INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
sustainable solutions for ending hunger and poverty
Supported by the CGIAR

ETHIOPIAN DEVELOPMENT RESEARCH INSTITUTE

ESSP II Working Paper 19

Sources of Inefficiency and Growth in Agricultural Output in Subsistence Agriculture: A Stochastic Frontier Analysis

Fantu Nisrane, Guush Berhane, Sinafikeh Asrat, Gerawork Getachew, Alemayehu Seyoum Taffesse, and John Hoddinott

Development Strategy and Governance Division, International Food Policy Research Institute – Ethiopia Strategy Support Program II, Ethiopia

Ethiopia Strategy Support Program II (ESSP II)

ESSP II Working Paper 019

March 2011

IFPRI-ADDIS ABABA
P.O. Box 5689
Addis Ababa, Ethiopia
Tel: +251-11-646-2921
Fax: +251-11-646-2318
E-mail: ifpri-addis@cgiar.org

IFPRI HEADQUARTERS
International Food Policy Research Institute
2033 K Street, NW • Washington, DC 20006-1002
USA
Tel: +1-202-862-5600
Skype: IFPRIhomeoffice
Fax: +1-202-467-4439
E-mail: ifpri@cgiar.org
www.ifpri.org

THE ETHIOPIA STRATEGY SUPPORT PROGRAM II (ESSP II) WORKING PAPERS

ABOUT ESSP II

The Ethiopia Strategy Support Program II is an initiative to strengthen evidence-based policymaking in Ethiopia in the areas of rural and agricultural development. Facilitated by the International Food Policy Research Institute (IFPRI), ESSP II works closely with the government of Ethiopia, the Ethiopian Development Research Institute (EDRI), and other development partners to provide information relevant for the design and implementation of Ethiopia's agricultural and rural development strategies. For more information, see http://www.ifpri.org/book-757/ourwork/program/ethiopia-strategy-support-program or http://www.edri.org.et/.

.

ABOUT THESE WORKING PAPERS

The Ethiopia Strategy Support Program II (ESSP II) Working Papers contain preliminary material and research results from IFPRI and/or its partners in Ethiopia. The papers are not subject to a formal peer review. They are circulated in order to stimulate discussion and critical comment. The opinions are those of the authors and do not necessarily reflect those of their home institutions or supporting organizations.

About the Author(s)

Fantu Nisrane, Consultant, International Food Policy Research Institute

Guush Berhane, Post Doctoral Fellow, Development Strategy and Governance Division, IFPRI

Sinafikeh Asrat, Research Officer, International Food Policy Research Institute, ESSP-II

Gerawork Getachew, Research Officer, International Food Policy Research Institute, ESSP-II

Alemayehu Seyoum Taffesse, Research Fellow, Development Strategy and Governance Division, IFPRI

John Hoddinott, Senior Research Fellow, Poverty, Health, and Nutrition Division, IFPRI

Sources of Inefficiency and Growth in Agricultural Output in Subsistence Agriculture: A Stochastic Frontier Analysis

Fantu Bachewe, Guush Berhane, Sinafikeh Asrat, Gerawork Getachew, Alemayehu Seyoum Taffesse, and John Hoddinott

Development Strategy and Governance Division, International Food Policy Research Institute – Ethiopia Strategy Support Program II, Ethiopia

Keywords: Efficiency; Agricultural Growth; Subsistence Agriculture; Stochastic Production Frontier; Ethiopia.

Table of Contents

Abstract 3
1. Introduction 4
2. Stochastic production frontier analysis..6
3. Empirical model specification and data description 9
3.1 Empirical model..9
3.2 Data description..11
3.2.1 Data used in the stochastic production frontier..12
3.2.2 Data used in the inefficiency equation 14
4. Results and discussion 18
4.1 Parameter estimates of the stochastic production frontier..18
4.1.1 The baseline model 18
4.1.2 Alternative specifications: Evolution of sources of output growth 21
4.1.3 Alternative specifications: Agroecological zone and crop production potential based frontiers 24
4.2 Determinants of efficiency 26
4.2.1 Baseline model 26
4.2.2 Alternative specifications 30
4.3 Trends in efficiency..31
4.3.1 Baseline model 31
4.3.2 Alternative specifications 35
5. Summary and key findings..41
References 43
Appendices..45

List of Tables

Table 3.1. Summary of input-output data used in stochastic production frontier......................15
Table 3.2. Mean values of household, FA, and agroecological specific variables used in the inefficiency equation 16
Table 4.1 Maximum likelihood estimates of parameters in the stochastic production frontier analysis 19
Table 4.2. Maximum likelihood estimates of the inefficiency function parameters 27
Table 4.3. Maximum likelihood estimates of time and agroecologal zone dummy variables included in the inefficiency equation 30
Table 4.4. Average efficiency estimates of farmers by agroecological zones and farmers' association 33
Table 4.5. Average efficiency scores of farmers from specifications using different criteria 36
Table 4.6. Average efficiency scores of farmers from specifications that used panel data formed across AEZ 38

List of Figures

Figure 3.1. Real value of output and input use among ERHS households, 1994–2009 13
Figure 4.1. Trends in average efficiency scores across AEZs, 1994–2009 32

Appendices

Appendix 1. Pearson’s correlation coefficients of selected variables used in the analysis 45
Appendix 2. Maximum likelihood coefficient estimates of frontiers comprising different rounds 46
Appendix 3. Maximum likelihood estimates of separate stochastic production frontiers that use agroecological zone level data...47
Appendix 4. Maximum likelihood parameter estimates of stochastic frontiers of those applying fertilizer, per-hectare specification, and high and low potential crop production regions 48
Appendix 5. Maximum likelihood coefficient estimates of the inefficiency equation parameters comprising different rounds 49
Appendix 6. Maximum likelihood coefficient estimates of the inefficiency equation parameters comprising different agroecological zones.......................................50
Appendix 7. Maximum likelihood parameter estimates of inefficiency equations of those applying fertilizer, per hectare specification, and high and low potential crop production regions 51
Appendix 8. OLS estimates of parameters of the production function 52

Abstract

Studying the sources of growth in agricultural production, examining the extent of inefficiency, and identifying the sources of such inefficiency, is an important step forward to improve the livelihood of subsistence farm households in developing countries. A stochastic frontier analysis is used because, in addition to accounting for sources of growth in agricultural output, this method explicitly incorporates efficiency differences in the analysis. The empirical analysis uses panel data from the Ethiopian Rural Household Survey collected during 1994 through 2009. The results indicate that most of the increase in agricultural output is attained by increased use of traditional inputs such as size and quality of cultivated land, labor, numbers of oxen and hoes, and was heavily influenced by amount of precipitation received. By contrast, the rate of fertilizer application contributed little for increase in output. Participation in the extension program made moderate contribution towards increases in output. Each agroecological zone included in the study gained from Hicks-neutral technological improvements during the 1994–2004 period. Nonetheless productivity levels in 2009 were not different from levels in 1994, and they had declined between 2004 and 2009. Average level of farming efficiency for the surveyed farmers across all the years was 0.46, indicating that an average farmer produces less than half of the value of output produced by the most efficient farmer using the same technology and inputs. However, average farming efficiency has improved during the 1995–2009 period. Farm households' level of farming efficiency is improved by reducing labor bottlenecks and increased education. Households that have diversified risk from plots that are located sufficiently apart appear more efficient. Households that own more animals both in terms of two or more ploughing oxen or total livestock units are more efficient. Drought affects efficiency adversely whenever it strikes. Farmers that live in close proximity to markets are less efficient. On average, farming inefficiency has consistently declined in the period considered. The results suggest that each agroecological zone is faced with different opportunities and obstacles.

1. Introduction

The debate on the role of agriculture in development, and particularly on whether or not the emphasis given to smallholder producers as engines of economic growth and poverty reduction is the right path to development, has recently resurfaced, partly following the renewed shift back to agriculture after the recent global food concerns (Christiaensen et al. 2010). There is ample evidence that productivity growth in developed countries' agriculture in the last few decades has been driven by rapid technological progress in agriculture itself (Ball 1985; Mullen and Cox 1995; Mullen 2007; Chavas 2008). The recent experiences of fast growing Asian economies also suggest that increases in agricultural productivity augmented by increases in labor productivity are an essential characteristic concomitant with success (Collier and Dercon 2009).

Where, within the agricultural sector, should resources be placed to increase productivity? Collier and Dercon (2009) question the relevance of maintaining the emphasis given to (African) smallholders as engines of growth. They argue that the organization and institutions surrounding smallholder producers are too weak to solely bring about needed efficiency in agriculture, and propose far 'more radical strategies that do not place all the bets of African agriculture in a single mode of production' (Collier and Dercon 2009). However, recent cross-country evidence indicates that agricultural growth is more effective than non-agriculture in reducing poverty among the poorest (Christiaensen et al. 2010), driven by a larger participation of poorer households in agriculture. Thus, the key issue in this debate is whether growth and poverty reduction is faster achieved when focusing on smallholder producers, on medium and large commercial farms, or on a combination of both.

Within the context of this debate, studying the sources of increased production, examining the extent of inefficiency, and identifying the sources of inefficiency is an important step forward to inform policy making in agrarian countries like Ethiopia. Using household level panel data from the Ethiopia Rural Household Survey (ERHS), the first objective of this paper is to answer the following questions: (1) given the prevailing input use intensity, how efficient are smallholder producers?; (2) What explains such (in)efficiencies?; and (3) which of these factors are amenable to policies that aim to improve efficiency?
Ethiopia's agriculture is largely characterized by rainfall dependent smallholder production where production is mainly for subsistence. For much of the last 20 years, output has fluctuated in response to variations in rainfall. However, data for the period 2004/05–2007/08 indicate expansion of cereal yield (Seyoum Taffesse 2008; Dercon and Vargas Hill 2009). However, given low level of modern input use (Dercon and Vargas Hill 2009) and hence low levels of intensification, the explanation for such yield increases is unclear. Bachewe (2009), using the Ethiopian Rural Household Survey (ERHS) panel data for the period 1994–2004, find evidence that such yield expansion moved farmers closer to the production possibility frontier, but find no evidence of pushing the frontier outwards, suggesting yield increases have come largely due to increased use of 'conventional' inputs, mainly labor, oxen, and land[1] and not from technological change or intensification (Dercon and Vargas Hill 2009). An important question is whether these recent increases are explained by intensification that pushes the frontier outwards, or are simply due to measurement problems. The second objective of this paper is therefore to systematically examine if and to what extent the recent reported expansion in cereal outputs exhibit structural breaks due to intensification.

[1] Land area cultivated increased by 44% of the size cultivated in 1996/97 (Taffesse 2009).

To address these issues, the paper takes advantage of recent developments in panel data Stochastic Frontier (SF) models (Battese and Coelli 1995; Kumbhakar and Lovell 2000) and long-term panel data for the Ethiopian villages found in the ERHS. SF approaches have often been used to measure firm-level technical efficiency. They have the attractive feature that under specified assumptions, in addition to accounting for the contribution to increased production of factors used in the agricultural production, they can be used to estimate farm-level relative inefficiency and to identify the sources of such inefficiency.

In the following section we discuss the theoretical model used in this study. Section 3 describes the baseline empirical model used in the study and briefly describes the data used in the stochastic frontier and inefficiency equation parts of the model. Section 4 presents results from the baseline model and other specifications estimated to investigate the robustness of the baseline model and to test other claims. Conclusions are reported in Section 5.

2. Stochastic production frontier analysis[2]

Let the stochastic production frontier for farmer i at period t be given by

$$Y_{it} = f(X_{it}, \beta) * \exp(V_{it} - U_{it}) \tag{1}$$

where $i \in (1,2,...,I)$ is an index for farm household i and $t \in (1,2,...,T)$ represents time period t. Y_{it} is output of farmer i at time period t while X_{it} is a (1 × k) vector of inputs of farmer i at time period t (and depending on the specification of $f(X_{it}, \beta)$, interaction terms of the inputs). β is a (k×1) vector of unknown parameters to be estimated. V_{it} and U_{it} are the idiosyncratic and inefficiency components of the "composed error term" of farmer i at time period t. The latter is assumed to take zero or positive values. Those producers that have positive values lie below the efficiency frontier while those that have zero values are efficient farmers that lie on the efficiency frontier. This component of the error term measures the departure of each producer from an efficiency frontier. Three assumptions are made:

i) V_{it} are identically and independently normally distributed with mean zero and standard deviation σ_v^2, that is V_{it}~N($0, \sigma_v^2$).

ii) U_{it} are independently distributed non-negative truncation of a normally distributed random variable with mean $Z_{it}\delta$ and standard deviation σ_u^2. Z_{it} is a (1 × m) vector of household and region specific variables that affect efficiency while δ is an (m × 1) vector of unknown parameters of the inefficiency equation.

iii) V_{it} and U_{it} are distributed independently of each other and are independently distributed of the X_{it}.

Given the stochastic production frontier specified by equation (6), the level of technical efficiency (TE_{it}) of each farm household i at period t is

$$TE_{it} = \frac{Y_{it}}{f(X_{it}, \beta) * \exp(V_{it})}$$
$$TE_{it} = \exp(-U_{it}) \tag{2}$$

Since U_{it} are a non-negative truncation of normally distributed random variable, TE_{it} can take a maximum value of one. The specification allows for efficiency to vary over time. This definition of technical efficiency follows from the idea that if a farm household's actual production level, Y_{it}, is less than the maximum achievable production level,

---

[2] Seminal studies on the development of this approach include Debreu (1951), Koopmans (1957), Farrell (1957), Shephard (1970), Winsten (1957), Aigner and Chu (1968), Afriat (1972), Aigner, Lovell and Schimidt (1977), Meeusen and Van den Broeck (1977), Pitt and Lee (1981), Battese and Coelli (1995), Kumabhakar and Lovell (2000).

$f(X_{it},\beta)*\exp(V_{it})$ that admits the existence of only idiosyncratic differences, and assuming no measurement error, then there is some inefficiency on the part of the farmer and this inefficiency is greater the lower Y_{it} is from $f(X_{it},\beta)*\exp(V_{it})$, or the higher is U_{it}. Note that the inefficiency effects, U_{it}, as well as the symmetric error terms, V_{it}, may carry the effects of errors of measurement in both the explanatory as well as the dependent variables, just as any other econometric model.

Technical inefficiency is assumed to be a function of farm household and region specific variables, Z_{it}, and a set of parameter values, δ, to be estimated along with the production function parameters. The inefficiency equation is specified as

$$U_{it} = Z_{it}\delta + W_{it} \qquad (3)$$

where W_{it} is a random variable that is assumed to be distributed with mean zero and variance σ_w^2. U_{it} is defined by the truncation of the normal distribution with the point of truncation given by $-Z_{it}\delta$. (Since $U_{it} = Z_{it}\delta + W_{it} \geq 0$, $W_{it} \geq -Z_{it}\delta$, so that W_{it} is truncated from below.) U_{it} is assumed to be the positive half of a normally distributed variable with mean zero ($U_{it} \sim N^{+}(0,\sigma_u^2)$).The truncated normal distribution for U_{it} is given by

$$g_u(U_{it}) = \frac{1}{\sqrt{2\pi}\sigma_u \Phi(Z_{it}\delta/\sigma_u)} \exp\left\{-\frac{(U_{it} - Z_{it}\delta)^2}{2\sigma_u^2}\right\}, \quad U_{it} \geq 0 \qquad (4)$$

where $\Phi(.)$ is the standard normal cumulative distribution. Thus $f(U_{it})$ is the density function of a normally distributed random variable with mean $Z_{it}\delta$ truncated below at zero.

The density function of the random variable V_{it} is given by

$$g_v(V_{it}) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left\{-\frac{V_{it}^2}{2\sigma_v^2}\right\}, \quad V_{it} \in (-\infty,\infty) \qquad (5)$$

Given V_{it} and U_{it} are assumed to be distributed independently and omitting subscripts, their joint distribution is given as

$$g_{uv}(U,V) = \frac{1}{2\pi\sigma_u\sigma_v \Phi(Z\delta/\sigma_u)} \exp\left\{-\frac{(U - Z\delta)^2}{2\sigma_u^2} - \frac{V^2}{2\sigma_v^2}\right\}, \; U \geq 0 \qquad (6)$$

Define the composite error term as $\varepsilon_{it} = V_{it} - U_{it} = Y_{it} - f(X_{it}, \beta)$. The joint distribution of ε_{it} and U_{it} is given by

$$f(U,\varepsilon) = \frac{1}{2\pi\sigma_u\sigma_v\Phi(Z\delta/\sigma_u)}\exp\left\{-\frac{(U-Z\delta)^2}{2\sigma_u^2} - \frac{(U+\varepsilon)^2}{2\sigma_v^2}\right\} \tag{7}$$

The marginal density function of ε is given by

$$g_\varepsilon(\varepsilon) = \int_0^\infty f(U,\varepsilon)dU$$

$$g_\varepsilon(\varepsilon) = \frac{1}{\sqrt{2\pi}(\sigma_v^2+\sigma_u^2)^{1/2}\left[\Phi(Z\delta/\sigma_u)/\Phi(\mu^*/\sigma_*)\right]}\exp\left\{-\frac{(U+\varepsilon)^2}{2(\sigma_v^2+\sigma_u^2)}\right\} \tag{8}$$

where $\mu^* = (\sigma_v^2 Z\delta - \sigma_u^2\varepsilon)\big/(\sigma_v^2+\sigma_u^2)$ and $\sigma_*^2 = (\sigma_v^2 \times \sigma_u^2)\big/(\sigma_v^2+\sigma_u^2)$. Accordingly, the density function of Y_{it} is

$$g_y(Y_{it}) = \frac{1}{\sqrt{2\pi}(\sigma_v^2+\sigma_u^2)^{1/2}\left[\Phi(\tilde{\mu}_{it})/\Phi(\tilde{\mu}_{it}^*)\right]}\exp\left\{-\frac{(Y_{it}-f(X_{it},\beta)+Z_{it}\delta)^2}{2(\sigma_v^2+\sigma_u^2)}\right\} \tag{9}$$

where $\tilde{\mu}_{it} = Z_{it}\delta/\sigma_u$, $\tilde{\mu}_{it}^* = \mu_{it}^*/\sigma_*$, and $\mu_{it}^* = [\sigma_v^2 Z_{it}\delta - \sigma_u^2(Y_{it} - f(X_{it},\beta))]/(\sigma_v^2+\sigma_u^2)$.

Let us define: $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\gamma = \sigma_u^2/\sigma^2$. Note that $\gamma \in (0,1)$; if $\gamma \to 0$ then either $\sigma_u^2 \to 0$ or $\sigma_v^2 \to \infty$ which occurs if the symmetric disturbance term V_{it} dominates the truncated efficiency component U_{it} which in turn indicates that the idiosyncratic error component dominates the inefficiency effects and that OLS estimation techniques are more appropriate than stochastic frontier analysis. As $\gamma \to 1$ either $\sigma_u^2 \to (\sigma_u^2 + \sigma_v^2)$ or $\sigma_v^2 \to 0$ which means that if the variation in the inefficiency component increasingly dominates the variation in ε_{it} and indicates estimating a stochastic production frontier is appropriate. Given the above reparameterizations and that we have observations for $t \in (1,2,...,T)$ and $i \in (1,2,...,I)$ the log likelihood equation is given by

$$L(\Theta,Y) = -\frac{1}{2}\left\{\left[\sum_{i=1}^{I}\sum_{t=1}^{T}\ln 2\pi + \ln\sigma^2\right] + \left[\sum_{i=1}^{I}\sum_{t=1}^{T}(Y_{it} - f(X_{it},\beta) + Z_{it}\delta)/\sigma^2\right] + \left[\ln\Phi(\tilde{\mu}_{it}) - \ln(\tilde{\mu}_{it}^*)\right]\right\}$$

(10)

where $\Theta' = (\beta', \delta', \sigma_v^2, \sigma_u^2)'$ is the parameter set. First order derivatives of this last equation with respect to the parameter set provide an expression which when solved yield the estimates for the parameters.[3]

3. Empirical model specification and data description

3.1 Empirical model

The stochastic production frontier

We use a Cobb-Douglas specification in our baseline estimation of the stochastic production frontier. Area, labor, and mean annual rain are in logarithms while the remaining variables are in levels.

$$\begin{aligned}\ln Y_{it} = {} & \beta_0 + \beta_1 \ln Area_{it} + \beta_2 \ln over\,16\; members_{it} + \beta_3 \ln annaul\; rain_{it} + \beta_4 FertUse_{it} + \beta_5\; oxen_{it} + \\ & \beta_6 Av\; land\; quality_{it} + \beta_7 hoe_{it} + \beta_8 Plough_{it} + \beta_9 Partnep_{it} + \beta_{10} CentralHL_{it} + \ldots + \\ & \beta_{13} Enset_{it} + \beta_{14} 1999dummy_i + \ldots + \beta_{16} 2009dummy_i + V_{it} - U_{it}\end{aligned} \tag{11}$$

where $t \in (1, 6, 11, 16)$ is the period for which data are available for the 16 year period extending from 1994 through 2009 and where $i \in (1, 2, 3, ..., 1516)$ represents farmer i. β_j, j = 1, 2, …,16 are coefficients of the production function to be estimated. $\ln Y_{it}$ is the logarithm of real value of output of household i in period t. $\ln Area_{it}$ is the logarithm of the total area of land cultivated by the household, $\ln over 16\, members_{it}$ is the logarithm of number of household members 16 years of age and older in household i at time t. The logarithm of rain received in millimeters in the FA where household i resided during the 12-month period prior to the survey at period t is given by $\ln annual rain_{it}$. $FertUse_{it}$ is fertilizer used in kilograms by farmer i in period t. $Oxen_{it}$ is the number of oxen that household i used for ploughing at time t. $Avlandqualty_{it}$ is average land quality of the plots cultivated by household i at time t. hoe_{it} and $Plough_{it}$ stand for the number of hoes and ploughs owned by household i at time t. $Partnep_{it}$ is a dummy variable that takes a value of 1 if household i participated in a new extension program at time t. Also included are agroecological zone and time dummy variables. These account for productivity differences that could result from variations in weather and overall agro-climatic conditions that vary between periods and agroecological zones, and are not captured by the remaining factors of production included in the model.[4]

---

[3] The theoretical model and empirical specification heavily draws from Bachewe (2009).

[4] The calculation of elasticities depends on the way in which these variables are specified. For those that are specified in logarithms, the coefficient estimates themselves are the elasticities. For those that enter the equation in a linear fashion, the coefficient estimates of these variables do not represent the elasticity; instead they represent the change in the logarithm of the

The inefficiency equation

Household specific factors are assumed to enter linearly in the model examining the correlates of inefficiency. The empirical counter-part of the inefficiency equation is specified as

$$
\begin{aligned}
U_{it} = {} & \delta_0 + \delta_1 Sex_{it} + \delta_2 Age_{it} + \delta_3 Education_{it} + \delta_4 Femaledummy_{it} + \delta_5 Householdsize_{it} + \\
& \delta_6 (Noplots * \ln area)_{it} + \delta_7 (areaha / over16members)_{it} + \delta_8 Oxendummy_{it} + \\
& \delta_9 Livestockunits_{it} + \delta_{10} Noagext_{it} + \delta_{11} Partnep + \delta_{12} Drought_{it} + \delta_{13} Surveymonth_{it} + \delta_{14} Elevation_{it} \\
& + \delta_{15} Dst\_healthctr_{it} + \delta_{16} Dst\_clos\_market_{it} + \delta_{17} Dst\_FActr_{it} + \sum_{k=18}^{36} \delta_k (AEZ_l * Year_t) \\
& + W_{it}
\end{aligned}
\tag{12}
$$

where $Education_{it}$ is education level of head of household i at time t. $Femaledummy_{it}$ is a dummy variable equaling 1 for household i if it had no male household member 16 years of age and older at time t. We include this dummy variable to see the effect of the gender composition of labor force on farming efficiency. $Oxendummy_{it}$ is a dummy variable that assigns a value of 1 for household i if it owns 2 or more plowing oxen at time t. $Noagext_{it}$ is the number of agricultural extension agents in the FA that household i resided at time t. $Drought_{it}$ is a dummy variable that takes a value of 1 if crop output of household i suffered from drought at time t. Since *Meher* season crops are harvested between August and October, farmers surveyed during this period and in the months that immediately follow the Meher season may find it easier to answer questions on production compared to those that are surveyed between February and July. As a result, there may be differences in data quality between the two groups of farmers due to potential measurement error. To control for this measurement error, a dummy variable $Surveymonth_{it}$ that assigns a value of 1 if household i was surveyed in the months of August through January and zero otherwise is included. The dummy variables associated with parameters δ_{18} through δ_{36} , $AEZ_l * Year_t$, are interactions variables created by multiplying the dummy variables on agroecological zone *l* and time t and dropping the product of Northern highlands and 1994. Battese and Coelli (1995) included time variables in stochastic production frontier and inefficiency equations to account for both technical change and time varying technical inefficiency effects, respectively. They argue that the year variable in equation (11) accounts for Hicks neutral technological change while the year variable in inefficiency equation (12) takes into

---

real value of output for a unit change in the respective inputs. That is, for these variables, $\beta_j = \partial \ln Y_{it} / \partial X_j$, and the elasticity of value of output with respect to these inputs is calculated as $E_{YX} = (\partial \ln Y_{it} / \partial X_{it}) * X_{it}$, where Y_{it} is the real value of output, and X_{it} is mean value of input X, where X is an input entering the equation linearly. For dummy variables, such as participation in the extension package and time and agroecological zone dummies, $\partial \ln Y_{it} / \partial X_j$ is not defined because it is discontinuous. Halvorson and Palmquist (1980) show that the elasticity of value of output with respect to a dummy variable is given by $E_{YX_{Dl}} = Exp(\beta_{Dl}) - 1$, where X_{Dl} represents the dummy variable, β_{Dl} is its estimated coefficient and Y is real value of output.

account inefficiency changes that occur during the period considered. The dummy variables are specified in similar ways for the stochastic frontier, and inefficiency parts of the model. However, the interpretation of the resulting estimates differs, as discussed below. These dummy variables capture regional, socio-economic, and administrative differences that may affect farming efficiency and parameter estimates of these dummy variables that measure efficiency gains of a zone over time.

3.2 Data description

Data from four rounds of the Ethiopian Rural Household Survey (ERHS) conducted in 1994, 1999, 2004, and 2009 are used in the baseline analysis, while additional data from ERHS 1995 and 1997 were used to check the robustness of the baseline analysis. The ERHS is a longitudinal household data set that includes households in 15 Farmers' Associations (FAs) of rural Ethiopia[5]. The surveys span 4 of the 9 administrative regions in Ethiopia. The largest proportion of the country's predominantly settled farmers are located in these 4 regions. The surveys cover 15 of the 389 woredas (districts) in the 4 regions. One FA was selected from each of the woredas, except for one large woreda in the Amhara region, Debre Birhan, from which four FAs were included in the sample. The surveys are conducted on a sample that is stratified over the country's three major agricultural systems found in five agroecological zones (Dercon and Hoddinott 2004). The first agroecological zone is known as northern highlands. This zone includes two villages in the Tigray region, Geblen and Harresaw, and one from the Amhara region, Shumsheha. The northern highlands are characterized by poor resource endowments, adverse climatic conditions, and frequent drought. The central highlands agroecological zone is represented by the villages of Dinki, Yetmen, and Debre Birhan, all located in the Amhara region, and Turufe Ketchema in the Oromia region. The Arussi/Bale agroecological zone includes the villages of Koro Degaga and Sirbana Godeti, both found in Oromiya. Adele Keke is the sole survey site found in the Hararghe agroecological zone of Oromiya. The remaining five villages of Imdibir, Aze Deboa, Gara Godo, Adado, and Doma are found in the Enset growing agroecological zone located in the Southern Nations, Nationalities and People region.

The 1994 survey round included approximately 1,470 households. Sample sizes in each village were chosen so as to approximate a self-weighting sample when considered in terms of farming system: each person (approximately) represents the same number of persons found in the main farming systems as of 1994. Sample attrition between 1994 and 2004 is low, with a loss of only 12.4 percent (or 1.3 percent per year) of the sample over this ten-year period.[6] Households that attrited were not replaced. In addition to those that attrited, we dropped about 600 households for which we have data as they, in a given survey year, did not cultivate any land.

[5] Although three more FAs were included in ERHS rounds 1999, 2004, and 2009 we did not include these additional data because complete data is available for these FAs only for 1999 and 2009 and given that these FAs are generally high productivity areas we believe that it is better to study them separately.

[6] Over the period 1994–2004, t tests of mean values for attriters and non-attriters showed no statistically significant differences in terms of initial levels of characteristics of the head (age, sex), assets (fertile land, all land holdings, cattle), or consumption. However, attriting households were, at baseline, smaller than non-attriting households. Between 1999 and 2004, there are some significant differences by village with one village, Shumsha, having a higher attrition rate than others in the sample. Our survey supervisors recorded the reason why a household could not be traced. Using these data, we examined attrition in Shumsha on a case-by-case basis, but could not find any dominant reason why households attrited. This result is also obtained when we estimate a probit model where the likelihood of attrition is the dependent variable.

3.2.1 Data used in the stochastic production frontier

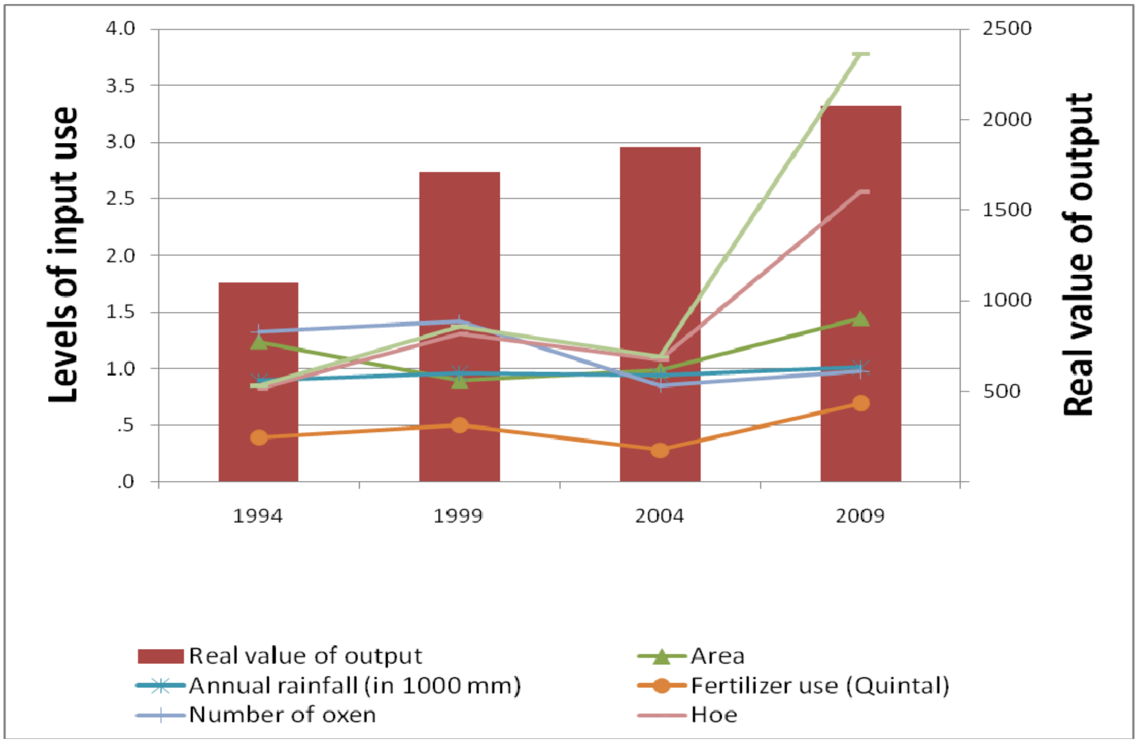
Crop production in Ethiopia is dominated by small-scale subsistence farm households that on average cultivate less than a hectare of land. Cultivated crops and cultivation mechanisms vary across regions, due to heterogeneous agroecology and food culture. Cereal production dominates the North and Central highlands, Arussi/Bale, which span from north to central Ethiopia, and Hararghe agroecological zones in the east, where oxen are the main source of ploughing power. The Enset agroecological zone, named after the crop most commonly grown in the region, is mainly dominated by a hoe-culture.[7] However, there are common characteristics among these regions. Household members are the major source of labor while application of modern inputs is minimal. Most importantly, agriculture in Ethiopia, and among surveyed households, is rain-fed with less than 2 percent of cultivated land irrigated.

The panel data used in this study included household and plot level information, mainly household characteristics, crop production techniques, and household level input uses and outputs.[8] Data on measured inputs and outputs were collected mainly in local units. Conversion of household level measured inputs and outputs into standard units was conducted using farmer association (FA) specific conversion factors. The conversion factors take into consideration FA level differences when converting local area and weight units into standard hectare and kilogram units. After converting measured inputs and outputs in to standard units, nominal prices collected at each round of the survey were used to convert the output produced on each plot into nominal value. As a third step, the area of plots cultivated by the household and the value of output produced on each plot were aggregated at a household level. This aggregation is required because most of the ERHS rounds provide data on the remaining measured inputs only at a household level, making it difficult to associate the factors of production used to produce the output on each plot of land. Finally, nominal value of output of each household was converted into real value output. For that purpose we used FA level food price indices to deflate the value of output of farm households included in the surveys. Data on the food prices were purposefully collected and food price indices were calculated as a Laspeyers index, based on FA prices using average shares in 1994 as the weights.[9]

Real value of household crop production is used as the regressand. Table 3.1 shows that average value of output at constant prices grew consistently starting from the lowest value of 1,099 birr in 1994 to the largest 2,077 in 2009, which has grown at an average annual rate of about 5 percent. Growth was fast between 1994 and 1999 at 11 percent while it was relatively slow during the 1999–2004 and 2004–2009 periods which saw respectively 1.5 and 2.5 percent annual growth in real value of output. Mean and median real values of output of those households applying fertilizer were 54 and 45 percent larger than their respective values by an average household.

[7] Farmers' associations included in ERHS are classified along lines of the five agroecological zones referred to in the text. We will frequently mention and present results along these divisions.
[8] See Bachewe (2009) for further description of different rounds of surveys.
[9] We used weighted output prices to convert nominal value of output produced by each household. We calculate the average price received by farmers who sold any output for each of the following crops: maize, barley, wheat, black teff, white teff, enset, chat, coffee, and sorghum, then constructed the median price for each crop and lastly constructed a weighted average of these medians, where the weights were based on the share of production (in value terms) of each crop.

Figure 3.1. Real value of output and input use among ERHS households, 1994–2009



During the 1994–2009 period each household had, on average, 3 household members of 16 years or older, which is used as proxy for labor. Households applying fertilizer had 12 percent larger supply of labor. Labor supply decreased during 1994–2009 at an average annual rate of 0.8 percent. Each household used an average of 1.1 hoes and 1.3 ploughs during the 1994 through 2004 period, while the number of hoes and ploughs used by an average household increased significantly to 2.6 and 3.8 in 2009. On average, each household had 1.1 oxen used for ploughing, while this number was 33 percent larger among those that used fertilizer. Average cultivated area per household ranged from 0.9 hectares in 1999 to 1.44 hectares in 2009. An average farmer cultivated 1.1 hectares of land over the years while those that applied fertilizer cultivated 29 percent more land than the average farmer.

Average fertilizer use per household ranged from 28 kilograms in 2004 to 69 kilograms in 2009. Indeed, fertilizer application rate in 2009 was about 39 percent larger than the average in 1999, which had the second largest average application rate of 50 kilograms per household. It is interesting to note that fertilizer application among those that actually apply fertilizer has consistently grown, except for the slight decline between 1994 and 1999, at an average annual rate of 0.14 percent. Unlike the case of the average farmer who experienced an average annual decline in fertilizer application rate of 9 percent, application rate among those that actually apply fertilizer grew the fastest between 1999 and 2004 at 3.9 percent. During the 2004–2009 period fertilizer application of an average farmer grew at the fastest rate of 30 percent while those that actually apply fertilizer increased their application at an average annual rate of only 1.8 percent. The reason for this can be gleaned from Table 3.1.

The number of households that apply fertilizer was the smallest in 2004 at 300 while it was the largest in 2009. So, even if average application rate among those that apply fertilizer is comparable, average application rate of an average farmer diverges. Access to extension was largest in 2009 at 28 percent while it was lowest in 1994 at 6 percent[10.] Disregarding 2009, participation averaged 9 percent over the period considered.

Average annual rainfall varied from 894 millimeters in 1994 to 1,011 millimeters in 2009. Except for 1994, it stayed in a narrow band of 943 and 1011 millimeters. Self-reported index of average land quality fell consistently over the period covered by these surveys, which is to say land quality has increased over the years with those that apply fertilizer reporting better quality land.[11]

3.2.2 Data used in the inefficiency equation

Variables used in the inefficiency equation are presented in Table 3.2. Data on age and sex of the head of the household are included in the inefficiency equation to determine if these factors contribute to differences in efficiency among farm households. The average age of the head of household hovered around 49 years for the first three rounds of ERHS used in the baseline, while the round in 2009 has a markedly older set of household heads at an average age of 53 years. Education of household heads was included in the inefficiency equation to see if human capital contributed to farming efficiency. Specifically, we generate a dummy variable which is equal to one if the head has at least three grades of schooling (corresponding to a minimal attainment of literacy), zero otherwise. About 17 percent of household heads had education level of grades 3 and above. The level of education of household heads saw an improvement of 1.6 percent over the 1994–2009 period.

An interaction variable created by multiplying the number of plots and the sum of the area of these plots was included to determine if fragmented land holdings affect farming efficiency. In general, inefficiency was reduced by a smaller number of plots, a lower fragmentation level and a shorter distance to the plots. On the other hand, plots that are sufficiently apart can reduce risks associated with severe weather and land quality. The number of plots increases over time, from 3.3 in 1994 to 4.5 in 2004 and declined to 4.1 plots in 2009. Cultivated area was divided by the number of household members of 16 years and older, and was used in the analysis to see the effect of cultivated land per working member of the household on farming efficiency. Scaled livestock units are also included in the analysis. This variable converts each type of farm animal owned by each household into livestock units. This variable is a proxy for household wealth as in most rural areas of Ethiopia animals are a store of value and a ready means of acquiring cash in times of need, which can be viewed as an insurance against any perceived production or input use risk.

[10] Since participation in the extension package is approximated using number of visits by an extension agent, it is likely an overestimation of participation.
[11] The index of land quality was created by multiplying the value for the slope of the land (1=flat, 2=hill, or 3=high-hill/high land) and the value for the fertility of the land (1=fertile, 2=moderately fertile, and 3=less fertile). A plot with a value of 1 is deemed "ideal" land while plots with the least desirable quality have a value of 9 for land quality.

Table 3.1. Summary of input-output data used in stochastic production frontier

Year	Group	Variable	Number	Real value of output	Cultivated area (ha)	Household members 16 or older	Annual rain fall	Fertilizer used (KG)	Number of oxen	Average land quality	Number of hoes	Number of ploughs
1994	Average farmer	Mean		1099	1.2	3.3	894	39.5	1.3	2.5	.8	.9
		Median	1317	529	.9	3.0	889	.0	1.2	2.0	1.0	1.0
		Maximum		23433	9.9	16.0	1604	3000.0	8.0	9.0	20.0	5.0
	Applying Fertilizer	Mean		1716	1.4	3.8	935	98.4	1.5	2.2	.9	1.1
		Median	529	1143	1.1	3.0	889	50.0	1.2	2.0	1.0	1.0
		Maximum		23433	9.9	13.0	1541	3000.0	8.0	9.0	9.0	5.0
1999	Average farmer	Mean		1712	.9	3.1	959	50.1	1.4	2.2	1.3	1.4
		Median	1211	1206	.7	3.0	953	7.0	1.0	2.0	1.0	1.0
		Maximum		39553	7.0	11.0	1480	504.0	6.0	9.0	24.0	15.0
	Applying Fertilizer	Mean		2181	1.2	3.2	959	97.7	1.7	1.8	1.4	1.6
		Median	621	1787	1.0	3.0	953	70.0	1.9	1.6	1.0	1.0
		Maximum		12825	5.5	11.0	1480	504.0	6.0	9.0	6.0	12.0
2004	Average farmer	Mean		1846	1.0	2.7	943	27.9	.9	2.2	1.1	1.1
		Median	1254	1305	.7	2.0	982	.0	1.0	2.0	1.0	1.0
		Maximum		28092	17.0	10.0	1541	1400.0	9.0	9.0	24.0	22.0
	Applying Fertilizer	Mean		3061	1.5	3.2	925	116.6	1.5	2.0	1.3	1.7
		Median	300	2440	1.2	3.0	982	79.0	1.0	2.0	1.0	1.0
		Maximum		16525	17.0	9.2	1541	1400.0	9.0	6.0	4.0	22.0
2009	Average farmer	Mean		2077	1.4	2.9	1011	69.5	1.0	2.1	2.6	3.8
		Median	1251	1386	1.0	3.0	970	12.0	1.0	1.8	2.0	2.0
		Maximum		28019	7.0	9.0	1534	1045.0	11.0	9.0	17.0	24.0
	Applying Fertilizer	Mean		2954	1.8	3.1	1072	125.7	1.3	1.8	3.0	4.3
		Median	692	2322	1.5	3.0	1086	87.0	1.0	1.7	2.0	2.0
		Maximum		28019	7.0	9.0	1534	1045.0	11.0	9.0	17.0	24.0

Source: ERHS rounds 1994, 1999, 2004, and 2009.

Table 3.2. Mean values of household, FA, and agroecological specific variables used in the inefficiency equation

Variable		Average value across 1994, 1999, 2004, and 2009	Average annual rate of change across 1994–2009
Type	Units		
Sex of head of household	0 if female, 1 if male	.776	-0.60
Age of head of household	Years	50	0.46
Education level of head	0 if illiterate, 1 if literate	.173	1.57
Household size	Count	5.7	-1.74
Number of plots cultivated	Count	4.2	2.11
Livestock units per household	Index	2.97	1.17
Number of extension officers in FA	Count	.82	1.50
Was crop damaged by drought	0 if no, 1 if yes	.18	--
Mean elevation	Meters	2088	--
Distance to nearest health center	Kilometers	16.9	-5.95
Distance to closest market	Kilometers	20.6	-5.13
Distance to nearest FA center	Kilometers	19.3	-5.85

Source: Calculated from ERHS panel data.

Farmers traveled about 17 kilometers on average to get medical services, while they traveled 21 kilometers to the nearest market and 19 kilometers to the nearest FA center. These distances declined by over 5 percent through the 1994–2009 period. We include these variables as a farmer had to devote his/her farming time and resources to acquire the services provided by these facilities. These variables might also reflect variations in the on-farm cost of purchased inputs. The farm households lived on villages located at an altitude of on average 2,088 meters above sea level. The altitude variable was included as a climatic indicator of the surveyed regions along with dummy variables that are assigned for different agro-climatic zones. In the inefficiency equation, we include the number of agricultural extension agents in the FAs. On average, there was fewer than one extension agent per FAs in all rounds, though the two most recent survey years that had data on this variable had higher averages than the first three.[12]

The dummy variable on drought was included in the analysis to control for the adverse effect of climatic factors on production efficiency of farmers which otherwise could have higher efficiency scores than reported. In 1994, drought adversely affected 30 percent of the surveyed farmers. The second dry year, as was reported by the farmers, was 2009 with 27 percent of the sampled households reporting drought. However, the year 2009 witnessed an amount of rainfall similar to the one in 1997 and larger than the average in 1999, when no farmer experienced drought. That all households in Geblen and Haresaw, and 84 percent of the households in Doma reported suffering from drought in 2009 is perhaps consistent with the fact that these three FAs received the three lowest amounts of rainfall in 2009. What probably has resulted in these paradoxical statistics is the fact that 99 percent of the farmers in Aze Deboa and 40 percent in Imdibir reported to have experienced drought while the respective FAs received the 4th and 5th largest amounts of rainfall.

[12] Since ERHS 2009 does not have data on this variable we use data from 2004. If the number of extension agents increased by 2009, as it did between 1999 and 2004, our assumption will just be a conservative one.

Estimation issues

Maximum likelihood estimates of parameters of the two-equation system given by equations (11) and (12) are reported in Table 4.1.[13] We also estimated the OLS version of the empirical model provided by equation (11) above (that is, by dropping all the variables in the inefficiency equation) and checked for negative skewness. All parameter estimates of the OLS regression have the expected sign and all except number of ploughs and participation in the extension package are significant. Moreover, OLS parameter estimates, tended to be larger (see Appendix 8). The residuals of the OLS estimate are negatively skewed with a skewness value of –2.88. As a rough guide, a ratio of the skewness value to its standard error that exceeds 2 is taken to indicate a departure from symmetry. In this data set this ratio is about 84, clearly indicating that the error terms are negatively skewed and that, holding other factors constant, OLS is less likely to be the appropriate approach to follow.

Four sets of analyses were conducted to serve different purposes. The first analysis, which is taken as the baseline, jointly estimates parameters of the production frontier and inefficiency equation given by equations (11) and (12) using data from ERHS rounds 1994, 1999, 2004, and 2009. The second set estimates these equations by grouping different rounds together. In particular, three specifications are included in this group. The first equation includes ERHS rounds of 1995 and 1997 in addition to those included in the baseline. The second uses data only from ERHS 2009. The third group of equations uses three panel data formed by pooling together ERHS conducted in 1994 and 1999, in 1999 and 2004, and in 2004 and 2009. The purpose of forming panel data that pertain to certain years is to see if and how sources of inefficiency and output growth have evolved over time. The third set of analyses estimate equations (11) and (12) but now averaging among households grouped by agroecological zone, which mainly is climate and altitude dependent, and by a broader category that classifies FAs into high and low potential regions, which categorizes FAs based on their potential of crop production. This was done to determine if aggregate production frontiers exist with agro-climatic zone and production potential-specific differences, and as a basis for investigating the policy implications of such differences. The analyses in the fourth set redeploy the household level data using two alternative specifications. The first one uses per-hectare values of factors of production and the second one uses the baseline data, but by dropping those that did not apply fertilizer at one period or another. Results of the baseline specification are discussed in section 4.1.1 while the rest are discussed in sections 4.1.2 and 4.1.3. We briefly discuss some of the caveats associated with this particular analysis at the end of section 4.1.3.

[13] The software package Frontier 4.1 was used for the analysis. The parameters are estimated in a three-step procedure. First OLS estimates of the frontier are calculated. These estimates are unbiased except for the intercept term. Then a two-phased grid search of γ is conducted with the β parameters set to the OLS estimates obtained in the first step. In addition, the intercept and σ^2 are adjusted using a corrected ordinary least squares method, and δ parameters are set to zero [Recall that on page 9 we defined σ^2 as the sum of the variances of the inefficiency and random components of the error term.]. The third step involves using the values selected from the grid search as starting values in a Davidson-Fletcher-Powell Quasi-Newton iterative procedure to obtain the final maximum likelihood

4. Results and discussion

4.1 Parameter estimates of the stochastic production frontier

4.1.1 *The baseline model*

All parameter estimates of the inputs included in the production frontier given by equation (4.5) have the expected sign and all are significantly different from zero at 1 percent. Out of the time and agroecological zone dummy variables, the time dummy representing 2009 is not significant. An important implication gleaned from the relative magnitudes of the coefficient estimates in Table 4.1 is that most of the measured increase in output was attained by increased use of traditional inputs. The elasticity results show that the value of output is highly elastic for changes in the size and quality of cultivated land per household, amount of rainfall received in the region, for changes in the numbers of hoes and oxen used for cultivation, and for changes in labor use. While the calculated elasticity of value of output for changes in the rate of fertilizer application is among the lowest, the estimated coefficient as well as the calculated elasticity associated with participation in the extension program is one of the largest. The fact that modern inputs (such as fertilizer) on average contribute little for increased output shows the extent to which agriculture among the surveyed subsistence farmers relies on such traditional factors as size of cultivated land, amount of rainfall, and numbers of implements, and this explains why crop production in Ethiopia is sensitive to changes in the level of use of these inputs.

The elasticity of value of output with respect to the number of household members 16 years and older, which is used as a proxy for labor, probably the most abundant resource in rural Ethiopia, is the fifth largest and its magnitude is significantly lower than the elasticities for the relatively scarce inputs, like land and amount of rainfall, consistent with expected decline in marginal value product with increased level of use of an input.

The number of hoes, ploughs, and oxen used for farming were included in the analysis to represent each household's capital stock. While the calculated elasticity of value of output with respect to hoes was the fourth largest at 0.106, the elasticity with respect to ploughs and ploughing oxen are significantly lower, less than a third of the elasticity with respect to hoes. With the exception of 2009, which saw a large increase in capital stock, the remaining rounds were marked with little availability of these inputs, resulting in little contribution to value of output by the inputs.

Parameter estimates of size and quality of cultivated land are respectively the first and third largest. This result is indicative of the crucial role traditional inputs play for increases in output in such subsistence agriculture. However, viewed together with the fast population growth in Ethiopia, which averaged 2.85 during the survey period and most of which happened in rural areas (WB 2008), with a relatively large young rural population, and with limited amount of land that can be brought under cultivation, the result also implies that future growth in output from such factors is unsustainable.[14]

[14] Population growth data comprises years 1994 through 2007.

Table 4.1 Maximum likelihood estimates of parameters in the stochastic production frontier analysis

Variable	Estimated coefficient	t-ratio	Calculated elasticity
Constant	5.056	12.795	
Area of cultivated land	0.409	21.088	0.409
Household members 16 years and older (labor)	0.100	5.391	0.100
Amount of rainfall 12 months before the survey	0.321	5.359	0.321
Amount of fertilizer used	0.002	10.375	0.074
Number of ploughing oxen	0.031[a]	1.733	0.035
Average land quality	-0.109	-9.236	-0.245
Number of hoes used	0.074	5.955	0.106
Number of ploughs used	0.019	2.780	0.034
Participated in new extension program	0.116	2.806	0.123
Central highlands	0.408	7.835	0.503
Arussi/Bale	0.607	10.083	0.836
Hararghe	0.860	13.536	1.364
Enset	0.719	11.837	1.053
1999 dummy	0.228	4.771	0.256
2004 dummy	0.193	4.419	0.213
2009 dummy	-0.047 [b]	-0.901	-0.046

Notes: 1. The dependent variable is household level real value of output.
2. All parameter estimates are significant at 1 percent of level of significance except those with superscript [a] and [b], which are significant at 10 percent and not significant, respectively.

At 0.074, the elasticity of real value of output for increased application of fertilizer is one of the lowest. However, fertilizer application rates have to be accompanied with sufficient use of complimentary inputs such as water and most importantly, with improved seeds to achieve the desired results. Nevertheless, the data indicate that although the rate of fertilizer application is positively correlated with average annual rainfall, the correlation is weak at 3 percent (Appendix 1). This justifies the argument that encourages a synchronized and increased application of modern inputs such as fertilizer with other complimentary inputs such as wider application of improved seeds and irrigation. However, this result is an underestimation of the contribution of fertilizer to output growth of those that actually apply fertilizer because the estimated coefficient as well as the computed elasticity considers the mean value of fertilizer application rates of farmers that do and do not apply fertilizers. In an effort to see the effect of fertilizer application rates only on those that apply fertilizer we estimated the production frontier and inefficiency equations excluding households that do not apply fertilizer. The analysis was conducted on 925 of the farmers that applied fertilizer at one or more of the survey rounds; the total number of data points was 2,142, about 43 percent of the aggregate, 5,033. The estimated coefficients are different from the baseline estimates in three respects (Appendix 4, columns 2 and 3). First of all, although the magnitudes of the parameter estimates are similar to the baseline estimates, the elasticity with respect to rainfall has the wrong sign. Secondly, the estimated coefficients of number of ploughing oxen and participation in the extension program are insignificant. Moreover, the estimated elasticity of value of output with respect to rate of fertilizer application is the second largest, next to elasticity with respect to area cultivated.

Along with the rate of fertilizer application, participation in the extension package by the farm household is included in the analysis to represent the extent of modern input adoption. At calculated elasticity of 0.12, utilization of modern inputs and methods of cultivation introduced through the extension package made the fourth largest contribution, next to quality of cultivated land. At only 9 percent during 1994–2004 and 28 percent in 2009, the proportion of farmers utilizing the extension package leaves much to be desired.[15] However, the fact that applying the extension package made such large contributions encourages efforts towards wider participation of farm households.

Although the amount of precipitation received is beyond farmers' control it plays a crucial role in determining the magnitude of total output and agricultural production in Ethiopia is sensitive to the amount and distribution of rainfall. The elasticity of value of output for changes in the amount of rainfall received 12 months prior to the survey is the second largest, underlining the crucial role that rainfall plays in Ethiopian agriculture. The extent of rainfall's importance is manifested also through the frequent famines that occur in the country during years of low rainfall.

Another important aspect of this part of the estimated model is the implications of the parameter estimates associated with the dummy variables on time and agroecological zones. Bear in mind that these dummy variables are specified so as to compare the production frontier of the Northern highlands agroecological zone in 1994 with frontiers of other agroecological zones at different time periods, and with its own frontier in other time periods. For instance, the production frontier of Central highlands is higher than the frontier of Northern highlands by 0.601, indicating that the production frontier of the Central highlands agroecological zone in 2004 is relatively superior to the frontier that Northern highlands had in 1994. [16] Estimated coefficients of the dummy variables representing 1999 and 2004 are positive and significant, implying that households in each agroecological zone had production frontiers about 25 and 21 percent superior to 1994, respectively, while the estimate representing 2009 is insignificant implying that the frontier in 2009 was not different from what it was in 1994.

To understand the distinct frontiers faced by different agroecological zones and time periods we need to look into the inherent features of each of the villages relative to the ones in the Northern highlands agroecological zone. Socio-economic studies conducted on the 15 peasant associations included in the first round of the surveys describe Geblen, one of the three villages in the Northern highlands agroecological zone, as a region where rainfall is erratic and inadequate, soils are eroded, and poverty level is among the highest (Gebre Egziabher and Tegegne 1996). The other two villages, Harresaw and Shumsheha, face similar climatic conditions and difficulties. Therefore, this agroecological region is the poorest in terms of agricultural resource endowment, resulting in its poor performance relative to other agroecological zones.

The time dummy variables are meant to capture the Hicksian neutral technological change that occurred during the 1994 to 2009 period. Parameter estimates of the coefficients of these dummy variables imply that there have been technical improvements among Ethiopian farmers during the 1994–2004 period, while productivity levels in 2009 were the same as in 1994, which means productivity levels in 2009 were lower than those achieved in 2004. The fact that a significant proportion of farmers reported to have experienced drought in 2009 may partly explain the relatively lower productivity in 2009. More importantly, while cultivated area increased by an average annual rate of 9.2 percent, fertilizer use of an average farmer

[15] Recall that in the 2009 ERHS participation in the extension package was approximated by the number of visits made by extension agents to farm households and as such it overestimates participation.
[16] To arrive at this result we need to insert a value of 1 for the 2004 and Central highlands dummies.

increased more than 2 fold, rainfall increased by an average annual rate of 1.5 percent, and the rest of the inputs, including numbers of oxen, hoes, and ploughs, and average land quality increased between 2004 and 2009. Total value of output increased at an average annual rate of only 2.5 percent during the same period. That is, even if value of output was the largest during this year the magnitude of the inputs used to attain that value is even larger, resulting in inferior performance in terms of productivity.[17] In any case, further investigation into the issue using nationally representative data is imperative. Unless such investigations reveal otherwise, the decline in productivity observed between 2004 and 2009, implies that future increases in output cannot be sustained only by intensified use of inputs that was witnessed between these two years but by changing the structure of production or by an intervention that basically changes the techniques of production that currently exist.

To investigate the type of returns to scale that existed among the surveyed farmers we tested the null hypothesis of constant returns to scale against the alternative hypothesis that the production function does not exhibit constant returns to scale. That is, the null hypothesis $H_0 : \beta_1 + \beta_2 + ... + \beta_9 = 1$ was tested. The test concludes that the data do not contradict the hypothesis of constant returns to scale. This result is important to farm households and policy makers alike as it implies expansion in production is achievable by intensive application of inputs. However, the estimated coefficient of the 2009 dummy, taken together with the increased application of inputs in 2009, contradict this conclusion, again calling for a closer look at the issue.

In conclusion, there are three important observations that can be deduced from this part of the analysis. First, the majority of output increases among subsistence farmers stem from increased use of conventional inputs because most farmers do not use modern inputs, including 57 percent of the cases where fertilizer was not applied. Second, in the long run, increased output levels can be realized only by increased application of modern inputs, as decreasing marginal returns to conventional inputs will set in given fixed land size, as is already being witnessed for labor. Third, current low levels of contributions of modern inputs towards increased output can be improved by a synchronized application of modern and conventional inputs, as implied by the correlation of such inputs and by the higher elasticity for application of the extension package relative to the low elasticity for fertilizer application. Policymakers can use these observations to prioritize the support they provide in a timely manner. In the short run, seeking ways to increase farmers' entitlements of such traditional inputs as land and oxen, and the construction of small-scale irrigation schemes and water wells, will help increase farm output. Such efforts should simultaneously be undertaken together with efforts that have longer-term effects. This includes, but is not limited to, improving the availability of expanded extension services, and of such social infrastructure as educational institutions, better roads, health facilities, and more importantly, large-scale irrigation schemes that can reduce the output shocks that place farmers at risk during seasons of low rainfall.

4.1.2 *Alternative specifications: Evolution of sources of output growth*

In order to investigate if sources of output growth have evolved in some discernable pattern, we estimated the stochastic production frontier by grouping different rounds together and by assigning dummy variables for different years. Appendix 2 lists coefficient estimates of six such exercises that essentially are estimated to investigate four issues. The first and second

[17] We also checked if the loss in productivity by 2009 was due to decreasing returns to scale. The test statistic that used parameter estimates and design matrix of the 2009 data is almost zero, implying that the data does not contradict the hypothesis of constant returns to scale.

specifications, which respectively assign a dummy variable for 1994 instead of 2009 and include rounds from 1994, 1995, 1997, 1999, 2004, and 2009 (2nd through 5th columns), are estimated to both check the robustness of the baseline model and to examine if the same production frontier for 2009, as in 1994 implied by the baseline model, holds in other specifications. The 3rd specification is estimated to investigate if the production frontier that uses data from 2009, which were collected during a period of fast economic and food price growth and relatively fast increase in agricultural productivity implied by CSA data, is structurally any different from the remaining rounds. Specifications 4 through 6 (8th through 13th columns) divide the data into 3 intersecting periods of five-years and are estimated to study how the production frontiers, the importance of the inputs in output growth, and production efficiency evolved through the years, in addition to investigating if agricultural production underwent any structural change. We also discuss determinants and patterns of efficiency implied by these specifications in sections 4.3 and 4.4.

Parameter estimates of the specification that attached a time dummy for 1994 instead of 2009 are given in column 2 of Appendix 2. Except for the time dummies most of the estimates are either identical or similar with coefficient estimates of the baseline model. The estimated coefficients of the time dummies compare each year's performance with level of productivity in 2009, which is normalized to zero. The time dummies on 1999 and 2004 are positive and significant and the one on 1994 is insignificant. Taken together, the sign and magnitude of the time dummies have the same implication as those in the baseline model, that 1999 had the highest levels of productivity followed by 2004, while 1994 and 2009 performed inferior and were indistinguishable productivity-wise. This same conclusion can also be derived from the 2nd specification, which uses data from 1995 and 1997, in addition to those included in the baseline model. As can be gathered by comparing column 4 of Appendix 2 and column 2 of Table 4.1, the estimated coefficients of the 2nd specification have the same implication in terms of importance of inputs for increases in output and productivity performance across years, which together with the 1st specification confirm the robustness of the baseline specification. This specification also implies that technical improvements during the 1994–1995 period were the strongest.

There are subtle differences in the implication of the magnitudes of the elasticities, though, which can also be enhanced by including the 3rd specification that comprises only 2009 data into the discussion. Take, for instance, the elasticities with respect to land, labor, and water, the magnitudes of which support our previous conclusion that increased use of these inputs contributed the most for increased value of output. Including estimated coefficients and calculated elasticities of the remaining inputs also adds other dimensions. Firstly, periods that saw increased use of scarce traditional inputs such as land, rainfall, and oxen were followed by a larger contribution of the inputs towards total value of output while this was not always the case for others such as hoes and ploughs. Perhaps this was because no more ploughs than a pair of oxen can pull or hoes that can be used by available labor are needed while there was a large increase in ownership of implements in 2009, which had the lowest estimated coefficients and calculated elasticities for these inputs. The possibility that households can own but not use some input may also be an addition to the list of explanations as to why productivity in 2009 was lower than in 2004. Calculated elasticities for these implements were the largest for the baseline model, which had mean values midway between the 2nd and 3rd specifications. Secondly, calculated elasticities of modern inputs, fertilizer application, and participation in the extension package, had generally increased with growth in application.[18] Average land quality increased throughout the period, contributing positively to increases in output.

[18] As was noted earlier, care is needed when interpreting estimates that include participation in the extension package in 2009, because this variable was approximated by number of visits to a household by extension agents, which we believe overestimated participation, leading to less precise coefficient estimates.

Elasticities of the specification that uses data from only the 2009 ERHS are different from the baseline specification in other respects. Although the mean size of cultivated land in 2009 was 35 percent larger than the baseline case, the elasticity was even lower. Mean labor availability was lower by 4 percent and precipitation larger by 6 percent in 2009 than the baseline scenario, while the elasticities were lower by about 22 and 3 percent, respectively, relative to the baseline case. Perhaps the only consistent change is the increase in fertilizer application by 49 percent in 2009 relative to the baseline scenario, which was accompanied with an elasticity that is 26 percent larger than the baseline case. Added with previous observations about implements, one may speculate that the estimates from different specifications may be implying that, although on average or in a static sense the current production technology has constant returns to scale, it may be leading towards decreasing returns to scale dynamically.

Comparing estimated coefficients of the last three specifications, each of which represent panel data spanning 5-years, with each other, with the baseline model, and with the specification that used only 2009 data, together with the descriptive statistics of each of the specifications provides added information regarding the evolution of the importance of inputs for increases in output, patterns of productivity, and efficiency. The following observations were made from such a comparison. First, mean values of input uses in the 1999–2004 panel were lower than their counterpart both in 1994–1999 (except hoes, ploughs, rainfall, and participation in the extension package) and in 2004–2009 (except for ploughing oxen and labor) panel data. However, since the proportion in which the inputs declined varied, calculated elasticities varied accordingly. Those with relatively large decline have low elasticity values, although the elasticity with respect to land increased, in spite of decline in average land holdings during this period.

Secondly, percentage changes in fertilizer application elasticities are large due to their small magnitude, which cause large swings in calculated elasticities. The general trend observed from results reported here is that average contribution of fertilizer to increased output grew during periods of large fertilizer application. Elasticities of annual rainfall imply that the contribution of rainfall seems to have climaxed between 1999 and 2004. This trend to have a maximum or a relatively fast growth at or before the mid-five-year period seems to be a common phenomenon. Although these specifications use panel data collected at five-year intervals and it seems reasonable to assume they represent different economic policy, price and marketing, and possibly different climatic settings, we refrain from making the conclusion in broader terms as it requires investigation of a larger number of cross section as well as short-spanned panels to generalize the pattern. Lastly, time dummies of the first specification are positive and significant, indicating that 1999 witnessed significant improvements in productivity over 1994. The time dummy of 2004 is insignificant in the second specification, implying little change in productivity during 1999–2004, while the time dummy representing 2009 was negative and significant, which, added with the first two, implies that 2009 was indeed inferior in terms of productivity relative to 2004. Although the outcome from the three five-year panels was also inferred from the baseline specification, indicating the robustness of the results, it also implies that further investigation into causes of such productivity decline is imperative if one is to come up with meaningful policy implications that will help introduce consistent improvements in productivity.

4.1.3 *Alternative specifications: Agroecological zone and crop production potential based frontiers*

Two groups of estimations are discussed in this section. The first groups of estimates use panel data pooled across agroecological zones (AEZs) while the second groups use two data sets pooled across production potential based differences. Both groups are necessitated to investigate three common questions. The first is to check if the agroecological zone differences indicated by the dummy variables imply more than just a shift in the frontier. Secondly and perhaps more importantly, it helps to investigate if different traditional inputs played different roles for increased production; this consequently, may have different policy implications. Thirdly, we will use results from these specifications in section 4.2 to investigate if the low average efficiency from the pooled data resulted from heterogeneity in types of crops grown in regions that vary across agroecology and production potential and to study the factors that explain levels of efficiency in each region. Estimated coefficients of stochastic frontiers for the 5 agroecological zones are attached in Appendix 3 while estimates of the 2 frontiers that use data pooled across production potential are attached in the last four columns of Appendix 4.

Generally, the estimated parameters have similar implications with the baseline estimates with one qualification. The similarity is that most of the output increases during the study period are obtained by increased use of traditional inputs, while the qualification is that different inputs played the key role in different regions. The importance of cultivated area is even more pronounced across all AEZs, while labor contributed the most in Hararghe and the least in Arussi/Bale. Ploughing oxen were more important relative to the baseline case in four of the AEZs where the estimates were significant. Although the extent varied across AEZs, hoes were less important in the separate estimates relative to the baseline case, while the estimates on ploughs were always insignificant. In AEZs where the parameter estimate on fertilizer application was significant, it, on average, was close to the baseline case. In the three AEZs where the estimated coefficient of participation in the extension package was significant one had larger, another smaller, while the third was the same as the baseline scenario. While average rainfall is an important factor in this rain-fed agriculture, the fact that the two relatively dry regions of Northern highlands and Hararghe have elasticities of 3.9 and 3.6 for this variable is indicative of the importance of this input for increasing output in such regions.[19]

Roughly, the Enset growing agroecological zone has parameter estimates closer to the national average, while the rest can be roughly categorized into three groups. Regions such as Northern highlands which have acute shortage of rainfall, receiving 59 percent of the overall average precipitation over the years; those that have high land productivity and could use expansion or intensification in land use, such as Arussi/Bale; and those that could use both, such as Hararghe. From the difference in estimates of the separate production frontiers one can conclude that each AEZ has different factors as its major sources of output growth,

[19] To obtain more variability across agroecological zones we used annual rainfall in levels rather than in logs while estimating agroecological based frontiers. The calculated elasticities of value of output with respect to rainfall for Northern highlands, Central highlands, Arussi/Bale, Hararghe, and Enset are 3.9, -0.14, -4.7, 3.6, and 0.76, respectively. While the estimate for Central highlands is not significant, it is hard to defend the negative coefficient of Arussi/Bale. Bachewe (2009) reported elasticity of rainfall larger than 1 for all agroecological zones except for Central highlands. Perhaps, an important factor to consider is rainfall variability; data which we do not have for the 1994–2004 period. A case in point is Aze Deboa, an FA in Enset agroecological zone, which received the fourth best rain in the country in 2009 but in which 99 percent of the farmers reportedly experienced drought during the 2008/9 meher season. This could also partly explain the decline in productivity between 2004 and 2009. That is, while per hectare factor use levels have in general increased yield has not increased by as much. Rain shortfall during a critical season could be a factor that could explain part of the low growth in yield.

is faced with different constraints, and consequently calls for different policies to improve production and productivity.

Appropriately created FA dummy variables, all of which except one are significant, indicate that even within each AEZ, there are significant variations in the production frontiers faced by the FAs in a given AEZ. Each of the five agroecological zones reflect different patterns of productivity over the years based on estimates of the time dummy variables of the separate frontiers, particularly considering their performance in 2009 relative to the remaining years. Productivity in 2009 in Northern highlands agroecological zone was superior relative to 1994, 1999, and 2004, while 1999 and 2004 saw improvements relative to 1994. Productivity of farmers in Central highlands was superior in 2009 relative to 1994, but inferior to 1999 and 2004. Arussi/Bale AEZ experienced increased productivity in 2009 relative to 1994 and 1999, while performance in 2009 was inferior to 2004. Hararghe experienced inferior performance in productivity in 1999, 2004, and 2009 relative to 1994, while performance in 2009 was superior to the other two. The Enset AEZ experienced the pattern in the baseline case. Given that productivity in 1999 and 2004 were superior relative to 1994 in all AEZs, with the exception of Hararghe, added with the superior performance in 1994 relative to 2009 in 2 AEZs, superior performance in 2009 relative to 1994 in the other 2, and similar levels in the last one, seems to be consistent with similarity in levels of productivity in 1994 and 2009 that was implied by the baseline specification.

Production frontiers that group farm households along production potential differences have implications similar with AEZ based and baseline estimates with the exception that annual rainfall has the wrong sign in high potential areas, which on average received 10 and 5 percent more precipitation than in low potential areas and the baseline scenario. The most important factor affecting output is area of cultivated land. By contrast in low production potential areas the elasticity with respect to precipitation is more than 2 times larger than the baseline scenario. Moreover, the elasticity with respect to cultivated land is 3 percent larger in low potential than in high potential areas. More generally, the elasticity of value of output is higher in low potential areas for all inputs except for participation in the extension package and number of hoes. This finding is in line with the fact that the classification is resource/potential based and that low potential regions could use increases in almost any input. Agroecological zone dummies indicate that different AEZs in each of the high and low potential regions are faced with different production frontiers. Time dummy variables of both areas are similar with the baseline scenario except that productivity levels of low potential areas in 1999 were not significantly different from 1994.

To conclude, estimates that grouped farmers across agroecological zones and production potential have two implications. First, they ascertain that each category has a unique production frontier that is generally structurally different. The second implication, which results from the first, is that long-term efforts to improve agricultural production and productivity of the respective group should take into account long-term production potential and constraint faced by each category.

4.2 Determinants of efficiency

4.2.1 *Baseline model*

Results of the baseline estimation conducted to get insights into what explains efficiency or lack of it among subsistence farmers in Ethiopia are presented in Table 4.2. All of the parameter estimates are significant and all except one have the expected sign. Estimated coefficients of time-agroecological zone interaction dummies are listed in Table 4.3.

Age of head of the household is included in the inefficiency equation to examine the effect of experience and physical strength on efficiency. The positive sign of the estimated value of this variable supports the argument that farmers become less efficient as they get older. This could result not only from efficiency loss as farmers get old but also because younger farmers tend to be more open and likely to be exposed to methods and techniques that cannot be captured by variables included in the analysis. The education variable is included in the inefficiency equation to see if it contributes positively towards efficiency, as it is believed that education increases human capital and contributes positively to change farmers' attitudes towards modern technology. The results show that farmers that are literate are more efficient. While the direct effect of education on efficiency is captured by the education variable included in the analysis, age, simultaneously with level of education, may be capturing the indirect effect of education such as better administration skills. The data indicate that younger farmers tend to be better educated, since age and level of education have significant negative correlation. Given that various specifications consistently result in a positive contribution of education, it could be used as a basis for a policy favoring expanding education in rural areas and to encourage parents to send their children to school. Currently, the Ethiopian government is working in both areas.

Sex of the head of the household was included to examine if gender has any bearing on efficiency; the parameter estimate of this variable is negative. However, it is arguable that the negative sign of the estimated coefficient implies that female headed households are less efficient per se than male headed households. Rather it may imply something that is inherent in the family system of rural Ethiopia. Females become head of a household only when males are deceased or not around, therefore when females are head of the household they take on farming in addition to their traditional homemaker role. In male-headed households females participate in agriculture especially in removing weeds, in addition to homemaking. The parameter estimates are, therefore, most probably the implications of scarcity of labor in female-headed households and the reduced attention they could afford to allocate for farming, as they also have to take care of the household. Moreover, such households may have poorer access to the timely application of inputs. The data indicate that on average female headed households cultivated 24 percent smaller area, applied 34 percent less fertilizer, had 26 percent less ploughing oxen, owned 30 percent fewer implements relative to male-headed households, and only 68 applied the extension package per 100 male-headed households that applied the package.

The claim above is also supported by the significant negative correlation of the female dummy with both number of household members 16 years and older and family size of the household.[20] Low levels of farming experience could also have a negative effect on efficiency as females start farming after males are deceased. Associated with this, the estimated coefficient on the female dummy variable is positive and the fourth largest. This indicates that, holding other factors constant, households with one or more male members

[20] Recall that the female dummy variable assigns a value of 1 for households that do not have male members 16 years and older.

are more efficient than households with all-female working members. The justification given above for the estimated coefficient on sex of head of the household applies here too. We also included family size in the inefficiency equation to determine if it plays a role in affecting efficiency. Households with fewer members are less efficient compared with larger sized households. In the absence of well-functioning labor markets, larger households face fewer labor bottlenecks at critical points in the farming cycle such as land preparation and harvest.

The interaction variable created by multiplying the number of plots that farmers cultivate with the logarithm of size of cultivated land is included in the analysis to assess the effect on farming efficiency of dissected plots for a given size of cultivated land. The negative coefficient on this parameter implies that for a given number of plots, cultivating larger plots reduces inefficiency. The sign on this coefficient may also represent the reduced risk that different plots provide if the plots are located sufficiently disbursed, such that farmers face different degrees of weather-induced variation and mineral content on the different plots.

Table 4.2. Maximum likelihood estimates of the inefficiency function parameters

Variable	Coefficient	t-ratio
Constant	-36.550	-20.45
Sex	-2.488	-4.74
Age	0.042	4.92
Level of education	-1.418	-2.86
Female dummy	2.595	3.95
Household size	-0.136	-2.82
Number of plots* log of cultivated area	-0.207	-6.32
Cultivated area/ number of members 16 years and older	-0.004[a]	-1.93
Oxen dummy	-1.670	-3.50
Livestock units	-0.346	-17.20
Number of agricultural extension agents in FA	2.852	11.77
Crop affected by drought	6.327	17.17
Survey month	9.413	12.88
Elevation	0.002	14.18
Distance to health center	0.080	9.67
Distance to closest market	-0.152	-14.71
Distance to nearest FA center	0.101	15.08
Sigma-squared	25	37
Gamma	0.99	1793
Average farming efficiency	0.457	
Log likelihood	-7708	

Note: All parameter estimates are significant at 1 percent except the one with superscript a, which is significant at 10 percent.

Another interaction variable of the ratio of the size of cultivated land over the number of household members 16 years and older was used in the inefficiency equation to investigate the claim that congested agricultural land holdings adversely affect efficiency and to quantify the magnitude of this effect. The result implies that for a given amount of labor, increase in the size of cultivated land leads to increased efficiency, which is to say that households that have little land per household member of ages 16 and older are less efficient, although the magnitude of the adverse effect is small.

The estimated coefficient on oxen dummy implies that owning two or more ploughing oxen substantially reduces inefficiency. Oxen are the major source of traction power in Ethiopia. Typically, two oxen are needed to pull a plough, which puts households that own one or no ox at a relative disadvantage. Traditionally, farmers with one ox use the arrangement that is known as 'Mekenago', whereby farmers borrow or exchange oxen in return for ploughing services provided by their own animals. Although we do not have data on oxen rentals, it is common practice for farmers to rent animals for ploughing. Therefore, the estimated coefficient of this dummy variable indicates that farmers that do not have at least two oxen suffer from insufficient ploughing power. The variable used as proxy for wealth, livestock unit, is significant and has the expected sign. Farmers with more livestock units, which can readily be converted to money, can be able to buy modern inputs that were not included in the list of measured inputs (such as pesticides and improved seeds), than those that own fewer livestock units. Moreover, families with more animals are more likely to have larger protein intake than those with fewer animals, which helps improve their working efficiency.

The analysis indicates that farmers residing in FA with larger numbers of agricultural extension agents are less efficient. The positive sign on this variable is not expected as extension agents serve as a bridge in transferring government's efforts to modernize agriculture.[21] The specification that included data from the 1995 and 1997 rounds, in addition to those used in the baseline model, has a negative sign while other specifications have mixed signs, which implies further investigation is needed to explain how the number of extension agents adversely affects efficiency. Added with the fact that the number of agricultural agents is uncorrelated with rate of fertilizer application and participation in the extension package (Appendix 1), one plausible explanation could be that FAs with fewer extension agents have been provided the services sufficiently that they do not need as many agents as those that are less efficient, but the fact that on average fewer than one agent existed makes this explanation suspect.

The dummy variable on drought is included to control for one of the factors that affect farming efficiency but is beyond farmers' control. On average, farmers that did not suffer from drought earned 96 and 128 percent larger value of output and value of output per hectare, respectively, than those that went through a dry season. This result, together with the parameter estimate on mean annual rainfall included in the production frontier, indicates the extent to which agriculture among the surveyed subsistence farmers relies on rainfall and explains why crop production in Ethiopia is sensitive to variation in the amount of rainfall. The magnitude of the estimate and the fact that those that did not suffer from drought were relatively more efficient implies that government and concerned agencies should pay significant attention to alleviate problems associated with water scarcity through such measures as construction of small-scale irrigation schemes and water wells.

The estimated coefficient of the dummy variable that takes a value of 1 if a household was surveyed in the months of August through January indicates that farmers surveyed in the months farther away from the harvest period significantly overstate the amount of output they produced or understate the amount of inputs they used. This variable is included in the analysis to control for the recall or measurement error that could arise from delayed interviewing. The magnitude of the estimated coefficient implies that such errors were of considerable magnitude.

[21] Farmers' association level data on the number of agricultural extension agents from 2004 is used for 2009, as we lack this data for 2009 and dropping the variable altogether was thought to be wasting important information that it provided for the remaining rounds. Moreover, since it is more likely that the number of agents has increased between 2004 and 2009, the coefficient estimate can be viewed as a lower bound for the positive contribution that agricultural extension agents had towards increased efficiency.

Three variables on distances that households had to travel to centers where they access essential services were included in the analysis to investigate the effect of the availability and proximity of social infrastructure on farming efficiency. The estimated coefficient of the distance to the nearest health center implies that farmers that have to travel longer distances to access health services are less efficient, which is consistent with the argument made to incorporate this variable. The justification for arguing that longer distances to the nearest health centers affect efficiency adversely is that farmers that lived closer to health centers are not only healthier because they can visit the centers frequently but also because they save travel time to the centers that they otherwise could use to tend their farms. The estimated coefficient of the distance to markets indicates that farmers living farther away from markets are more efficient relative to those that are closer to markets. This could be because most of the villages that are farther away from markets are high potential areas as compared to those early-settled areas whose production potential may have been exhausted from years of cultivation and are now close to markets. So, this variable may be capturing the effect of other factors on efficiency rather than access to markets. The analysis indicates that farmers that lived closer to FA centers, where support from the Ministry of Agriculture is provided, were more efficient than those that were located farther away. Farmers that are located closer to FA centers can have easier access to services provided by the centers and extension agents can visit them frequently. The sign on this variable clearly contradicts what was implied by the coefficient estimate of number of extension agents. The estimated coefficient on elevation indicates that lowland farmers are marginally more efficient.

As discussed above, the relative value of the estimated coefficients of the time-zone interaction dummy variables in the inefficiency equation of two time periods are interpreted as efficiency gains or losses of farmers in a given agroecological zone at a given period relative to their performance in the other period, given the different production frontiers faced by the zone at each period. Table 4.3 summarizes estimates of parameters δ_{18} through δ_{36}. Seventeen of the 19 estimates are significant at 1 percent. The estimated coefficients imply that farmers in the Northern highlands were more efficient in 1999 and 2004 than in 1994. They were about as efficient in 2004 as they were in 1999, and they were relatively less efficient in 2009 than they were throughout 1994–2004. Farmers in Central highlands experienced efficiency gains between 1994 and 2004. Although efficiency levels in 2009 were inferior to 1999 and 2004, they were superior to 1994. While a similar comparison of time-agroecological zone dummies reveals different patterns across AEZs there is one similarity. Efficiency levels in 2009 were superior to 1994, although they were inferior to 2004 in all AEZs. Performance in 2009 relative to 1999 was inferior in Northern and Central highlands and Hararghe. To appreciate the relative values the reader is advised to comparatively study the sign and magnitude of these estimates with the discussion on average efficiency scores provided in section 4.3.

To check for the joint explanatory power of the variables in the inefficiency equation we estimated the model without these variables - that is, the hypothesis $H_0 : \delta_1 = \delta_2 = ... = \delta_{36} = 0$ was tested against the alternative hypothesis that these variables jointly explain inefficiency. The test statistic of this claim is calculated using the formula –2(log likelihood under H_0 - log likelihood under H_A). This statistic has a mixed χ^2 distribution with degrees of freedom equal to the number of restrictions, in this case 36. The null hypothesis was rejected as the test statistic is 6,984, implying that the data reject the null that these variables do not explain farming inefficiency. In the following section we will briefly discuss determinants of efficiency of specifications that grouped households across time, agroecological zone, and production potential criterion.

Table 4.3. Maximum likelihood estimates of time and agroecologal zone dummy variables included in the inefficiency equation

Agroecological zone	Variable	Year 1994	1999	2004	2009
Northern highlands	Coefficient	--	-26.82	-26.94	14.62
	t-ratio	--	-28.60	-15.23	17.04
Central highlands	Coefficient	2.83	-16.68	-18.29	-3.83
	t-ratio	3.21	-15.28	-19.72	-3.42
Arussi/Bale	Coefficient	5.84	3.26	-7.95	-1.63[a]
	t-ratio	6.04	3.43	-7.60	-1.49
Hararghe	Coefficient	1.05[a]	-16.12	-17.79	-14.90
	t-ratio	0.74	-10.77	-21.67	-16.97
Enset	Coefficient	14.35	8.00	-9.41	7.49
	t-ratio	17.43	10.48	-9.58	6.56

Note: All parameter estimates are significant at 1 percent except ones with superscripted a, which are not significant.

4.2.2 *Alternative specifications*

We used the separate production frontiers that used data in which households were grouped along different periods, agroecological zones, and production potential differences to investigate if period, agroecological zone, and production potential specific production frontiers existed. Similarly, estimates of the simultaneously estimated inefficiency equations are meant to investigate if different factors played different roles during different periods, agroecological zones, and areas with different production potentials. Parameter estimates of the inefficiency equation that grouped households across periods are provided in Appendix 5, estimated coefficients of the specifications that grouped households across AEZ classification are provided in Appendix 6. Appendix 7 comprises estimated coefficients of inefficiency equations of households in high and low potential FAs, of those households applying fertilizer, and the baseline specification converted into per hectare values. Since estimated coefficients of these specifications are mostly similar with the baseline model, implying similarity of factors affecting efficiency, we will discuss these results only briefly, emphasizing the differences.

The set of estimates for the specification that includes data from ERHS rounds of 1995 and 1997, in addition to those in the baseline, is different from the baseline estimates in two respects. In this specification, increased number of extension agents improves efficiency, while efficiency increases marginally with elevation. Estimated coefficients of the specification that uses data from only ERHS 2009 imply that unlike the baseline case efficiency increased among those farmers farther away from health centers and closer to market centers. With the exception that two of the estimates are insignificant, there is no difference between the baseline and the specification that uses ERHS 1994 and 1999 data. The specification that uses data from ERHS 1999 and 2004 is different in both interaction of number of plots and cultivated area and survey month, while the specification that uses data from ERHS 2004 and 2009 has different signs for survey month and for the three distance variables, although one is not significant.

Parameter estimates of the inefficiency equations that used both agroecology and production potential as a basis to form each group within each category are qualitatively similar with the baseline estimates for the most part, with some exceptions, and with quantitative variation. Among these differences, those notable are increases in efficiency with increased distance to health centers and decreased distance to market centers in Northern and Central highlands, and with increased distance to farmers' association centers in Arussi/Bale agroecological regions. Differences in magnitudes of estimated coefficients of various specifications imply differences in importance of different factors in determining efficiency levels in each group.

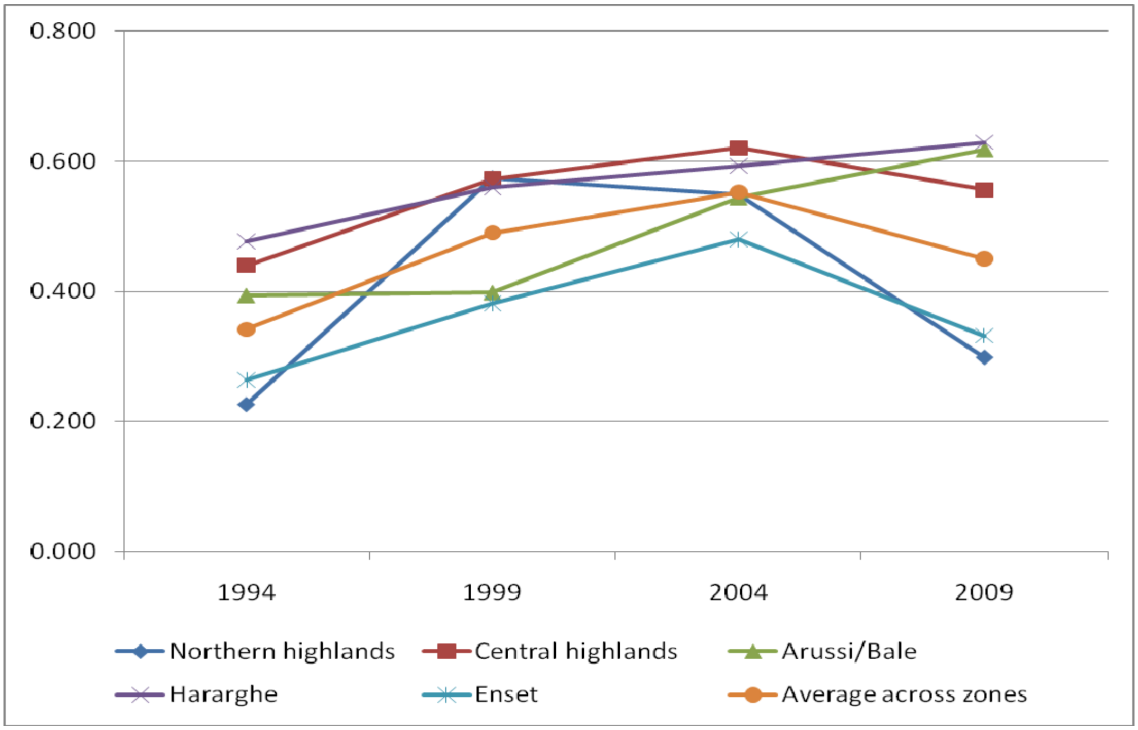
4.3 Trends in efficiency

4.3.1 *Baseline model*

Each farmer's degree of farming efficiency is calculated using the formula $TE_{it} = \exp(-U_{it})$, which is given by equation (2). The best farmer that lies on the frontier scores a value of 1 while the value gets closer to zero as farming efficiency falls. Since this way of measuring efficiency compares farmers with their own peers, it is justifiable to say that farmers can achieve the highest possible efficiency level if what constrains them is solved. Average farming efficiency of the surveyed farmers across the years included in the survey was 0.457, indicating that most farmers were less than one-half efficient relative to those operating at the frontier. In addition to the factors discussed above, stochastic factors implausible for control by farmers may have impinged on this substantial difference in efficiency.

Table 4.4 lists average efficiency levels of each agroecological zone and farmers' associations within each AEZ. In addition to the table, Figure 4.1 helps discern overall and agroecological zone level trends in efficiency. Average efficiency grew between 1994 and 1999 across all agroecological zones. Average annual growth rates ranged from 0.2 percent in Arussi/Bale to 31 percent in Northern highlands, while nationwide this number was about 10 percent. Although there was a slight decline in efficiency in Northern highlands between 1999 and 2004, mainly due to efficiency losses in Haresaw and Shumsheha FAs, gains in efficiency were recorded across the remaining AEZs. Average annual growth rates between 1999 and 2004 ranged between 1.2 percent in Hararghe to 7.3 percent in Arussi/Bale and the average across the sample was 2.9 percent. For most, efficiency declined between 2004 and 2009. North and Central highlands and Enset AEZs suffered average annual efficiency losses of 9.1, 2.1, and 6.2 percent, respectively, while Arussi/Bale and Hararghe gained in efficiency at average annual rate of 2.7 and 1.3 percent. As a result of large efficiency losses that outweighed gains, overall average efficiency declined by 2.7 percent annually between 2004 and 2009. Recall that this overall average trend in efficiency is accompanied with levels of productivity that increased between 1994 and 2004 and that declined between 2004 and 2009.

Figure 4.1. Trends in average efficiency scores across AEZs, 1994–2009



Despite the significant decline in efficiency in Northern highlands between 2004 and 2009, it had the fastest growth in efficiency during the period as average efficiency grew by an average annual rate of over 6.9 percent during the entire period, mainly due to the large efficiency gains between 1994 and 1999, particularly in Haresaw. The average efficiency scores of Arussi/Bale, Enset, Hararghe, and Central highlands augmented annually with 3.4, 2.6, 2.0 and 1.9 percent, respectively.

The time-zone dummy coefficient estimates, summarized in Table 4.3, can be read together with average efficiency estimates given in Table 4.4. Recall that estimated coefficients of time-AEZ dummies in Table 4.3 indicate how that period fared in terms of efficiency relative to another period in the same AEZ. For instance, comparing coefficient estimates of 1999 and 2009 of Northern highlands indicates that 1999, which, with a large negative value, came with a large decline in *inefficiency*, should naturally lead to superior performance of an average farmer, as is the case in almost all cases. But not all AEZ-time dummy and average performance in efficiency combinations may have a linear relationship, when, even if the year came with efficiency gains in an AEZ, performance by farmers may not measure up to the gains. However, only three such cases were observed in this sample: Northern highlands between 1999 and 2004, Arussi/Bale between 2004 and 2009, and Hararghe between 2004 and 2009. The strong and significant negative correlation of about –0.9 between average efficiency levels and the estimated agroecological zone dummy variables reflects that the inverse relationship between average efficiency levels and the parameter estimates of the time-agroecological zone dummy variables holds for other zones as well.

Table 4.4. Average efficiency estimates of farmers by agroecological zones and farmers' association

Agroecological zone/Farmers association[a]	1994	1999	2004	2009	Average across Years
Northern highlands	**0.226**	**0.574**	**0.548**	**0.298**	**0.405**
Haresaw	0.003	0.584	0.452	0.077	0.264
Geblen	0.214	0.468	0.555	0.102	0.332
Shumsheha	0.378	0.635	0.612	0.592	0.544
Central highlands	**0.440**	**0.573**	**0.620**	**0.556**	**0.547**
Dinki	0.238	0.420	0.541	0.485	0.424
Debre Berhan Milki	0.431	0.651	0.668	0.535	0.568
Debre Berhan Kormargefia	0.550	0.625	0.525	0.541	0.560
Debre Berhan Karafino	0.357	0.585	0.679	0.549	0.541
Debre Berhan Bokafia	0.531	0.665	0.578	0.536	0.577
Yetemen	0.411	0.566	0.679	0.577	0.560
Turufe ketchema	0.540	0.583	0.667	0.631	0.604
Arussi/Bale	**0.394**	**0.399**	**0.544**	**0.618**	**0.486**
Sirbana Godeti	0.690	0.552	0.531	0.579	0.592
Korodegaga	0.164	0.273	0.555	0.644	0.406
Hararghe	**0.477**	**0.560**	**0.593**	**0.629**	**0.564**
Adele Keke	0.477	0.560	0.593	0.629	0.564
Enset	**0.264**	**0.381**	**0.480**	**0.332**	**0.363**
Imdibir	0.266	0.220	0.503	0.321	0.332
Aze-Deboa	0.333	0.497	0.575	0.361	0.440
Adado	0.462	0.575	0.371	0.488	0.472
Gara-Godo	0.094	0.319	0.504	0.159	0.263
Do'oma	0.144	0.132	0.512	0.254	0.252
Average across zones	**0.341**	**0.490**	**0.553**	**0.451**	**0.457**

Source: Calculated from efficiency scores of farm households reported in the SFA analysis.
Note: Agroecological zones are in bold while FAs included in each zone are listed under each zone.

Two observations can be made about average trends in FA and AEZ level average efficiency scores in Table 4.4. First, farmers in each of the AEZs experienced a steady increase in efficiency between 1994 and 2004, with the exception of Northern highlands between 1999 and 2004. Most AEZs suffered efficiency losses between 2004 and 2009. If we were to disaggregate this into three five-year periods all AEZs, except Arussi/Bale, experienced the fastest average annual growth in the first five years of 1994–1999. The second five-year period of 1999–2004 was accompanied with the second fastest growth in all, except Arussi/Bale AEZ, which had its fastest growth during this period, and Northern highlands, which had a negative growth rate then. The third five-year period of 2004–2009 witnessed a decline in efficiency in Northern and Central highlands and Enset AEZs while growth rates in Arussi/Bale and Hararghe AEZs were the second fastest. Disaggregating the trends in performance across FA-time lines, except two FAs in the Enset and one FA in Arussi/Bale AEZ all FAs had efficiency gains between 1994 and 1999. Two FAs lost in average efficiency in each of Northern and Central highlands and one in each of Arussi/Bale and Enset between 1999 and 2004. During the 2004–2009 period, all FAs in Northern highlands, all but one in Central highlands and Enset AEZs had lower performance in 2004 compared to 2009, while all FAs in Arussi/Bale and Hararghe had efficiency gains.

The second observation, which is the result of the first added with the fact that some AEZs gained in efficiency between 2004 and 2009 while others lost, is that poorly performing AEZs and FAs at the beginning of the period were able to narrow the gap in performance relative to best performing AEZs and FAs between 1994 and 2004. AEZ level range in efficiency was 0.25, 0.19, and 0.14 in 1994, 1999, and 2004, respectively, while it was 0.33 in 2009. This was a decline in average efficiency gap of 4.6 and 5.7 percent between 1994 and 1999, and 1999 and 2004, but an increase in average annual efficiency gap of 27 percent between 2004 and 2009. FA level range in efficiency declined annually by about 4.5 percent from 0.69 in 1994 to 0.53 in 1999, and by 8.4 percent during the 1999–2004 period, to 0.31 in 2004, while it increased annually by 16 percent between 2004 and 2009 to get to 0.57 in 2009.

Explanations of divergence in efficiency levels are rooted mainly on the premise that they result from differences in farmers' ability to manage inputs at the given technology, availability and composition of labor, differences in risk faced, and wealth, among others. However, we also alluded to region specific factors such as availability of services and to such stochastic factors as elevation and occurrence of drought. We would like to concentrate on the last factor due to the all pervasive influence of rainfall and lack of it, manifested as drought, on output growth and efficiency, respectively. We believe that in addition to farmers' managerial skills and resource constraints that determine efficiency, observed trends in efficiency are affected significantly by availability of rainfall. The fact that the coefficient estimate of drought is one of the largest in the inefficiency equation and that the amount of precipitation is one of the largest contributors to output growth support this claim. Add to this the fact that drought free areas produced almost twice the value of output relative to drought-prone areas. Moreover, disaggregating the average efficiency performances across AEZ/FA-time lines provides support to the claim that stochastic factors, such as drought, play a sufficient role in explaining observed trends in efficiency, and are of important concern.

Added with the previous explanation meant to elaborate on both observations, improved farming efficiency observed during 1994–1999 may be the result of multiple factors. In 1995, the Ethiopian government launched what is called the National Extension Intensification Program (NEIP), with the intention of enhancing the availability of inputs and access to credit for over 32,000 half-hectare plots throughout the country, by adopting the methods that were originally introduced by Sasakawa-Global 2000 (SG2000 1995). In 1996 the NEIP expanded to 320,000 plots. During this period of relative peace and stability the government increased the budget share allocated to economic services and focused its attention on strengthening the economy. This increased focus on economic services was downgraded during the Ethiopia-Eritrea war, which became particularly problematic in 2000. On average the military spending to GDP ratio was about 8.5 percent during the 1998 to 2001 period, more than 3 times the average of the previous four-year period, which was 2.7 percent. This probably has resulted in reduced budget for agricultural services, or at least has shifted away the focus and reduced the momentum from agriculture, leading to lower levels of growth in production efficiency by 2004. Moreover, the NEIP has reduced its program while SG 2000 had altogether abandoned its extension program in 2000. In addition to these, proportionately more effort and resources are needed to improve upon the higher efficiency level achieved in 1999 (SG2000 2007).

The military build-up explanation provided for the relatively slow growth in average efficiency observed between 1999 and 2004 is less applicable for the relatively peaceful period of 2004 through 2009, which was accompanied with sample-wide decline in average efficiency, as it is also corroborated with the average military spending to GDP ratio of 3.3 percent. The plausible explanation for the decline in average efficiency between 2004 and 2009 can be the fact that per hectare real value of output grew by a lower rate relative to rate of growth of inputs utilized on each hectare. While per hectare average utilization rates of fertilizer, hoes, and ploughs grew by 20, 35, and 31 percents, respectively, and on average 11 percent more rain was received in 2009 than in 2004, and the average cultivated area grew by about 9.2 percent, the real value of output produced on each hectare actually declined by 5.4 percent, although average household level value of output increased annually by 2.5 percent. This indicates that although availability of inputs have expanded in 2009 relative to what they were in 2004 actual output levels have not expanded as they should, which is what time dummies and the relative efficiency estimates capture.

4.3.2 *Alternative specifications*

Efficiency estimates accompanying alternative specifications of production frontiers and inefficiency equations are meant to reveal if efficiency performance takes a specific pattern when farm households are compared with superior performing farm households in a relatively narrowly defined group. Indeed the results show that average efficiency improves when farmers are compared with others in the same agroecological zone and production potential as they most likely grow similar crops and have comparable production potential. Similarly, comparing farmers in a narrowly defined time gap will reduce the variation in efficiency as it removes differences that arise from variation in weather over time. Of course, there are exceptions to this general case when, for instance, a significant number of farmers in an AEZ perform worse in a narrowly defined group compared with superior performing farmers, especially if the best performing farmers of the entire sample resided in that AEZ.

Average efficiency scores of specification that use data aggregating farmers in different periods, production potentials, those applying fertilizer, and the baseline converted into per hectare value are given in Table 4.5. Similarly, average efficiency scores based on estimations that aggregate households across agroecological lines are provided in Table 4.6. Results of three estimations using the baseline data dropping the year dummy of 2009, data from six rounds of ERHS (1994, 1995, 1997, 1999, 2004, and 2009), and ERHS 2009 are not included in these tables. However, average efficiency scores of these estimations are listed in columns 2, 4 and 6 of Appendix 5, respectively. Since the only difference of the specification in which the 2009 dummy was dropped from the baseline is the arrangement of time dummies, it has exactly the same average efficiency score. Average efficiency performance of the panel data that adds data from 1995 and 1997 in addition to those included in the baseline is lower relative to the baseline average, as it comprises a larger group of farm households. Similarly, average efficiency performance of farmers in the estimation that uses only ERHS 2009 is larger than the baseline average, as efficiency performance of farmers is compared within a smaller group.

Average efficiency scores of farmers that used panel data constituting three five-year periods of 1994–1999, 1999–2004, and 2004–2009 are about 0.4, 0.530, and 0.534. There are two observations that can be discerned from the first panel of Table 4.5, which disaggregates these averages by year and agroecology. First, comparing the entries in Table 4.5 with those in Table 4.4, average efficiency of farmers in many AEZ-years is higher, supporting the general statement made above. Household number weighted overall average of the three estimates is 0.538, which is 18 percent better than the baseline scenario. Secondly, the same pattern of efficiency growth observed in the base line case is apparent from these estimates: efficiency grew the fastest between 1994 and 1999.

Table 4.5. Average efficiency scores of farmers from specifications using different criteria

Panel forming criteria	Period	Northern highlands	Central highlands	Arussi/ Bale	Hararghe	Enset	Average across AEZs
	Panel data formed from rounds that span 5 years						
1994 - 1999 panel	1994	.2088	.4517	.2558	.3513	.5404	.3920
	1999	.3669	.5091	.3776	.0612	.4344	.4097
	Average	**.2798**	**.4795**	**.3132**	**.2063**	**.4883**	**.4005**
1999 - 2004 panel	1999	.4711	.5383	.4426	.4369	.6015	.5242
	2004	.5118	.5879	.5583	.2161	.5602	.5355
	Average	**.4927**	**.5637**	**.4984**	**.3284**	**.5806**	**.5300**
2004 - 2009 panel	2004	.5226	.4132	.6180	.5147	.5836	.5187
	2009	.6083	.4195	.6189	.6095	.5857	.5490
	Average	**.5640**	**.4162**	**.6185**	**.5624**	**.5847**	**.5338**
	Panel data formed across production potential						
High production potential	1994		.5292	.4131		.5166	.4910
	1999		.5196	.4263		.5351	.4973
	2004		.5224	.4203		.5142	.4927
	2009		.5155	.4524		.5438	.5062
	Average		**.5219**	**.4280**		**.5277**	**.4968**
Low production potential	1994	.3760	.4767		.5908	.4267	.3793
	1999	.3689	.4995		.6186	.4496	.4845
	2004	.3782	.4847		.6241	.4215	.6110
	2009	.3943	.4762		.6112	.4222	.4299
	Average	**.4363**	**.4547**		**.6112**	**.4469**	**.4451**
	Panel data of those applying fertilizer and per hectare values of inputs and output						
Those applying fertilizer	1994	.5252	.6165	.5130	.5858	.5917	.5880
	1999	.4888	.5549	.3128	.7377	.6715	.5373
	2004	.5438	.6866	.4420	.7334	.7002	.6305
	2009	.7137	.6374	.6230	.7185	.7123	.6606
	Average	**.5716**	**.6177**	**.4842**	**.6981**	**.6631**	**.6027**
Per hectare specification	1994	.1658	.4347	.1422	.4670	.2562	.3327
	1999	.6126	.6150	.2213	.5918	.3749	.4846
	2004	.6172	.6036	.5234	.5783	.4800	.5529
	2009	.5798	.5672	.5986	.6222	.2835	.4390
	Average	**.4045**	**.5476**	**.4726**	**.5659**	**.3473**	**.4505**

Source: Calculated from efficiency scores of respective specifications.

As can be seen from the middle panel of Table 4.5 and Appendix 7 the first observation applies with some modification for efficiency scores of estimations that aggregate farmers across production potential. The modifications are that increases in average efficiency in these relatively homogenous groups are not uniform just as in the previous case and that most AEZ-year entries of low production potential areas are actually lower than their baseline counterparts, resulting in a lower average efficiency score of 0.445, which is even lower than the baseline average of 0.457 and the average in high potential areas of about 0.5. Still farmer number weighted average of the two estimates, 0.47, is larger than the baseline case. Moreover, comparing average efficiency performance of the two AEZs entirely composed of low potential FAs, Northern highlands and Hararghe, with their counterparts in the baseline case reveals improvement in both, while comparing the two AEZs comprised of both high and low potential FAs, Central highlands and Enset, reveals a decline in the first and an increase in the latter. The second observation applies here too, again with a modification. While the growth pattern in efficiency witnessed in the baseline case holds in low production potential areas, growth in efficiency in high potential areas was the fastest between 1999 and 2004, the 2nd fastest growth happened between 1994 and 1999, and although growth between 2004 and 2009 was slow, it was not negative.

In addition to having similar growth pattern as the baseline case, overall and AEZ-year average efficiency scores of farmers in the specification that used per hectare values are almost the same as the baseline case, up to two decimal places. The specification that used the panel data tailored to those applying fertilizers also satisfies the first observation. Overall efficiency averaged at 0.603, which is about 31 percent larger than the baseline case. This is an interesting result since it implies that, relative to an average farmer, most farmers that apply fertilizer not only gain in terms of increased output as a result but were able to use other inputs more efficiently.

Farmer association and AEZ level average farming efficiency of five AEZ based estimates are given in Table 4.6. Average efficiency improved when farmers are compared in a relatively smaller group and efficiency grew fast between 1994 and 1999, slowed down between 1999 and 2004, and mostly declined between 2004 and 2009.

Table 4.6. Average efficiency scores of farmers from specifications that used panel data formed across AEZ

Agroecological zone	Farmers association	1994	1999	2004	2009	Average across years
Northern highlands	Haresaw	.273	.270	.301	.347	**.298**
	Geblen	.617	.589	.601	.581	**.597**
	Shumsheha	.408	.357	.432	.386	**.398**
	Northern highlands average	**.416**	**.397**	**.432**	**.427**	**.418**
Central highlands	Dinki	.478	.484	.492	.474	**.482**
	Debre Berhan Milki	.555	.579	.523	.614	**.567**
	Debre Berhan Kormargefia	.539	.571	.587	.602	**.573**
	Debre Berhan Karafino	.566	.549	.553	.509	**.544**
	Debre Berhan Bokafia	.561	.512	.587	.535	**.550**
	Yetemen	.570	.627	.573	.608	**.595**
	Turufe ketchema	.619	.620	.602	.659	**.624**
	Central highlands average	**.559**	**.572**	**.558**	**.582**	**.567**
Arussi/ Bale	Sirbana Godeti	.574	.521	.549	.535	**.546**
	Korodegaga	.595	.616	.596	.575	**.595**
	Arussi/Bale Average	**.586**	**.573**	**.576**	**.559**	**.574**
Hararghe	**Adele Keke**	.473	.669	.549	.650	**.585**
Enset	Imdibir	.416	.464	.419	.440	**.435**
	Aze-Deboa	.335	.306	.333	.285	**.315**
	Adado	.431	.374	.420	.356	**.393**
	Gara-Godo	.477	.501	.497	.488	**.490**
	Do'oma	.534	.497	.456	.458	**.488**
	Enset average	**.440**	**.423**	**.426**	**.400**	**.422**
Average across FAs		**0.494**	**0.503**	**0.495**	**0.497**	**0.497**

Source: Calculated from efficiency scores of respective specifications

Caveats of variables used and the application of SFA on ERHS data

In addition to the usual problems associated with measurement error and omitted variable bias that other methods suffer from, the SFA approach may have other shortcomings. These shortcomings warrant that the results and implications of the stochastic frontier analysis must be interpreted with some qualifications. Two caveats stand out.

The first one questions the basic structure of the SFA approach as it tries to explain unobservable levels of farmers' inefficiency with observable factors. The basic difference between those that propose and apply SFA and those who question this approach seems to be in how they define a production function/frontier. While the first group sees a production function as a technical relation between factors that are directly put into the production process and the output produced, the second group suggests that if a factor affects levels of efficiency and thereby the magnitude of output it should be included in the set of factors that determine levels of production even if its effect is not direct. If the latter approach is followed, variables in the inefficiency equation are to be included in the frontier, eliminating the idea of composed error terms, which SFA is based on, as the factors that explain the inefficiency component of the error term have now become part of the frontier. Although we do not try to settle the argument here, we would like to indicate that both approaches should theoretically lead to the same conclusion in identifying sources of inefficiency, with the only difference being how we define a production process, which dictates whether or not we estimate a simultaneous equation that treats the two parts separately. We follow the SFA approach not only because it follows the standard definition of production functions but also because it enables us to measure levels and evolution of efficiency of individual decision making units.

The second caveat concerns the variables utilized in this study, which could have resulted in the large divergence in efficiency between those operating on and under the estimated production frontiers. Although we used these variables out of necessity we would like to admit the possible shortcomings that could result from their utilization. Our use of the real value of output instead of volume of production, the substitution of labor use by a proxy variable, and the fact that implements ownership may not necessarily reflect level of use are notable ones.

Although plot level data on measured output was collected in all rounds of ERHS and was more preferred as a regressand, the analysis used household level aggregations of plot level real value of output because data on most factors put into production were collected at the household level. In addition to the detailed information that was lost because of the aggregation, this variable may carry imperfections associated with the price data collected separately from the surveys.

The consequence of using real value of output as a regressand is that higher efficiency scores will be assigned for those farmers that produce crops generating higher monetary value, such as cash crops. Although one can argue the choice of what to produce as a component of production decision, the researchers admit that this could be a significant handicap, in particular in interpreting the efficiency scores, as farmers' decision on what to produce could be dictated by factors beyond their control, such as climatic conditions, and thus a farmer should be compared with other farmers belonging to the subset that produce the same crop.

To take this into account we recalculated two additional average efficiency scores by first dropping those households that had efficiency scores in the lowest and highest quintiles. The second set of efficiency scores were recalculated by dropping those that had efficiency scores in the lowest and highest quintiles of the sub-sample (or by dropping those above and below two quintiles in the original data set). Average efficiency of farm households in the first sub-sample is 0.47, which is about 3 percent larger than 0.4, the overall average. The average efficiency for the second sub-sample is 0.478, which is about 5 percent higher than the overall average.

Due to differences in the details of survey instruments implemented to collect data on labor use in different rounds we decided to substitute this variable with the number of household members of 16 years and older. Although the number of household members of 16 years and older better approximate labor use than any other variable in the data set, which constitutes subsistence farmers that largely depend on family labor, it may also lead to wrong estimates and implications. Specifically, a household with few working members that hired laborers to work on the household farm may appear to have superior efficiency than a household that was unable to hire extra labor. A similar problem may arise if a household used some inputs that were not included in the data, such as pesticides.

The problems just discussed may have led to imperfect parameter estimates of the variables in all approaches that estimate production functions. However, the problem is compounded by using the stochastic production frontier approach because the imperfections will pass on to the inefficiency equation that is estimated simultaneously with the production frontier, and the efficiency score of a household, which is derived from the relative position of its real value of output from those that are deemed efficient and operate at the frontier. Notwithstanding these shortcomings, we used the SFA approach because its advantages outweigh these disadvantages.

5. Summary and key findings

A number of stochastic production frontiers with time-varying inefficiency effects were estimated for about 1,400 subsistence farm households residing in five agroecological regions in Ethiopia. The data included in the analysis was collected at four points during a sixteen-year period spanning 1994 to 2009. The stochastic production frontier analysis was selected over other options such as the ordinary least squares (OLS) or the data envelopment analysis (DEA) as it enables one to disentangle the idiosyncratic effects that farmers face from household-specific inefficiency effects.

The baseline results indicate that among the Ethiopian farmers in this panel, most of the increase in output was attributable to increased use of traditional inputs. The value of output was highly elastic with respect to the amount of rainfall received in each region. Increases in the size and quality of cultivated land, changes in labor use and/or human capital, and changes in the numbers of oxen and hoes used for cultivation also significantly contributed to increased output. By contrast, the calculated elasticities with respect to the rate of fertilizer application and participation in the official extension program were among the lowest, implying that, on average, these factors contributed little to agricultural output. Moreover, models that estimated the frontiers and inefficiency equations using differently grouped years indicate that elasticities of scarce factors, like land, are increasing, consistent with theoretical predictions of production functions, while elasticities of inputs available in large supply, such as labor, are declining. The elasticities from the baseline model, as well as other specifications that investigated if there is any discernable pattern, imply not only that Ethiopian agriculture relies heavily on traditional factors of production and that crop production in Ethiopia is especially sensitive to changes in the amount of these traditional inputs but also that future growth in agricultural production cannot be achieved by intensive use of only some inputs. Thus, they essentially imply that the structure has to change.

The magnitude of the effect of rainfall on output suggests that the government and concerned agencies, in collaboration with farmers, should put a premium on finding ways to reduce the shocks faced during periods of low rainfall. The fact that area of cultivated land plays such an increasingly important role for increased output is a warning that such increased output cannot be attained into the future given that land size is fixed and the rate of population growth that mostly relies on agriculture is high. In Ethiopia most farmers use either a pair of oxen or simple hand tools to till their land. Due to a shortage of capital and the prospects that the extensive use of (heavy) machinery can exacerbate wind and water erosion it is hard to argue in support of the use of heavy machinery. However, small and medium scale machinery used in similarly dry agricultural areas such as parts of India could be productively deployed in Ethiopia.

The finding that the elasticity of such modern inputs as fertilizer is insignificantly small on average implies that there is an untapped opportunity for expanding agricultural production and increasing agricultural productivity by intensification of modern inputs including fertilizers. The results indicate that each of the agroecological zones had gained from Hicksian-neutral technological improvements during the 1994-2004 period, while there were no changes in 2009 relative to 1994 and there were declines in productivity between 2004 and 2009, as was corroborated with several specifications. One can draw from these results that increases in traditional input use do not translate to increased productivity, which is crucial for sustained long-term growth and plans should be in place for increased application of modern inputs. Descriptive analyses show some of the fixed inputs, such as land, are already fragmented and hence increased application of other traditional inputs, such as labor, will eventually have decreasing marginal returns. In this respect, policy makers need to pay attention to factors that facilitate extension package adoption and increased fertilizer application. This requires rethinking the services that agricultural agents provide, introducing simple agricultural machinery that helps alleviate labor bottlenecks and facilitating the availability of credits that will help farmers acquire ploughing oxen and other inputs.

Education contributes towards reduced inefficiency. Female headed households or households that had labor bottlenecks suffer from increased inefficiency due to the multiple roles played by women. Households that are faced with diversified risk from plots that are located sufficiently apart appear more efficient while households that had less land per working member are more inefficient. Households that own more animals either in terms of two or more ploughing oxen or total livestock units are less inefficient. Households residing in FAs with expanded agricultural extension services are considerably less inefficient. Drought affects efficiency adversely whenever it strikes. Farmers that live in close proximity to markets are more inefficient. On average farming inefficiency has consistently declined in the period considered. The results suggest that each agroecological zone faces different opportunities and obstacles, and as such policy makers need to take a close look at each of the zones before implementing uniform changes that concern all of the zones.

References

Afriat, S. N. 1972. Efficiency Estimation of Production Functions. *International Economic Review* 13 (3): 568–598.

Aigner, D. J., and S. F. Chu,.1968. On Estimating the Industry Production Function. *American Economic Review* 58 (September): 826–839.

Aigner, D. J., C. A. Lovell, and P. Schimidt. 1977. Formulation and Estimation of Stochastic Production Function Models. *Journal of Econometrics* 6 (1): 21–37.

Bachewe, F. N. 2009. The State of Subsistence Agriculture in Ethiopia: Sources of Output Growth and Agricultural Inefficiency. Ph.D. Dissertation. University of Minnesota.

Ball, V.E. 1985. Output, Input, and Productivity Measurement in U.S. Agriculture, 1948–79. *American Journal of Agricultural Economics* 67 (3): 475–486.

Battese, G. E., and T. J. Coelli. 1995. A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics* 20: 325–332.

Collier, P., and S. Dercon. 2009. African Agriculture in 50 Years: Smallholders in a rapidly changing world? Expert Meeting on How to Feed the World in 2050. FAO, Economic and Social Development Department.

Chavas, JP. 2008. On the economics of agricultural production. *Australian Journal of Agricultural and Resource Economics* 52 (4): 365–380.

Christiaensen, L., P. Lanjouw, J. Luoto, and D. Stifel. 2010. *The Reliability of Small Area Estimation Prediction Methods to Track Poverty.* UNU-Wider Working Paper 2010/99.

Debreu, G. 1951. The Coefficient of Resource Utilization. *Econometrica,* 19 (3): 273–292.

Dercon and Hoddinott. 2004. *The Ethiopian Rural Household Surveys: Introduction.* Washington, D.C.: International Food Policy Research Institute.

Dercon, S., and R.V. Hill. 2009. Growth from agriculture in Ethiopia: Identifying key constraints. Paper prepared as part of a study on Agriculture and Growth in Ethiopia. Department for International Development, London.

Farrel, M. J. 1957. The Measurement of Productive Efficiency. *Journal of Royal Statistical Society*, Series A, CXX, Part 3: 253–290.

Gebre Egziabher, K., and S. Tegegne.1996. Ethiopian Village Studies: Designed and edited by Philippa Bevan and Alula Pankhurst. Ethiopian Rural Household Survey Village Studies, International Food Policy Research Institute.

Halvorson, R. and R. Palmquist. 1980. The Interpretation of Dummy Variables in Semilogarithmic Equations. *The American Economic Review* 70 (3): 474–475.

Koopmans, T. C. 1957. An Analysis of Production as an efficient Combination of Activities. In *Activity Analysis of Production and Allocation*, ed. T.C. Koopmans. Cowles Commission for Research in Economics, Monograph No. 13, Wiley, New York.

Kumbhakar, S. C., and A. K. Lovell. 2000. *Stochastic Frontier Analysis.* Cambridge: Cambridge University Press.

Meeusen, W., and J. van den Broeck. 1977. Efficiency Estimation from Cobb_douglas Production Functions with Composed Error. *International Economic Review* 18 (2): 435–444.

Mullen, J.D., and T.L. Cox 1995. The returns from research in Australian broadacre agriculture. *Australian Journal of Agricultural Economics* 39 (2): 105–128.

Mullen, J. 2007. Productivity growth and the returns from public investment in R&D in Australian broadacre agriculture. *Australian Journal of Agricultural and Resource Economics* 53 (4): 359–384.

Pitt, M., and L. Lee. 1981. The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry. *Journal of Development Economics* 9: 43–64.

SG 2000. 1995. The SasakawaGlobal 2000 Agricultural Project in Ethiopia. SG 2000. Addis Ababa, Ethiopia.

SG 2000. 2007. web page: http://www.saa–tokyo.org/english/sg2000/crop.shtml. Accessed on 2 November 2007.

Shepard, R. W. 1970. Theory of Cost and Production Functions. Princeton: Princeton University Press.

Taffesse, A.S. 2008. Decomposition of Growth in Cereal Production in Ethiopia. Paper prepared for the UK Department for International Development funded study *Understanding the constraints to continued rapid growth in Ethiopia: the role of agriculture.* Ethiopian Economics Association.

Taffesse, A.S. 2009. Cereal Production in Ethiopia. Unpublished manuscript. ESSP II Memo.

Winsten, C. B. 1957. Discussion on Mr. Farrell's Paper. *Journal of Royal Statistical Society* Series A, CXX, Part 3: 282–284.

World Bank. 2008. World Development Indicators. Washington, D.C.: The World Bank.

Appendices

Appendix 1. Pearson's correlation coefficients of selected variables used in the analysis

Variable	Value of output	Area	Labor	Oxen	Land quality	Fertilizer use	Rainfall	Education	Number of extension agents	Participated in NEP	Livestock units
Area	.458**										
Labor	.138**	.121**									
Oxen	.321**	.401**	.168**								
Land quality	-.184**	-.147**	-.012	-.061**							
Fertilizer use	.450**	.365**	.134**	.389**	-.110**						
Rainfall	.072**	.006	.054**	-.083**	-.274**	.029*					
Education	.075**	.011	-.021	-.009	-.064**	.028*	.074**				
Number of extension agents	.020	.079**	-.070**	.148**	.008	.023	-.159**	-.044**			
Participated in NEP	.117**	.102**	.035*	.036**	-.057**	.063**	.046**	.067**	-.021		
Livestock units	.368**	.486**	.215**	.628**	-.062**	.411**	-.036*	-.002	.012	.071**	
Household size	.168**	.109**	.652**	.196**	-.011	.108**	.022	.047**	-.109**	.012	.237**

Note: Coefficients with ** and * are significant at 1 and 5 percent level of significance.

Appendix 2. Maximum likelihood coefficient estimates of frontiers comprising different rounds

Variable	ERHS in 1994, 1999,2004, and 2009		All rounds		ERHS in 2009		ERHS in 1994 and 1999		ERHS in 1999 and 2004		ERHS in 2004 and 2009	
	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio
Constant	4.924	14.304	6.437	20.830	5.360	9.827	6.686	8.934	4.768	10.657	5.767	14.487
Area of cultivated land	0.409	22.151	0.415	26.064	0.357	12.703	0.440	15.207	0.469	18.033	0.395	17.346
Household members 16 years and older	0.100	5.266	0.087	4.154	0.077b	1.769	0.160	3.618	0.079	3.804	0.088	4.734
Amount of rainfall 12 months before the survey	0.333	6.484	0.101a	2.213	0.344	4.120	0.068c	0.604	0.376	5.593	0.243	4.045
Amount of fertilizer used	0.002	10.952	0.002	11.790	0.001	5.690	0.003	11.661	0.002	7.024	0.001	5.701
Number of ploughing oxen	0.031a	2.042	0.083	5.759	0.046a	2.446	0.011c	0.350	0.063	2.951	0.087	4.626
Average land quality	-0.109	-9.367	-0.095	-9.783	-0.032c	-1.166	-0.137	-8.212	-0.108	-8.221	-0.082	-5.356
Number of hoes used	0.074	6.512	0.066	6.086	0.039	4.104	0.112	4.273	0.110	5.843	0.044	4.114
Number of ploughs used	0.019	3.118	0.016	2.726	0.001c	0.266	0.006c	0.248	0.030b	1.821	0.016	3.067
Participated in new extension program	0.117	3.062	0.126	3.116	0.033c	0.623	0.182a	2.481	0.220	3.966	0.041c	0.967
Central highlands	0.404	8.415	0.407	8.947	0.004c	0.036	0.379	5.794	0.432	7.245	0.395	6.759
Arussi/Bale	0.610	10.805	0.540	9.697	0.307	2.775	0.501	6.182	0.332	4.615	0.537	7.487
Hararghe	0.852	14.656	0.770	13.238	0.417	3.718	0.759	9.699	0.841	10.950	0.907	10.737
Enset	0.714	11.873	0.911	17.372	-0.050c	-0.435	1.332	14.434	1.038	13.970	0.388	5.860
1994 dummy	0.053c	0.998										
1995 dummy			0.346	6.532								
1997 dummy			0.206	4.069								
1999 dummy	0.280	5.564	0.261	5.401			0.190	4.163				
2004 dummy	0.236	5.220	0.287	6.236					-0.026c	-0.665		
2009 dummy			0.044c	0.803							-0.142	-3.596

Note: All estimates are significant at 1 percent except those with superscript a and b, which are significant at 5 and10 percent, respectively, and b, which are not significant.

Appendix 3. Maximum likelihood estimates of separate stochastic production frontiers that use agroecological zone level data

Note: All estimates are significant at 1 percent except those with superscripts a and b, which are significant at 5 and10 percent, respectively, and c, which are not significant

Variable	Northern Highlands		Central highlands		Arussi/Bale		Hararghe		Enset	
	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio
Constant	2.811	4.367	7.219	35.510	11.996	29.880	4.871	6.276	6.651	34.500
Area of cultivated land	0.332	8.339	0.425	17.049	0.731	14.746	0.400	4.428	0.435	14.989
Household members 16 years of age and older	0.106	4.072	0.071[a]	2.390	0.044[c]	1.165	0.164	3.452	0.070[b]	1.864
Amount of rainfall 12 months before the survey	0.007	5.488	0.000[c]	-1.023	-0.005	-10.471	0.005	3.492	0.001	4.638
Amount of fertilizer used	-0.002[c]	-0.85	0.001	5.813	0.001	3.596	0.002[a]	2.222	0.000[c]	-0.529
Number of ploughing oxen	0.172	4.209	0.067	3.270	0.020[c]	0.877	0.167[b]	1.821	0.151	3.591
Average land quality	-0.036[c]	-2.28	-0.078	-4.412	-0.029[c]	-0.776	-0.031[c]	-0.82	-0.037[c]	-1.451
Number of hoes used	0.038[b]	1.930	0.039	3.617	0.008[c]	0.430	0.049[c]	1.628	0.068	2.563
Number of ploughs used	0.004[c]	0.285	0.005[c]	0.782	0.012[c]	0.773	0.000[c]	-0.01	0.027[c]	1.308
Participated in new extension program	0.154[a]	1.998	0.104[b]	1.945	0.062[c]	0.855	-0.036[c]	-0.31	0.116[b]	1.710
Geblen	0.231[c]	1.146								
Shumsheha	-1.508	-3.92								
Yetmen			0.345	4.136						
Turufe Ketchema			0.216[a]	2.548						
DB-Milki			0.402	4.359						
DB-Kormargefia			0.322	3.340						
DB-Karafino			0.639	10.230						
DB-Bokafia			0.562	6.163						
Korodegaga					-0.368	-3.997				
Aze Deboa									-0.166[c]	-1.507
Adado									0.404	3.501
Gara-Godo									-0.917	-7.941
Do'oma									-0.278[b]	-1.901
1999 dummy	0.394	4.341	0.357	7.757	0.489	4.885	-0.856	-2.72	1.176	12.735
2004 dummy	0.837	9.528	0.549	11.927	1.131	9.990	-1.122[a]	-2.09	0.976	10.603
2009 dummy	1.298	5.549	0.352	6.047	1.008	7.629	-0.798a	-2.15	-0.083c	-0.765

Appendix 4. Maximum likelihood parameter estimates of stochastic frontiers of those applying fertilizer, per-hectare specification, and high and low potential crop production regions

Variable	ERHS in 1994, 1999, 2004, and 2009; those applying fertilizer		Per hectare values		High potential crop production FAs		Low potential crop production FAs	
	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio
Constant	8.536	15.397	5.944	15.575	10.811	11.109	2.727	6.119
Area of cultivated land	0.429	17.949			0.334	16.368	0.344	13.218
Household members 16 years of age and older	0.085	2.786	0.260	18.236	0.067	2.664	0.126	4.385
Amount of rainfall 12 months before the survey	-0.218	-2.612	0.188	3.243	-0.466	-3.241	0.651	9.722
Amount of fertilizer used	0.001	8.983	0.001	11.521	0.002	10.918	0.002	4.747
Number of ploughing oxen	0.024c	1.419	0.005b	2.134	0.018c	0.883	0.147	4.992
Average land quality	-0.080	-4.905	-0.108	-9.340	-0.056	-2.818	-0.098	-6.550
Number of hoes used	0.044	4.512	0.019	9.460	0.071	5.358	0.044	2.839
Number of ploughs used	0.011b	1.906	0.000c	0.174	0.011c	1.410	0.040	3.781
Participated in new extension program	0.063c	1.614	0.101a	2.456	0.150	3.269	0.075c	0.998
Central Highlands	0.536	5.141	0.400	8.234			-0.075c	-0.940
Arussi/Bale	0.712	6.477	0.558	9.875	0.224	4.671		
Hararghe	0.937	8.698	0.733	11.933			0.783	10.639
Enset	0.512	4.343	0.891	14.233	0.613	8.444	0.231	3.205
1999 dummy	0.150	3.112	0.364	7.753	0.367	6.043	-0.010c	-0.144
2004 dummy	0.249	4.635	0.265	6.327	0.133a	2.424	0.315	4.600
2009 dummy	0.090c	1.630	0.143	3.000	-0.047c	-0.751	0.059c	0.696

Note: All estimates are significant at 1 percent except those with superscripts a and b, which are significant at 5 and10 percent, respectively, and c, which are not significant.

Appendix 5. Maximum likelihood coefficient estimates of the inefficiency equation parameters comprising different rounds

Note: All estimates are significant at 1 percent except those with superscripts a and b, which are significant at 5 and 10 percent, respectively, and c, which are not significant

Variable	ERHS in 1994, 1999, 2004 and 2009		All rounds		ERHS in 2009		ERHS in 1994 and 1999		ERHS in 1999 and 2004		ERHS in 2004 and 2009	
	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio
Constant	-36.550	-20.449	-14.315	-7.929	-35.614	-10.563	-33.594	-8.579	-20.354	-18.243	-28.639	-25.07
Sex	-2.488	-4.741	-1.501	-2.628	-0.323[c]	-0.515	-2.850	-4.958	-2.125	-2.988	-1.171[a]	-2.374
Age	0.042	4.924	0.030	3.199	0.043	2.667	0.015[c]	1.277	0.006[c]	0.575	0.047	3.919
Level of education	-1.418	-2.865	-1.533	-3.725	-1.998 [c]	-2.493	-1.812	-2.713	-0.282[c]	-0.663	-2.228	-3.779
Female dummy	2.595	3.949	4.692	5.944	2.650	3.095	3.366	3.741	5.558	7.144	2.235	3.669
Household size	-0.136	-2.825	-0.290	-4.875	-0.402	-3.720	-0.135[a]	-1.977	0.020[c]	0.401	-0.134[c]	-1.641
Number of plots* log of cultivated area	-0.207	-6.323	-0.201	-7.018	0.017[c]	0.254	0.049[c]	0.669	0.147	3.239	-0.177	-5.173
Cultivated area/ number of members 16 years and older	-0.004[b]	-1.931	-0.004	-5.647	-1.598	-4.910	-5.577	-8.200	-0.002[a]	-2.444	-0.002[a]	-2.405
Oxen dummy	-1.670	-3.495	-2.097	-4.002	-4.461	-5.097	-1.750[a]	-2.559	-0.336[c]	-0.595	-2.634	-4.089
Livestock units	-0.346	-17.205	-0.254	-9.814	-0.169	-2.938	-0.360	-13.477	-0.067[c]	-1.329	-0.124	-2.679
Number of agricultural extension agents in FA	2.852	11.767	-0.545[c]	-1.554	5.599	10.273	2.539	5.109	0.964	2.993	3.450	14.61
Crop affected by drought	6.327	17.168	4.522	10.387	9.705	11.897	2.479	4.627			11.599	16.65
Survey month	9.413	12.877	9.445	21.778	-1.914[b]	-1.740	7.438	6.709	-20.354	-18.243	-4.868	-6.711
Elevation	0.002	14.179	-0.001	-6.729	0.003	10.288	0.002	8.006	0.001	4.223	0.002	11.36
Distance to health center	0.080	9.670	0.134	12.628	-0.732	-4.345	0.139	10.509	0.005[c]	0.568	-0.004[c]	-0.386
Distance to closest market	-0.152	-14.707	-0.094	-7.811	0.325	6.511	-0.212	-12.579	-0.101	-8.547	0.066	4.445
Distance to nearest FA center	0.101	15.082	0.084	7.865	0.586	3.560	0.159	10.688	0.042	5.127	-0.035	-3.570
Sigma-squared	25	32	45	38	19	16	27	23	26	28	16	25
Gamma	0.991	1998	0.995	4003	0.994	1124	0.993	1444	0.991	1523	0.987	892
Average Efficiency	0.457		0.399		0.458		0.400		0.530		0.534	

Appendix 6. Maximum likelihood coefficient estimates of the inefficiency equation parameters comprising different agroecological zones

Variable	Northern Highlands		Central highlands		Arussi/Bale		Hararghe		Enset	
	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio
Constant	-19.114	-12.869	-1.853[b]	-1.858	-45.196	-5.087	-0.050[c]	-0.050	-4.250[a]	-2.562
Sex	-0.727[c]	-1.173	-4.698	-5.769	-3.933	-4.751	-2.172	-3.868	-1.887[b]	-1.923
Age	0.038	3.109	0.006[c]	0.560	0.061	2.827	0.021[c]	1.521	0.001[c]	0.049
Level of education	-0.120[c]	-0.125	-3.086	-5.178	-1.813	-2.986	-2.425[a]	-2.093	-1.162[c]	-1.189
Female dummy	0.136[c]	0.224	5.719	6.695	4.634	4.308	0.004[c]	0.005	4.922	4.907
Household size	-0.052[c]	-0.508	0.131[b]	1.664	-0.252[a]	-1.974	-0.016[c]	-0.193	-0.230[a]	-2.123
Number of plots* log of cultivated area	-0.289	-3.146	-0.449	-14.623	-0.277	-3.901	0.044[c]	0.433	0.417	5.949
Cultivated area/ number of members 16 years and older	-0.057	-5.031	-0.010	-4.502	-0.021	-15.286	-0.008	-7.285	-0.008[b]	-1.714
Oxen dummy	-0.882[c]	-1.158	0.326[c]	0.486	0.857[c]	1.633	1.193[c]	1.179	-2.231[a]	-2.312
Livestock units	-0.800	-4.774	-0.166	-8.790	-0.159	-5.063	-1.358	-4.542	-0.817	-9.724
Number of agricultural extension agents in FA	-19.114	-12.869	-3.160	-9.287	2.811[c]	0.666	-0.050[c]	-0.050	-0.286[c]	-0.319
Crop affected by drought	0.887[c]	1.481	4.216	4.944	5.882	6.039	-0.019[c]	-0.026	7.999	9.322
Survey month	-8.489[c]	-0.863	-5.580	-9.157	-3.319[c]	-0.797	-1.340[c]	-1.588	17.594	7.466
Elevation	0.003	19.716	0.000[b]	-1.920	0.006[a]	2.177	0.000[c]	-0.132	-0.002	-5.709
Distance to health center	-0.917	-3.855	-0.668	-6.253	6.945	4.910	-0.364[c]	-0.519	-0.246	-8.502
Distance to closest market	1.011	5.134	0.197[b]	1.654	-0.357	-4.305	0.400[c]	0.572	0.068[a]	2.044
Distance to nearest FA center	0.363[a]	2.412	0.062	4.300	-6.428	-4.636	-0.027[c]	-0.090	-0.097[a]	-2.073
Sigma-squared	12.457	17.617	14.308	13.463	12.107	15.842	2.467	7.143	44.53	18.6
Gamma	0.991	866.114	0.992	1192.458	0.985	464.168	0.907	50.215	0.994	1492.0
Average Efficiency	0.418		0.567		0.574		0.585		0.422	
Log likelihood	-1493		-1688		-817		-388		-2536	

Note: All estimates are significant at 1 percent except those with superscripts a and b, which are significant at 5 and10 percent, respectively, and c, which are not significant.

Appendix 7. Maximum likelihood parameter estimates of inefficiency equations of those applying fertilizer, per hectare specification, and high and low potential crop production regions

Variable	ERHS in 1994, 1999, 2004 and 2009; those applying fertilizer		Per hectare values		High potential crop production FAs		Low potential crop production FAs	
	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio	coefficient	t-ratio
Constant	-5.836	-3.451	-38.309	-29.546	-30.339	-11.04	-37.010	-19.453
Sex	-0.969	-3.670	-2.655	-4.919	-4.683	-6.199	-2.159	-3.998
Age	0.027	3.844	0.045	5.324	0.024[a]	2.035	0.040	3.760
Level of education	-1.618	-4.474	-1.458	-2.798	-0.904[c]	-1.603	-0.754[c]	-1.052
Female dummy	-0.175[c]	-0.348	2.128	3.237	4.736	5.160	1.480[a]	2.279
Household size	-0.016[c]	-0.423	0.048[c]	0.886	-0.271	-2.779	-0.069[c]	-0.887
Number of plots* log of cultivated area	-0.025[c]	-0.839	-0.125	-3.315	-0.014[c]	-0.435	-0.073[b]	-1.769
Cultivated area/ number of members 16 years and older	-0.727	-5.510	-0.006	-6.628	-0.009	-4.957	-0.005[a]	-2.445
Oxen dummy	-1.953	-6.530	-1.820	-3.353	-2.760	-3.513	0.266[c]	0.469
Livestock units	-0.181	-11.638	-0.378	-16.146	-0.081[a]	-2.423	-0.597	-7.253
Number of agricultural extension agents in FA	0.879	3.708	3.188	13.673	2.009	3.028	3.733	9.171
Crop affected by drought	2.953	9.603	6.292	13.855	8.243	7.791	3.389	8.243
Survey month	-0.673[c]	-1.312	9.591	15.124	12.121	11.682	15.149	11.595
Elevation	0.001	9.449	0.002	19.382	0.000[c]	1.168	0.003	19.912
Distance to health center	0.189	12.898	0.086	10.850	-0.090[a]	-2.260	-0.003[c]	-0.224
Distance to closest market	-0.208	-12.057	-0.165	-17.561	-0.064[a]	-2.075	0.000[c]	0.035
Distance to nearest FA center	0.017[c]	1.627	0.102	13.866	0.262	5.357	-0.059	-4.899
Sigma-squared	6.639	14.498	25.720	26.667	32.753	16.792	17.468	23.506
Gamma	0.976	369	0.990	1426	0.995	1966	0.985	774
Average Efficiency	0.603		0.451		0.497		0.445	
Log likelihood	-2208		-7912		-3426		-4160	

Note: All estimates are significant at 1 percent except those with superscripts a and b, which are significant at 5 and 10 percent, and c, which are not significant

Appendix 8. OLS estimates of parameters of the production function

Variable	Estimated Coefficient	t-ratio
Constant	-7.828	-10.479
Area of cultivated land	.752	20.910
Household members 16 years and older	.209	4.000
Amount of rainfall 12 months before the survey	1.820	15.723
Amount of fertilizer used	.002	5.996
Number of ploughing oxen	.270	6.713
Average land quality	-.104	-4.379
Number of hoes used	.164	7.030
Number of ploughs used	-.004[b]	-.240
Participated in new extension program	.132[b]	1.353
Central highlands	.626	4.925
Arussi/Bale	.620	4.339
Hararghe	2.111	14.233
Enset	.539	4.073
1999 dummy	1.087	11.822
2004 dummy	1.784	19.475
2009 dummy	.183[a]	1.688

Notes: 1. The dependent variable is household level real value of output
2. All parameter estimates are significant at 1 percent of level of significance except the ones with superscript a and b, which are significant at 10 percent and not significant, respectively.