

Afghanistan's foreign trade has been growing largely due to better linkages[3] and access to regional and global markets[4]. Commercialization and diversifying production at the farm level may be an important step in the long run to stabilize food supply and positively respond to the changing market demand. It will also improve household's cash income which will enable them to purchase sufficient production inputs and improve technical efficiencies. Therefore, this study aims to measure the overall efficiency levels and assess the determinants of farm level efficiency, particularly the impact of crop diversification on production efficiency.

The remainder of this study proceeds as follows. Section II provides a brief review of literature relating to the concept and measurement of technical efficiency and crop diversification. Section III introduces research strategy and methods to estimate the levels of farm-specific technical efficiency and its determinants, section IV describes the data and briefly explains the variables used in the analysis. Section V provides the empirical results and Section VI summarises the findings and discusses the policy implications. A detailed mathematical derivation of model equations will be presented in Annex I. Some descriptive analysis of the data will be presented in Annex II.

1.2. Motivation and Relevance of the Study

The problem of measuring technical efficiency and subsequently the economic performance of the farming sector is important to both the households and policy makers. Their primary concern is to understand how far the output for the agriculture sector can be expected to increase by simply increasing the levels of efficiency, without absorbing further resources. Empirical evidence of farmer specific efficiency analysis and identification of the potential factors affecting it can help address productivity gains simply by improving socio-economic characteristics and farm management practices. Improving

[2] 1 Jerib =0.2 hectare
[3] For instance, Chabahar strategic port agreement was signed in may 2016 between India, Iran, and Afghanistan which will enhance bilateral trade in the region
[4] Afghanistan became a member of WTO in July 2016

efficiencies without increasing the level of inputs usage can lead to saving unnecessary production costs.

Farm level technical efficiency requires rational input allocation and improved farm management techniques to achieve the optimum output levels. This is vital for producers who intend to optimize their production decisions particularly under changing market conditions, high input costs, economic hardship and rapid technological progress. It is also relevant for policy makers interested in enhancing the farming sector's economic performance and competitiveness, promoting economic development and sustainable economic practices.

The contribution of this study is twofold. Although the subject of technical efficiency is important, to the best of the author's knowledge there are no published studies that have investigated technical efficiency at the farm level using nationally representative data that consist of a large sample of households across all eight agro-ecological zones of Afghanistan. Therefore, the contribution of this study is unique and is aimed to greatly help households and policy makers in decision making related to production and productivity. Secondly, there is limited research available that explicitly evaluates the impact of crop diversification on technical efficiencies. Therefore, this study is intended not only to analyse the effect of crop diversification on the level of technical efficiency but also identify and evaluate the impact of other important factors, such as access to extension services, off-farm employment, agro-ecological zones, and other farm and household socio-economic characteristics.

1.3. Objectives and Research Questions

In being a useful tool to diagnose farm economic problems, assessment of technical efficiency has drawn broad research interest. The assessment of farm level technical efficiency and the factors that affect it provides valuable information to improve farm management and economic performance. Avoiding sources of inefficiency and waste of resources is necessary for economic viability and sustainability in the long run. In this regard, analysing and measuring technical efficiency has important implications for economic performance, commercialization, technological innovation and the overall input use in the farming sector.

This study focuses on the farmers' decision-making processes that are required to incentivize farmers to cultivate a diverse portfolio of crops and reduce dependence on staple crops. The primary intention of this study is to estimate the level of technical efficiency and identify the potential factors determining it by answering the following empirical questions:

a) What are the levels of aggregate technical efficiency among Afghan crop farmers?
b) What is the status and extent of crop diversification among smallholder farmers and how does it affect the level of technical efficiency?
c) What are the implications of other important external factors such as access to extension services, agro-ecological zones, off-farm employment, education, and other farm and household characteristics?
d) Are there scale inefficiencies in the farming sector in Afghanistan?

1.4. Scope and Limitations

The analysis in this study is based on the information generated from the household survey during a single year. Using cross-section data to analyse production decisions makes it difficult to draw concrete policy inferences on the level of technical efficiency that might be subject to change over time. However, a strength of the data is that it covers multiple seasons throughout the same year. A limitation of the data is that information is at the farm level and cannot be disaggregated by plot level or, in the case of inputs, by crop. Therefore, the analysis is limited to the estimation of an aggregate level production function.

The frontier techniques used in this study assume that all inputs required to produce output have been measured and included. However, as with other studies, it is possible to raise questions about whether all inputs have actually been accounted for, since farms that are apparently inefficient may just use less of certain unmeasured inputs. A more general problem is errors or inaccuracies in the measures of inputs, but we assume these are not systematic.

II. Literature Review

2.1. Concept and Measures of Technical Efficiency

Since the pioneering work by Farrell (1957), a number of approaches to efficiency measurement have emerged. The two main approaches that have been extensively used in the efficiency literature are: 1) parametric Stochastic Frontier Analysis (SFA) initially proposed by Aigner et al. (1977) and Meeusen and Van den Broeck (1977); and 2) non-parametric Data Envelopment Analysis (DEA) initially developed by Charnes et al. (1978).

Choosing between the SFA and DEA approaches to measure efficiency has been controversial and depends upon the objective of the research, the type of industry and the availably of data (Wadud and White, 2000). The nonparametric approach (DEA) does not rely on the definition of a functional form characterizing the underlying technology and therefore avoids misspecification problems. However, a drawback of this technique is that it is deterministic and ignores the stochastic error term which implies that deviations from the frontier are entirely attributed to inefficiency effects. As a result, technical efficiency ratings obtained from the nonparametric approach are generally lower than those obtained under the parametric SFA alternative (Coelli, 2005; Kumbhakar and Lovell, 2000; Wadud and White, 2000).

In contrast, the main advantage of the econometric or parametric SFA approach is that it incorporates a composed error structure with a two-sided symmetric term and a one-sided component which permits to distinguish between inefficiency and exogenous shocks. The one-sided component reflects inefficiency, while the two--sided error captures the random effects and exogenous shocks outside the control of the production unit, including measurement errors and other statistical noise typical of empirical relationships (Aigner et al., 1977; Meeusen and Van den Broeck, 1977). In addition, it allows hypothesis testing and construction of confidence intervals (Wadud and White, 2000). The disadvantages of this approach are the need to assume a functional form for the frontier technology and for the distribution of technical inefficiency term of the composite error term.

This study adopts the stochastic frontier function approach since agricultural crop production exhibits random shocks and there is a need to separate the influence of stochastic factors (random shocks and measurement errors) from the effects of other inefficiency factors by assuming that deviation from the production frontier may not be entirely under the control of farmers.

Production efficiency is widely used in agricultural economics to assess the performance of farmers. Efficiency can be divided into two concepts, the technical efficiency (also called output oriented efficiency), and allocative efficiency (also referred to as the input-oriented efficiency). Allocative efficiency can be viewed as the ability of a farm to use the inputs in optimal proportions given their respective prices and technology (i.e. obtaining optimal output or profits with the least cost of production). Technical efficiency, on the other hand, is the ability of a farming unit to produce a maximum level of output given the level of inputs (Farrell, 1957). In measuring output-oriented technical efficiency, the inputs are exogenously given and the objective is to maximize output as the only choice variable.

To illustrate, assume the case where production involves two outputs (*q1* and *q2*) and a single input (*x*) as depicted by Figure 2(a). Given the CRS property of the production function and assuming the input (*x*) quantity is fixed, the technology can be represented in two dimensions where the curve *ZZ′* is the unit production possibility curve. Point A, located below the possibility curve, corresponds to an inefficient producer because curve *ZZ′* represents the upper bound of the production possibilities. Alternatively, all points along the production possibility curve represent farmers that are 100 percent technically efficient.

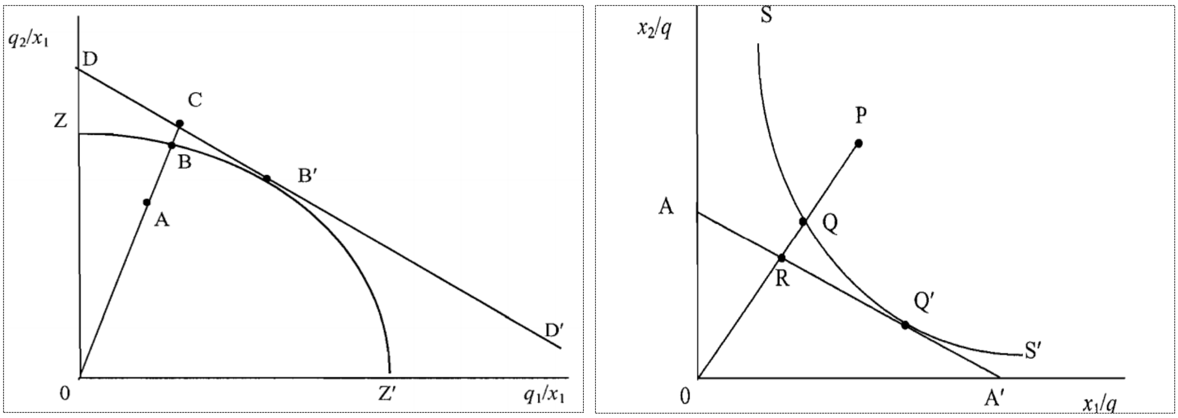

Figure 2: Technical and allocative efficiencies from an output (a) and input (b) orientation

Source: Adopted from Coelli et al., (2005)

In microeconomics of production, technical efficiency is defined as the maximum attainable level of output for given level of inputs, given the current range of alternative technology available to the farmer. In Figure 2(a), the distance *AB* represents technical inefficiency which is the amount by which output could be increased without requiring extra input Thus, considering a farm producing at point A, the Farrell (1957) output-oriented technical efficiency can then be calculated as *TE=0A/0B*. Allocative efficiency, on the other hand, can be calculated as *AE=0B/0C*. Both measures have an output/revenue-increasing interpretation (similar to cost-reducing interpretation of allocative inefficiency in the input-oriented case). The overall efficiency could be defined as the product of these two measures *(0A/0B) x (0B/0C)* or *TE X AE*.

Similarly figure 2(b) illustrates input-oriented efficiency. Assume a farm uses quantities of two inputs (X_1 and X_2) defined by point *P* (where point P represents an inefficient), to produce an output, the technical efficiency is represented by the distance *QP* which is the amount by which all inputs could be proportionally reduced without reduction in output to achieve the technically efficient level of production (point Q which is located on the isoquant curve represented by *SS'*). If the input prices and input price ratio are represented by the slope of the isocost line (*AA'),* the allocative and technical efficiency measures can be calculated as: *AE=0R/0Q* and *TE=0Q/0P*.

Given that this study is concerned with Afghanistan which is a developing country, the main concern maybe output shortfall rather than input over usage, therefore output oriented approach is preferred. Moreover, the lack of price data implies that this study will not address allocative inefficiency.

2.2. Crop Diversification

The notion of diversification might have different inferences within the farming sector. Diversification might imply a shift away from monoculture to producing multiple crops on a single farm throughout the season/year. It could also be viewed as having many enterprises at the farm, for instance a larger mix of crops or a combination of livestock and crop units. This study is concerned with the first concept where diversification is defined as adding multiple crops (especially high value crops such as vegetables, fruits, potato, etc.) to the planting practice at the farm.

Wheat is a major staple food crop in Afghanistan. Summary statistics of the ALCS data used in this study suggest that wheat production is about 54% of the total quantity of crops a farmer produces annually. However, the share of wheat in total revenue (when quantities of crops are weighted by their respective prices) is somewhat lower. This difference might indicate that adding high value horticulture crops to the production portfolio may be positively associated with the farmer's income. A study in Punjab of India confirms that incorporating horticultural crops in the production mix increases net expected returns while increasing the labour and working capital requirements. (Chhatre et al., 2016). Van den Berg et al. (2007) concluded that diversification into high value vegetable crops would enable Chinese farms to sustain a reasonable income level. Guvele (2001) found that crop diversification reduces variability in income in Sudan.

Crop diversification could be viewed as a hedge against risks due to shocks such as extreme weather conditions, crop diseases and pests, and unexpected fall of market prices. The inherent characteristics of crop diversification that are widely accepted in the literature is that it reduces potential risk against uncertainty by reducing high dependency on monoculture, reduces economic losses due to diseases, weed and infestation, and increases soil fertility through crop rotation (Krupinsky et al., 2002).

Crop diversification is an environmentally sound and viable climate smart agriculture practice that is widely perceived to significantly enhance farm productivity and increases resilience in rural farming systems. According to Lin, (2011) crop diversification improves soil fertility, controls for pests and diseases, and brings about yield stability, nutrition diversity, and health. It can also serve as a superior substitute for the use of chemicals to maintain soil fertility and control pests. Thus, crop diversification is considered as one of the most feasible, cost-effective, and ecologically sound practices that improves farm productivity and increases sustainability and resilience in farming systems.

Nevertheless, there is limited empirical evidence that explicitly studies the impact of crop diversification on technical efficiency, with mixed conclusions. For instance Nguyen (2014), Manjunatha et al. (2013), Ogundari (2013), Rahman (2009), and Coelli and Fleming (2004) concluded that crop diversification significantly improves technical efficiency of the farms in Vietnam, India, Bangladesh, Nigeria, and Papa New Guinea, respectively. On the other hand, Haji (2007) found no significance relationship between crop diversification and TE but has found that crop diversification significantly reduced allocative and economic efficiencies in Ethiopia. In addition, Llewelyn and Williams (1996) found that crop diversification significantly reduces technical efficiency in Indonesia. They argued that it is possible that the increased inefficiency with diversification may be transitory as farmers improve their ability to grow new crops as both the age and diversification variables are statistically significant.

Given the mixed empirical evidence presented, it is important to evaluate the impact of crop diversification on technical efficiency especially in the case of Afghanistan where investment in the farm sector has substantially increased to transform farming from a subsistence to a diversified and commercialized system.

III. Methodology and Theoretical Framework

Since efficiency varies across producers, it is natural to seek determinants of efficiency variation. Early studies adopted a two-stage methodological approach, in which efficiency scores are derived from the estimation of a stochastic frontier function in the first stage, and estimated efficiencies are regressed against a vector of explanatory variables (Z_i) using OLS or Tobit regression in the second stage. However, the two-step approach has been criticized on the grounds that the household's knowledge of its level of technical efficiency or exogenous determinants of inefficiency (Z_i) might affect its input choices (X_i), hence efficiency might be dependent on the explanatory variables (Wang and Schmidt, 2002). Furthermore, even if X_i and Z_i are uncorrelated, ignoring the dependence between them and of the inefficiency with Z_i will cause the first-step technical efficiency index to be underdispered, so the results of the second-stage regression are likely to be downward biased (Kumbhakar and Wang, 2015).

Kumbhakar and Lovell (2000) and Battese and Coelli (1995) have advocated a single-stage simultaneous estimation approach in which explanatory variables are incorporated directly into the inefficiency error component. In this approach, either the mean or the variance of the inefficiency error component is hypothesized to be a function of the explanatory variables.

Following Aigner et al. (1977) and Meeusen and Van den Broeck (1977), the formulation of stochastic frontier model in terms of general production function could be specified as:

$$Y_i = f(X_i, \beta) + v_i - u_i = f(X_i, \beta) + \varepsilon_i \qquad (1)$$

Where Y_i is a scalar output of the i^{th} farmer, X_i is the vector that collects direct inputs, and β is a vector of paramters to be estimated. ε_i is a composed error term where v_i is a two-sided "noise" component assumed to be independently and identically distributed (iid), symmetric, and distributed independently from U_i. It captures the effects of random shocks beyond the countrol of farmers (i.e. measurment errors as well as other noise). u_i is a non-negative ($u_i \geq 0$) technical inefficiency component of the error term that captures the factors that are under the control of the producer (i.e. determinants of ineffiency to be defined in the inefficiency model). u_i is assumed to be independently and identically distributed as normal-half-normal distribution (Aigner et al. 1977). There are other possible specifications of the distributional assumptions on u_i (i.e. truncated-normal distribution) suggested by Greene (1980) and Lee (1983) which are still being used in empirical work. Jondrow et al. (1982), Battese and Coelli (1992, 1995), suggest that the half-normal model is the most useful formulation. Other variants such as the truncated-normal model with heterogeneity in the mean allow for great flexibility in the modelling tools.

Since $u_i \geq 0$, $\varepsilon_i = v_i - u_i$ is not symmetric, and v_i, u_i are distributed independently of X_i, estimation of equation (1) by Ordinary Least Square (OLS) provides consistent estimates of the parameters except for the constant (β_0) since $E(\varepsilon_i) \neq 0$ (Kumbhakar and Wang, 2015). Further, OLS does not provide estimates for the farm-specific techical efficiency. In addtion to obtaining estimates of the production techonolgy parameters (β's) from (Xi, β), the farmer-specific inefficiency u_i is the ultimate objective of the efficiency estimation techniques. To estimate the farmer-specific efficiency, it is required that separate estimates of statistical noise v_i and technical inefficiency u_i are extracted from ε_i for each producer.

In equation (1) the inefficiency component (u_i) of the error term is the log difference between the maximum and the actual output (i.e. $u_i = ln\, Y_i^* - ln\, Y_i$), therefore $u_i\, x100\%$ is the percentage by which actual output can be increased using the same inputs if production is fully efficient (Kumbhakar and Wang, 2015). In other words, it is the percentage of output that is lost due to technical inefficiency. The estimated value of u_i is referred to as the output-oriented (technical) inefficiency, with a value close to 0 implying fully efficient. Rearranging (1), we can derive the following equation for technical efficiency:

$$TE_i = \exp(-u_i) = \frac{Y_i}{Y_i^*} = \frac{Y_i}{f(x_i\, ;\, \beta)\; exp\{v_i\}} \qquad (2)$$

Which defines the farm-specific technical efficiency as the ratio of observed output (Y_i) to the frontier output $f(x_i\, ;\, \beta)\; exp\{v_i\}$ which is a maximum output feasible (under the current techology used) in an enviornment characterized by the stochastic elements specified by (v_i). Because $u_i \geq 0$, the ratio is bounded between 0 and 1, therefore a farm achieves maximum efficiency if, and only if, $TE_i = 1$. Otherwise $TE_i \leq 1$ is a shortfall of observed output from the the maximum feasible output in an environment characterized by v_i that is stochastic and varies across farmers(Kumbhakar and Lovell, 2000).

Using the conditional mean function, Jondrow et al. (1982) showed estimation of observation-specific technical efficiency (u_i) conditional on the error term (ε_i) as:

$$TE_i = \mathrm{E}(-u_i\,|\varepsilon_i) = \sigma^* \left[\frac{f^*(\varepsilon_i\,\lambda/\sigma)}{1-F^*(\varepsilon_i\,\lambda/\sigma)} - \frac{\varepsilon_i \lambda}{\sigma}\right] \qquad (3)$$

Where: $\sigma^{*2} = \sigma_u^2\sigma_v^2/\sigma^2$, $\lambda = \sigma_u/\sigma_v$, $f^*(.)$ is the standard normal density function, and $F^*(.)$ is the distribution function, both functions being estimated at $\varepsilon\lambda/\sigma$. TE_i can be obtained by the method of Maximum Likelihood Estimation (MLE) which will simultaneously produce estimates of the variance parameters.

Using λ parameterization, the log likelihood function for the Aigner et al. (1977) model specified in equation (1) assuming a half-normal distribution on *ui* is given as:

$$\ln(L) = -\left(\frac{N}{2}\right)(ln2\pi + ln\sigma^2) + \sum_{i=1}^{N}\left[\ln\phi\left[-\epsilon_i\,{}^{\lambda}/_{\sigma}\right] - \frac{1}{2}\,({}^{\varepsilon_i}/_{\sigma})^2\right] \qquad (4)$$

Where $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\lambda = \sigma_u/\sigma_v$ are the variance parameters which measures the fitness and correctness of the model. The variance parameter σ^2 indicates whether conventional production function would be a satisfactory representation of the data used or not. The ratio of standard errors λ, is an indicator of relative variability of the sources of variation (i.e. inefficiency and statistical noise). A value of $\lambda>1$ implies that the discrepancy between the observed and maximum attainable level of output is dominated by variability emanating from technical inefficiency. A detailed mathematical derivation of (3) and (4) are presented in Annex I at the end of this chapter.

Battese and Corra (1977) used the gamma parameterization in formulating the likelihood function, instead of λ. They argued that λ could take any non-negative value, thus the gamma parameterization has an advantage in the numerical maximization process as it takes value between 0 and 1 and therefore it searches if the maximizing value are conveniently restricted to this (tight) parameter space. The log likelihood function for equation (1) using gamma parameterization by Battese and Corra, (1977)is given by:

$$\ln(L) = -\left(\frac{N}{2}\right)\left(ln\left(\frac{\pi}{2}\right)\right) + ln\sigma^2\,) + \sum_{i=1}^{N}\ln\left[1-\phi\left(\frac{\epsilon_i\sqrt{\gamma}}{\sigma^2}\sqrt{\frac{\gamma}{1-\gamma}}\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\epsilon_i^2 \qquad (5)$$

Where $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\gamma = \sigma_u^2/\sigma^2$ are the variance parameters. The gamma parameter could be used to test the presence of inefficiency in the model because it measures relative proportion of variability due to inefficiency (u_i) in total variability. In other words, it shows the percentage of the variation in output that is due to technical efficiency, ranging from 0 to 1, where a value close to 1 implies that a random component of the inefficiency significantly contributes to the production system (Battese and Corra, 1977; Coelli, 1995).

IV. Description of Data and Variables

This study uses data from the Afghanistan Living Condition Survey (ALCS) conducted by the Central Statistics Organization (CSO) in 2013-14. CSO is collecting these data about the country for more than 10 years (previously known as the National Risk and Vulnerability Assessment).

The data is disaggregated for residential populations (urban, rural and nomad). Geographically the survey covered all 34 provinces of the country. In total 35 strata were identified, 34 for the provinces of Afghanistan and one for the nomadic (Kuchi) population. The sampling frame used for the resident population in the ALCS 2013-14 was the pre-census household listing conducted by CSO in 2003-05, updated in 2009. Households were selected on the basis of a two-stage cluster design within each stratum. In the first stage Enumeration Areas (EAs) were selected as Primary Sampling Units (PSUs) with probability proportional to Enumeration Area (EA) size. Subsequently, in the second stage ten households were selected as the Ultimate Sampling Unit (USU). Figure (3) shows the geographical coverage of the survey across the country.

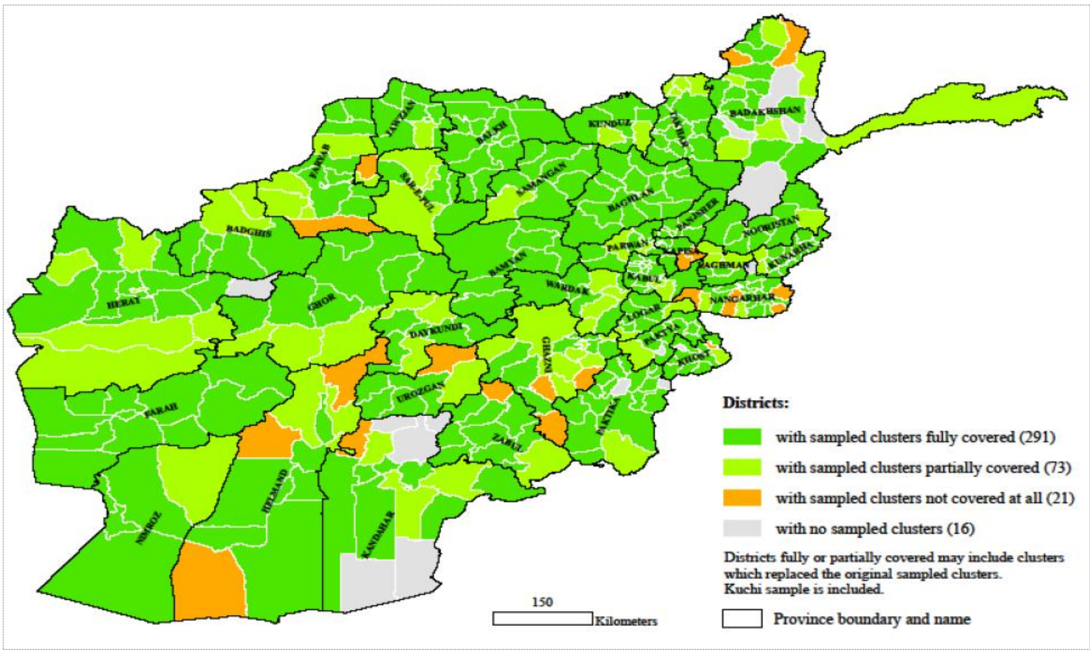

Figure 3: ALCS coverage by districts
Source: ALCS Survey Report

The reality of survey taking in Afghanistan imposed a number of deviations from the sampling design. In view of sustained levels of insecurity, clusters in inaccessible areas were replaced by clusters drawn from a reserve sampling frame that excluded insecure districts. In 182 out of 2,100 clusters (8.7 percent), originally sampled clusters could not be covered, in most cases due to security reasons. For a total of 182 clusters the coverage shifted in time or replacement clusters were selected. In addition, 19 clusters, representing 190 households, were not implemented and not replaced. Non-response within clusters was very limited. Only 845 (4.1 percent) of the households in the visited clusters were not available or refused or were unable to participate. In 841 of these non-response cases, households were replaced by reserve households listed in the cluster reserve list, leaving 4 (0.02 percent) households unaccounted for.

The data are representative at national and provincial level. It covered 20,786 households and 157,262 persons across the country. The data are unique in the sense that it also includes the nomadic (Kuchi) population of Afghanistan. Another distinguishing feature of the survey is the continuous data collection during a cycle of 12 months, which captures important seasonal variation in a range of indicators including agriculture. Using a structured questionnaire, data were collected on a number of indictors including agriculture production, labour market, household assets, education, and other household characteristics.

A limitation of the data, particularly for the purpose of this study is that the data could not be disaggregated by plot and crop. Therefore, the analysis is restricted to the estimation of an aggregate production function.

Initial descriptive analysis of the data showed that as many as 9,642 households reported some involvement in agriculture. However, after accounting for missing values on key variables, the total number of usable observations was 7,016 households.

Referring to equation (1), the dependent variable is aggregate physical output of crops weighted by prices of the respective crops. The dependent variable is the sum of revenue (measured in Afghan currency) of individual crops aggregated throughout the year for each farm household. Summary statistics of the dependent variable are presented in Table 1. The dependent variable was checked for potential outliers, and there seems to be no extreme values that influence the results. The price data used to weigh physical output comes from the NRVA 2011-12 survey. Lack of price data on some crops and unavailability of price data at the same year in which the ALCS survey was conducted is a limitation. However, for the purpose of this study, the price data were only used to weight physical quantities of crops and to calculate annual aggregate revenues in Afghan currency (Afghani). The input variables in the frontier and the variables in the inefficiency models are briefly described in the following sections.

4.1. Description of the Input Variables

Table (1) provides summary statistics for all the variables used in the analysis. The variable Land, measured in Jeribs is the total land cultivated by the household in various seasons throughout the year. This includes both irrigated and rain-fed land owned or leased by the household that was actually cultivated throughout the year. The size of the agriculture holding is small in Afghanistan, and therefore availability of agriculture land is an important factor for production.

Farm labour is another variable included in the analysis. Agriculture labour is coded based on the occupations and sub-categories in the survey including farm workers (those who are directly involved in production of crops or animal keeping), fishers, hunters, government extension workers, etc. Since this study deals with production, only the first type of labour is included, for three sources of labour supply involved in production: family labour, child labour and hired labour. Persons aged 14 and over are adult labourers and those below this threshold are child labour. However, the productivity of one unit of the child labour used in production may vary as compared to the productivity of adult labour, therefore households that reported child labour's involvement in production were not included in the analysis. Hired labour includes only those who were hired in by the farm. Labour is treated as a variable input which is measured in hours. Household labour hours and hired labour hours were added. It is important to note that majority of the households reporting hired labour did not report household labour and vice versa.

Table 1: Summary statistics for variables used in the analysis

Variable	Description	Mean	SD	Min	Max
Dependant Variable					
Y	Aggregate Annual Revenue (AFN)	58,252	90,431	110	1,280,000
Inputs					
X_1	Land (Jeribs)	7.035	9.12	0.10	90.00
X_2	Labour (hours)	63.79	62.37	1.00	417.85
X_3	Seed Expenditure (AFN)	2,354	3,617	0.00	45,000
X_4	Fertilizer Expenditure (AFN)	4,763	8,298	0.00	99,000
X_5	Chemicals Expenditure (AFN)	365.0	1,197	0.00	10,000
X_6	Tractor Rental (AFN)	2,504	4,296	0.00	60,000
X_7	Other Expenditure (AFN)	2,178	5,766	0.00	90,000
Sources or Factors of Efficiency/Inefficiency					
Z_1	Diversification Index (1-hhindex)	0.296	0.23	0.00	0.82
Z_2	Opium Share by Province (%)	0.033	0.10	0.00	0.48
Z_3	Extension Services (1=access, 0=otherwise)	0.209	0.41	0.00	1.00
Z_4	Land Quality (0=irrigated, 1=rain fed)	0.231	0.42	0.00	1.00
Z_5	Household Size (Persons)	8.335	3.46	1.00	36.0
Z_6	Household Head Age (Years)	44.38	13.8	14.00	95.0
Z_7	HH Head Sex (0=female, 1=male)	0.996	0.06	0.00	1.00
Z_8	HH Head Literacy (0=no, 1=literacy)	0.321	0.47	0.00	1.00
Z_9	HH Education (no formal schooling)	0.834	0.37	0.00	1.00
	HH Education (lower secondary)	0.052	0.22	0.00	1.00
	HH Education (upper secondary)	0.079	0.27	0.00	1.00
	HH Education (technical & teacher Collage)	0.021	0.14	0.00	1.00
	HH Education (university & postgrad)	0.013	0.12	0.00	1.00
					(Continued)
Z_{10}	Off-farm Employment (0=no, 1=yes)	0.125	0.33	0.00	1.00
Z_{11}	Own Cattle (heads)	1.601	2.04	0.00	31.0
Z_{12}	Own Tractor (number)	0.053	0.23	0.00	3.00
Z_{13}	Own Oxen (number)	0.232	0.62	0.00	9.00
Z_{14}	Farm Size 1 (0.1-2 Jeribs)	0.318	0.47	0.00	1.00
	Farm Size 2 (>2-5 Jeribs)	0.293	0.46	0.00	1.00
	Farm Size 3 (>5-10 Jeribs)	0.226	0.42	0.00	1.00
	Farm Size 4 (>10-20 Jeribs)	0.103	0.30	0.00	1.00
	Farm Size 5 (>20 Jeribs & above)	0.060	0.24	0.00	1.00
Z_{15}	Agro-ecological Zone 1 (NEM)	0.023	0.15	0.00	1.00
	Agro-ecological Zone 2 (CM)	0.137	0.34	0.00	1.00
	Agro-ecological Zone 3 (HFL)	0.043	0.20	0.00	1.00
	Agro-ecological Zone 4 (SMF)	0.202	0.40	0.00	1.00
	Agro-ecological Zone 5 (HVSB)	0.121	0.33	0.00	1.00
	Agro-ecological Zone 6 (TP)	0.064	0.24	0.00	1.00
	Agro-ecological Zone 7 (NMF)	0.169	0.37	0.00	1.00
	Agro-ecological Zone 8 (EMF)	0.241	0.43	0.00	1.00
	N				7,052

Source: Author's calculations of the ALCS 2013-14 data

Other inputs include expenditures on seed, chemical fertilizers, chemicals (i.e. pesticides and herbicides), tractor rental, and other expenditures measured (i.e. irrigation water) in Afghan currency (Afghani[5] symbolized as AFN throughout this study).

Table 1 provides summary statistics of the variables included in the model. It is important to note that some variables have shown wider variation across households leading to potential outliers. This study checked whether the inclusion or removal of these outliers has impacts on the results, it was found that results are slightly driven by outliners particularly in some input variables (including Land, labour, chemicals and other expenditures). Therefore 1% of the largest values of the labour and 0.5% in the other two variables (namely land and chemicals) were dropped. Percent of zero values in input variables are reported in Table A4 in Annex II.

4.2. The Determinants of Technical Efficiency

The objective of stochastic frontier models is not only to serve as a benchmark against which technical efficiency of producers is estimated, but also to explore how external variables such as farm and household characteristics exert influence on the farmer's performance (Kumbhakar and Lovell, 2000). A number of potential sources of efficiency or inefficiency were identified and are briefly described in this section.

4.2.1. **Crop Diversification**

Crop diversification is the main variable of interest in this study. The concept of crop diversification implies production of multiple crops on the farm throughout the year by an individual household. The Herfindahl index is used as a measure for crop diversification or specialization. The index captures the degree or extent of diversification for an individual farm household. In other words, the Herfindahl index is calculated for each farm separately to measure the degree of diversification using the following equation:

$$CD_i = 1 - HH_i \ = 1 - \sum_{j=1}^{J} \left(\frac{Y_j}{\sum_{j=1}^{J} Y_j} \right)^2 \quad 0 \leq HH_i \leq 1 \qquad (6)$$

Where Y_j is represents the revenue share occupied by the j^{th} crop (for $j = 1, 2, \ldots, J$) in total revenue earned by households annually. The computed HH_i index ranges from (close to) zero, reflecting complete diversification (i.e. maximum number of crops), to one, reflecting complete specialization (i.e. just one crop). In order to help ease the interpretation of the results, a direct measure for crop diversification was constructed by subtracting the Herfindahl index from 1 (to create a Diversification index $CD_i = 1\text{-}HH_i$) which ranges between 0 (specialization) and 1 (complete diversification). Any value above zero signifies diversification.

The average value of index for crop diversification (CD_i) for the sample farms is 0.30 (equivalent to HH_i=0.70) with a standard deviation of 0.233 (Table 1 above), implying presence of a relatively low level of crop diversification in the sample. The numbers equivalent or effective number which is the inverse of Herfindahl-Hirschman Index ($1/HH_i$)

[5] 100 Afghani are equivalent to 1.43 US dollars.

is useful in indicating the number of equal share crops consistent with the concentration. The effective number helps show the number or group of farmer with crop production that is equally-concentrated or diversified. The distribution of the index for crop diversification and effective or equivalent number are shown in figure 4.

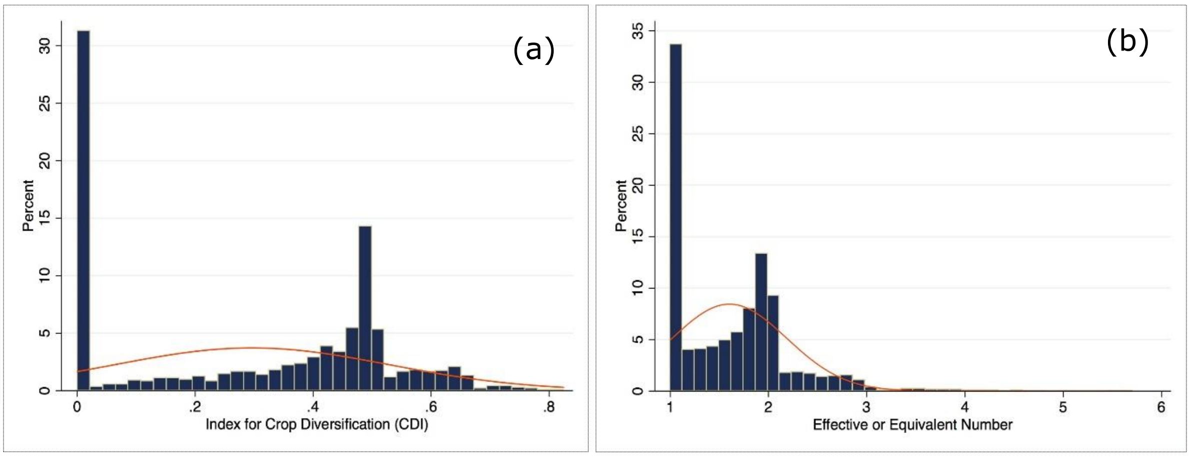

Figure 4: Distribution of Herfindahl Index (a) and effective number (b)
Source: Author's calculations of the ALCS 2013-14 data

The Herfindahl-Hirschman index has been widely used as a measure of crop diversification (Lakner et al. 2015; Ogundari, 2013; Manjunatha et al., 2013; Rahman, 2009; and Weiss et al. 2002). Nguyen, (2014) reported the average Herfindahl index of 0.75 for Vietnam which is slightly higher than the estimated average of 0.70 Afghanistan (corresponding to mean CD_i of 0.30) whereas Rahman (2009) reported the average Herfindahl index of 0.60 for Bangladesh, Ogundari (2013) Herfindahl index of 0.46 in Nigeria, and Manjunatha et al. (2013) reported 0.55 in India.

Summary statistics and characteristics of relatively more diversified farms and less diversified farms are reported in Tables A1 in Annex II. The farms in the sample ware divided in two sub-categories; those above the median level of Herfindahl index and below. The summary statistics show relatively higher total revenue for more diversified farms than those less diversified farms.

4.2.2. **Agro-Ecological Zones**

Afghanistan has a continental climate that is arid to semi-arid and is generally characterized by hot summers and cold winters. The wide range of altitude in Afghanistan leads to a great variation in climate within relatively small distances, which in turn affects the availability of water (rainfall), average annual temperature, and number of growing days. Temperature regimes are greatly modified by altitude – low sites are almost frost-free with very hot summers; the higher areas are arctic in winter (Thieme, 2006).

The climatic types as listed by Khaurin (1996) which is also quoted by (Thieme, 2006) are continental desert climate in the extreme north, Sub-tropical desert climate in the south, continental semi-arid Mediterranean climate in the north west, warm semi-arid Mediterranean climate in the lower central and north west, continental semiarid to moist

Mediterranean with no winter frost in the north east central, dry steppe climate in the lower Kabul valley, alpine in high mountains, centre and north east.

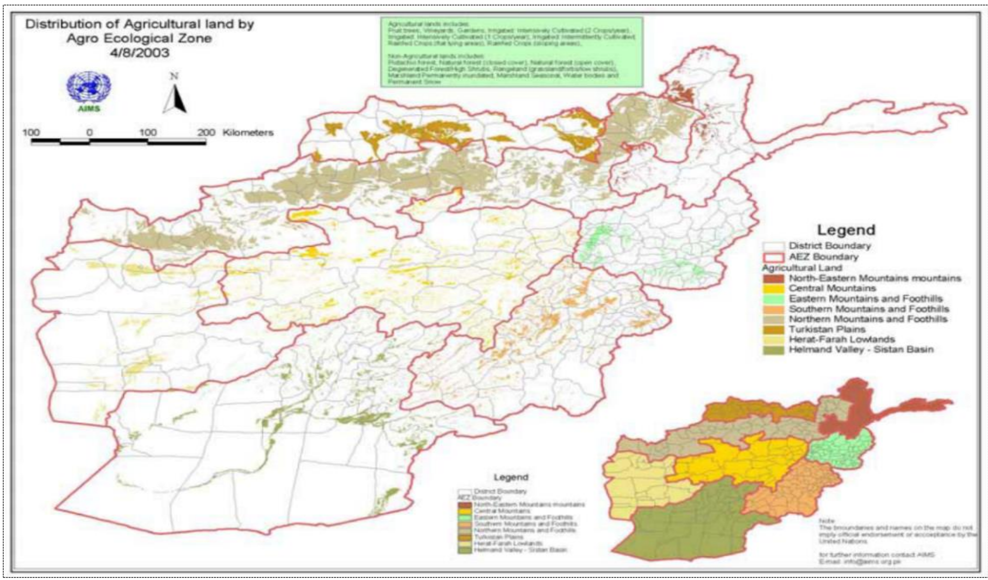

Figure 5: Agro-Ecological zones of Afghanistan
Source: Adopted from Maletta and Favre (2003)

Based on early work by Humlum (1959) later on revived by Dupree (1980), Afghanistan was divided into 11 geographical zones. However, recently a study by Maletta and Favre (2003) concluded that not all the 11 mentioned geographical zones have agricultural significance (i.e. some zones were classified as deserts). Based on ecological properties of land and climate, and some supplementary criteria about accessibility and prevailing agricultural activities, Maletta and Favre (2003) adopted the 8 agro-ecological zones scheme. These zones were constructed in the form of whole districts aggregations (Figure 5).

Annual participation, dry months and frost period across these 8 zones varies greatly. These variations, particularly the amount of annual rainfall may have potential effects on yield and the type of crops being grown. Table (2) summaries these climatic variations across the 8 zones.

Table 2:Agro-ecological zones of Afghanistan

Agro-ecological Zone	Annual Precipitation (mm)	Dry months	Frost Months
North-Eastern Mountains	200-800	2-6	1-9
Central Mountains	200-800	2-6	1-9
Heart-Farah Lowlands	<100-300	6-12	0-3
Eastern Mountains & Foothills	100-700	2-9	0-10
Turkistan Plains	<100-400	5-8	0-2
Helmand Valley-Sistan Basin	<100-300	6-12	0-3
Southern Mountains & Foothills	100-700	2-9	0-10
Northern Mountains & Foothills	200-800	2-9	0-8

Source: Adopted from Maletta and Favre, (2003) *and* Thieme (2006)

This study uses the eight agro-ecological zoning scheme to control for variation in crop production attributed to agro-climatic conditions. Afghanistan is generally categorized as a dry country where frequent droughts adversely affect farm production. Availability of irrigation water is important for crop production and varies greatly by agro-ecological zones (Table 2). These differences across agro-ecological zones are hypothesized to affect crop yields. In addition, the type and number of crops grown in each zone might have an impact on the extent of crop diversification.

4.2.3. **Access to Extension Services**

Access to extension services is vital in assisting farmers in the production decision making process since it can be a reliable source of information, technical advice, trainings and improved farm management practices. Access to extension services is broadly believed in the literature to have a positive impact on the farm output and on the level of crop diversification. Tables A2 in the Annex II provide summary statistics and characteristics of farms with respect to access to extension services. The summary statistics show that farm revenues for farmers who have availed themselves of extension services are slightly higher than those who did not have contact with the extension services. In addition, farmers with access to extension services adopted a relatively diversified farming system than those who had no access.

About 21% of the sample farmers have access to extension services. Although relatively few farmers can avail of them, extension visits and training provided are important sources of information, farm management techniques, use and dissipation of innovation and technology. The survey directly provides data on whether farmers have had access to extension services or not. A binary variable was constructed which is equal to 1 if farmers have access and zero otherwise.

4.2.4. **Farm Size**

Farm size in Jeribs is the measure of the land variable. Impact of farm size on technical efficiency is investigated in the literature with mixed conclusions. Most of the empirical evidence suggests inverse relationship between the farm size and technical efficiency (i.e. smaller farm size is associated positively with the level of technical efficiency). Therefore, in context of Afghanistan where agriculture holding is relatively small, it is important to account for potential variability due to the farm size.

4.2.5. **Off-farm employment**

There are a number of recent studies that have identified and included off-farm employment in the inefficiency effect model. The impact of off-farm employment on technical efficiency is ambiguous. On one hand, off-farm employment shrinks the availability of labour for on-farm activities, especially if hiring agricultural labour incurs transaction costs, and therefore may negatively affect technical efficiency. On the other hand, off-farm employment enables households to increase their incomes, to overcome credit and insurance constraints and to increase their use of industrial inputs. Studies such as Essilfie et al. (2011) in Ghana, Haji (2007) in Ethiopia, Yang et al. (2016) and Zhang et al. (2016) in China, found that off-farm employment positively contributed to technical efficiency. On the other hand, studies conducted in North America and Europe concluded that technical efficiency is negatively related to off-farm employment due to reduction in labour supply to farm activities (Goodwin and Mishra, 2004; O'Neill et al., 2001).

4.2.6. **Cattle Ownership at the Farm**

Cattle ownership is used as a proxy for availability of animal manure at the farm. Animal manure is an important source of organic fertilizer, especially in the context of Afghanistan, and is generally believed to improve soil fertility. It is treated as a continuous variable being measured as the number of cattle heads owned by the farm at the time of the survey.

4.2.7. **Oxen and Tractor Ownership at the Farm**

Oxen and tractors are the two main sources of traction power used on the farm for ploughing and other farming activities. A dummy variable on whether a household owns a tractor, oxen or both was included in the model. It is generally believed that households who own a tractor or oxen or both might be cost effective, and therefore might have influence on the technical efficiency. On the other hand, oxen or tractor ownership may substitute for farm labour especially since some of the activities that are traditionally carried out by labour maybe completed by oxen or tractor.

4.2.8. **Quality of Land**

As mentioned earlier, households own and cultivate either irrigated, rain-fed or a combination of both irrigated and rain-fed land to produce crops. Based on the descriptive statistics of the survey data, annual aggregate revenue for those household who cultivate irrigated land alone is much higher than those who operate a combination of both irrigated and rain-fed land. Therefore, it is *a priori* expected that households who own and operate rain-fed land may be less efficient compared to those who have access to irrigated land. To capture this variation attributed to the quality of land, a binary variable (equal to 0 for those who cultivated irrigated land alone, and 1 if the household cultivated rain-fed or a combination of both irrigated and rain-fed) was included in the analysis.

4.2.9. **Opium Cultivation**

Another important source of (in)efficiency, especially in the context of Afghanistan, might be opium production. Using the Afghanistan ministry of counter-narcotics annual data, an intensity variable is constructed to capture opium cultivation by province. The ALCS survey used in this study have also collected information on opium production from the households, however the reliability of the data might be a concern as production and trade of narcotics is illegal by the constitution, therefore households who actually produce opium might refrain from provision of data or provide misleading information. About 97.9% of household in the ALCS survey did not report growing opium.

In general, there are certain zones and provinces where production of opium is relatively more common than other areas. Largely, opium production may have connection with the security situation in the country (i.e. provinces that are opium free are relatively secure). Therefore, inclusion of this variable might also proxy for insecurity following that most of the opium infected areas are likely to be insecure. It may also capture unreported access to revenue as opium is a cash crop.

4.2.10. **Household Socio-economic Characteristics**

Household socio-economic characteristics such as household size, household head literacy and education (formal schooling), and household head sex, are generally included in the inefficiency effect model. Household's socio economic characteristics are widely believed in the literature to affect efficiency. For instance, household size may affect labour supply. Household head education is used a proxy for farming experience and necessary skills of management. In the context of this study, in addition to the formal education by the household head, literacy rate is important and therefore was also included.

Table 3: Household head literacy and education levels

Literacy/Education Level	Number of Heads	Percent
Literacy Rate		
Can't Read & Write	4,791	67.94
Can Read & Write	2,261	32.06
Formal Schooling		
No Formal Schooling	5,326	83.42
Lower Secondary	364	5.16
Upper Secondary	560	7.94
Teacher College	150	2.13
University & Postgraduate	95	1.35
N		7,052

The descriptive statistics as reported in Table (3) of the household head literacy rate and formal education attendance shows that literacy rate or level is important as 68% of the household's heads were reported to have no skill to read and write while 83.4% of them have not attended any type of formal schooling. This requires that these two aspects should be controlled for separately, especially since literacy rate is of more importance given the data.

V. Empirical Model Specification

Based on equation (1), the translog stochastic frontier model initially developed by Aigner et al. (1977) and Meeusen and Van den Broeck (1977), can be specified as below:

$$lnY_i = \sum_{k=1}^{7} \beta_k \, lnX_{ik} + \frac{1}{2}\sum_{k=1}^{7}\sum_{j=1}^{7} \beta_{jk} \, lnX_{ik} \, lnX_{ij} \; + \; v_i - u_i \qquad (7)$$

Where *ln* denotes the natural logarithm, Y_i represents aggregate revenue of the i^{th} producer, *k* represents the number of inputs used, X_{ij} represents a set of 7 input categories (mainly land, labour, seed, fertilizer, chemicals, tractor rental, and other expenditures) used by the i^{th} farmer, and β is a vector that collects unknown parameters to be estimated. In addition, ε_i is the composed error term where $\varepsilon_i = v_i - u_i$ with $u_i \geq 0$. The random error v_i accounts for the stochastic effects beyond the producers control, measurement errors as well as other statistical noise, and u_i captures production inefficiency due to factors that are in the control of the producer.

There are a number of distributional assumptions that could be made on the composed error term as explained section 3.1. In this study, two distributional assumptions on the inefficiency component of the error term are made and results are cross checked and tested.

a) The Half-Normal Distribution: the normal-half normal case imposes the following restrictions on the error term:
$v_i = iid\ N(0, {\sigma_v}^2)$
$u_i = iid\ N^+(0, {\sigma_u}^2)$
v_i and u_i are distributed independently of each other and of the regressors.
b) Truncated–Normal distribution: the truncated-normal case imposes the following restrictions on the error term:
$v_i = iid\ N(0, {\sigma_v}^2)$
$u_i = iid\ N^+(\mu, {\sigma_u}^2)$
v_i and u_i are distributed independently of each other and of the regressors, and μ is nonzero mean for u_i

All input and output variables were transformed to their corresponding log values. Given equation (5) and the distributional assumption on the inefficiency component (u_i) of the composed error term, the Battese and Coelli (1995) inefficiency model could be specified as:

$$u_i = \delta_0 + \sum_{i=1}^{14} \delta_i Z_i + w_i \qquad (8)$$

Where u_i is the inefficiency, Z_i is the vector of exogenous variables (namely sex, age, literacy, education of the household head, household size, index for crop diversification, access to extension services, cattle ownership, oxen ownership, tractor ownership, off-farm employment, land quality, opium share by province, farm size, and agro-ecological zones) that are likely to affect efficiency, δ's are the parameters to be estimated, and w_i is the error term of the efficiency model. As the dependent variable in equation (8) is defined in terms of technical inefficiency, a farm-specific variable associated with the negative (positive) coefficient will have a positive (negative) impact on technical efficiency.

5.1. Maximum Likelihood Estimator (MLE)

The estimation of the model involves (i) estimating the parameters of the frontier function, and (ii) estimating inefficiency. There are various methods of estimation depending on the distributional assumptions for the error components. Early methods include Corrected Ordinary Least Square (COLS) and Corrected Mean Absolute Deviation (CMAD) which estimates technical efficiency without imposing any assumptions on the inefficiency component of the error term. However, these methods assume that the frontier function is deterministic, and the randomness of the model comes entirely from the variation in inefficiency. Therefore, deviations from the estimated frontier are entirely attributed to inefficiency, and there is no role for other randomness such as data errors (Kumbhakar and Wang, 2015).

On the other hand, the choice of distributional assumptions on the components of the error term is central to the ML estimation approach of the stochastic frontier model. After these distributional assumptions are imposed, the log-likelihood function of the model is derived and numerical maximization procedures are used to obtain the ML estimates of the model parameters. Consequently, the maximum likelihood estimate of an unknown parameter is defined to be the value of the parameter that maximizes the probability (or likelihood) of randomly drawing a particular sample of observations. Aigner et al. (1977) focused on the implicit assumption that the likelihood of inefficient behaviour monotonically decreases for increasing levels of inefficiency. They parameterized the log-likelihood function for the half-normal model in terms of the variance parameters.

Maximizing a log likelihood function usually involves taking first derivatives with respect to the unknown parameters and setting them to zero. However, since these first order conditions are highly nonlinear and cannot be solved analytically for parameters the likelihood function is maximized using an iterative optimization procedure.

5.2. Robustness and Hypothesis Tests

Prior to undertaking the maximum likelihood estimation, it is important to check the validity of the stochastic frontier specification. Schmidt and Lin (1984) and Coelli (1995) proposed that in specifying the stochastic frontier model, a pre-test of the skewness of the OLS residual based on the third moment (M3T) should be carried out to test the null hypothesis of no skewness. The theory behind the test is that, for a production-type stochastic frontier model with the composed error $\varepsilon_i = v_i - u_i$ with $u_i \geq 0$ and v_i distributed symmetrically around zero, the residuals from the corresponding OLS estimation should skew to the left (i.e. negative skewness). Thus, a negative skew of the third moment is an indication of the existence of efficiency effects. The Coelli (1995) test is given by:

$$M3T = m_3 / \sqrt{\frac{6m_2^3}{N}}$$

Where m_2 and m_3 are the second and the third sample moments of the OLS residuals, respectively. If the value of M3T is statistically significant at the 1% level the frontier framework is supported. In our case, the computed value of the test statistic is -6.51. Because it has a normal distribution, the critical value is 1.96, so the result confirms the rejection of the null hypothesis of no skewness in the OLS residuals. This result is further confirmed by significance of variance parameters (γ and σ^2) in Table 5 where results of the stochastic frontier model are presented and the generalized log-likelihood ratio test for γ presented in Table 4.

One of the drawbacks of the parametric SFA approach is having to specify functional form representing the production technology and imposing assumptions on the error components of the model. In addition, the stochastic frontier model imposes certain assumptions on the inefficiency term of the composed error term. It is important to ensure that the model specification correctly represents the data. It is therefore of interest to test the following hypothesis before presenting the results.

- Hypothesis 1: H_0:β_{jk} =0 the null hypothesis that identifies an appropriate functional form between the restrictive Cobb-Douglas and the translog production function. It specifies that the coefficients on square and interaction terms of input variables in equation (7) are not statistically different from zero. The Cobb-Douglas production frontier is a special case of the translog frontier in which the coefficients of the second-order terms are zero, i.e., $\beta_{jk} = 0, j \leq k$ = 1, 2,..., 7.

- Hypothesis 2: H_0: γ=0 in equation (8) the null hypothesis that the inefficiencies are not stochastic and that the technical inefficiency effects are not present in the model at every level, so the joint effect of these variables on technical inefficiency is statistically insignificant. If this null hypothesis is not rejected, the Stochastic frontier model could be reduced to the OLS specification. In this case, if there is output difference among farmers given equal inputs, this difference is purely due to the difference in random shocks that are outside of the control of the farmer.

- Hypothesis 3: H_0: $\delta_0=\delta_1= \delta_2\ldots \delta_n=0$ in equation (8) the null hypothesis specifies that the influence of identified inefficiency factors (i.e. household socio-economic, farm-specific, and geographical factors) is zero.

- Hypothesis 4: H_0: $u_i = iid\, N^{+}(0, {\sigma_u}^2)$ in equation (7) the null specifying that half normal distribution better fits the model as opposed to the alternative case which assumes truncated normal distribution for the u_i.

- Hypothesis 5: H_0: $\sum_1^7 \beta_i = 1$ in equation (7) the null hypothesis specifying that there exists constant return to scale in the production function. A Wald test will be used to test whether the production function exhibits a constant, increasing, or decreasing returns to scale.

A Generalized log-likelihood ratio (LR) test can be used to test which specification better fits the data. The Generalized log-likelihood ratio test is given by:

$$LR = -2[\ln\{L(H_0)\} / \ln\{L(H_1)\}] = -2[\ln\{L(H_0)\} - \ln\{L(H_1)\}]$$

Where $L(H_0)$ and $L(H_1)$ are the values of the likelihood functions under the null (H_0) and alternative (H_1) hypothesis respectively. The computed test statistics should be compared with critical values of the mixed chi-square distribution proposed by Kodde and Palm (1986). The LR and Wald tests are applied using the *lrtest and test* command in STATA.

Table 4: Hypothesis testing

Null Hypotheses	Test Statistic	P-Value	Decision
Functional Form (Translog vs Cobb-Douglas)			
H_0: $\beta_1=\beta_2= \beta_3=\ldots\beta_n=0$	LR= 593.43	0.000	Reject H_0
Specification of Frontier Model			
H_0: $\gamma=\delta_0=\delta_1=\ldots\delta_n=0$	LR= 171.49	0.000	Reject H_0
H_0: $\delta_0=\delta_1= \delta_2=\ldots \delta_n=0$	LR= 1,151.06	0.000	Reject H_0
H_0: $u_i = iid\, N^{+}(0, {\sigma_u}^2)$	LR= 8.89	0.004	Reject H_0
Testing for Constant Return to Scale			
H_0: $\sum_1^7 \beta_i = 1$	Wald(χ^2)= 0.14	0.711	Fail to reject H_0

The results of the stochastic frontier model can be significantly affected by the choice of the functional form. The most widely used functional forms in estimating the production function are the Cobb-Douglas (restricted) and the translog (relatively more flexible).

These two specification of the stochastic frontier function were therefore selected and compared. The first hypothesis aims to test the choice of functional form using the generalized log-likelihood ratio test. The calculated LR statistic is 593.4 at 28 degrees of freedom which is greater than the χ^2 critical value of 47.67 at 1% significance level, therefore the test rejects the Cobb-Douglas functional form in favour of translog production functional form. The test indicates that square and interaction terms in the translog model specified in equation (7) are significantly different from zero, thus the translog model could not be reduced to the Cobb-Douglas specification.

The second null hypothesis can be tested using the generalized likelihood ratio test based on the value of log likelihood function under OLS and maximum likelihood estimation of stochastic frontier model. The computed LR test statistic is 171.5 at 1 degree of freedom which is greater than the χ^2 crucial value of 5.41 at 1% significance level. Therefore, the null hypothesis that technical inefficiency effects are not present in the data is rejected at 1% significance level. Thus, the traditional average (OLS) production function is not an appropriate representation of the sample data. These findings confirm the results of M3T test presented earlier.

The third hypothesis is that the explanatory variables in the inefficiency model are simultaneously equal to zero. The LR test is used to calculate the test statistic using the log likelihood value of stochastic frontier model without explanatory variables of inefficiency effect model (H_0) and the full frontier model with all explanatory variables of inefficiency effect model (H_1). The computed LR test statistic is 1,151.06 at 27 degree of freedom which is greater than the χ^2 crucial value of 46.35 at 1% significance level. Based on the calculated LR test statistic, the null hypothesis is rejected at 1% level of significance. Therefore, the explanatory variables associated with inefficiency effect model are jointly different from zero.

The fourth hypothesis was tested to validate the distributional assumption of the inefficiency term (u_i). Two models were constructed corresponding to the distributional assumptions of half normal and truncated normal for the one-sided error term as specified in section 3.5. The LR test is used to calculate the test statistic using the log likelihood value of stochastic frontier model assuming half normal destruction on the inefficiency term (H_0) and the frontier model assuming truncated normal distribution on inefficiency term (H_1). The calculated LR statistic is 8.9 at 1 degree of freedom which is greater than the χ^2 critical value of 5.41 at 1% significance. Therefore, H_0 is rejected implying that the truncated normal model is preferred to half-normal, however the results of both models are presented in Table 5.

The fifth hypothesis tests whether the production function exhibits constant return to scale. The computed Wald test statistic is 0.14 with a p-value of 0.711. Thus, the null hypothesis of constant return to scale cannot be rejected, implying that the specified production function exhibits constant return to scale.

VI. Empirical Results and Discussion

6.1. Results of the Stochastic Frontier Model:

The maximum likelihood estimates of parameters of the stochastic frontier production function (SFPF) and inefficiency model given by two-equation system (7) and (8) are simultaneously obtained using STATA and are reported in Tables 5 and 6. Both half normal (first column) and truncated normal (second column) specification of the inefficiency term (u_i) were assumed and estimated. All seven inputs have the expected positive impact on the farm revenues.

The estimated value of σ^2 is positive and 3.83 which is statistically significant at 1% level. These values indicate that there exists sufficient evidence to suggest that technical inefficiencies are present in the data and that the differences between the observed (actual) and frontier (potential) output are due to inefficiency and not chance alone. Theoretically, this implies that the estimated model and distributional assumptions for the error terms are appropriate.

Gamma *(γ)* is the variance ratio, explaining the total variation in output from the frontier level of output attributed to technical efficiency. The estimated value of γ (the ratio of the variance of output due to technical efficiency) is 0.902 for the preferred truncated normal model, indicating that about 90 percent of the difference between the observed and frontier output are primarily due to the inefficiency factors which are under the control of farms in districts (Table 5).

The square terms (particularly labour, fertilizer, chemical and tractor rental squared) and several of the interaction terms are significantly different from zero indicating the rejection of the Cobb-Douglas model as an adequate representation of the data. It therefore justifies the non–linear functional form and that there exists important interaction among the variables.

6.2. Output Elasticities and Return to Scale

Since inputs and output variables were transformed to their corresponding log values, and were normalized by their respective sample means, therefore the estimated parameters are directly interpreted as partial elasticities at the sample mean. All slope coefficients or output elasticities of inputs had the expected signs and were found to be highly significant except for the variable of other farm expenditures. Coefficient estimates are quite similar for half- and truncated-normal specification of the SFA.

The results in Table 5 for the preferred truncated normal model show that land is the most important variable; the estimated coefficient is large and statistically significant at 1% with a positive sign which confirms the *priori* expectation. Expenditures on fertilizer and seed exhibits the second and third largest partial elasticities so is an important determinant of revenue. Other expenditures variable turned out to be insignificant at 5% level. Since farming is mostly subsistence and the farm size is small, other extra expenditures are quite uncommon and may not be a viable option especially for farmers that generate low cash income. All other purchased inputs are significant with the expected positive signs.

Table 5: MLE estimates for the stochastic frontier model

	Truncated-Normal		Half-Normal	
	Coefficient	SE	Coefficient	SE
Dependent Variable (Total Aggregate Revenue in AFN)				
Constant	0.147***	0.041	0.179***	0.041
Ln Land (X_1)	0.433***	0.025	0.431***	0.025
Ln Labour (X_2)	0.051**	0.021	0.050**	0.021
Ln Seed Expenditures (X_3)	0.131***	0.015	0.131***	0.015
Ln Fertilizer Expenditures (X_4)	0.199***	0.015	0.200***	0.015
Ln Chemical Expenditures (X_5)	0.038**	0.015	0.039**	0.015
Ln Tractor Rental (X_6)	0.117***	0.017	0.116***	0.017
Ln other Expenditures (X_7)	0.021*	0.012	0.021*	0.012
0.5 x Ln Land $(X_1)^2$	0.014	0.016	0.009	0.016
0.5 x Ln Labour $(X_2)^2$	0.057***	0.014	0.057***	0.014
0.5 x Ln Seed Expenditures $(X_3)^2$	0.035***	0.004	0.034***	0.004
0.5 x Ln Fertilizer Expenditures $(X_4)^2$	0.037***	0.004	0.038***	0.004
0.5 x Ln Chemical Expenditures $(X_5)^2$	0.018***	0.006	0.018***	0.006
0.5 Ln Tractor Rental $(X_6)^2$	0.032***	0.005	0.032***	0.005
0.5 Ln other Expenditures $(X_7)^2$	0.006*	0.003	0.006*	0.003
Ln Land x Ln Labour	-0.021**	0.010	-0.021**	0.010
Ln Land x Ln Seed	-0.007**	0.003	-0.007**	0.003
Ln Land x Ln Fertilizer	0.006**	0.003	0.007**	0.003
Ln Land x Ln Chemicals	0.001	0.004	0.001	0.004
Ln Land x Ln Tractor Rental	-0.005	0.003	-0.005	0.003
Ln Land x Ln Other Expenses	0.009***	0.003	0.009***	0.003
Ln Labour x Ln Seed	0.008***	0.003	0.008***	0.003
Ln Labour x Ln Fertilizer	-0.010***	0.003	-0.010***	0.003
Ln Labour x Ln Chemicals	-0.001	0.004	-0.002	0.004
				(Continued)
Ln Labour x Ln Tractor Rental	0.003	0.003	0.003	0.003
Ln Labour x Ln Other Expenses	-0.004*	0.003	-0.005*	0.003
Ln Seed x Ln Fertilizer	0.001	0.001	0.001*	0.001
Ln Seed x Ln Chemicals	-0.003**	0.001	-0.002**	0.001
Ln Seed x Ln Tractor Rental	-0.001**	0.001	-0.001**	0.001
Ln Seed x Ln Other Expenses	-0.002***	0.001	-0.002***	0.001
Ln Fertilizer x Ln Chemicals	-0.000	0.001	-0.000	0.001
Ln Fertilizer x Ln Tractor Rental	-0.003***	0.001	-0.003***	0.001
Ln Fertilizer x Ln Other Expenses	0.002***	0.001	0.002***	0.001
Ln Chemicals x Ln Tractor Rental	-0.002*	0.001	-0.002*	0.001
Ln Chemicals x Ln Other Expenses	-0.001	0.001	-0.001	0.001
Ln Tractor Rental x Ln Other Expenses	-0.002***	0.001	-0.002***	0.001
$(\sigma)^2$	0.371***	0.011	0.371***	0.012
γ	0.902***	0.038	0.889***	0.045
Log-Likelihood		-7,500.13		-7,504.22
Chi2		4,675.44		4,784.92
Prob Chi2		0.000		0.000
N		7,052		7,052

Note: Significances is indicated by * p<0.10, ** p<0.05, *** p<0.010

Returns to scale can be used to measure total resource productivity. The concept of returns to scale shows how output responds to increase in all inputs together. The sum of the partial elasticities with respect to every input estimated by the maximum likelihood estimator of the translog stochastic production function is 0.99. This is roughly consistent with constant returns to scale which implies that an increase in all available inputs leads to an equal proportional increase in farm revenues.

Marginal effects of the explanatory variables at the mean could be obtained by:

$$Marginal\ Effect\ of\ X_i = \frac{dy}{dx_i}\frac{\overline{x_i}}{\overline{y}} = b\frac{\overline{x_i}}{\overline{y}}$$

where, b =parameter estimate (partial elasticity associated with each independent variable), $\bar{x}$ = Mean of independent variable, $\bar{y}$ = Mean of dependent variable. The computed marginal effects of the input variables are reported in Table 6. Land and labour have the largest impact as they are regarded as the most important factors for crop production.

Table 6: Marginal Effects

Variable	Elasticity	Marginal Effect
Land (X1)	0.43	3,584.01
Labour (X2)	0.05	46.73
Seed Expenditures (X3)	0.13	3.25
Fertilizer Expenditures (X4)	0.20	2.44
Chemical Expenditures (X5)	0.04	6.04
Tractor Rental (X6)	0.12	2.73
Other Expenditures (X7)	0.02	0.55

Labour turns out to have marginal effect of 46% (i.e. a unit change in labour will change the revenues by 46%). In comparison to the computed average hourly wage for agricultural labour which is about 38.5 per hour (or 256 Afghan daily), marginal effect of labour is fairly higher.

6.3. The Inefficiency Effect Model

Maximum likelihood estimator is used to estimate the δ coefficients of equation (8) for technical inefficiency (Table 6). A negative sign of the estimated parameters indicates a reduction in technical inefficiency or an increase in technical efficiency. For the case of the inefficiency effect model, all variables are significant except sex, age, and education of household head, household size and off farm employment.

The estimated coefficient for the index of crop diversification is negative and statistically significant at the 1 percent level. This indicates that greater crop diversification (lower *HHi*) is associated with higher level of technical efficiency. The finding that less diversified farms are more inefficient is consistent with Nguyen (2014), Manjunatha et al. (2013), Ogundari (2013), Rahman (2009), and Coelli and Fleming (2004) for Vietnam, India, Bangladesh, Nigeria, and Papa New Guinea, respectively. Figure 6 illustrates this effect; the higher the degree or extent of diversification, the higher the level of technical efficiency. Although diversifying crops may require additional management skills, it has advantages of greater utilization of inputs, producing marketable crops and reducing reliance on production of a single staple crop mainly for home consumption.

Table 7: Maximum likelihood estimation of the inefficiency model

	Truncated-Normal		Half-Normal	
Variable	Coefficient	SE	Coefficient	SE
Constant	1.226***	0.430	1.089**	0.459
Head Sex (male)	-0.451	0.377	-0.485	0.408
Head Age (years)	0.001	0.002	0.002	0.003
Head Education (lower secondary)	0.095	0.163	0.104	0.181
Head Education (upper secondary)	0.172	0.143	0.198	0.158
Head Education (teacher collage)	0.102	0.243	0.113	0.272
Head Education (university & postgrad)	-0.282	0.320	-0.312	0.358
Head Literacy (can read & write)	-0.026	0.093	-0.028	0.103
Household Size (persons)	0.001	0.011	0.001	0.012
Diversification Index	-3.754***	0.347	-4.334***	0.354
Extension Services (1=yes)	-0.337***	0.098	-0.388***	0.109
Oxen and Yaks (number)	-0.169***	0.062	-0.193***	0.069
Tractor/Threshers (number)	-0.822***	0.255	-0.930***	0.291
Cattles (number)	-0.107***	0.022	-0.120***	0.024
Off-farm Employment (1=yes)	-0.027	0.100	-0.041	0.111
Opium share by province (%)	-0.914	0.724	-1.024	0.849
Farm Size (>2 to 5 Jeribs)	-0.380***	0.094	-0.409***	0.102
Farm Size (>5 to 10 Jeribs)	-0.340***	0.117	-0.375***	0.126
Farm Size (>10 to 20 Jeribs)	-0.054	0.157	-0.072	0.173
Farm Size (>20 & above Jeribs)	0.353*	0.203	0.371*	0.225
Land Quality (Low)	0.236***	0.090	0.229**	0.100
Agro-ecological Zone 1 (CM)	-0.066	0.191	-0.040	0.207
				(Continued)
Agro-ecological Zone 2 (HFL)	-0.054	0.225	-0.026	0.246
Agro-ecological Zone 3 (SMF)	-0.903***	0.209	-0.977***	0.227
Agro-ecological Zone 4 (HVSB)	-0.823***	0.234	-0.878***	0.256
Agro-ecological Zone 5 (TP)	-0.010	0.207	0.032	0.225
Agro-ecological Zone 6 (NMF)	-0.231	0.183	-0.216	0.199
Agro-ecological Zone 7 (EMF)	-0.477**	0.197	-0.484**	0.215
N		7,052		7,052

Note: Table reports estimates of equation (7). The omitted categories are: none for education level, <2 Jeribs for farm size and agro-ecological zone 8, none for extension services, , none for literacy, none for off-farm employment, and poor quality of land; significance levels indicated by * p<0.10, ** p<0.05, *** p<0.010

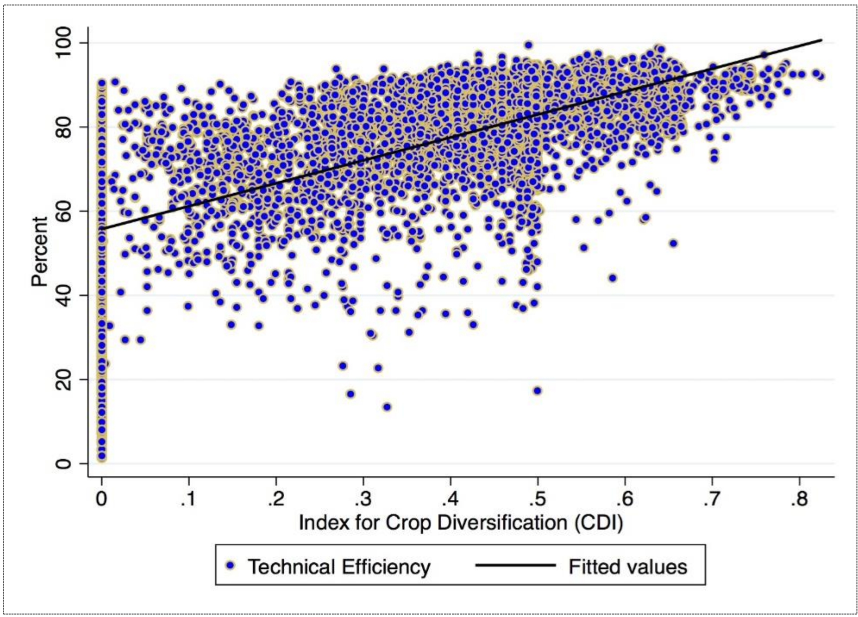

Figure 6: Distribution of TE by the index of crop diversification
Source: Author's calculations of the ALCS 2013-14 data

In addition, Table 7 summarizes the estimated ranges of technical efficiency and number of farms that fall in each specific range of crop diversification. The majority of the farms (about 50%) are experiencing the degree of diversification in the middle range (between 0.3 and 0.6).

Table 8: Number of farms their TE by crop diversification

Index of Diversification	Mean TE (%)	No. of Farms	Percent of Farms
0-<0.1	54.98	2,346	33.27%
0.1-<0.2	67.42	316	4.48%
0.2-<0.3	71.94	452	6.41%
0.3-<0.4	76.97	637	9.03%
0.4-<0.5	82.08	2,236	31.71%
0.5-<0.6	84.60	562	7.97%
0.6-<0.7	86.84	415	5.88%
0.7-1	89.76	88	1.25%
N			7,052

Source: Author's calculations of the ALCS 2013-14 data

The negative and significant effect of access to extension services on technical inefficiency implies that famers who have had contact with extension services have higher technical efficiency, perhaps because they are helped to diversify. The descriptive analysis of diversification and extension services reveal that farmers who have access to extension services have implemented relatively more crop diversification than those who did not have access to extension services (Table A2 in appendix II). In a recent study, Makate et al. (2016) found that farmers with access to extension services had 38.4 % more chance of adopting a diversified cropping system than their counterparts (those without access to extension). Extension workers have technical knowledge on crop production and improved production management practices that can assist farmers to implement their crop diversification decisions. Elias et al. (2013) concluded that extension services increases farm productivity by 20% in Ethiopia. Mango et al. (2015) and Bozoğlu and Ceyhan (2007) found a positive impact of extension services on technical efficiency in Zimbabwe and Turkey respectively.

There seems to be an inverted U-shaped relationship between farm size and technical efficiency. Efficiency level rises initially with farm size (inefficiency is lower in farms with 2-10 Jeribs compared to <2 Jeribs) but appears to fall when farm size exceeds 20 Jeribs. Figure 7 shows the distribution of technical efficiency and the index of crop diversification by the farm size. It is evident that both crop diversification and technical efficiency initially follow the same pattern; as the farm size initially increases, the levels of crop diversification and technical efficiency also increase, but eventually when farm size is 20 or above efficiency fall and crop diversification levels out as the as farm size increases beyond 20 Jeribs.

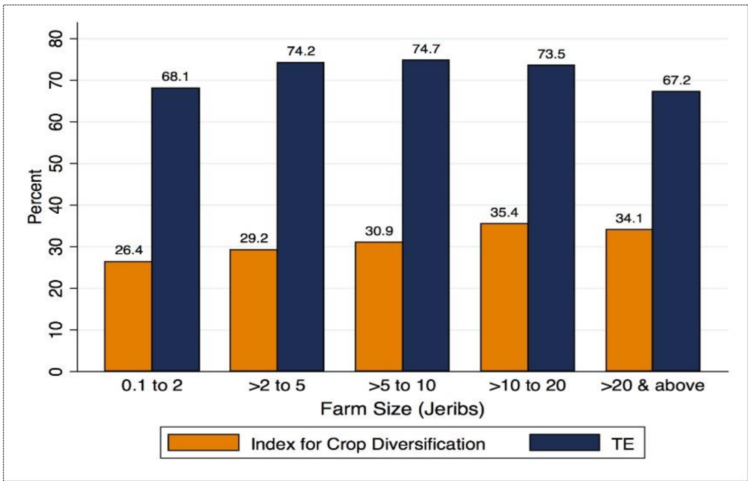

Figure 7: TE and index of crop diversification by farm size
Source: Author’s calculations of the ALCS 2013-14 data

The computed average efficiency scores imply that medium sized farms are relatively more efficient. This may be due to the fact that medium level farms are more diversified. Findings on the relationship between farm size and efficiency vary in the literature. Oladeebo and Oyetunde (2013) and Bhatt and Bhat (2014) find an inverse relationship between farm size and technical efficiency. Manjunatha et al. (2013) and Mburu et al. (2014) concludes that increased farm size improves technical efficiencies Helfand and Levine (2004) concluded that the relationship between farm size and efficiency is non-linear, with efficiency first falling and then rising with size and fall again when farm size is too large. Adhikari and Bjorndal (2012) concluded that medium size farmers achieve a higher technical efficiency than large and small farm sizes, suggesting that productive efficiency can be increased with the encouragement of creating medium size holdings. Narala and Zala (2010) found that medium size farms are the most efficient in rice farming in Gujrat India, presumably due to medium farmers having agriculture as their main occupation and allocating their resources more effectively.

Another possible explanation of the observed inverted U-shaped relationship between farm size and technical efficacy may be due to the fact that the small farms may be incurring higher fixed costs. On the other hand, large firms are more likely to operate in diseconomies of scale and are more likely to suffer from resource misallocation and monitoring production activities.

While the mean technical efficiency across the entire country is estimated to be 72.64%, it varies across agro-ecological zones (Figure 8). Southern Mountains and Foothills (SMF) records the highest average level of 81%, followed by Helmand Valley and Sistan Basins (HVSB) of 79%, Eastern Mountain and Foothills (EMF) of 77%, Central Mountains (CM) of 66%, Heart-Farah Lowlands (HFL) of 65%, Northern Mountains and Foothills (NMF) 64%, Turkistan Plains (TP) 61%, and North Eastern Mountains (NEM) experienced the lowest level of 58%. However, there is only a statistically significant difference for 3 zones including SMF, HVSB and EMF having higher efficiency than the NEM zone (Table 5). The distribution of technical efficiency and degree of crop diversification across all 8 agro-ecological zone is shown in Figure 8 below which confirms that the most efficient agro-ecological zones (particularly SMF, HBVS, and EMF) are relatively more diversified on average as compared to those relatively less inefficient zones.

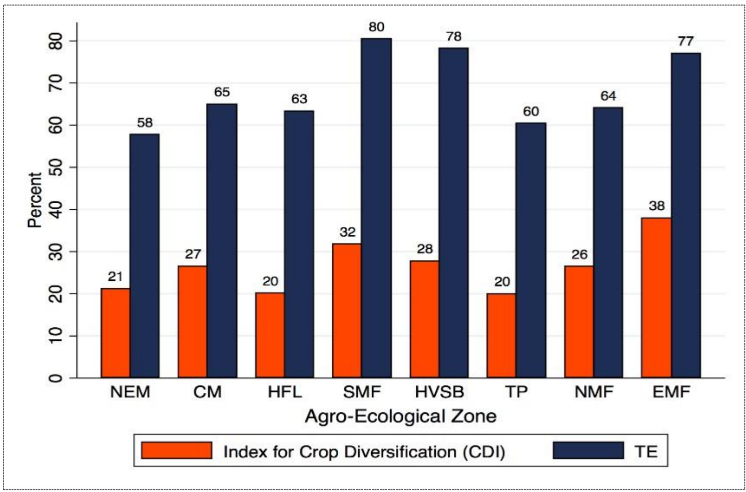

Figure 8: TE and crop diversification by agro-ecological zones
Source: Author's calculations of the ALCS 2013-14 data

Ownership of cattle, oxen and tractors by the households are positively correlated to the level of technical efficiency. Cattle and oxen ownership might imply availability of animal manure which is an important and cheap source of organic fertilizer (particularly in small-scale farming system) in soil that is widely believed to have a positive impact on soil fertility. To a certain degree, animal manure is considered as a good substitute for chemical fertilizers. Oxen and tractors ownership are also considered as important sources of the cheaper traction power available to farmers than those who hire tractor power and therefore farmers with greater number of oxen and tractors/thrashers maybe more efficient. In addition, tractor use may also indicate farm mechanization that ensures timely land preparation, planting and weeding.

Farmers who are operating a combination of rain-fed land and irrigated land were found to be more inefficient as compared to those who cultivated irrigated land alone. This suggests that rain-fed land is associated with lower crop yields. In addition, farmers who operate rain-fed land are less likely to diversify their production as was found in the descriptive statistics. This is because most crops, especially high value vegetables crops, require more water and therefore are not commonly produced on rain-fed land.

Opium intensity by province was found to be positively but insignificantly linked with technical efficiency. Insignificance may be due to a trade-off between effects of access to cash and insecurity. Production in provinces where farmers grow opium may be relatively more efficient compared to other regions, because farmers can purchase inputs (and sales of opium may inflate reported revenue), but opium affected provinces are likely to be more insecure. The descriptive statistics given in Table A3 in Annex II confirms the higher farm revenues in provinces where more than 1% of opium production is reported as compared to those which are opium free or producing less than 1% of opium.

The insignificant efficiency factors include household head age, sex, literacy, and education levels, the size of household, and off-farm employment, indicating that these factors may not have significant impact on the farm technical efficiency.

6.4. Estimation of Technical Efficiency:

Based on equations (2) and (3), farm-specific indices of technical efficiency were estimated assuming both half normal and truncated normal specification on the inefficiency component of the composed error term. It is evident from the results that the estimated technical efficiency estimates from the preferred truncated normal distribution range from 1.5% to 99.29%, with a sample mean of 71.9%. This reveals that there is substantial technical inefficiency in the Afghan farming sector. The main implication of this result is that farmers could increase their output by 29.1% on average without using additional resources, simply by improving technical efficiency. These estimates of technical efficiency are comparable with findings of other recent studies, for instance, Mwajombe and Mlozi (2015), Elias et al. (2013), Alam et al. (2012), Amaza et al. (2006), and Kudaligama et al. (2000) have estimated average efficiency levels of 72% in Tanzania, 78% in Bangladesh, 72% in Ethiopia, 65% in Nigeria, and 72% in India respectively. The frequency distributions of the technical efficiency estimates are presented in Table 8. Moreover, distribution of the estimated technical efficiency and crop diversification with respect to the effective or equivalent number (i.e. the households with equal share of crops) is reported in Table A5 in Annex II.

Table 9: Range and frequency of technical efficiency

Efficiency Range (%)	Truncated-Normal		Half-Normal	
	Number of farms	Percentage	Number of farms	Percentage
<25	153	2.17	163	2.31
25-<50	761	10.79	846	12.00
50-<60	646	9.16	695	9.86
60-<70	925	13.12	1,010	14.32
70-<80	1,560	22.12	1,786	25.33
80-<90	2,416	34.26	2,365	33.54
90-100	591	8.38	187	2.65
Mean		71.88		69.87
SD		17.47		17.27
Minimum		1.32		1.39
Median		77.20		75.27
Maximum		99.29		97.53
N		7,052		7052

Source: Author's calculations of the ALCS 2013-14 data

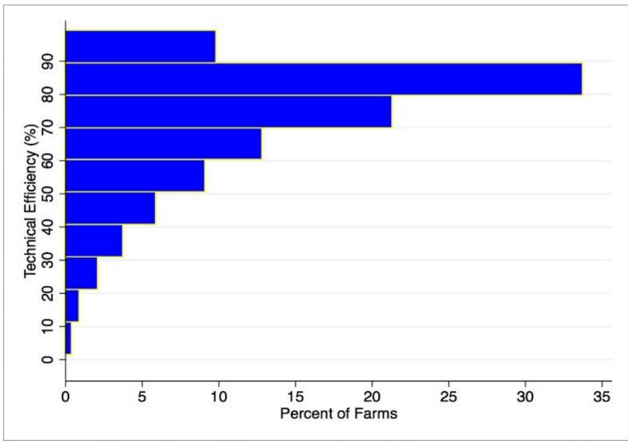

Figure 9: Distribution of Technical Efficiency

Source: Author's calculations of the ALCS 2013-14 data

Distribution of the estimated efficiency indices estimated by the preferred truncated-normal model reveals that about 13% of the sample farmers realized less than 50% of the potential revenues, whereas about 43% of farms have achieved more than 80% technical efficiency. The remaining farmers were operating between the levels of 50% and 80% (Table 8 and Figure 9). The distribution of efficiency indices derived from the half-normal is quite similar to those of the truncated-normal case (Table 6).

6.5. Endogeneity in Crop Diversification:

In the previous sections, crop diversification was assumed to be exogenous, however the decision to diversify may be an endogenous variable because the decision to adopt a diversified production system is likely to depend on unobservable variables. This means that failing to account for potential endogeneity may lead to endogeneity bias and consequently result in estimating an inconsistent effects of crop diversification on technical efficiency in the model presented in the previous section. The use of other input variable might also be endogenous, but since the main focus of this study is on analysing the impact of crop diversification on technical efficiency, thus this study will focus on the endogeneity issue in the crop diversification variable.

Due to its voluntary nature, the farmers self-select or choose whether to produce a single crop or a number of different crops. For instance, farmers who are relatively wealthier and have more technical knowledge on crop diversification as a viable strategy might be more likely to adopt crop diversification than their counterparts (such as those without access to extension), thus this unobserved selection bias may overstate the impact of crop diversification on technical efficiency. On the other hand, to the extent that CD is measured with error there may be attenuation bias so the basic SFA model may underestimate the impact of crop diversification on technical efficiency. In either case, there are unobserved factors in the error term (u_i) that are correlated with the endogenous variable (CD_i) that may result in biased estimates of the impact of crop diversification on technical efficiency in the basic SFA model.

To account for this potential selection bias due to endogeneity, the instrumental variables method is used. The IV for crop diversification used in this study is the mean value of the

Crop Diversification Index (CD_i) for other farm households in the district which is calculated as follow:

$$IV = \left(\frac{Sum\, X_d - X_i}{N_d - 1}\right)$$

Where sum X_d is the sum of the Diversification Index ($CDi=1\text{-}HH_i$)) in the district, X_i is the CD value of the i^{th} farm in the respective district, N_d is the number of observations in the respective district (so n-1 for the district is the number of observations in the district excluding the i^{th} farm itself). This means that the IV will differ slightly for each farm depending on the variance relative to that farm. On average, there are 40 farm households in each district. While constructing the IV, 14 observations were dropped because there were too few observations in a few districts which made it impossible to calculate the average value of CD, as a result the sample was reduced from 7,052 to 7,038.

The extent or degree of crop diversification may be magnified through social interactions between farmers in the local neighbourhood. Farms that face similar demographic characteristics and preferences are likely to adopt similar production systems. For instance, a farm household located in a district where farmers have greater access to information and markets, and are therefore more likely to diversify, is more likely to adopt a diversified production system than a farm in a less diversified district. Observing that neighbours diversify would encourage a farmer to follow the example, so even relatively 'low ability' farmers are more likely to diversify. However, the fact that neighbours diversify should not in itself affect the efficiency of the farmer as factors associated with efficiency, such as farm size or ownership of livestock, are not affected by neighbours' diversification.

Although there is a growing concern about the endogeneity issues in the stochastic frontier models, there is still limited work available in the literature to address it. Addressing the endogeneity issue is relatively more complicated in the stochastic frontier models due to the special nature of the error term. Standard Instrumental variable (IV) approaches cannot be used and the literature has yet to develop a strategy for addressing endogeneity with respect to the scaling factors in the one-sided error of a stochastic frontier model (Gronberg et al., 2015).

The instrumental variable estimator used in this study follows recent work of Karakaplan and Kutlu (2015) who developed a general maximum likelihood based framework to handle the endogeneity problem in the stochastic frontier models. The analysis are conducted using the *sfkk* command in STATA (Karakaplan, 2017). For further discussion and mathematical derivation see Karakaplan and Kutlu (2015) and (Karakaplan, 2017).

Endogeneity is investigated by applying the Durbin and Wu-Hausman test. The calculated test statistic is 39.82 and rejects the null hypothesis of no endogeneity in crop diversification at 1% level. In addition, Karakaplan and Kutlu (2015) provides an endogeneity test similar to the standard Durbin-Wu-Hausman test for endogeneity as part of sfkk estimation. This test was also carried out and the computed test statistic is 24.5 with a p-value of (0.00) which rejects the null hypothesis of no endogeneity in CD with 1% significance.

The stochastic frontier and inefficiency effect models were simultaneously estimated. For comparison, Table 9 provides estimates of both exogenous (column 1) and endogenous model (column 2) assuming a half-normal distribution for the inefficiency component (u_i) of the error term in equation (6). The *sfkk* command is still under development and only allows half-normal distribution for the u_i term. The parameter estimates of the two are quite similar.

Table 10: Estimation of Endogenous Stochastic Frontier model

	Endogenous		**Exogenous**	
variable	**b**	**se**	**b**	**se**
Dependent Variable (Total Aggregate Revenue in AFN)				
Constant	0.127***	0.045	0.159***	0.041
Ln Land (X_1)	0.413***	0.025	0.442***	0.025
Ln Labour (X_2)	0.052**	0.021	0.053**	0.021
Ln Seed Expenditures (X_3)	0.126***	0.015	0.131***	0.015
Ln Fertilizer Expenditures (X_4)	0.202***	0.015	0.201***	0.015
Ln Chemical Expenditures (X_5)	0.043***	0.015	0.039***	0.015
Ln Tractor Rental (X_6)	0.112***	0.017	0.116***	0.017
Ln other Expenditures (X_7)	0.035***	0.012	0.022*	0.012
0.5 x Ln Land $(X_1)^2$	-0.008	0.016	0.009	0.016
0.5 x Ln Labour $(X_2)^2$	0.057***	0.014	0.065***	0.016
0.5 x Ln Seed Expenditures $(X_3)^2$	0.033***	0.004	0.034***	0.004
0.5 x Ln Fertilizer Expenditures $(X_4)^2$	0.038***	0.004	0.038***	0.004
0.5 x Ln Chemical Expenditures $(X_5)^2$	0.019***	0.006	0.018***	0.006
0.5 Ln Tractor Rental $(X_6)^2$	0.031***	0.005	0.032***	0.005
0.5 Ln other Expenditures $(X_7)^2$	0.010***	0.003	0.006*	0.003
Ln Land x Ln Labour	-0.019*	0.010	-0.018*	0.010
Ln Land x Ln Seed	-0.006**	0.003	-0.007**	0.003
Ln Land x Ln Fertilizer	0.009***	0.003	0.007**	0.003
Ln Land x Ln Chemicals	0.001	0.004	0.001	0.004
Ln Land x Ln Tractor Rental	-0.004	0.003	-0.004	0.003
Ln Land x Ln Other Expenses	0.008***	0.003	0.009***	0.003
Ln Labour x Ln Seed	0.009***	0.003	0.009***	0.003
Ln Labour x Ln Fertilizer	-0.010***	0.003	-0.009***	0.003
Ln Labour x Ln Chemicals	-0.003	0.004	-0.003	0.004
Ln Labour x Ln Tractor Rental	0.004	0.003	0.002	0.003
Ln Labour x Ln Other Expenses	-0.005*	0.003	-0.006**	0.003
Ln Seed x Ln Fertilizer	0.001	0.001	0.001	0.001
Ln Seed x Ln Chemicals	-0.002**	0.001	-0.002**	0.001
Ln Seed x Ln Tractor Rental	-0.001**	0.001	-0.001*	0.001
				(Continued)
Ln Seed x Ln Other Expenses	-0.002***	0.001	-0.002***	0.001
Ln Fertilizer x Ln Chemicals	0.000	0.001	-0.000	0.001
Ln Fertilizer x Ln Tractor Rental	-0.003***	0.001	-0.003***	0.001
Ln Fertilizer x Ln Other Expenses	0.002***	0.001	0.002***	0.001
Ln Chemicals x Ln Tractor Rental	-0.002*	0.001	-0.001	0.001
Ln Chemicals x Ln Other Expenses	-0.001	0.001	-0.001	0.001
Ln Tractor Rental x Ln Other Expenses	-0.003***	0.001	-0.002***	0.001
The Inefficiency Model				
Constant	1.052*	0.492	1.025**	0.465
Head Sex (male)	-0.447	0.437	-0.486	0.412
Head Age (years)	0.001	0.003	0.002	0.003
Head Education (lower secondary)	0.114	0.197	0.088	0.186
Head Education (upper secondary)	0.226	0.173	0.210	0.163
Head Education (teacher collage)	0.168	0.301	0.160	0.279

Head Education (uni & postgrad)	-0.204	0.382	-0.369	0.370
Head Literacy (can read & write)	-0.074	0.114	-0.036	0.106
Household Size (persons)	0.004	0.013	0.003	0.013
Diversification Index	-6.806***	1.027	-4.481***	0.383
Extension Services (1=yes)	-0.473***	0.123	-0.421***	0.113
Oxen & Yaks (number)	-0.201***	0.077	-0.210***	0.070
Tractor/Threshers (number)	-0.785***	0.303	-1.051***	0.312
Cattles (number)	-0.101***	0.027	-0.117***	0.024
Off-farm Employment (1=yes)	-0.009	0.122	-0.044	0.117
Opium share by province (%)	-0.060	0.870	-0.876	0.860
Farm Size (>2 to 5 Jeribs)	-0.306***	0.110	-0.412***	0.104
Farm Size (>5 to 10 Jeribs)	-0.295**	0.133	-0.324**	0.128
Farm Size (>10 to 20 Jeribs)	-0.032	0.183	-0.017	0.175
Farm Size (>20 & above Jeribs)	0.276	0.263	0.404*	0.227
Land Quality (Low)	0.038**	0.122	0.220**	0.102
Agro-ecological Zone 1 (CM)	0.029	0.220	0.004	0.210
Agro-ecological Zone 2 (HFL)	-0.120	0.260	-0.038	0.250
Agro-ecological Zone 3 (SMF)	-0.851***	0.239	-0.997***	0.231
Agro-ecological Zone 4 (HVSB)	-0.953***	0.269	-0.884***	0.260
Agro-ecological Zone 5 (TP)	0.050	0.242	0.037	0.229
Agro-ecological Zone 6 (NMF)	-0.106	0.213	-0.229	0.203
Agro-ecological Zone 7 (EMF)	-0.315**	0.228	-0.483**	0.219**
Log-Likelihood		-5,795.52		-7,541.77
Wald Chi2		5,474.54		4,801.98
Prob. Chi2		0.000		0.000
Mean Efficiency (%)		73.88%		69.87%
N		7,038		7,053

Note: The omitted categories are: none for education level, <2 Jeribs for farm size and agro-ecological zone 8, none for extension services, , none for literacy, none for off-farm employment, and poor quality of land; significance levels indicated by * p<0.10, ** p<0.05, *** p<0.010

Correcting for the potential endogeneity of the variable of crop diversification (CD_i) decreases slightly its coefficient (from -4.95 to -6.81) in the inefficiency model (Table 9). Failing to account for the endogeneity issue underestimated the effect of crop diversification on technical efficiency in the standard exogenous stochastic frontier model presented in column 1 of Table 9. This is consistent with attenuation bias due to measurement error in CD so there was a downward bias in the estimation of the coefficient on CD in the basic SFA model. As a result, the average estimated level of technical efficiency by the endogenous model is also 4% (i.e. mean efficiency for exogenous model is 69.9% and 0.73.9% from endogenous model) higher than the estimated efficiency by the standard model assuming that crop diversification is exogenous.

The validity of the instruments is also tested (results are reported in Table A6 in Annex II). The instrument is strongly correlated with the endogenous variable *(CD_i)*, conditional on the other covariates. This correlation is highly statistically significant (at 1%) indicating that the instrument is valid and strongly correlated with the endogenous variable. The endogenous variable was regressed on the instrument and all other covariates (i.e. all covariates included in the basic inefficiency model). The estimated coefficient for the instrumental variable is large (0.71) and statically significant at 1% level. A test of the joint significance of the instrument rejected the null hypothesis of weak instruments with an F-statistic of 1,075.57 (well above 10, the minimum value for an instrument to be strong) with ($Prob > F = 0.00$). The instrument is sufficiently correlated with the diversification index but appears uncorrelated with the error term (u_i). This means that the average value of *CD* for neighbouring farms in the district is likely to affect technical efficiency of the i^{th} farm only through its impact on the crop diversification.

6.6. Findings and Conclusions

The results of this study reveal that farming sector in Afghanistan experiences significant technical inefficiencies, indicating that farm revenue could be increased by about 29.1% with better utilization of existing inputs, i.e., without employing further resources that raise production costs. On average Afghan farmers achieve 71.9% of potential farm revenues (mean technical efficiency is about 0.71.9), which implies the existence of potential for improving revenue with proper management practices and without using extra input resources. This finding is particularly desirable as the study found constant returns to scale, signifying that an increase in inputs leads to equal proportional increase in revenue. Among inputs, land, household labour, fertilizer, seeds, tractor rental, and other expenditures were found to be positively and significantly contributing to production.

This study identified and examined the impact of a number of factors on technical efficiency. Crop diversification was found to have a positive influence on technical efficiencies. Other important determinants of efficiency were access to extension services; cattle, oxen and tractor ownership; and farm size (with an inverted U relationship). Other factors were not consistently significant factors: agro-ecological zones, opium production and household characteristics.

Robustness checks and hypotheses tests confirm the appropriateness of employing stochastic frontier modelling techniques with a translog production function. Statistical noise and random shocks are not the only reasons for a short-fall in farm revenues given the frontier; other factors under the control of farmers are responsible for technical inefficiencies. Tests on the distributional assumptions for the composed error term validated that truncated normal distribution better fits the specification of the frontier model. Stochastic production frontier function was modelled using two different functional forms (Cobb-Douglas and Translog) and tested using the general log-likelihood ratio test. The result of the test revealed that translog was the appropriate functional form.

The basic model was investigated for endogeneity issues. The test indicated that there is potential endogeneity in crop diversification. The issue was addressed using a recently developed instrumental variable estimator in the stochastic frontier analysis framework. The result of the endogenous models showed that the endogeneity problem resulted in underestimating the impact of crop diversification on technical efficiency. Thus, estimated technical efficiency by the endogenous model was slightly higher (mean of 0.73.9% compared to 69.9%) than the standard SFA with half-normal destruction on u_i model assuming that crop diversification is exogenous.

Given the problem of aggregating production across crops with revenue, and the implication for interpretation of results, this study also conducted similar analysis using the volume of physical output as the dependent variable. Wheat is a major crop; 85% of the household grow wheat alone or combined with other crops (ALCS Household Survey Report, 2013-14). Therefore, similar analyses were carried out on wheat alone as well as grains (wheat and other grains), the estimated results appear to be quite comparable and not significantly different from the revenue-based model presented in this study.

6.7. Further research

This study can be extended to examine the trends of crop diversification and how crop diversification evolved over time. In addition, it is important to investigate the drivers of crop diversification. Crop Diversification might be restricted by farm size, agro-ecological zones and even the household habits of food consumption. For instance, wheat is the main staple food crop that accounts for 60% of the caloric intake in the Afghan diet, thus replacing wheat might not be a choice for some households. Thus, it is worthwhile to investigate how sensitive the efficiency estimates are to the farm size and agro-ecological zones.

Allocative efficiency as mentioned earlier is another important part of the total productivity of farms. Optimal use and allocation of inputs may potentially be an aspect that could improve overall productivity of farms. This could not be addressed given the absence of price data for inputs.

Production of high value cash crops with the basic objective of improving household cash income might require improved local and regional market opportunity. In fact, lack of access to markets maybe another restriction for diversifying farm production. Further, lack of a well-developed farm to market supply chain for the high value crops may make it difficult to move away from single crop production. Access to credit and other institutional aspects of farming might also effect both crop diversification and technical efficiencies.

If crop diversification is a desired strategy for farmers in Afghanistan, as was found in this sturdy, another line of research can focus on what drives or restricts crop diversification and investigate the crop choices and optimum combinations of annual crops to inform farmers on better crop mixes or enterprises to ensure productivity gains as well as food security.

References

Adhikari, C.B., Bjorndal, T., 2012. Analyses of technical efficiency using SDF and DEA models: evidence from Nepalese agriculture. *Applied Economics*. 44, 3297–3308.

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

Alam, M.F., Khan, M.A., Huq, A.S.M.A., 2012. Technical efficiency in tilapia farming of Bangladesh: a stochastic frontier production approach. Aquaculture International 20, 619–634.

ALCS Household Survey Report, 2013. ALCS 2013-14 Main Report-English - 2015 12 21.pdf.

Amaza, P.S., Bila, Y., Iheanacho, A.C., 2006. Identification of factors that influence technical efficiency of food crop production in West Africa: Empirical evidence from Borno State, Nigeria. J. *Agriculture and Rural Development in Tropics and Subtropics*. JARTS 107, 139–147.

Battese, G.E., Coelli, T.J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*. 20, 325–332.

Battese, G.E., Corra, G.S., 1977. Estimation of a production frontier model: with application to the pastoral zone of Eastern Australia. *Australian J. of Agricultural Economics.* 21, 169–179.

Bhatt, M.S., Bhat, S.A., 2014. Technical Efficiency and Farm Size Productivity-Micro Level Evidence from Jammu & Kashmir. *International Journal of Food and Agricultural Economics*. 2, 27.

Bozoğlu, M., Ceyhan, V., 2007. Measuring the technical efficiency and exploring the inefficiency determinants of vegetable farms in Samsun province, Turkey. *Agricultural Systems*. 94, 649–656.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *European J. of Operational Research*. 2, 429–444.

Chhatre, A., Devalkar, S., Seshadri, S., 2016. Crop diversification and risk management in Indian agriculture. *DECISION* 43, 167–179.

Coelli, T., 1995. Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. *J. Productivity Analysis*. 6, 247–268.

Coelli, T., Fleming, E., 2004. Diversification economies and specialization efficiencies in a mixed food and coffee smallholder farming system in Papua New Guinea. *Agricultural Economics*. 31, 229–239.

Coelli, T.J., 1995. Recent developments in frontier modelling and efficiency measurement. *Australian J. Agricultural. Economics*. 39, 219–245.

Coelli, T., Rao, O'Donnell, Battese (Eds.), 2005. An introduction to efficiency and productivity analysis, 2nd ed. ed. Springer, New York.

Dupree, L., 1980. Afghanistan, Oxford Pakistan paperbacks. Oxford University Press.

Elias, A., Nohmi, M., Yasunobu, K., Ishida, A., 2013. Effect of Agricultural Extension Program on Smallholders' Farm Productivity: Evidence from Three Peasant Associations in the Highlands of Ethiopia. *J. Agricultural Science*. 5.

Farrell, M.J., 1957. The Measurement of Productive Efficiency. *J. Royal Statistical Society*. Series A (General).

Felix Larry Essilfie, 2011. Estimation of farm level technical efficiency in small scale maize production in the Mfantseman Municipality in the Central Region of Ghana: A stochastic frontier approach. *J. Development and Agricultural Economics*. 3.

Goodwin, B.K., Mishra, A.K., 2004. Farming efficiency and the determinants of multiple job holding by farm operators. *American J. Agricultural Economics*. 86, 722–729.

Greene, W.H., 1980. Maximum likelihood estimation of econometric frontier functions. *J. Econometrics*. 13, 27–56.

Gronberg, T.J., Jansen, D.W., Karakaplan, M.U., Taylor, L.L., 2015. School district consolidation: Market concentration and the scale-efficiency tradeoff: School Consolidation: Scale & Efficiency. *Southern Economics J*. 82, 580–597.

Guvele, C.A., 2001. Gains from crop diversification in the Sudan Gezira scheme. *Agricultural Systems*. 70, 319–333.

Haji, J., 2007. Production Efficiency of Smallholders' Vegetable-dominated Mixed Farming System in Eastern Ethiopia: A Non-Parametric Approach. *J. African Economics*. 16, 1–27.
Helfand, S., Levine, E., 2004. Farm size and the determinants of productive efficiency in the Brazilian Center-West. *Agricultural Economics*. 31, 241–249.
Jondrow, J., Lovell, C.K., Materov, I.S., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *J. Econometrics*. 19, 233–238.
Karakaplan, M.U., 2017. Fitting endogenous stochastic frontier Models in Stata. The Stata Journal.
Karakaplan, M.U., Kutlu, L., 2015. Handling Endogeneity in Stochastic Frontier Analysis. Unpubl. Manuscr.
K. Mwajombe, K., R. S. Mlozi, M., 2015. Measuring Farm-level Technical Efficiency of Urban Agriculture in Tanzanian Towns: The Policy Implications. *World J. Social Sciences*. 2.
Krupinsky, J.M., Bailey, K.L., McMullen, M.P., Gossen, B.D., Turkington, T.K., 2002. Managing plant disease risk in diversified cropping systems. *Agronomy. J*. 94, 198–209.
Kudaligama, V.P., Yanagida, J.F., others, 2000. A comparison of intercountry agricultural production functions: A frontier function approach. *J. Economic Development.* 25, 57–74.
Kumbhakar, S.C., Lovell, C.A.K., 2000. Stochastic Frontier Analysis. Cambridge University Press.
Kumbhakar, S., Wang, H.-J., 2015. A practitioner's guide to stochastic frontier analysis using Stata. Cambridge University Press, New York, NY.
Lakner, S., Kirchweger, S., Hoop, D., Brümmer, B., Kantelhardt, J., 2015. Impact of diversification on technical efficiency of organic farming in Switzerland, Austria and Southern Germany. Diskussionspapiere, Department für Agrarökonomie und Rurale Entwicklung.
Lee, L.-F., 1983. A test for distributional assumptions for the stochastic frontier functions. *J. Econometrics*. 22, 245–267.
Lin, B.B., 2011. Resilience in Agriculture through Crop Diversification: Adaptive Management for Environmental Change. *BioScience* 61, 183–193.
Lin, T.-F., Schmidt, P., 1984. A Test of the Tobit Specification Against an Alternative Suggested by Cragg. The *Review of Economics and Statistics*. 66, 174.
Llewelyn, R.V., Williams, J.R., 1996. Nonparametric analysis of technical, pure technical, and scale efficiencies for food crop production in East Java, Indonesia. *Agricultural Economics*. 15, 113–126.
Makate, C., Wang, R., Makate, M., Mango, N., 2016. Crop diversification and livelihoods of smallholder farmers in Zimbabwe: adaptive management for environmental change. SpringerPlus 5.
Maletta, H., Favre, R., 2003. Agro-ecological zones. Food and Agriculture Organization of the United Nations
Mango, N., Makate, C., Hanyani-Mlambo, B., Siziba, S., Lundy, M., 2015. A stochastic frontier analysis of technical efficiency in smallholder maize production in Zimbabwe: The post-fast-track land reform outlook. *Cogent Economics and Finance* 3.
Manjunatha, A.V., Anik, A.R., Speelman, S., Nuppenau, E.A., 2013. Impact of land fragmentation, farm size, land ownership and crop diversity on profit and efficiency of irrigated farms in India. *Land Use Policy* 31, 397–405.
Mburu, S., Ackello-Ogutu, C., Mulwa, R., 2014. Analysis of Economic Efficiency and Farm Size: A Case Study of Wheat Farmers in Nakuru District, Kenya. *Economic Research International.* 2014, 1–10.
Meeusen, W., Van den Broeck, 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *Int. Economic Review*. 435–444.
Narala, A., Zala, Y.C., 2010. Technical Efficiency of Rice Farms under Irrigated Conditions in Central Gujarat. *Agricultural Economic Research Review*. Vol. 23, pp 375–381.

Nguyen, H.Q., 2014. Crop Diversification, Economic Performance and Household's Behaviours: Evidence from Vietnam. Econ. Perform. Househ. Behav. Evid. Vietnam July 2 2014.

Ogundari, K., 2013. Crop diversification and technical efficiency in food crop production: A study of peasant farmers in Nigeria. *Int. J. Social Economics.* 40, 267–287.

Oladeebo, J.O., Oyetunde, O.T., 2013. Relationship Between Plot Size and Technical Efficiency of Cassava Farms in Oyo State of Nigeria: A Stochastic Frontier Analysis. *Int. J. Current Research*. 5, 2782–2786.

O'Neill, S., Matthews, A., others, 2001. Technical change and efficiency in Irish agriculture. Econ. Soc. Rev. 32, 263–284.

Rahman, S., 2009. Whether crop diversification is a desired strategy for agricultural growth in Bangladesh?. *Food Policy.* 34, 340–349.

Thieme, O., 2006. Country Pasture/Forage Resource Profiles. Forest 4, 7–0.

Van den Berg, M.M., Hengsdijk, H., Wolf, J., Van Ittersum, M.K., Guanghuo, W., Roetter, R.P., 2007. The impact of increasing farm size and mechanization on rural income and rice production in Zhejiang province, China. *Agricultural Systems*. 94, 841–850.

Wadud, A., White, B., 2000. Farm household efficiency in Bangladesh: a comparison of stochastic frontier and DEA methods. *Applied Economics*. 32, 1665–1673

Wang, H.-J., Schmidt, P., 2002. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *J. Productivity Analysis*. 18, 129–144.

Weiss, C., Thiele, H., others, 2002. Diversifikation und Wachstum landwirtschaftlicher Unternehmen. Agrarwirtschaft 51, 156–163.

Yang, J., Wang, H., Jin, S., Chen, K., Riedinger, J., Peng, C., 2016. Migration, local off-farm employment, and agricultural production efficiency: evidence from China. *J. Productivity Analysis*. 45, 247–259.

Zhang, L., Su, W., Eriksson, T., Liu, C., 2016. How Off-farm Employment Affects Technical Efficiency of China's Farms: The Case of Jiangsu. *China & World Economy*. 24, 37–51.

Appendices

6.8. Annex I: Detailed Mathematical Derivation of SFA

The literature on stochastic frontier models begins with Aigner et al. (1977) normal-half normal model which assumes the following distribution for the components of the composite error term:

- $fv(v_{i)} = iid\ N(0, \sigma_v{}^2)$
- $u_i = iid\ N^+(0, \sigma_u{}^2)$
- V_i and u_i are distributed independently of each, and of the repressors.

The density function of half-normal destruction for the u_i can be further illustrated by figure (10).

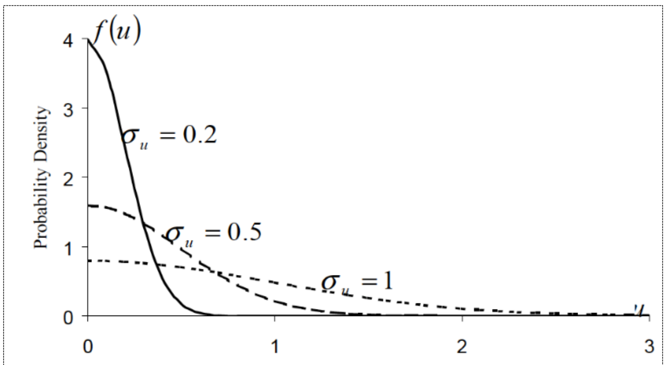

Figure 10: Half-Normal distribution

Assuming a half normal distribution for the inefficiency term of the composed error, the density function of $u_i \geq 0$ and v_i are given by

$$f(u) = \frac{2}{\sqrt{2\pi}\ \sigma_u} exp\left\{-\frac{u^2}{2\sigma_u^2}\right\}$$

$$f(v) = \frac{2}{2\pi\sigma_v} exp\left\{-\frac{v^2}{2\sigma_v^2}\right\}$$

Given the basic assumption of the stochastic frontier models that ui and v_i are independent from each other, the join density function of u_i and v_i is the product of their individual density function, is given by

$$f(u,v) = \frac{2}{2\pi\sigma_u\sigma_v} exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right\}$$

And because $\varepsilon_i = v_i - u_i$, the join density function of u and ε can be specified as:

$$f(u,\varepsilon) = \frac{2}{\sqrt{2\pi}\ \sigma_u\ \sigma_v}\ exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon+u)^2}{2\sigma_v^2}\right\}$$

The marginal density function of ε can be obtained by integrating u out of $f(u,\varepsilon)$ which yields:

$$f(\varepsilon) = \int_0^\infty \int (u,\varepsilon)du$$

$$= \frac{2}{\sqrt{2\pi}\ \sigma_u}\left[1 - \Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right] exp\left\{-\frac{\varepsilon^2}{2\sigma^2}\right\}$$

$$= \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left\{-\frac{\varepsilon\lambda}{\sigma}\right\}$$

Where $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\lambda = {}^{\sigma_u}/_{\sigma_v}$, and $\Phi(.)$ and $\phi(.)$ are the standard normal cumulative distribution density functions. The parameter of λ represents degree of asymmetry of the distribution of the error term. The larger λ is, the more pronounced the asymmetry will be. On the other hand, if λ is equal to zero, then the symmetric error component dominates the one-side error component in the determination of ε_i. Therefore, the complete error term is explained by the random disturbance v_i, which follows a normal distribution. ε_i therefore has a normal distribution.

The marginal density function $f(\varepsilon)$ is asymmetrically distributed with mean and variance of:

$$E(\varepsilon) = -E(u) = -\sigma_u\sqrt{\frac{2}{\pi}}$$

$$V(\varepsilon) = \frac{\pi - 2}{\pi}\sigma_u^2 + \sigma_v^2$$

The log-likelihood function for the normal - half normal stochastic frontier model is:

$$\ln(L) = -\left(\frac{N}{2}\right)(ln2\pi + ln\sigma^2) + \sum_{i=1}^{N}\left[\ln\phi\left[-\epsilon_i\,{}^{\lambda}/_{\sigma}\right] - \frac{1}{2}\,({}^{\varepsilon_i}/_{\sigma})^2\right]$$

Meanwhile, Jondrow et al. (1982)also computed the expected value of ui's conditional on the composed error term for the case in which the asymmetric error term follows an exponential distribution. They provided the following result:

$$f(u,\varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)}$$
$$= \frac{1}{\sqrt{2\pi}\ \sigma_*} exp\left\{-\frac{(u-\mu_*)^2}{2\sigma_*^2}\right\} \Big/ exp\left\{-\frac{(u-\mu_*)^2}{2\sigma_*^2}\right\}$$

Where $\mu_* = -\varepsilon\sigma_u^2/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$. Since $f(u|\varepsilon)$ is distributes as $N^+(\mu_*, \sigma_*^2)$ the mean of this distribution can serve as; point estimated of u_i which is given by:

$$E(u_i|\varepsilon_i) = \mu_{*i} + \sigma_* \left[\frac{\phi\left(\mu_{*i}/\sigma_*\right)}{1-\Phi\left(-\mu_{*i}/\sigma_*\right)}\right]$$

$$= \sigma_* \left[\frac{\phi(\varepsilon_i\lambda/\sigma)}{1-\Phi\left((\varepsilon_i\lambda/\sigma)\right)} - \frac{\varepsilon_i\lambda}{\sigma}\right]$$

Therefore, the estimates of u_i can be obtained from the following specification:

$$TE_i = exp\{\hat{u}\} = exp\{-E(u_i|\varepsilon_i)\}$$

6.9. Annex II: Summary Statistics & Characteristics of farms

A 1: Characteristics of households with CDI below and above the median (0.37)

	Specialized	Diversified	Two-Tailed T Test	
	Mean	Mean	Difference	SE
Aggregate Annual Revenue (AFN)	46,884	69,621	-22,737***	-10.6
Land (Jeribs)	6.77	7.30	-0.530**	-2.44
Labour (hours)	60.16	67.41	-7.253***	-4.89
Seed Expenditure (AFN)	2,389	2,320	69.33	-0.8
Fertilizer Expenditure (AFN)	3,582	5,945	-2,363***	-12.1
Chemicals Expenditure (AFN)	228.0	502.1	-274***	-9.68
Tractor Rental (AFN)	2,460	2,549	-88.7	-0.87
Other Expenditure (AFN)	797	749	47.98	-0.91
Herfindahl Index (HHI)	0.91	0.50	0.414***	-163
Household Size (persons)	7.91	8.76	-0.842***	-10.3
Head Age (years)	44.22	44.55	-0.325	-0.99
Head Sex (1=male, 0=female)	1.00	1.00	-0.0014	-1.00
Extension Services (1=access, 0=No)	0.19	0.23	-0.043***	-4.49
Head Literacy (1=yes, 0=otherwise)	0.31	0.34	-0.0315***	-2.83
Off-farm Employment (1=yes, 0=No)	0.12	0.13	-0.0199**	-2.53
Cattle (number)	1.33	1.87	-0.533***	-11.1
Tractors (number)	0.05	0.06	-0.0102	-1.85
Oxen (number)	0.23	0.24	-0.00851	-0.58
N				7,052

significance levels indicated by * p<0.10, ** p<0.05, *** p<0.010

A 2: Characteristics of farms by access to extension services

	Access		No Access	
Variable	Mean	SD	Mean	SD
Aggregate Annual Revenue (AFN)	59,238	94,208	57,992	89,415
Land (Jeribs)	6.25	9.15	7.24	9.10

Labour (hours)	61.54	65.10	64.38	61.63
Seed Expenditure (AFN)	2,289	4,157	2,372	3,460
Fertilizer Expenditure (AFN)	4,969	7,057	4,709	8,596
Chemicals Expenditure (AFN)	284.1	876.7	386.4	1,267
Tractor Rental (AFN)	2,443	4,293	2,520	4,297
Other Expenditure (AFN)	716.1	2,151	787.8	2,237
Herfindahl Index (HHI)	0.69	0.23	0.71	0.23
Crop Diversification Index (CDI)	0.31	0.23	0.29	0.23
Household Size (persons)	9.06	3.87	8.14	3.31
Head Age (years)	45.28	14.20	44.15	13.63
Head Sex (1=male, 0=female)	1.00	0.03	1.00	0.07
Head Literacy (1=yes, 0=otherwise)	0.44	0.50	0.29	0.45
Off-farm Employment (1=yes, 0=No)	0.20	0.40	0.11	0.31
Cattle (number)	1.91	1.85	1.52	2.08
Tractors (number)	0.06	0.25	0.05	0.23
Oxen (number)	0.15	0.55	0.25	0.63
N	1,473		5,579	

A 3: Characteristics of farm households in provinces with less and more than 1% of opium production

	Less Than 1%		1% & more than 1%	
Variable	Mean	SD	Mean	SD
Aggregate Annual Revenue (AFN)	51,902	82,492	82,181	112,000
Land (Jeribs)	6.88	9.41	7.61	7.90
Labour (hours)	62.44	62.44	68.88	61.89
Seed Expenditure (AFN)	2,236	3,660	2,800	3,415
Fertilizer Expenditure (AFN)	3,497	5,609	9,536	13,457
Chemicals Expenditure (AFN)	146.5	531.2	1,188	2,216
Tractor Rental (AFN)	2,130	4,036	3,916	4,909
Other Expenditure (AFN)	758	2,115	829	2,577
Herfindahl Index (HHI)	0.70	0.23	0.71	0.24
Crop Diversification Index (CDI)	0.30	0.23	0.29	0.24
Household Size (persons)	8.20	3.37	8.85	3.73
Head Age (years)	44.9	13.80	42.4	13.4
Head Sex (1=male, 0=female)	1.00	0.07	1.00	0.03
Extension Services (1=access, 0=No)	0.22	0.42	0.17	0.37
Head Literacy (1=yes, 0=otherwise)	0.35	0.48	0.21	0.41
Off-farm Employment (1=yes, 0=No)	0.13	0.33	0.11	0.32
Cattle (number)	1.69	2.13	1.27	1.60
Tractors (number)	0.06	0.24	0.05	0.21
Oxen (number)	0.28	0.68	0.04	0.25
N	5,573		1,479	

A 3: Percent of zero values in input variables

Variable	Non-zero values	Zeros	%of Zero Values
Total Revenue (Y)	7,052	None	0%
Land (Jeribs)	7,052	None	0%
Labour (hours)	7,052	None	0%

Seed Expenditures (AFN)	4,582	2,470	35%
Fertilizer Expenditures (AFN)	4,935	2,117	30%
Chemicals Expenditures (AFN)	1,594	5,458	77%
Tractor Rental Expenditures (AFN)	4,078	2,974	42%
Other Cost Expenditures (AFN)	3,048	4,004	57%
N			7,052

Table A 5: Destribution of TE and CD by equivalent number (1/HHi)

Equivalent Number	Mean TE (%)	Mean CD	No. of Farms
1	54.12	-	2,198
1-<2	78.33	0.59	3,843
2-<3	85.51	0.59	689
3-<4	88.76	0.71	93
4-<5	91.22	0.77	19
5-<6	92.28	0.82	3
N			7,052

Table A6: Testing the correlation and valadity of IV

Variable	Coefficient	SE
Dependent Variable - Crop Diversification Index (CDI=1-HHI)		
Instrument (IV)	0.710***	0.022
Ln Land (X_1)	0.015	0.012
Ln Labour (X_2)	-0.005	0.006
Ln Seed Expenditures (X_3)	-0.005	0.004
Ln Fertilizer Expenditures (X_4)	0.007*	0.004
Ln Chemical Expenditures (X_5)	-0.008	0.005
Ln Tractor Rental (X_6)	-0.001	0.005
Ln other Expenditures (X_7)	-0.023***	0.003
0.5 x Ln Land $(X_1)^2$	-0.014**	0.006
0.5 x Ln Labour $(X_2)^2$	0.001	0.004
0.5 x Ln Seed Expenditures $(X_3)^2$	-0.001	0.001
0.5 x Ln Fertilizer Expenditures $(X_4)^2$	0.001	0.001
0.5 x Ln Chemical Expenditures $(X_5)^2$	-0.004*	0.002
0.5 Ln Tractor Rental $(X_6)^2$	-0.000	0.001
0.5 Ln other Expenditures $(X_7)^2$	-0.006***	0.001
Ln Land x Ln Labour	-0.008***	0.003
Ln Land x Ln Seed	-0.002**	0.001
Ln Land x Ln Fertilizer	-0.002***	0.001
Ln Land x Ln Chemicals	-0.001	0.001
Ln Land x Ln Tractor Rental	0.000	0.001
Ln Land x Ln Other Expenses	0.002**	0.001
Ln Labour x Ln Seed	-0.000	0.001
Ln Labour x Ln Fertilizer	-0.001	0.001
Ln Labour x Ln Chemicals	0.001	0.001
Ln Labour x Ln Tractor Rental	-0.002**	0.001
Ln Labour x Ln Other Expenses	-0.001	0.001
Ln Seed x Ln Fertilizer	0.000*	0.000
Ln Seed x Ln Chemicals	-0.001**	0.000
Ln Seed x Ln Tractor Rental	-0.000	0.000

Ln Seed x Ln Other Expenses	-0.000	0.000
Ln Fertilizer x Ln Chemicals	-0.000	0.000
Ln Fertilizer x Ln Tractor Rental	0.000**	0.000
Ln Fertilizer x Ln Other Expenses	0.001***	0.000
Ln Chemicals x Ln Tractor Rental	-0.000	0.000
Ln Chemicals x Ln Other Expenses	-0.001**	0.000
Ln Tractor Rental x Ln Other Expenses	0.000	0.000
Head Sex (male)	0.038	0.039
Head Age (years)	-0.000	0.000
Head Education (lower secondary)	-0.013	0.012
Head Education (upper secondary)	0.004	0.010
Head Education (teacher collage)	-0.012	0.017
Head Education (Uni & postgrad)	0.010	0.021
Head Literacy (can read & write)	0.004	0.007
Household Size (persons)	0.001	0.001
Extension Services (1=yes)	-0.014**	0.006
Oxen and Yaks (number)	0.021***	0.004
Tractor/Threshers (number)	0.007	0.011
Number of Cattles (number)	0.001	0.001
		(Continued)
Off-farm Employment (1=yes)	0.031***	0.007
Opium share by province (%)	0.108***	0.042
Farm Size (>2 to 5 Jeribs)	0.010	0.009
Farm Size (>5 to 10 Jeribs)	0.023	0.014
Farm Size (>10 to 20 Jeribs)	0.058***	0.022
Farm Size (>20 & above Jeribs)	0.085**	0.034
Land Quality (Low)	-0.080***	0.008
Agro-ecological Zone 1 (CM)	0.036**	0.018
Agro-ecological Zone 2 (HFL)	-0.035*	0.021
Agro-ecological Zone 3 (SMF)	0.021	0.018
Agro-ecological Zone 4 (HVSB)	-0.041**	0.020
Agro-ecological Zone 5 (TP)	-0.074***	0.019
Agro-ecological Zone 6 (NMF)	0.019	0.017
Agro-ecological Zone 7 (EMF)	0.083***	0.018
Constant	0.088*	0.046
Log-Likelihood		1,682.54
R2		0.331
Test of Endogeneity-Durbin (score) chi2(1)[a]		39.82
Test of Endogeneity-Wu-Hausman F(1,7009)[a]		39.68
Test of Weak IV- F statistic		1,075.57
N		7,038

*Note: The omitted categories are: none for education level, <2 Jeribs for farm size and agro-ecological zone 8, none for extension services, , none for literacy, none for off-farm employment, and poor quality of land; significance levels indicated by * p<0.10, ** p<0.05, *** p<0.010*

[a] H0: CDi is exogenous, rejected. The P-value is (p=0.000) *[b] H0: Instrument is weak, rejected. The P-value is (p=0.000)*