

# Ethiopia Stochastic Frontiers Analysis Draft 2

*Tomas Morley*

*30 March 2017*

## Abstract

Maize is an important staple crop in Ethiopia with wide implications for food security and policy. In this paper we adopt the framework developed by Dijk et al. (2016) to decompose the maize yield gap in Ethiopia into (1) the technical efficiency yield gap, (2) the allocative yield gap, (3) the economic yield gap and (4) the technology yield gap. Our dataset is constructed using plot level data covering Ethiopian maize plots derived from a survey of Ethiopian households. This is combined with maize specific growing information derived from crop simulation models. The results of the decomposition and econometric analysis suggest that the technology yield gap makes up the largest share of the total yield gap followed by the feasible yield gap, the technical efficiency yield gap and the economic yield gap. The results from this paper can be used to design targeted policy and farming recommendations at the plot, farm household, local and national level.

## Introduction

Ethiopia has great agricultural potential because of its vast areas of fertile land, diverse climate, generally adequate rainfall, and large labour pool. Consequently, agriculture has remained as the foundation of Ethiopia's economy, accounting for around 40 percent of the national income and around 80 percent of the total employment (EEA, 2015). It also supplies food to the urban areas and raw materials to the manufacturing sector. Many other economic activities depend on agriculture, including marketing, processing, and export of agricultural products. Among the major drivers of GDP growth crop production had been the second most important contributor next to construction. The agricultural sector has, therefore, been crucial for the overall performance of the Ethiopian economy and will likely continue to be the backbone of the Ethiopian economy for years to come. The sector is expected to play a central role in decreasing poverty and increasing growth in Ethiopia.

Agricultural production in Ethiopia is overwhelmingly of a subsistence nature where millions of smallholder farmers cultivate on average less than one hectare of land. About 80 per cent of smallholder production is destined for home consumption and use as seed for subsequent harvests. The smallholder sector is the main source of employment, agricultural GDP, export earnings, and food and raw materials supplies to urban areas and food industries. Crop production is by far the most dominant sub-sector, accounting for over 97 per cent of agricultural output and more than 60% of agricultural GDP (RATES, 2003). It is estimated that crop production and livestock husbandry account for over 86 per cent of the agricultural GDP (WB, 2013).

Despite this potential, agricultural productivity in Ethiopia is very low compared to other Sub-Saharan African countries. Ethiopian agriculture has remained underdeveloped due to periodic drought, soil degradation, deforestation, high population density, poor infrastructure (making it difficult and expensive to get goods to market) and low level of technology use among others.

Maize is one of the most important staple crops in terms of calorie intake for millions of Ethiopians and is grown in many parts of the country. Because of its large geographical coverage and the scale of production, it plays a critical role in the food security of the country and particularly the poor. According to the CSA 2004/5 national survey of consumption expenditure, maize accounted for about 17 percent of the national calorie intake followed by sorghum (14 percent) and wheat (13 percent) among the major cereals (Taffesse et al., 2012). Maize is second to teff in terms of area, but it ranks first among cereals in production. According to CSA (2010) maize accounted for around 30 percent of all cereals produced in the country and more than 20 percent of the land cultivated by all cereal crops. According to the Agricultural Growth Program (AGP)

Ethiopia is the fifth largest producer of maize in Africa in terms of the area devoted to maize but is second, only to South Africa, in yield and third, after South Africa and Nigeria, in production and smallholder farmers and around 9 million smallholder farmers account for more than 95 % of the total maize area and production in the country (<http://ethioagp.org/maize-2/>).

Maize is grown in the mid highland areas of the country and there are huge tracts of land in all regions suitable for maize farming. Maize is grown primarily in the Amhara, Oromia and SNNPR regions of the country from which 5 zones namely West Wellega, Jimma, East Shewa, East Wellega, and West Gojjam account for over 44 percent of total smallholder production. Marketable surplus is mainly produced in these high potential zones, and the production in the other zones is mainly for subsistence. Maize production has increased in Ethiopia in recent years, due in large part to an improvement in the supply of seeds to farmers. While there have been significant gains made in maize production over the past decade, there is still a significant opportunity to further increase productivity. Ethiopia can be considered a success story for maize production as, apart from South Africa, it is the only country in Sub-Saharan Africa that shows substantial progress in maize productivity and input use (Abate et al., 2015). After a period of limited growth, yield more than doubled from around 1.5 ton/ha in 2000 to over 3 ton/ha in 2013 (FAOSTAT). The government has worked hard to ensure timely delivery of seeds and fertilizer, as well as to improve agricultural extension services and trainings for farmers. Maize yield levels are also slightly above the regional average. Despite the recent progress in productivity, yield levels in Ethiopia are still very low relative to what could be produced. According to the Central Statistical Agency of Ethiopia (CSA, 2010), the low level of crop productivity in the country is attributed to inefficient farming methods that utilize outdated crop production and protection technologies, and fragmented pieces of land holdings. Estimates indicate that the current maize yield could be doubled if farmers adopt higher quality inputs and proven agronomy best practices. Data from the Global Yield Gap Atlas (GYGA) shows that the average yield gap is on average 82% in Ethiopia. This is in strong contrast with for example Latin American countries which are able to reach around 45% of potential yield.

Yield gap measurements are a very powerful metric to determine the absolute potential to increase production and are therefore frequently used in policy studies (Sumberg, 2012). This makes them particularly useful as a basis for forward looking scenario studies to assess future agricultural production and food security. By ‘controlling’ for the biophysical production environment, yield gaps can easily be compared across regions and countries and, therefore, provide a good indicator of relative (potential) performance. A limitation is that yield gap measurements offer only limited guidance in explaining the observed difference between actual and potential yield. It has been argued that it is simply not cost-efficient for farmers to use the amount of inputs that are needed to reach potential. This is underscored by Lobell et al. (Lobell et al., 2009), who compared a large number of yield gap studies and find that (best-practice) farmers reach only around 80% of potential yield. The ‘exploitable’ yield gap is therefore lower than the total yield gap and is governed by economic constraints such as input prices and market access. Other important causes for the existence of yield gaps are differences in farm-level characteristics (e.g. size, education and off farm activities) and the socio-economic context (e.g. subsidies, rural infrastructure and access to finance).

The aim of this paper is to provide a detailed analysis of the maize yield gap in Ethiopia. We decompose the conventional yield gap into four elements that capture, technical efficiency, economic efficiency, availability of technologies and an ‘unexploitable’ yield gap element.

The structure of this paper is as follows. Section 2 describes the econometric procedure used to decompose the yield gaps. Section 3 describes the data used for this study. Section 4 presents the results of our estimations and yield gap calculations. Section 5 discusses the results. Section 7 concludes.

## Data

The main data source is the Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) Ethiopian Socioeconomic survey (ESS). This survey was implemented by the Central Statistical Agency of Ethiopia (CSA) and funded by the World Bank through a grant from the Bill and Melinda Gates foundation. Although part of a panel, the first wave of the ESS did not record crop harvested quantities for a

majority of plots making it unsuitable for this analysis. However, the second wave (2013-2014) of the ESS provides detailed information on crop production, input use, environmental and socioeconomic variables.

In both waves of the data respondents were visited twice as part of a post planting and post harvesting questionnaire. Through these questionnaires, detailed information at the plot level was gathered including the size of the harvest for each crop. The vast majority of plots were GPS measured, however, in some cases this was not the case and area measurements were imputed using multiple imputation and following LSMS-ISA World Bank (2014). Respondents were asked to record key production inputs including seed application, fertilizer and labour rates. Moreover, important variables regarding environmental variables, crop management practices, and the socioeconomic status of the household were recorded at the plot, household and community levels.

The longitude and latitude of each household was recorded and households were matched as part of the LSMS-ISA with climate, soil and geographical from sources including IFPRI, Worlclim, NASA and the Ethiopian Roads Agency among others. Using the longitude and latitude of the households we were able to further augment the dataset with more granular information from the Africa Soil Information Service (AfSIS, <http://africasoils.net>) and the Global Yield Gap Atlas (GYGA, [www.yieldgap.org](http://www.yieldgap.org)). AfSIS provides soil quality maps for Africa at 250m spatial resolution and various depths based on 28 thousand sampling locations (Hengl et al., 2015). We use AfSIS data to derive the soil organic carbon stock and pH for the top 200cm soil layer. We also derive the water-limited potential yield from GYGA which acts as a cap on the highest yield levels that can be attained in Ethiopia from a biophysical standpoint.

Table 1 provides summary statistics for the key variables in our study.

**Table 1: Descriptive Statistics**

Variable	Mean	Median	SD	Skewness	Min	Max
yld	1,698	1,046	2,140	3.344	2.956	17,978
N	25.26	0	55.18	4.655	0	691.3
area	0.17	0.093	0.327	12.03	0	8.125
lab	194.1	79.17	327.7	3.895	0	2,950
seedha	55.78	34.27	76.81	5.265	0	975.6
phdum55_2_70	0.667	1	0.471	-0.709	0	1
crop_count2	0.564	1	0.496	-0.257	0	1
dumoxen	0.625	1	0.484	-0.515	0	1
SOC2	16.67	16.11	4.77	0.294	3.912	37.21
logslope	2.249	2.272	0.799	-0.057	0	4.439
elevation	1,797	1,841	435.1	-0.87	345	2,909
GGD	7.209	7.135	0.894	0.683	4.922	10.42
AI	6.947	6.644	2.447	0.239	0.765	12.85
TS	1,137	1,062	355.1	1.013	558	3,026

## Estimation procedures

The analytical framework in this paper follows Dijk et al. (2016), and decomposes the yield gap into its constituent parts. Central to this framework is the estimation of the frontier yield response curve. There are competing methods for estimating this function including data envelopments analysis, corrected ordinary least squares and stochastic frontiers maximum likelihood. We follow the SF ML method (Aigner, Lovell, and Schmidt 1977; ???) which involves specification of the production technology and a composite error term that reflects both statistical error in the model and an asymmetric inefficiency term.

$$y_i = f(x_i, \beta) + \epsilon_i \quad (1)$$

$$\epsilon_i = v_i - u_i \quad (2)$$

$$v_i \sim \mathbb{N}(0, \sigma_v^2) \quad (3)$$

$$u_i \sim \mathbb{N}^+(0, \sigma_u^2) \quad (4)$$

Where  $y_i$  is the log of maize yield and  $x_i$  is the log of the inputs including the nitrogen, labour and seed rates. The composite error term  $\epsilon_i$  includes a truncated normal inefficiency term  $u_i$  and a statistical error term  $v_i$  where  $u$  and  $v$  are independent. With these specifications (Aigner, Lovell, and Schmidt 1977) show that the composite error term  $\epsilon_i$  follows the following density

$$f_\epsilon(\epsilon_i) = \frac{2}{\sigma} \phi\left(\frac{\epsilon_i}{\sigma}\right) \Phi\left(-\frac{\lambda \epsilon_i}{\sigma}\right) \quad (5)$$

Given a suitable form for the production technology, the parameters of the stochastic frontiers model can be estimated using maximum likelihood (ML) estimation. Common production functions used in crop yield models include the Cobb Douglas and translog production functions. The translog is the more flexible of the two and nests the Cobb Douglass form allowing model selection based on a likelihood ratio (LR) test. Viewed as an extension of the cobb douglass or as a second order expansion of an unknown production function we can write

$$Y_i = F(X_i, \beta) = \exp\left(\alpha + \sum_{k=1}^K \beta_k \ln X_{ik} + \sum_{k=1}^K \sum_{j=1}^K \gamma_{jk} \ln X_{ik} \ln X_{ij}\right) \quad (6)$$

Where  $Y_i$  is the output and the  $X_{ik}$  are the production inputs. Taking the log of expression (6) yields:

$$y_i = f(x_i, \beta) = \alpha + \sum_{k=1}^K \beta_k x_{ik} + \sum_{k=1}^K \sum_{j=1}^K \gamma_{jk} x_{ik} x_{ij} + \epsilon_i \quad (7)$$

Where the log of the output is written as  $y_i$  and the log of the inputs are written as  $x_i$ . Expression (7) has the familiar translog form, or, setting setting  $\gamma_{jk} = 0$ , the Cobb Douglas form. Expression (7) is then augmented with a composite error term  $\epsilon_i$  comprised of the truncated normal technical inefficiency and the symmetric error term  $\epsilon_i = v_i - u_i$ . In this form it is possible to carry out SF ML estimation as described in (1) - (5) above. We can also include environmental variables such as the slope and elevation of the plot yielding the following model in the translog case:

$$y_i = \alpha + \sum_{k=1}^K \beta_k x_{ik} + \sum_{k=1}^K \sum_{j=1}^K \gamma_{jk} x_{ij} x_{ik} + W_i \theta + \epsilon_i \quad (8)$$

Where  $W_i$  are the environmental variables and  $\theta$  is a vector of parameters.

As our interest is also in the determinants of the inefficiency term we allow the mean of the pretruncated normal inefficiency term to depend on exogenous determinants of inefficiency such as the education, age or gender of the farmer or access to markets and extension services. In this case equation 4 becomes:

$$u_i \sim \mathbb{N}^+(z_i \delta, \sigma_u^2) \quad (9)$$

Where  $z_i$  is a vector of exogenous determinants of inefficiency. Estimation is carried out by maximum likelihood using the density for the error term  $\epsilon_i$  and incorporating this new term. The resulting coefficients  $\delta$  are not directly interpretable but the marginal effects of the exogenous determinants of technical inefficiency on the mean of technical inefficiency  $u_i$  can be derived. We follow Kumbhakar and Sun (2013) and calculate the marginal effects on the conditional mean of technical inefficiency in keeping with the Jondrow et al (1982) estimator.

Finally, our framework for decomposing maize yields includes a measure of the economically optimal level of maize yield, The yield at which profits are maximized. This level can be identified as the point at which the relative market price of the input to the output is tangent to the frontier yield response function. The tangent or marginal physical product (MPP) of (7) with respect to one of the inputs  $k$  is

$$MPP = \frac{Y}{N} \left( \beta_k + 2\beta_k + \sum_j \ln x_j \right) \quad (10)$$

which we equate with the relative price  $\frac{P_n}{P_m}$  and solve for  $N^*$

$$\frac{Y}{X_k} \left( \beta_0 + 2\beta_k x_k + \sum_j \ln x_j \right) - \frac{P_n}{P_m} = 0 \quad (11)$$

The value of nitrogen that solves this relationship is the economically optimal nitrogen input level. However, this value cannot be found analytically. Instead, it is necessary to use numerical methods to solve for the optimal use of inputs.

## Endogeneity

Still lots to add

### testing for endogeneity

easy using control function approach

- We note from Angrist and Krueger that an OLS (instead of tobit) first stage is OK, and Wooldridge also makes this point, although neither seem to explain why. Should also mention tobit + literature on average production functions arguing for tobit

### Accounting for endogeneity

Harder using LIML approach, or two step LIML or corrected 2sls

- in progress
- mainly based on Amsler et al. (2016)

## Results

We initially test the suitability of a The resulting SF ML estimates are presented in table 1 for four models. Model 1 shows the coefficient estimates for the core translog production function terms. Model 2 extend this to include the exogenous determinants of inefficiencies. Models 3 and 4 do likewise and also include the environmental variables.

**Table 2: Translog production function**

Variable	model 1	model 2	model 3	model 4
(Intercept)	4.761**	4.221**	4.828**	4.135**
logN	0.224**	0.223**	0.188**	0.149*
loglab	0.329**	0.286**	0.324**	0.281**
logseed	0.569**	0.456**	0.507**	0.392**
logNsqr	0.046**	0.027**	0.046**	0.029**

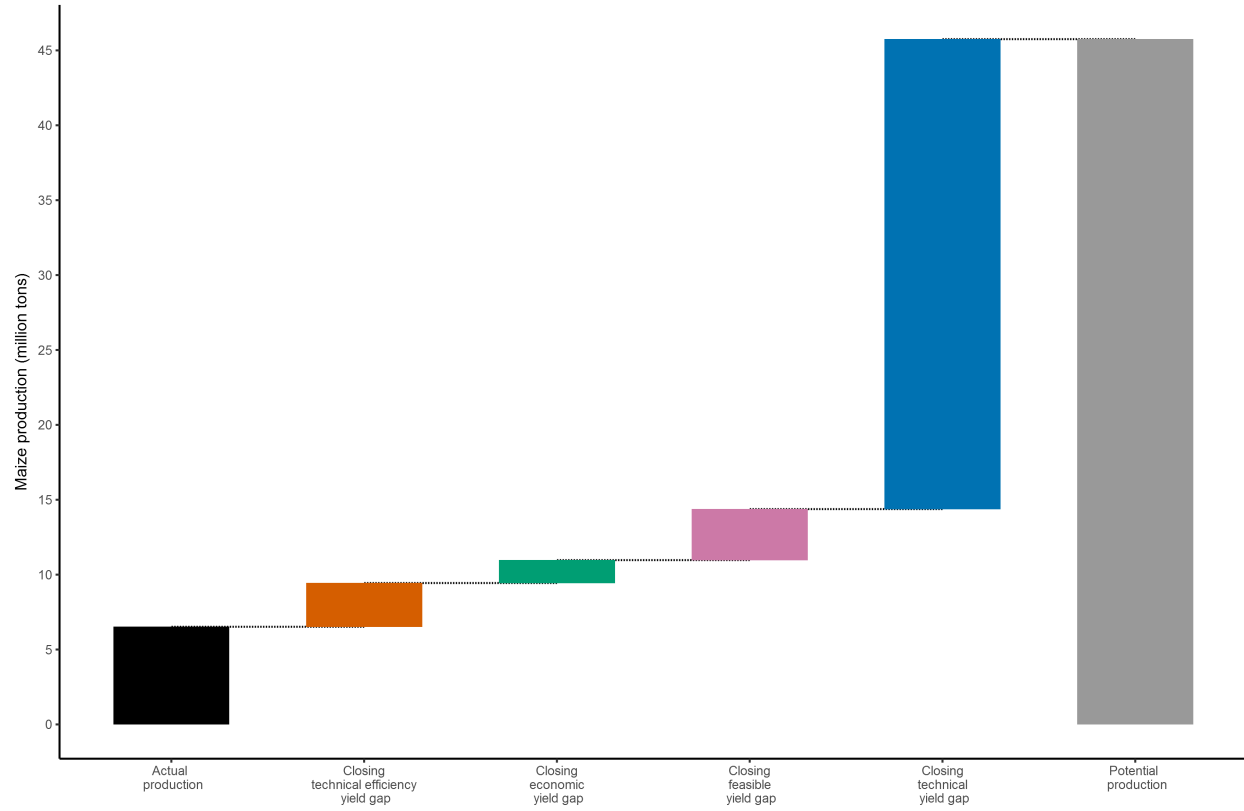
Variable	model 1	model 2	model 3	model 4
loglabsq	0.018**	0.020**	0.018**	0.019**
logseedsq	0.011	0.009	0.017	0.014
logN:loglab	-0.022**	-0.023**	-0.023**	-0.023**
logN:logseed	-0.059**	-0.036**	-0.053**	-0.026
loglab:logseed	-0.063**	-0.058**	-0.060**	-0.053**
logarea	NA	-0.099**	NA	-0.094**
phdum55_2_70	NA	0.118**	NA	0.106*
crop_count2	NA	0.421**	NA	0.426**
dumoxen	NA	0.065	NA	0.067
SOC2	NA	0.015**	NA	0.020**
logslope	NA	-0.216**	NA	-0.216**
elevation	NA	-0.000**	NA	-0.000**
GGD	NA	0.136**	NA	0.167**
AI	NA	0.035**	NA	0.038**
TS	NA	-0.000**	NA	-0.000**
Z_age	NA	NA	0.002	-0.002
Z_sex	NA	NA	-0.209	-0.047
Z_ed_any	NA	NA	0.030	-0.010
Z_title	NA	NA	0.139	0.112
Z_extension	NA	NA	-0.681**	-0.752**
Z_credit	NA	NA	-0.495*	-0.524**
Z_dist_market	NA	NA	-0.003*	0.002
Z_popEA	NA	NA	-0.000*	-0.000*
Z_logarea_tot	NA	NA	-0.123*	-0.044
sigmaSq	2.254**	1.885**	2.716**	2.241**
gamma	0.822**	0.786**	0.840**	0.809**

The raw coefficients of the exogenous determinants of technical inefficiency cannot be interpreted as elasticities. However, the marginal effects of each variable can be computed following Kumbhakar & Sun (2013). These are observation specific and a concise estimate of the marginal effects is presented in table 2 in the form of the average partial effects (APE) for models 3 and 4 respectively.

**Table 3: Marginal effects (APE) of exogenous determinants of inefficiency**

	model 3	model 4
age	0.000	-0.000
sex	-0.022	-0.006
ed_any	0.003	-0.001
title	0.014	0.013
extension	-0.070	-0.089
credit	-0.051	-0.062
dist_market	-0.000	0.000
popEA	0.000	0.000
logarea_tot	-0.013	-0.005

**Figure 3: Potential yield**



**Table 4:**

Zone	TEYG	EYG	EUYG	TYG	YG
AFAR	7.366	8.996	5.950	77.69	100
AMHARA	8.927	2.661	8.301	80.11	100
BENSHANGULGUMUZ	6.958	7.194	9.067	76.78	100
DIRE DAWA	6.595	5.817	9.707	77.88	100
GAMBELLA	15.642	23.183	2.283	58.89	100
HARARI	9.923	2.145	15.476	72.46	100
OROMIYA	6.632	3.327	9.063	80.98	100
SNNP	8.060	5.684	9.068	77.19	100
SOMALI	5.608	5.387	5.258	83.75	100
TIGRAY	7.749	10.680	3.358	78.21	100
Total	7.645	5.148	8.353	78.85	100

Aigner, Dennis, C.A.Knox Lovell, and Peter Schmidt. 1977. "Formulation and estimation of stochastic frontier production function models." *Journal of Econometrics* 6 (1). North-Holland: 21–37. doi:10.1016/0304-4076(77)90052-5.

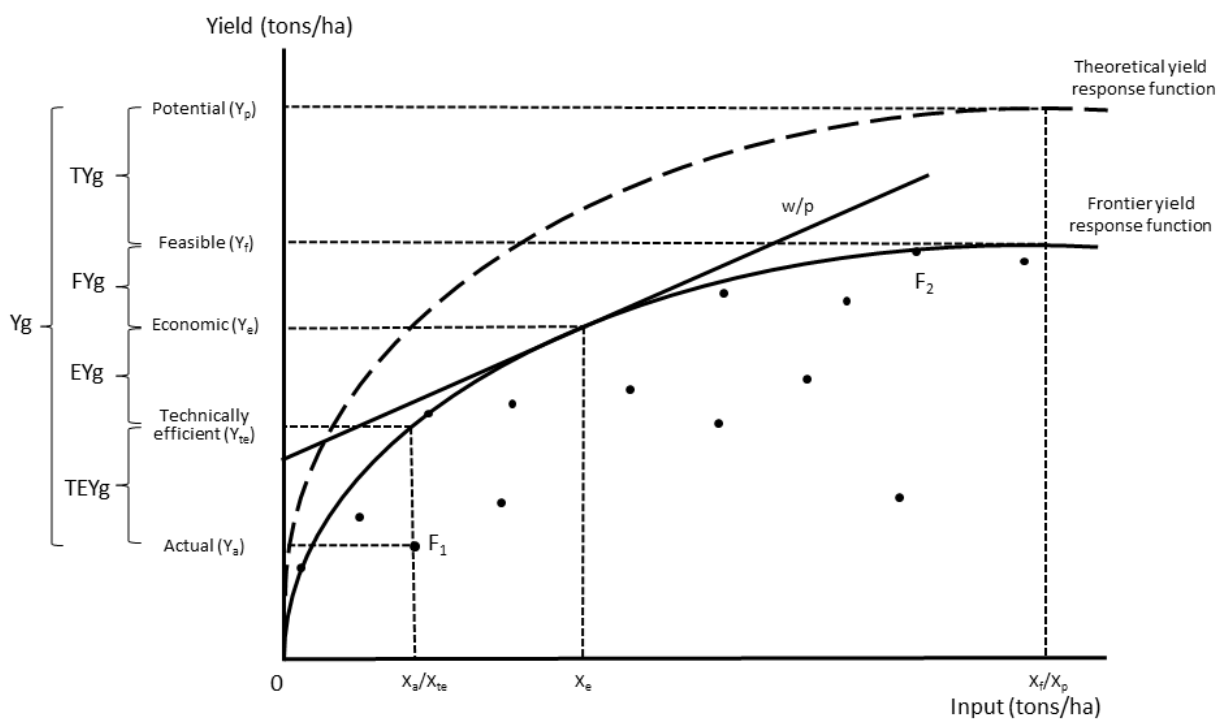


Figure 1: