



**University of Abomey-Calavi (Benin)**

\*\*\*\*\*

**Faculty of Agronomic Sciences (FSA)**

\*\*\*\*\*



Laboratoire de Biomathématiques &  
d'Estimations Forestières

**Master in Statistics with major in Biostatistics**

## **Production unit's efficiency analysis using metafrontier: Application to conventional and organic cotton in Benin**

**A thesis Submitted in Partial Fulfilment of the Requirements for the Degree of  
MASTER OF SCIENCE**

**Presented by**

**CHABI ADJOBO Ayédesso Marc Aurèle**

**Supervisors**

**Dr. Ir. Epiphane SODJINOU**  
Assistant Professor in Agricultural  
Economics and Applied Statistics

**Pr. Dr. Ir. Romain GLELE KAKAÏ,**  
Full Professor in Biomathematics and Forest  
Estimations

*Academic year: 2014-2015*

## **Certification**

We certify that this thesis has been written by Ayédesso Marc Aurèle CHABI ADJOBO under our supervision at the Faculty of Agronomic Sciences of the University of Abomey-Calavi (Benin) to obtain his Master of Science degree in Biostatistics.

Pr. Dr. Ir. Romain GLELE KAKAÏ,  
Full Professor in Biomathematics and Forest Estimations

## **Acknowledgements**

I am appreciative for the endless guidance and support received from my co-supervisor, Dr. Epiphane SODJINO. I am grateful for his help, providing me.

I am heartily thankful to my co-supervisor, Pr. Dr. Ir. Romain GLELE KAKAÏ, for his encouragement and support throughout my study and research at University of Abomey-Calavi.

I also appreciate my family members, including my father Nestor, my mother Débora, my brothers Ayéleroun and Lélayé; sisters Olouwatogni, Lébini and Oladjidé for their love, help, and support.

Finally, I would like to offer my regards and thanks to the faculty and staff of The Laboratory of Biomathematics and Forest Estimations and to all of those who supported me in any respect during the completion of my study at LaBEF especially friends of my promotion.

## **Abstract**

Comparisons of unit's efficiencies operate under different technologies remains an important issue in economic studies. This paper tries to shed light on the sensitivity of empirical results to the selection of the estimation method. The primary avenue of comparison in this analysis will be to assess the sensitivity of technical efficiency predictions to the choice of estimation method. Real data was used to compare metafrontier estimation methods. Three methods are compared in this study: stochastic frontier using all data, stochastic metafrontier and Data Envelopment Analysis (DEA) metafrontier. Data used to compare methods were collected in the framework of the Project SYPROBIO (SYstèmes de PROduction BIOlogique diversifiés) during the agricultural campaign 2012–2013. To handle performance of MF, the percentage of underestimated firm's efficiency was calculate using pooled data SF, SMF and DEA metafrontier. The paired-t.test and Spearman's rank order correlation were used to compare efficiency and TGR derived from the approaches. Results showed that the stochastic frontier using pooled data did not return the optimum output set. For both the stochastic metafrontier production function and the data envelopment analysis metafrontier, there are very large differences in the technical efficiencies. While the predicted technical efficiencies vary widely across farms for both estimations, the variations, across the two estimation methods, are significant. But no methods can be considered as significantly better than other. Difference between two methods depends only on linear programming function's used to obtain metafrontier, technologies gap ratio and the fact that the DEA frontier is not stochastic.

Keywords: DEA metafrontier, stochastic metafrontier, technical efficiency, technologies gap ratio.

## Table of content

Acknowledgements .....	i
Abstract .....	ii
Table of content.....	iii
List of tables .....	vi
List of figures .....	vii
List of Appendices .....	viii
1. Introduction.....	1
2. Literature review .....	3
2.1. Metafrontier efficiency.....	3
2.1.1. Introduction.....	3
2.1.2. Assumption .....	3
2.1.3. Principles of stochastic metafrontier.....	3
2.1.3.1. Stochastic frontier efficiency .....	4
2.1.3.2. Estimating the parameters of stochastic frontier efficiency .....	5
2.1.3.3. Technical efficiency (TE) based on stochastic frontier .....	7
2.1.3.4. Stochastic metafrontier .....	8
2.1.3.5. Estimating the parameters of stochastic metafrontier efficiency.....	9
2.1.3.6. Technical efficiency (TE) based on stochastic metafrontier .....	11
2.2. Data Envelopment Analysis (DEA) approach to Metafrontier .....	12
2.2.1. Construction of technology's DEA frontier.....	13
2.2.2. Construction of DEA metafrontier.....	13
2.3. Matching methods .....	14
2.3.1. Introduction.....	14
2.3.2. Parameters estimate in causal effect evaluation.....	15
2.3.3. Matching methods.....	16

2.3.3.1.	Variables to include .....	16
2.3.3.2.	Distance measure $D_{ij}$ .....	17
2.3.4.	Matching algorithms .....	18
3.	Methods.....	21
3.1.	Presentation of implementation example .....	21
3.2.	Estimation of SF efficiencies in each group and pooled data .....	21
3.3.	DEA frontier and DEA metafrontier .....	23
3.4.	Metafrontier.....	23
3.4.1.	Stochastic metafrontier .....	23
3.4.2.	DEA metafrontier.....	24
3.4.3.	Comparison of metafrontier methods .....	24
3.4.4.	Consistency of Propensity Scores Matching for efficiency comparison .....	24
4.	Results and discussion .....	25
4.1.	Efficiencies under each technology frontier.....	25
4.1.1.	Data Envelopment Analysis.....	25
4.1.2.	Stochastic Frontier Model.....	25
4.2.	Stochastic frontier for pooled sample and stochastic metafrontier efficiency .....	26
4.3.	DEA metafrontier efficiency.....	29
4.4.	Comparison of DEA metafrontier and Stochastic metafrontier analysis .....	31
4.4.1.	Result of paired t-test .....	31
4.4.2.	Matching methods to efficiencies comparison .....	32
4.4.2.1.	Result of logistic regression .....	32
4.4.2.2.	Characteristics of organic cotton producers versus conventional cotton producers after matching.....	33
4.4.2.3.	Support region .....	33
4.4.2.4.	Variance of treatment effects.....	34
4.5.	Discussion .....	35

5. Conclusion .....	37
6. References .....	39
APPENDIX .....	43

## List of tables

Table 1: descriptive statistics of efficiencies on literature example.....	7
Table 2: Descriptive statistics of metafrontier efficiencies for literature example .....	12
Table 3: T-test Hypothesis .....	23
Table 4: Percentage of farmer's efficiency out of frontier of all groups (%).....	27
Table 5: Descriptive statistics on efficiency using parametric method.....	27
Table 6: Technologies Gap Ratio with Stochastic Metafrontier .....	28
Table 7: Descriptive statistics on efficiency using non-parametric method .....	29
Table 8: Technologies Gap Ratio with DEA Metafrontier .....	30
Table 9: Logistic regression .....	33
Table 10: Descriptive statistics of socioeconomic characteristics of farmer's after matching:	33
Table 11: Variance of treatment effects .....	34



## List of figures

Figure 1: Distribution of technical efficiencies estimate with DEA .....	25
Figure 2: Distribution of technical efficiencies estimate with stochastic frontier.....	26
Figure 3: Distribution of Stochastic metafrontier efficiencies .....	28
Figure 4: Distribution of Technologies Gap Ratio with Stochastic Metafrontier .....	29
Figure 5: Distribution of stochastic metafrontier efficiency .....	30
Figure 6: Distribution of Technologies Gap Ratio with DEA metafrontier.....	31
Figure 7: Box plot of stochastic and DEA approaches .....	32
Figure 8: The distribution of the propensity score .....	34

## List of Appendices

Appendix 1: Categorization of the commonly used impact evaluation methods according to the available data (Sodjinou, 2012).....	44
Appendix 2: Conventional cotton SFA .....	45
Appendix 3: Organic cotton SFA .....	46
Appendix 4: Pooled data SFA .....	47
Appendix 5: Stochastic Metafrontier parameters .....	48
Appendix 6: Comparison between SFA and SM .....	49
Appendix 7: Summary of T.test result for return to scale .....	50
Appendix 8: Comparison of TGR between DEA and Stochastic Metafrontier .....	51
Appendix 9: Comparison of efficiencies between DEA and Stochastic Metafrontier.....	52
Appendix 10: Sensitivity analysis .....	53
Appendix 11: R code for Metafrontier analysis .....	54
Appendix 12: Stata code for Matching methods .....	64

# 1. Introduction

Firms intervene in a competitive economic system within which it must take a cost-benefit constraint. Indeed the firm is economic and social organizations which combine productive resources (human, material, financial ...) to produce goods or services. Therefore, in this environment, the managers or producers have to find growing point and source of inefficiency (Demsetz, 1995). Generally, a firm is efficient if it produce outputs (services) as far as feasible using amount of inputs (resources). A firm's efficiency analysis is an important area of economic analysis that has attracted the attention of economists, especially in the last three decades (Ajibefun, 2008). Several approaches have been developed and aim at analysing firm efficiency.

A problem for this type of analysis is that efficiency scores are not directly comparable across farms if they have been computed taking different technological frontiers as a reference (Beltrán-Esteve and al, 2014). Thus authors, like Hayami (1969) suggested a so called metaproduction function that examines the causes of agricultural productivity under different technologies. The framework of the metaproduction function does not necessarily imply that all producers operate on a universal production function. Each firm may choose to operate on different part of the production possibility curve, depending on its resource endowments, adoption and diffusion of technology, and economic environments. Lau and Yotopoulos (1989) employ the meta-production function approach to compare agricultural productivity across countries. This approach is econometrically advantageous due to its competence to collect data from different countries so that the scope of variations of the dependent and independent variables and the number of observations can be dramatically increased. Moreover, it reduces the possibility of multicollinearity among inputs, as the key inputs are usually changing together.

Several shortcomings exist inherent to this approach. Mainly this approach does not take into account "technology gap" characterized by the incomparability of data; the differences in the basic economic environment and the specification of an appropriate production function pose some difficulties (Villano and al, 2010).

Traditional methods attribute these "technology gaps" to technical inefficiency between farms, underlined the need to undertake methods that allow technology gaps to be distinguished from technical inefficiency. Some methods could be used to take into account "technology gaps" in estimating efficiency of farms: the stochastic metafrontier framework

(Battese and Rao, 2002; Battese and al., 2004), Data Envelopment Analysis (DEA) metafrontier framework (Rao and al., 2003), latent class models (Greene, 2004), random parameter model (Greene, 2004) and switching regression model (Sriboonchitta and Wiboonpongse, 2004). Among these methods, stochastic metafrontier and DEA metafrontier aim to compute the metafrontier of all groups frontier and have ability to estimate the technology-gap ratios, in addition to technical inefficiencies (Villano and al, 2010).

Thus, with metafrontier, comparison of efficiency is possible between farmer's that come from different technologies. The metafrontier model enables the computation of comparable technical efficiencies for farms operating under different technologies (Ali and al, 2013). It is easy, with metafrontier, to know if technologies influence efficiency of producers. An example for this situation is comparison between organic and conventional agriculture. This comparison play central role in the contribution of organic agriculture to the future of the agriculture world (Ponti and al, 2012).

While stochastic metafrontier includes production, cost, profit and perhaps revenue functions as alternative methods of describing the production technology and estimating efficiency, the Data Envelopment Analysis metafrontier involves mainly the use of linear programming techniques. The two methods have a range of advantages and disadvantages, which may influence their choice in practice. The main advantage of parametric frontier analysis, which is the stochastic metafrontier, is that it allows the test of hypothesis concerning the goodness of fit of the model. However, the major disadvantage is that it requires specification of technology distribution, which may be restrictive in most cases ().

Furthermore, the major advantage of the nonparametric frontier analysis, which is the Data Envelopment Analysis (DEA), is that it does not require the specification of a particular functional form for the technology (Rao and al., 2003). Majority of applications of frontier methodology in efficiency analysis utilize only one of the above methods at a time to estimate the production function and technical efficiency of production.

This study tries to shed light on the sensitivity of empirical results to the selection of the estimation method. The primary avenue of comparison in this analysis will be to assess the sensitivity of technical efficiency predictions to the choice of estimation method.

Specifically, this study aims to:

- compare Stochastic Frontier of pooled data and stochastic metafrontier;
- compare, using empirical example, stochastic and Data Envelopment Analysis approaches to estimate of efficiency under metafrontier;
- examine ability of PSM to compare efficiency of farms using two technologies.

## 2. Literature review

### 2.1. Metafrontier efficiency

#### 2.1.1. Introduction

The study covers 97 countries that are major agricultural producers of the world. These countries account for roughly 99 per cent of the world's agricultural output as well as 99 per cent of the world's population. The countries included in the data set are evenly distributed over all the regions of the world. The 97 countries in the data set are grouped into four regions (Africa, Americas, Asia and Europe). The aim was to identify the best performing countries, known as peers, on the one hand for each of the regions and on the others hand for the world (RAO and al, 2003). The first part of objective is handle with frontier estimation methods (SFA and DEA) for each region, implying that four different frontier must be compute. But for second part of objective, frontier for overall regions is need. Metafrontier estimation methods are then needed.

The present literature reviews is based on this example.

Since technology is a representation of the state of knowledge pertaining to the transformation of  $N$  inputs into  $M$  outputs, it is possible to conceptualise the existence of an over-arching technology, referred to as the metatechnology, which we represent by  $T^*$  (RAO and al, 2003). Therefore, the metatechnology is the totality of the technologies. For example, if a particular output,  $y$ , can be produced using a given input vector,  $X$ , in any one of the technologies, we consider that  $(X, Y)$  belongs to the metatechnology,  $T^*$ , that is defined by:

$T^* = \{(x, y): x \geq 0 \text{ and } y \geq 0, \text{ such that } x \text{ can produce } y \text{ in at least one regional technology, } T_1, T_2, \dots, T_K\}.$

$$T^* = \{(x,y):x \geq 0 \text{ and } y \geq 0, \text{ such that } x \text{ can produce } y \text{ in at one technolog y}\}$$

#### 2.1.2. Assumption

All of farmers must have potential access to all set of technologies. There is not condition on noise and inefficiency distribution in Meta frontier efficiency (Villano, 2010).

#### 2.1.3. Principles of stochastic metafrontier

Let assumes that there are  $J$  well-defined groups for the same industry, such as different regions within a country, different ethnic groups involved in production or different technologies.

### 2.1.3.1. Stochastic frontier efficiency

Suppose that, for the  $j^{\text{th}}$  group, there are sample data on  $N_j$  firms that produce one output from the various inputs. The stochastic frontier model for this group is defined by:

$$Y_{ij} = f(x_{ij}, \beta) e^{V_{ij} - U_{ij}}, \quad i = 1, 2, \dots, N_j \quad (1)$$

where  $Y_{ij}$  denotes the output for the  $i^{\text{th}}$  firm in the  $j^{\text{th}}$  group;  $x_{ij}$  denotes a vector of functions of the inputs used by the  $i^{\text{th}}$  firm in the  $j^{\text{th}}$  group;  $V_{ij}$  represents statistical noise assumed to be independently and identically distributed as  $N(0, s_{V_j})$  random variables; and  $U_{ij}$  are non-negative random variables assumed to account for technical inefficiency in production and assumed to be independently distributed as truncations at zero of the  $N(\mu_j, \sigma_{U_j})$  distribution.

The next step in the estimation of the relationship between the output and the explanatory variables is to specify the algebraic (or functional) form of the function  $f(\cdot)$  of equation 1.

The expression of equation (1) assumes that the exponent of the frontier production function is linear in the parameter vector  $\beta$ , so that  $x_{ij}$  is a vector of functions of (logarithms of) the inputs for the  $i^{\text{th}}$  firm in the  $j^{\text{th}}$  group involved. The Cobb-Douglas or translog production functions are convenient for the presentation of the decomposition below:

Cobb-Douglas:

$$\ln y_{ij} = \beta_{0j} + \sum_{k=1}^K \beta_{kj} \ln x_{kij} + V_{ij} - U_{ij} \quad (2)$$

Translog:

$$\ln y_{ij} = \beta_{0j} + \sum_{k=1}^K \beta_{kj} \ln x_{kij} + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{klij} \ln x_{kij} \ln x_{lij} + V_{ij} - U_{ij} \quad (3)$$

Where  $k=1=1, 2, \dots, K$  and  $x_{kij}$  or  $x_{lij}$  denote the value of  $k^{\text{th}}$  or  $l^{\text{th}}$  input for the  $i^{\text{th}}$  firm in the  $j^{\text{th}}$  group.

Consider example in introduction of literature, and using the translog production function, the fourth regional frontiers were compute using:

- Africa

$$\begin{aligned} \ln \text{YIELD}_{i1} = & \beta_{01} + \beta_{11} \ln X1_{i1} + \beta_{21} \ln X2_{i1} + \beta_{31} \ln X3_{i1} + \beta_{41} \ln X4_{i1} + \beta_{51} \ln X5_{i1} + \\ & 0.5 * \beta_{61} \ln X1_{i1} * \ln X1_{i1} + 0.5 * \beta_{71} \ln X1_{i1} * \ln X2_{i1} + 0.5 * \beta_{81} \ln X1_{i1} * \ln X3_{i1} + \\ & 0.5 * \beta_{91} \ln X1_{i1} * \ln X4_{i1} + 0.5 * \beta_{101} \ln X1_{i1} * \ln X5_{i1} + 0.5 * \beta_{111} \ln X2_{i1} * \ln X2_{i1} + \\ & 0.5 * \beta_{121} \ln X2_{i1} * \ln X3_{i1} + 0.5 * \beta_{131} \ln X2_{i1} * \ln X4_{i1} + 0.5 * \beta_{141} \ln X2_{i1} * \ln X5_{i1} + \\ & 0.5 * \beta_{151} \ln X3_{i1} * \ln X2_{i1} + 0.5 * \beta_{161} \ln X3_{i1} * \ln X4_{i1} + 0.5 * \beta_{171} \ln X3_{i1} * \ln X5_{i1} + \\ & 0.5 * \beta_{181} \ln X4_{i1} * \ln X4_{i1} + 0.5 * \beta_{191} \ln X4_{i1} * \ln X5_{i1} + 0.5 * \beta_{201} \ln X5_{i1} * \ln X5_{i1} + \\ & V_{i1} - U_{i1} \end{aligned}$$

- Americas

$$\begin{aligned} \ln \text{YIELD}_{i1} = & \beta_{02} + \beta_{12} \ln X1_{i2} + \beta_{22} \ln X2_{i2} + \beta_{32} \ln X3_{i2} + \beta_{42} \ln X4_{i2} + \beta_{52} \ln X5_{i2} + \\ & 0.5 * \beta_{62} \ln X1_{i2} * \ln X1_{i2} + 0.5 * \beta_{72} \ln X1_{i2} * \ln X2_{i2} + 0.5 * \beta_{82} \ln X1_{i2} * \ln X3_{i2} + \\ & 0.5 * \beta_{92} \ln X1_{i2} * \ln X4_{i2} + 0.5 * \beta_{102} \ln X1_{i2} * \ln X5_{i2} + 0.5 * \beta_{112} \ln X2_{i2} * \ln X2_{i2} + \\ & 0.5 * \beta_{122} \ln X2_{i2} * \ln X3_{i2} + 0.5 * \beta_{132} \ln X2_{i2} * \ln X4_{i2} + 0.5 * \beta_{142} \ln X2_{i2} * \ln X5_{i2} + \\ & 0.5 * \beta_{152} \ln X3_{i2} * \ln X2_{i2} + 0.5 * \beta_{162} \ln X3_{i2} * \ln X4_{i2} + 0.5 * \beta_{172} \ln X3_{i2} * \ln X5_{i2} + \\ & 0.5 * \beta_{182} \ln X4_{i2} * \ln X4_{i2} + 0.5 * \beta_{192} \ln X4_{i2} * \ln X5_{i2} + 0.5 * \beta_{202} \ln X5_{i2} * \ln X5_{i2} + \\ & V_{i1} - U_{i1} \end{aligned}$$

OUTPUT (Y): This paper considers two output aggregates, viz., crops and livestock outputs.

LAND (X1): This variable includes the arable land, land under permanent crops as well as the area under permanent pasture, expressed in millions of hectares.

MACHINE(X2): This variable includes the total number of wheeled and crawler tractors used in agriculture, but exclude garden tractors.

LABOUR(X3): The labour variable used is the economically active population in agriculture, which is defined as all persons engaged in or seeking employment in the operation of a family farm or business

FERTILISER (X4): This input is the sum of the nitrogen (N), potassium (P2O2) and phosphate (K2O) contained in the commercial fertilizers that were applied as the measure of fertilizer input in this paper. This variable is expressed in thousands of tonnes.

LIVESTOCK(X5): The livestock input variable used in the study is the sheep-equivalent of five categories of animals. The categories of animals considered are buffaloes, cattle, pigs, sheep and goats.

### 2.1.3.2. Estimating the parameters of stochastic frontier efficiency

The estimation methods are underpinned by some assumptions on two random variables  $U_{ij}$  and  $V_{ij}$  (Greene, 1980):

- each  $U_{ij}$  is distributed independently to each  $V_{ij}$ ;
- $U_{ij}$  and  $V_{ij}$  are uncorrelated with explanatory variables  $x_{ij}$ ;
- $E(U_{ij}) = 0$
- homoscedastic :  $E(V_{ij}^2) = \sigma_{V_j}^2$ ;

Maximum Likelihood (ML) estimators are often used to estimate parameters because it is asymptotically more efficient than other methods (Greene, 1980). The ML method is based on the maximization of the log-likelihood function. For the efficiency models, under the assumption of normality of noise terms ( $V_{ij} \sim \text{iidN}(0, \sigma_v^2)$ ) and quasi-normality for inefficiency term ( $U_{ij} \sim \text{iidN}^+(0, \sigma_u^2)$ ), the log-likelihood function of Y is:

$$\ln L(y | \beta, \sigma, \lambda) = -\frac{1}{2} \ln \left( \frac{\pi \sigma^2}{2} \right) + \sum_{i=1}^I \ln \Phi \left( -\frac{\varepsilon_i \lambda}{\sigma} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^I \varepsilon_i \quad (4)$$

where  $y$  is a vector of log-output,  $\sigma^2 = \sigma_u^2 + \sigma_v^2$ ;  $\lambda^2 = \sigma_u^2 / \sigma_v^2 \geq 0$ ;  $\Phi(x)$  is a cumulative distribution function of a standard random variable evaluated  $x$ ;  $\varepsilon_i = V_i - U_i = \ln q_i - X_i \beta$

Maximizing  $\ln L(y | \beta, \sigma, \lambda)$  with respect to parameters  $\beta$ , variance components  $\sigma$  and  $\lambda$  yield equations for ML estimators of  $\beta$ ,  $\sigma$  and  $\lambda$ . These equations are solved iteratively.

Maximizing in example presented above allows obtaining:

- Africa

$$\begin{aligned} \ln \text{YIELD}_{i1} = & -0.2 + 1.03 \ln X1_{i1} + 0.03 \ln X2_{i1} + 10.15 \ln X3_{i1} - 0.04 \ln X4_{i1} - 0.39 \ln X5_{i1} + \\ & 0.5 * (-0.36) \ln X1_{i1} * \ln X1_{i1} + 0.5 * 0.075 \ln X1_{i1} * \ln X2_{i1} + 0.5 * 0.06 \ln X1_{i1} * \ln X3_{i1} + \\ & 0.5 * 0.013 \ln X1_{i1} * \ln X4_{i1} + 0.5 * (-0.08) \ln X1_{i1} * \ln X5_{i1} + 0.5 * 0.083 \ln X2_{i1} * \ln X2_{i1} + \\ & 0.5 * 0.22 \ln X2_{i1} * \ln X3_{i1} + 0.5 * (-0.014) \ln X2_{i1} * \ln X4_{i1} + 0.5 * (-0.121) \ln X2_{i1} * \ln X5_{i1} + \\ & 0.5 * 0.45 \ln X3_{i1} * \ln X2_{i1} + 0.5 * 0.025 \ln X3_{i1} * \ln X4_{i1} + 0.5 * (-0.27) \ln X3_{i1} * \ln X5_{i1} + \\ & 0.5 * 0.010 \ln X4_{i1} * \ln X4_{i1} + 0.5 * 0.009 \ln X4_{i1} * \ln X5_{i1} + 0.5 * 0.39 \ln X5_{i1} * \ln X5_{i1} + \\ & V_{i1} - U_{i1} \end{aligned}$$

- Americas

$$\begin{aligned} \ln \text{YIELD}_{i1} = & 1.27 - 0.76 \ln X1_{i2} + 1.13 \ln X2_{i2} + 0.44 \ln X3_{i2} + 0.15 \ln X4_{i2} - 0.72 \ln X5_{i2} + \\ & 0.5 * (-0.89) \ln X1_{i2} * \ln X1_{i2} + 0.5 * 0.267 \ln X1_{i2} * \ln X2_{i2} + 0.5 * 0.16 \ln X1_{i2} * \ln X3_{i2} + \\ & 0.5 * 0.045 \ln X1_{i2} * \ln X4_{i2} + 0.5 * 0.65 \ln X1_{i2} * \ln X5_{i2} + 0.5 * 0.05 \ln X2_{i2} * \ln X2_{i2} + \\ & 0.5 * (-0.03) \ln X2_{i2} * \ln X3_{i2} + 0.5 * (-0.004) \ln X2_{i2} * \ln X4_{i2} + 0.5 * (-0.45) \ln X2_{i2} * \ln X5_{i2} + \\ & 0.5 * (-0.18) \ln X3_{i2} * \ln X2_{i2} + 0.5 * 0.035 \ln X3_{i2} * \ln X4_{i2} + 0.5 * (-0.148) \ln X3_{i2} * \ln X5_{i2} + \\ & 0.5 * 0.021 \ln X4_{i2} * \ln X4_{i2} + 0.5 * (-0.035) \ln X4_{i2} * \ln X5_{i2} + 0.5 * 0.12 \ln X5_{i2} * \ln X5_{i2} + \\ & V_{i1} - U_{i1} \end{aligned}$$



### 2.1.3.3. Technical efficiency (TE) based on stochastic frontier

Under assumption making on inefficiency term  $U_{ij}$  ( $U_{ij} \sim \text{iidN}^+(0, \sigma_U^2)$ ), the information about likely and unlikely value of  $U_{ij}$  can be summarise in the form of truncated normal Probability Distribution Function (pdf) (Coelli, 2005):

$$p(U_{ij} | q_i) = \frac{1}{\sqrt{2\pi\sigma_*^2}} \exp\left\{-\frac{1}{2\sigma_*^2}(U_{ij} - U_{ij}^*)^2\right\} / \Phi\left(\frac{U_{ij}^*}{\sigma_*}\right) \quad (5)$$

Where  $U_{ij}^* = -(\ln q_{ij} - X_{ij}\beta_j)\sigma_{U_j}^2 / \sigma_j^2$  and  $\sigma_*^2 = \sigma_{V_j}^2 \sigma_{U_j}^2 / \sigma_j^2$

This pdf has been used by Jondrow and al (1982) to derive the following predictors of  $U_{ij}$  :

$$\hat{U}_{ij} \equiv E\{U_{ij} | q_{ij}\} = U_{ij}^* + \sigma_* \left[ \frac{\phi(U_{ij}^*/\sigma_*)}{\Phi(U_{ij}^*/\sigma_*)} \right] \quad (6)$$

where  $\phi(x)$  is the pdf of standard normal random variable evaluated at  $x$ . A technical efficiency can be obtain with  $\hat{TE}_{ij} = \exp(-\hat{U}_{ij})$

Battese and Coelli (1988) also derive alternative predictors of  $TE$  based on  $p(U_{ij} | q_i)$  :

$$\hat{TE}_{ij} = E\{\exp(-U_{ij}) | q_{ij}\} = \left[ \Phi\left(\frac{U_{ij}^*}{\sigma_*} - \sigma_*\right) / \Phi\left(\frac{U_{ij}^*}{\sigma_*}\right) \right] \exp\left\{\frac{\sigma_*^2}{2} - U_{ij}^*\right\} \quad (7)$$

This predictor minimizes the mean errors squares and then can be considered as the optimal predictors (Coelli, 2005).

On application data, descriptive statistics of efficiencies is

Table 1: descriptive statistics of efficiencies on literature example

	Mean	Standard Errors	Minimum	Maximum
Africa	0.499	0.253	0.200	0.971
Americas	0.751	0.170	0.504	0.970
Asia	0.706	0.194	0.348	0.981
Europe	0.756	0.154	0.451	0.980

This efficiency can't allow comparison of countries in different regions because there efficiencies are compute under different frontier. It is reason why Rao (2003) use metafrontier.

#### 2.1.3.4. Stochastic metafrontier

The metafrontier production function model for units in the industry is expressed by (Battese and al, 2004):

$$Y_i^* \equiv f(x_i, \beta^*) = e^{x_i \beta^*}, \quad i = 1, 2, \dots, N = \sum_{j=1}^J N_j \quad (8)$$

Where  $\beta^*$  denotes the vector of parameters for the metafrontier function such that

$$x_i \beta^* \geq x_{ij} \beta_j \quad (9)$$

The assumptions for the stochastic frontiers for the different groups are reasonable for given sample data, but these assumptions associated with the stochastic metafrontier model may not be appropriate (e.g., the  $U_{ij}$ 's may not be identically distributed over all groups) (RAO, 2003).

Parameters of the metafrontier model are estimated using data from firms in all groups (in the combined data set). The stochastic frontiers for the different groups and that of the metafrontier would generally be assumed to be of the same functional form (e.g., Cobb-Douglas or translog).

Cobb-Douglas:

$$\ln Y_i = \beta_0^* + \sum_{k=1}^K \beta_k^* \ln x_{ki} \quad (10)$$

$$\text{With } \beta_0^* + \sum_{k=1}^K \beta_k^* \ln x_{ki} \geq \beta_{0j} + \sum_{k=1}^K \beta_{kj} \ln x_{kij}$$

Translog

$$\ln Y_i^* = \beta_i^* + \sum_{k=1}^K \beta_k^* \ln x_i + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{kl}^* \ln x_{ki} \ln x_{li} \quad (11)$$

$$\text{With } \beta_i^* + \sum_{k=1}^K \beta_k^* \ln x_i + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{kl}^* \ln x_{ki} \ln x_{li} \geq \beta_{0j} + \sum_{k=1}^K \beta_{kj} \ln x_{ij} + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{kl}^* \ln x_{kij} \ln x_{lij} \quad (12)$$

Using application example , the translog functional form model is :

$$\begin{aligned} \ln \text{YIELD}_i = & \beta_0^* + \beta_1^* \ln X1_i + \beta_2^* \ln X2_i + \beta_3^* \ln X3_i + \beta_4^* \ln X4_i + \beta_5^* \ln X5_i + \\ & 0.5 * \beta_6^* \ln X1_{i1} * \ln X1_{i1} + 0.5 * \beta_7^* \ln X1_{i1} * \ln X2 + 0.5 * \beta_8^* \ln X1_{i1} * \ln X3_i + \\ & 0.5 * \beta_9^* \ln X1_i * \ln X4_i + 0.5 * \beta_{10}^* \ln X1_i * \ln X5_i + 0.5 * \beta_{11}^* \ln X2_i * \ln X2_i + \\ & 0.5 * \beta_{12}^* \ln X2_i * \ln X3_{i1} + 0.5 * \beta_{13}^* \ln X2_i * \ln X4_i + 0.5 * \beta_{14}^* \ln X2_i * \ln X5_i + \\ & 0.5 * \beta_{15}^* \ln X3_i * \ln X2_i + 0.5 * \beta_{16}^* \ln X3_i * \ln X4_i + 0.5 * \beta_7^* \ln X3_i * \ln X5_i + \\ & 0.5 * \beta_{18}^* \ln X4_i * \ln X4_i + 0.5 * \beta_{19}^* \ln X4_i * \ln X5_i + 0.5 * \beta_{20}^* \ln X5_i * \ln X5_i + \end{aligned}$$

where  $i = 1, 2, \dots, N = N_1 + N_2 + N_3 + N_4 = 97$  with

$$\left\{ \begin{array}{l}
\beta_0^* + \beta_1^* \ln X1_i + \beta_2^* \ln X2_i + \beta_3^* \ln X3_i + \beta_4^* \ln X4_i + \beta_5^* \ln X5_i + 0.5 * \beta_6^* \ln X1_{i1} * \ln X1_{i1} + \\
0.5 * \beta_7^* \ln X1_{i1} * \ln X2 + 0.5 * \beta_8^* \ln X1_{i1} * \ln X3_i + 0.5 * \beta_9^* \ln X1_i * \ln X4_i + 0.5 * \beta_{10}^* \ln X1_i * \ln X5_i + \\
0.5 * \beta_{11}^* \ln X2_i * \ln X2_i + 0.5 * \beta_{12}^* \ln X2_i * \ln X3_{i1} + 0.5 * \beta_{13}^* \ln X2_i * \ln X4_i + 0.5 * \beta_{14}^* \ln X2_i * \ln X5_i + \\
0.5 * \beta_{15}^* \ln X3_i * \ln X2_i + 0.5 * \beta_{16}^* \ln X3_i * \ln X4_i + 0.5 * \beta_7^* \ln X3_i * \ln X5_i + 0.5 * \beta_{18}^* \ln X4_i * \ln X4_i + \\
0.5 * \beta_{19}^* \ln X4_i * \ln X5_i + 0.5 * \beta_{20}^* \ln X5_i * \ln X5_i \geq \beta_{01} + \beta_{11} \ln X1_{i1} + \beta_{21} \ln X2_{i1} + \beta_{31} \ln X3_{i1} + \\
\beta_{41} \ln X4_{i1} + \beta_{51} \ln X5_{i1} + 0.5 * \beta_{61} \ln X1_{i1} * \ln X1_{i1} + 0.5 * \beta_{71} \ln X1_{i1} * \ln X2 + 0.5 * \beta_{81} \ln X1_{i1} * \ln X3_{i1} + \\
0.5 * \beta_{91} \ln X1_{i1} * \ln X4_{i1} + 0.5 * \beta_{10,1} \ln X1_{i1} * \ln X5_{i1} + 0.5 * \beta_{11,1} \ln X2_{i1} * \ln X2_{i1} + \\
0.5 * \beta_{12,1} \ln X2_{i1} * \ln X3_{i1} + 0.5 * \beta_{13,1} \ln X2_{i1} * \ln X4_{i1} + 0.5 * \beta_{14,1} \ln X2_{i1} * \ln X5_{i1} + \\
0.5 * \beta_{15,1} \ln X3_{i1} * \ln X2_{i1} + 0.5 * \beta_{16,1} \ln X3_{i1} * \ln X4_{i1} + 0.5 * \beta_{17,1} \ln X3_{i1} * \ln X5_{i1} + \\
0.5 * \beta_{18,1} \ln X4_{i1} * \ln X4_{i1} + 0.5 * \beta_{19,1} \ln X4_{i1} * \ln X5_{i1} + 0.5 * \beta_{20,1} \ln X5_{i1} * \ln X5_{i1} + \\
\text{and} \\
\beta_0^* + \beta_1^* \ln X1_i + \beta_2^* \ln X2_i + \beta_3^* \ln X3_i + \beta_4^* \ln X4_i + \beta_5^* \ln X5_i + 0.5 * \beta_6^* \ln X1_{i1} * \ln X1_{i1} + \\
0.5 * \beta_7^* \ln X1_{i1} * \ln X2 + 0.5 * \beta_8^* \ln X1_{i1} * \ln X3_i + 0.5 * \beta_9^* \ln X1_i * \ln X4_i + 0.5 * \beta_{10}^* \ln X1_i * \ln X5_i + \\
0.5 * \beta_{11}^* \ln X2_i * \ln X2_i + 0.5 * \beta_{12}^* \ln X2_i * \ln X3_{i1} + 0.5 * \beta_{13}^* \ln X2_i * \ln X4_i + 0.5 * \beta_{14}^* \ln X2_i * \ln X5_i + \\
0.5 * \beta_{15}^* \ln X3_i * \ln X2_i + 0.5 * \beta_{16}^* \ln X3_i * \ln X4_i + 0.5 * \beta_7^* \ln X3_i * \ln X5_i + 0.5 * \beta_{18}^* \ln X4_i * \ln X4_i + \\
0.5 * \beta_{19}^* \ln X4_i * \ln X5_i + 0.5 * \beta_{20}^* \ln X5_i * \ln X5_i \geq \beta_{02} + \beta_{12} \ln X1_{i2} + \beta_{22} \ln X2_{i2} + \beta_{32} \ln X3_{i2} + \\
\beta_{42} \ln X4_{i2} + \beta_{52} \ln X5_{i2} + 0.5 * \beta_{62} \ln X1_{i2} * \ln X1_{i2} + 0.5 * \beta_{72} \ln X1_{i2} * \ln X2_{i2} + 0.5 * \beta_{82} \ln X1_{i2} * \ln X3_{i2} + \\
0.5 * \beta_{92} \ln X1_{i2} * \ln X4_{i2} + 0.5 * \beta_{10,2} \ln X1_{i2} * \ln X5_{i2} + 0.5 * \beta_{11,2} \ln X2_{i2} * \ln X2_{i2} + \\
0.5 * \beta_{12,2} \ln X2_{i2} * \ln X3_{i1} + 0.5 * \beta_{13,2} \ln X2_{i2} * \ln X4_{i1} + 0.5 * \beta_{14,2} \ln X2_{i2} * \ln X5_{i1} + \\
0.5 * \beta_{15,2} \ln X3_{i1} * \ln X2_{i1} + 0.5 * \beta_{16,2} \ln X3_{i1} * \ln X4_{i1} + 0.5 * \beta_{17,2} \ln X3_{i1} * \ln X5_{i1} + \\
0.5 * \beta_{18,2} \ln X4_{i1} * \ln X4_{i1} + 0.5 * \beta_{19,2} \ln X4_{i1} * \ln X5_{i1} + 0.5 * \beta_{20,2} \ln X5_{i1} * \ln X5_{i1}
\end{array} \right.$$

### 2.1.3.5. Estimating the parameters of stochastic metafrontier efficiency

The MF model is a deterministic parametric frontier of specific functional form (e.g., Cobb Douglas or Translog) such that the predicted value for the MF is larger than or equal to the predicted value from the stochastic frontier for all firms and groups (Battese and Dalzell, 2004).

To identify the best envelope, it is necessary to specify criterion that can be used in judging the best. Two criteria can be considered, one based on the sum of absolute deviations and the other based on the sum of squares of deviations of the metafrontier values from those of the group frontiers.

#### ✓ Minimum Sum of Absolute Deviations

Given the estimates for the parameters of the group stochastic frontiers  $\beta_j$ ,  $j=1,2,\dots,J$ , the  $\beta^*$ -parameters can be estimated by solving the optimization problem below:

$$\text{Min}L = \sum_{i=1}^N |\ln f(x_i, \beta^*) - \ln f(x_i, \beta_j)| \quad (13)$$

$$\text{s.t. } \ln f(x_i, \beta^*) \geq \ln f(x_i, \beta_j) \quad (14)$$

There are several interesting features to the application of this criterion:

1. the deviations used here are essentially logarithms of  $f(x_i, \beta^*)/f(x_i, \beta_j)$ , which represent the radial distance between the metafrontier and the j-th group frontier; thus the use of (13) and (14) implies that the resulting metafrontier minimizes the sum of logarithmic radial distances between the metafrontier and group frontiers;
2. since the optimization is subject to the inequality restrictions in (14), all the deviations involved are positive and, therefore, the absolute deviations are simply equal to the deviations;
3. if  $f(x_i, \beta^*)$  in equation (8) is assumed to be log-linear in the parameters (as it is in these papers), the optimization problem in (15) and (14) simplifies to the following linear programming (LP) problem:

$$\text{MinL} \equiv \sum_{i=1}^N |x_i, \beta^* - x_i, \beta_j| \quad (16)$$

$$\text{s.t. } x_i, \beta^* \geq x_i, \beta_j \quad (17)$$

As the vector of estimated coefficients for the stochastic frontier for each j-th group, and the input vectors are assumed to be fixed, the above minimization can be solved by linear programming:

$$\min_{\beta^*} \bar{x}\beta^* \quad (18)$$

$$\text{s.t. } x_i\beta^* \geq x_i\beta_j$$

where  $\bar{x}$  is the row vector of means of the elements of the  $x_i$ -vectors for all observations in the data set.

#### ✓ **Minimum Sum of Squares of Deviations**

Minimization of the objective function in (13) and (14) assigns the same weight for all the radial distances for all the farms in the sample. An alternative approach is to estimate the parameters of the metafrontier function by minimizing the sum of squares of the deviations of the values on the metafrontier function from the group- specific stochastic frontiers. This method assigns higher weights to the deviations associated with farms that have larger technology& gap ratios. This leads to the following optimization problem:

$$\text{MinL} \equiv \sum_{i=1}^N (x_i, \beta^* - x_i, \beta_j)^2 \quad (19)$$

Subject to (16)

This approach is similar to the use of the least-squares criterion. The optimization problem in (19), identical to constrained least-squares estimation, is quadratic programming (QP) problem that minimizes the distances of the values on the metafrontier from those on the estimated stochastic frontier functions.

In this paper, the criterion of Minimum Sum of Absolute Deviations is considered.

#### 2.1.3.6. Technical efficiency (TE) based on stochastic metafrontier

The observed output defined by the stochastic frontier for  $J$ -th the group in Equation (1) can be alternatively expressed in terms of the meta-frontier function in Equation (8), such that:

$$y_{ij} = e^{-U_{ij}} \times \frac{e^{x_i\beta_j}}{e^{x_i\beta^*}} \times e^{x_i\beta^* + V_{ij}} \quad (20)$$

The first term on the right-hand side of equation (20) is the technical efficiency relative to the stochastic frontier for the  $J$ -th group.

$$TE_{ij} = e^{-U_{ij}} = \frac{y_{ij}}{e^{x_i\beta_j + V_{ij}}} \quad (21)$$

The second term is what Battese and Rao (2002) term the technology gap ratio (TGR), which is expressed as

$$TGR_{ij} = \frac{e^{x_i\beta_j}}{e^{x_i\beta^*}} \quad (22)$$

The TGR measures the ratio of the output for the frontier production function for the  $J$ -th group relative to the potential output that is defined by the meta-frontier function, given the observed inputs (Battese and Rao, 2002; Battese and al., 2004). The TGR has values between zero and one.

The technical efficiency of the  $i$ -th firm, relative to the metafrontier, denoted by  $TE_i^*$ , is defined in an analogous way to equation (21). It is the ratio of the observed output relative to the last term on the right-hand side of equation (20), which is the metafrontier output, adjusted for the corresponding random error, i.e.,

$$TE_i^* = \frac{y_{ij}}{e^{x_i\beta^* + V_{ij}}} \quad (23)$$

Accordingly, following Equations (20), (21), (22) and (23),  $TE_i^*$  can be expressed as:

$$TE_i^* = TE_{ij} \times TGR_{ij} \quad (24)$$

Using the following result is obtained. Based on results on table1, it is clear that the best country is in Africa but if the below result is considered, the best country is in America.

Table 2: Descriptive statistics of metafrontier efficiencies for literature example

	Mean	Standard Errors	Minimum	Maximum
Africa	0.365	0.176	0.155	0.868
Americas	0.559	0.164	0.336	0.970

## 2.2. Data Envelopment Analysis (DEA) approach to Metafrontier

DEA is a linear-programming methodology that uses data on inputs and outputs of firms, production entities or decision-making units (referred to as units in this and the next section) to construct a piece-wise linear surface over the data points. This frontier surface is constructed by the solution of a sequence of linear programming problems one for each unit in the sample. The degree of technical inefficiency of each unit (the distance between the observed data point and the frontier) is produced as a by-product of the frontier construction method.

In simple terms, DEA attempts to benchmark the performance of each unit against the best practice for all units. The “best-practice” is derived after taking into account the output structure (e.g., share of crops and livestock in total output) as well as the input structure of the unit under consideration. A measure of technical efficiency is then obtained by measuring the radial distance of the unit from its best practice. DEA can be either input-orientated or output-orientated. In the input-orientated case, DEA defines the frontier by seeking the maximum possible proportional reduction in input usage, with output levels held constant, for each unit. In the output-orientated case, the DEA method seeks the maximum possible proportional increase in outputs, with input levels held fixed. The two measures provide the same technical efficiency scores when a constant returns-to-scale (CRS) technology applies, but are unequal when variable returns-to-scale (VRS) technology holds.

Given that units are within technologies, it is possible to identify a “technology’s frontier” using DEA on the data for units from the given region. Thus, DEA can be used to construct K technologies frontiers. The metafrontier is then constructed by using DEA to analyse the data set obtained by pooling all the observations for units from all the regions.

### 2.2.1. Construction of technology's DEA frontier

Suppose that, for the  $j^{\text{th}}$  group, there are sample data on  $N_j$  firms that produce one output from the various inputs. The linear programming (LP) problem that is solved for the  $i^{\text{th}}$  unit in an output-orientated DEA model is as follows (RAO and al, 2003):

$$\begin{aligned} \max_{\phi, \lambda} \quad & \phi \\ & -\phi y_i + Y_j \lambda \geq 0 \\ \text{Such that} \quad & x_i - X_j \lambda \geq 0 \\ & \lambda \geq 0, \end{aligned}$$

where

$y_i$  is the  $M \times 1$  vector of output quantities for the  $i^{\text{th}}$  unit;

$x_i$  is the  $N \times 1$  vector of input quantities for the  $i^{\text{th}}$  unit;

$Y_j$  is the  $M \times N_j$  matrix of output quantities for all  $N_j$  units;

$X_j$  is the  $N \times N_j$  matrix of input quantities for all  $N_j$  units;

$\lambda$  is a  $N_j \times 1$  vector of weights; and

$\phi$  is a scalar.

Each LP produces  $\phi$  and  $\lambda$  vectors. The  $\phi$ -vector provides information on the technical efficiency score for the  $i^{\text{th}}$  unit and the  $\lambda$ -vector provides information on the *peers* of the (inefficient)  $i^{\text{th}}$  unit. Then

$$TE_j = \frac{1}{\phi}$$

### 2.2.2. Construction of DEA metafrontier

The metafrontier is constructed using a DEA model based on the pooled data for all the units in all the regions. Since we have a total of  $N = \sum N_j$  units, we re-run the above LP model with the input and output matrices with data for all units (RAO and al, 2003):

$$\begin{aligned} \max_{\phi^*, \lambda^*} \quad & \phi^* \\ & -\phi^* y_i + Y_j^* \lambda^* \geq 0 \\ \text{Such that} \quad & x_i - X_j^* \lambda^* \geq 0 \\ & \lambda^* \geq 0, \end{aligned}$$

where

$y_i$  is the  $M \times 1$  vector of output quantities for the  $i^{\text{th}}$  unit;

$x_i$  is the  $N \times 1$  vector of input quantities for the  $i^{th}$  unit;

$Y_j^*$  is the  $M \times N$  matrix of output quantities for all  $N$  units;

$X_j^*$  is the  $N \times N$  matrix of input quantities for all  $N$  units;

$\lambda^*$  is a  $N \times 1$  vector of weights; and

$\phi^*$  is a scalar.

The optimum solution provides a technical efficiency score for a given unit relative to the metafrontier identified using data from all the units in all regions. It is easy to see that, for any given unit,  $\phi^*$  is no larger than  $\phi$ , because the constraints in the regional LP problem are a subset of the constraints in the metafrontier LP problem. Units are shown to be not more technically efficient when they are assessed against the metafrontier than against the regional frontier.

## 2.3. Matching methods

### 2.3.1. Introduction

Supposed a study aims at assessing difference between efficiency of producers under different agricultural practice ( $r_1$  for organic cotton and  $r_0$  for conventional cotton). Let  $y_1$  metafrontier-based efficiency of the  $i$ th organic cotton producer and  $y_0$  metafrontier-based efficiency of the same producer in the case of conventional cotton. Then, the difference is:  $\Delta_i = y_1 - y_0$ .

The principal problem in this study is the missing data, called the counterfactual outcome (Rosenbaum and Rubin, 1983) due to fact that each producer uses only one technology ( $r_0$  or  $r_1$ ) but not a both. So some speculation is needed to obtain. The literature proposes several methods to determine effect of particular technology on output (in the case of this study, the outputs are technical inefficiency). It is important to notice that in impact evaluation and in comparison study, the parameter often used is: Average Treatment Effect on Treated (ATET or ATT).

The Matching methods are study in this research by supposed the strong ignorability of the treatment conditional on a set of covariates. The advantages of these methods are possibility to consider some pre-treatment covariates more “important” than others (Sodjinou, 2012). Matching usually refers to the methods that aim to identify for each treated individual, one or more similar individuals (in term of covariates  $x$ ) in the control group. In general case of binary and continuous outcomes, several studies have examined the performance of different propensity score methods for estimating treatment effects in medicine (Austin, 2008, Austin,



2010 and Austin, 2011). In particular case of efficiency outcome where values are doubly-censored in 0 (left censoring) and 1 (right censored), matching methods have frequently been used. But there is a paucity of research examining the relative performance of different matching methods for estimating ATET when outcomes were efficiency. The objective of the current study is to examine the ability of different matching methods to compare two technologies by estimating ATET when outcomes are technical efficiency.

When estimating causal effects using observational data, it is desirable to replicate a randomized experiment as closely as possible by obtaining treated and control groups with similar covariate distributions. This goal can often be achieved by choosing well-matched samples of the original treated and control groups, thereby reducing bias due to the covariates. We define “matching” broadly to be any method that aims to equate (or “balance”) the distribution of covariates in the treated and control groups (Harder and al, 2010). This may involve 1:1 matching, weighting or subclassification. The use of matching methods is in the broader context of the careful design of nonexperimental studies (Rosenbaum, 1999; Rubin, 2007). It intends to reduce the observed covariate differences between the treatment group and control group, under the *exogeneity* assumption (Sodjinou, 2012). This assumption has two parts.

The first is called *ignorability* of treatment assumption (given observed covariates  $x$ ), also known as no omitted variable bias, unconfoundedness (Rosenbaum and Rubin, 1983). The assumption means that conditional on observed pre-treatment covariate  $x$ , the treatment assignment is independent of potential outcomes. It also implies that systematic differences in outcomes between treated and comparison individuals with the same values for covariate are attributable to treatment (Caliendo and Kopeinig, 2008).

The second part of exogeneity assumption is called *common support* or *overlap*. A further requirement besides independence is the common support or overlap condition. It ensures that persons with the same  $X$  values have a positive probability of being both participants and non-participants (Heckman, LaLonde, and Smith, 1999). The overlap assumption, illustrated by Figure 1, is defined by:

$$0 < \Pr(d = 1 | x) < 1$$

### 2.3.2. Parameters estimate in causal effect evaluation

There are two major versions of aggregated treatment effects: the average treatment effect (ATE) and the average treatment effect on the treated group (ATT). A simple definition of the ATE can be written as:

$$ATE = E(Y_{1i} | T_i = 1, 0) - E(Y_{0i} | T_i = 1, 0)$$

where  $Y_{1i}$  the output of  $i$ th individuals when he use treatment 1,  $Y_{0i}$  the output of  $i$ th individuals when he use treatment 0 and  $T_i$  denote the treatment status (1 for treated and 0 for untreated).

In other words, the ATE can be defined as the average effect that would be observed if everyone in the treated and the control groups received treatment, compared with if no one in both groups received treatment (Harder and al, 2010).

The definition of ATT can be expressed as:

$$ATT = E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 1)$$

In contrast to the ATE, the ATT refers to the average difference that would be found if everyone in the treated group received treatment compared with if none of these individuals in the treated group received treatment. The value for the ATE will be the same as that for the ATT when the research design is experimental

Others parameters of interest are the effect of the treatment on the untreated (ATEU) and the local average treatment effect (LATE). ATEU can be estimated in a situation where the expansion of a program is under consideration and LATE means the impact of the program on those whose participation status changes due to a change in policy (Sodjinou, 2012).

### **2.3.3. Matching methods**

There are various methods that can be used in observational to assess the impact of a policy intervention. Those methods can be dividing in two groups according to ignorability of the treatment conditional on a set of covariates. Matching methods are among these categories. In the literature for impact assessment, matching usually refers to the methods that aim to identify for each treated individual, one or more similar individuals (in term of covariates  $x$ ) in the control group using distance between them.

There are two main aspects to determining the measure of distance (or “closeness”) to use in matching. The first involves which covariates to include, and the second involves combining those covariates into one distance measure.

#### **2.3.3.1. Variables to include**

The key concept in determining which covariates to include in the matching process is that of strong ignorability. As discussed above, matching methods, and in fact most non-experimental study methods, rely on ignorability, which assumes that there are no unobserved differences between the treatment and control groups, conditional on the observed covariates (Harder and al, 2010). To satisfy the assumption of ignorable treatment assignment, it is important to include in the matching procedure all variables known to be related to both

treatment assignment and the outcome (Rubin and Thomas, 1996). Generally poor performance is found of methods that use a relatively small set of “predictors of convenience,” such as demographics only (Shadishandal., 2008). Including variables that are actually unassociated with the outcome can yield slight increases in variance. However, excluding a potentially important confounder can be very costly in terms of increased bias. Researchers should thus be liberal in terms of including variables that may be associated with treatment assignment and/or the outcomes. Some examples of matching have 50 or even 100 covariates included in the procedure (e.g., Rubin, 2001). To avoid allegations of variable selection based on estimated effects, it is best if the variable selection process is done without using the observed outcomes, and instead is based on previous research and scientific understanding (Rubin, 2001). One type of variable that should not be included in the matching process are those that may have been affected by the treatment of interest (Rosenbaum, 1984; Frangakis and Rubin, 2002; Greenland, 2003). This is especially important when the covariates, treatment indicator, and outcomes are all collected at the same point in time.

### 2.3.3.2. Distance measure $D_{ij}$

In practice, to measure distance, various approaches are available in literature such as Exact Matching on Covariate (EMC), Mahalanobis Metric Matching (MMM) and Propensity Score Matching (PSM).

- **Exact Matching on Covariates (EMC)**

Let  $x$  a pre-treatment covariate for the treatment and the control group.

The EMC method aims to identify in the control group (group of nonparticipants), the individuals who have similar characteristics (in term of variables  $x$ ) with those of the participants. In other words, in the EMC method each treated unit/individual is matched with a control unit that has the exact same values for all covariates (Sodjinou, 2012). The distance  $D_{ij}$  between individuals  $i$  and  $j$  for exact matching is:

$$D_{ij} = \begin{cases} 0 & \text{if } x_i = x_j \\ \infty & \text{if } x_i \neq x_j \end{cases}$$

Ideally, exact matching on all confounding variables is the best matching approach because the sample distribution of all confounding variables would be identical in the treated and control groups. Unfortunately, exact matching on a single confounding variable will reduce the number of final matched cases. Supposing that there are  $k$  confounding variables and each variable has three levels, there will be  $3^k$  patterns of levels to get perfectly matched samples.

Thus, it is impractical to use the exact matching technique to get the identical distribution of confounding variables between the two groups.

- **Mahalanobis Metric Matching (MMM)**

The most common approach used to compute distance is the Mahalanobis distance function. The Mahalanobis distance function  $D_{ij}$  between individuals  $i$  and  $j$  is defined by:

$$D_{ij} = (x_i - x_j)' \Sigma^{-1} (x_i - x_j)$$

If interest is in the ATT,  $\Sigma$  is the variance covariance matrix of  $x$  in the full control group; if interest is in the ATE then  $\Sigma$  is the variance covariance matrix of  $x$  in the pooled treatment and full control groups.

If  $x$  contains categorical variables they should be converted to a series of binary indicators, although the distance works best with continuous variables. Like any other distance measure, the Mahalanobis distance function is arbitrary. In particular, it weights every variable equally, which may not be appropriate since some pre-treatment covariates are more “important” than others (Sodjinou, 2012). This problem can be partially overcome by using propensity score.

- **Propensity Score Matching (PSM)**

The propensity score for individual  $i$  is defined as the probability of receiving the treatment given the observed covariates:  $e_i(x_i) = P(T_i = 1 | x_i)$  Rosenbaum and Rubin (1983). There are two key properties of propensity scores.

- ✓ The propensity scores are balancing scores: At each value of the propensity score, the distribution of the covariates  $x$  defining the propensity score is the same in the treated and control groups. Thus, grouping individuals with similar propensity scores replicates a mini-randomized experiment, at least with respect to the observed covariates.
- ✓ If treatment assignment is ignorable given the covariates, then treatment assignment is also ignorable given the propensity score. This justifies matching based on the propensity score rather than on the full multivariate set of covariates.

In practice, and in the case of binary treatment, the propensity score can be estimated using Probit, Logit or discriminant score. The Probit or Logit models usually yield to similar results. The discriminant matching is equivalent to matching on the propensity score if the covariates  $x$  have a multivariate normal distribution.

#### **2.3.4. Matching algorithms**

After choosing the matching method, the following step is to choose the comparison unit(s) to match with each treatment unit.

#### **2.3.4.1. Nearest neighbour matching (NNM)**

The most straightforward matching estimator is k:1 nearest neighbour (NN) matching (Rubin, 1973).. The individual from the comparison group is chosen as a matching partner for a treated individual that is closest in terms of propensity score. Several variants of NN matching are proposed, e.g. NN matching with replacement and without replacement. In the former case, an untreated individual can be used more than once as a match, whereas in the latter case it is considered only once. NNM with replacement increase average quality of matching, decrease the bias and increase variance (Caliendo and Kopeinig, 2005).

In its simplest form, 1:1 nearest neighbor matching selects for each treated individual  $i$  the control individual with the smallest distance from individual  $i$ . A common complaint regarding 1:1 matching is that it can discard a large number of observations and thus would apparently lead to reduced power. Use more than one nearest neighbour (oversampling) involves a trade-off between variance and bias. It trades reduced variance, resulting from using more information to construct the counter-factual for each participant, with increased bias that results from on average poorer matches (Smith, 1997)

#### **2.3.4.2. Caliper matching (CM)**

Without any restrictions, k:1 matching can lead to some poor matches, if for example, there are no control individuals with propensity scores similar to a given treated individual. One strategy to avoid poor matches is to impose a caliper and only select a match if it is within the caliper. This can lead to difficulties in interpreting effects if many treated individuals do not receive a match, but can help avoid poor matches. CM is a variant of NNM that imposes a tolerance level on the maximum distance (caliper). For example, one could attempt to match treated and untreated subjects on the first five digits of the propensity score. For Smith and Todd (2005), CM is one way of imposing a common support condition. However, if fewer matches can be performed, the variance of the estimates increases.

#### **2.3.4.3. Radius matching**

Dehejia and Wahba (2002) suggest a variant of caliper matching which is called *radius matching*. The basic idea of this variant is to use not only the nearest neighbour within each caliper but all of the comparison members within the caliper. A benefit of this approach is that it uses only as many comparison units as are available within the caliper and therefore allows for usage of extra (fewer) units when good matches are (not) available. Hence, it shares the attractive feature of oversampling mentioned above, but avoids the risk of bad matches. The smaller the size of the neighbour the better is the quality of the matches.

#### **2.3.4.4. Stratification matching**

The stratification matching (also termed interval matching, blocking or sub-classification) consists of dividing the range of variation of the propensity score in intervals such that within each interval treated and control units have on average the same propensity score (Becker and Ichino, 2002). Within each interval, the difference between the average outcomes of the treated and the controls is computed. The ATET of interest is obtained as an average of the ATE of each block with weights given by the distribution of treated units across blocks. The problem in SM is to define the number of strata. Cochran (1968) states that five strata can remove 90% of the bias associated with one single covariate.

#### **2.3.4.5. Kernel Matching**

For the matching algorithms presented above, only few control individuals are used to build the potential outcome of treated units. With Kernel Matching all treated are matched with a weighted average of all controls with weights that are inversely proportional to the distance between the propensity scores of treated and controls (Becker and Ichino, 2002). KM can be seen as a weighted regression of the counterfactual outcome on an intercept with weights given by the kernel weights (Smith and Todd, 2005). One major advantage of KM is the lower variance which is achieved because more information is used (Caliendo and Kopeinig, 2008).

Finally, as noted by Becker and Ichino (2002), all the algorithms (NNM, RM, KM and SM) described above reach different points on the frontier of the trade-off between quality and quantity of the matches and none of them is a priori superior to the others. However, Becker and Ichino note that, their joint consideration offers a way to assess the robustness of the estimates. For Dolton (2002), with ideal data the choice between nearest neighbour and kernel may be fairly arbitrary but in some circumstances this is not so.

### 3. Methods

#### 3.1. Presentation of implementation example

Data were collected in the framework of the Project SYPROBIO (SYstèmes de PROduction BIOlogique diversifiés) during the agricultural campaign 2012–2013. This data have been subject to publication on “Agricultural and Food Economics” journals by Sodjinou and al (2015) on “Socioeconomic determinants of organic cotton adoption in Benin, West Africa”. The study was carried out in the Republic of Benin, a tropical West African country. The project promotes farmer-driven innovations to address food security, poverty alleviation and resilience to climate change. The project is led by FiBL (Research Institute of Organic Agriculture) and is implemented in Benin, Burkina Faso and Mali.

The three research sites are located respectively in the Centre of Benin, Northeast of Benin and in Northwest-Benin. In each of the three categories of areas, a village were selected, taking into account criteria such as the coexistence of organic farming (certified) and conventional cotton farming, the year of introduction of organic cotton (at least 3 years) and physical accessibility in all seasons of the year (Sodjinou and al, 2015). Thus, Aklampa in Glazoué District was selected for the Centre of Benin, Bonhanrou (District of Banikoara) for the Northeast-Benin, and Batia (in Tanguieta District) for the Northwest-Benin.

But in the present research only data on Aklampa and Bonhanrou was selected.

#### 3.2. Estimation of SF efficiencies in each group and pooled data

Most calculations and estimations were done within the statistical software environment “R” (R Core Team, 2013) using the add-on packages “frontier” (Tim Coelli and Arne Henningsen, 2013) for Maximum Likelihood Estimation of Stochastic Frontier Production and “lntest” (Torsten Hothorn and al, 2015).

- **Functional form**

The stochastic frontier is used to estimate farmer’s efficiencies. The stochastic frontier models were estimated both with the Cobb-Douglas and the Translog functional form. The likelihood ratio (LR) test has been used to choose significantly better functional form in each group and for pooled data. Suppose:

for conventional cotton ( $j=1$ ),  $N_1 = 88$  and  $i = 1, 2, \dots, 91$ .

for organic cotton ( $j = 2$ ),  $N_2 = 76$  and  $i = 1, 2, \dots, 76$ .

The Translog functional form is as following:

✓ for conventional:

$$\begin{aligned} \ln \text{YIELD}_{i1} = & \beta_{01} + \beta_{11} \ln \text{ENGR}_{i1} + \beta_{21} \ln \text{PEST}_{i1} + \beta_{31} \ln \text{MOT}_{i1} + \beta_{51} \ln \text{AMORT}_{i1} + \\ & 0.5 * \beta_{61} \ln \text{ENGR}_{i1} * \ln \text{ENGR}_{i1} + 0.5 * \beta_{71} \ln \text{ENGR}_{i1} * \ln \text{PEST}_{i1} + 0.5 * \beta_{81} \ln \text{ENGR}_{i1} * \ln \text{MOT}_{i1} + \\ & + 0.5 * \beta_{10,1} \ln \text{ENGR}_{i1} * \ln \text{AMORT}_{i1} + 0.5 * \beta_{11,1} \ln \text{PEST}_{i1} * \ln \text{PEST}_{i1} + 0.5 * \beta_{12,1} \ln \text{PEST}_{i1} * \ln \text{MOT}_{i1} + \\ & 0.5 * \beta_{14,1} \ln \text{PEST}_{i1} * \ln \text{AMORT}_{i1} + 0.5 * \beta_{15,1} \ln \text{MOT}_{i1} * \ln \text{MOT}_{i1} + 0.5 * \beta_{17,1} \ln \text{MOT}_{i1} * \ln \text{AMORT}_{i1} + \\ & + 0.5 * \beta_{20,1} \ln \text{AMORT}_{i1} * \ln \text{AMORT}_{i1} + V_{i1} - U_{i1} \end{aligned}$$

✓ for organic cotton

$$\begin{aligned} \ln \text{YIELD}_{i2} = & \beta_{02} + \beta_{12} \ln \text{ENGR}_{i2} + \beta_{22} \ln \text{PEST}_{i2} + \beta_{32} \ln \text{MOT}_{i2} + \beta_{52} \ln \text{AMORT}_{i2} + \\ & 0.5 * \beta_{62} \ln \text{ENGR}_{i2} * \ln \text{ENGR}_{i2} + 0.5 * \beta_{72} \ln \text{ENGR}_{i2} * \ln \text{PEST}_{i2} + 0.5 * \beta_{82} \ln \text{ENGR}_{i2} * \ln \text{MOT}_{i2} + \\ & 0.5 * \beta_{10,2} \ln \text{ENGR}_{i2} * \ln \text{AMORT}_{i2} + 0.5 * \beta_{11,2} \ln \text{PEST}_{i2} * \ln \text{PEST}_{i2} + 0.5 * \beta_{12,2} \ln \text{PEST}_{i2} * \ln \text{MOT}_{i2} + \\ & 0.5 * \beta_{14,2} \ln \text{PEST}_{i2} * \ln \text{AMORT}_{i2} + 0.5 * \beta_{15,2} \ln \text{MOT}_{i2} * \ln \text{MOT}_{i2} + 0.5 * \beta_{17,2} \ln \text{MOT}_{i2} * \ln \text{AMORT}_{i2} + \\ & + 0.5 * \beta_{20,2} \ln \text{AMORT}_{i2} * \ln \text{AMORT}_{i2} + V_{i2} - U_{i2} \end{aligned}$$

where: YIELD=land productivity in kg/ha;

ENGR= fertilizer cost in FCFA/ha, mineral fertilizer for conventional cotton and organic fertilizer for organic cotton;

PEST= pesticide cost in FCFA/ha, chemical pesticide for conventional cotton and biologic pesticide for organic cotton;

MOT= labour in men/day;

AMORT= depreciation of equipment

$V_{ij}$  = noise term

$U_{ij}$  = Inefficiencies term

• **Estimation of parameters and technical inefficiencies :**

The parameters  $\beta$ ,  $\sigma$  and  $\lambda$  has been obtain by maximization of Loglikelihood function of  $y$  :

$$\ln L(y | \beta, \sigma, \lambda) = -\frac{1}{2} \ln \left( \frac{\pi \sigma^2}{2} \right) + \sum_{i=1}^I \ln \Phi \left( -\frac{\varepsilon_i \lambda}{\sigma} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^I \varepsilon_i$$

A technical efficiency can be obtain with  $\hat{T\hat{E}}_{ij} = \exp(-\hat{U}_{ij})$ , knowing that :

$$\hat{U}_{ij} \equiv E\{U_{ij} | q_{ij}\} = U_{ij}^* + \sigma_* \left[ \frac{\phi(U_{ij}^*/\sigma_*)}{\Phi(U_{ij}^*/\sigma_*)} \right]$$

Shepard-type efficiencies are computed for output-oriented models. For example, if  $\hat{T\hat{E}}_{ij} = 0.8$  then the technical efficiency measure indicates that the output vector,  $y$ , is 80 per cent of the potential output, using the same input vector.



### 3.3. DEA frontier and DEA metafrontier

Most calculations and estimations were done within the statistical software environment “R” (R Core Team, 2013) using the add-on packages “Benchmarking” (Peter Bogetoft and Lars Otto, 2015) for Data Envelopment Analysis (DEA), under different technology assumptions and “lmtest” (Torsten Hothorn and al, 2015).

DEA involves the use of mathematical linear programming to construct a non-parametric frontier over data. This study used the output oriented DEA. Output-oriented DEA is configured to determine a firm’s potential output given its inputs if it operated efficiently as firms along the best practice frontier. This is more analogous to the stochastic production function approach, which estimated the potential output for a given set of inputs and measured capacity utilization as the ratio of the actual to potential output. Output-oriented models are “...very much in the spirit of neo-classical production functions defined as the maximum achievable output given input quantities” (Färe, Grosskopf and Lowell, 1994, p. 95).

Two scale assumptions are generally employed: constant returns to scale (CRS), and variable returns to scale (VRS). The latter encompasses both increasing returns to scale (IRS) and decreasing returns to scale (DRS). The choice between the third scales assumption (CRS, IRS and DRS) was doing using student T-test (Table 3).

Table 3: T-test Hypothesis

Type of test	Hypothesis	
	Null	Alternative
$\theta^{IRS} > \theta^{DRS}$	CRS	IRS
$\theta^{IRS} < \theta^{DRS}$	CRS	DRS

### 3.4. Metafrontier

#### 3.4.1. Stochastic metafrontier

Most calculations and estimations were done within the statistical software environment “R” (R Core Team, 2013) using the add-on packages “lpSolve” (Berkelaar and al, 2015) for linear programming and “lmtest” (Hothorn and al, 2014) for comparisons of models via asymptotic likelihood ratio tests.

The MF model is defined by Battese *and al* (2004) as a deterministic parametric frontier of specific functional form (e.g., Cobb Douglas or Translog) such that the predicted value for the MF is larger than or equal to the predicted value from the stochastic frontier for all firms and groups.

The translog functional form of MF is:

$$\begin{aligned} \ln \text{YIELD}_i = & \beta_0 + \beta_1 \ln \text{ENGR}_i + \beta_2 \ln \text{PEST}_i + \beta_3 \ln \text{MOT}_i + \beta_5 \ln \text{AMORT}_i + \\ & 0.5 * \beta_6 \ln \text{ENGR}_i * \ln \text{ENGR}_i + 0.5 * \beta_7 \ln \text{ENGR}_i * \ln \text{PEST}_i + 0.5 * \beta_8 \ln \text{ENGR}_i * \ln \text{MOT}_i + \\ & 0.5 * \beta_{10} \ln \text{ENGR}_i * \ln \text{AMORT}_i + 0.5 * \beta_{11} \ln \text{PEST}_i * \ln \text{PEST}_i + 0.5 * \beta_{12} \ln \text{PEST}_i * \ln \text{MOT}_i + \\ & 0.5 * \beta_{14} \ln \text{PEST}_i * \ln \text{AMORT}_i + 0.5 * \beta_{15} \ln \text{MOT}_i * \ln \text{MOT}_i + 0.5 * \beta_{17} \ln \text{MOT}_i * \ln \text{AMORT}_i + \\ & 0.5 * \beta_{20} \ln \text{AMORT}_i * \ln \text{AMORT}_i \end{aligned}$$

The stochastic frontiers for the different groups and that of the metafrontier would generally be assumed to be of the same functional form. The parameters of the metafrontier model are estimated using data from firms in all groups (in the combined data set). To identify the best envelope, the sum of absolute deviations has been the specify criterion. In addition, its standard deviations are obtained using bootstrap.

The TGR measures the ratio of the output for the frontier production function for the J – th group relative to the potential output that is defined by

$$\text{TGR}_{ij} = \frac{e^{x_i \beta_j}}{e^{x_i \beta^*}}$$

### 3.4.2. DEA metafrontier

The same method in DEA frontier was used in metafrontier. The difference was in nature of data. The DEA was applied using all data in metafrontier.

### 3.4.3. Comparison of metafrontier methods

To handle performance of MF, the percentage of underestimated firm's efficiency was calculate using poled data SF, SMF and DEA metafrontier. The paired-t.test was used to compare efficiency and TGR derived from the approaches. The effect of the different methodologies was also further examined using the Spearman's rank order correlation between the different efficiency and TGR measures.

### 3.4.4. Consistency of Propensity Scores Matching for efficiency comparison

The consistency of PSM was evaluated by analyse:

- Sample size of treated individual after matching (support region);
- Sensitivity analysis: to check if there is a hidden bias in the estimations, to inform the how robust is his estimation
- Variance of treatment effects related to the statistical significance of the treatment effects and the estimation of their standard errors; the bootstrapping was used to estimation.

## 4. Results and discussion

This section describes the results of the estimation of the technology frontiers and associated TE measures and results of matching methods. First, the SF results and specification tests are analysed for technologies. Second, results of SF for pooled sample and for stochastic metafrontier efficiency have been presented. Third, the performance of matching is presented. Fourth, the results obtain have been discussed.

### 4.1. Efficiencies under each technology frontier

#### 4.1.1. Data Envelopment Analysis

DEA results were obtained using an output orientation under the assumptions of Constant Return to Scale (CRS) and Variable Return to Scale. A t.test between efficiency mean under Increase Return to Scale (IRS) and efficiency mean under Decreasing Return to Scale (DRS) revealed that (Appendix 7):

- conventional cotton technology operate under assumption of DRS,
- while organic cotton technology well works under assumption of CRS.

The distributions of the technical efficiency estimates of the conventional cotton farmers and the organic cotton farmers are illustrated in Figure 2. The technical efficiency estimates of the conventional cotton and organic cotton producers have a very large variation with respectively an average value of 0.4528555 and 0.5290699

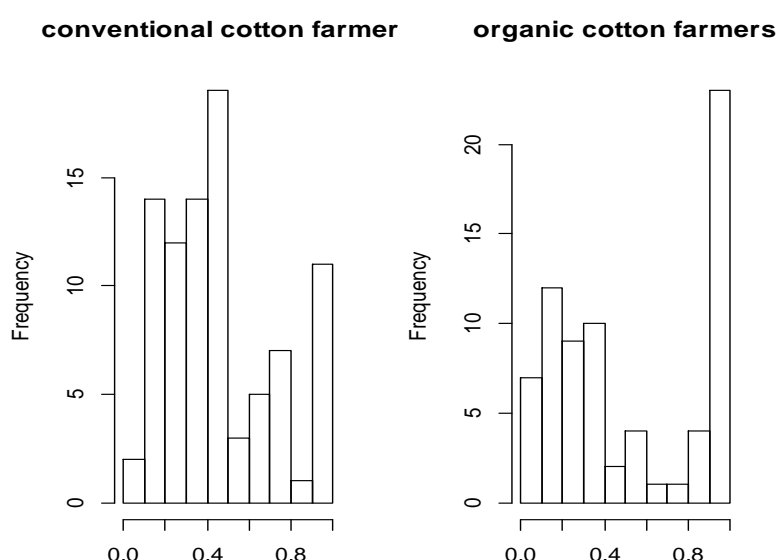


Figure 1: Distribution of technical efficiencies estimate with DEA

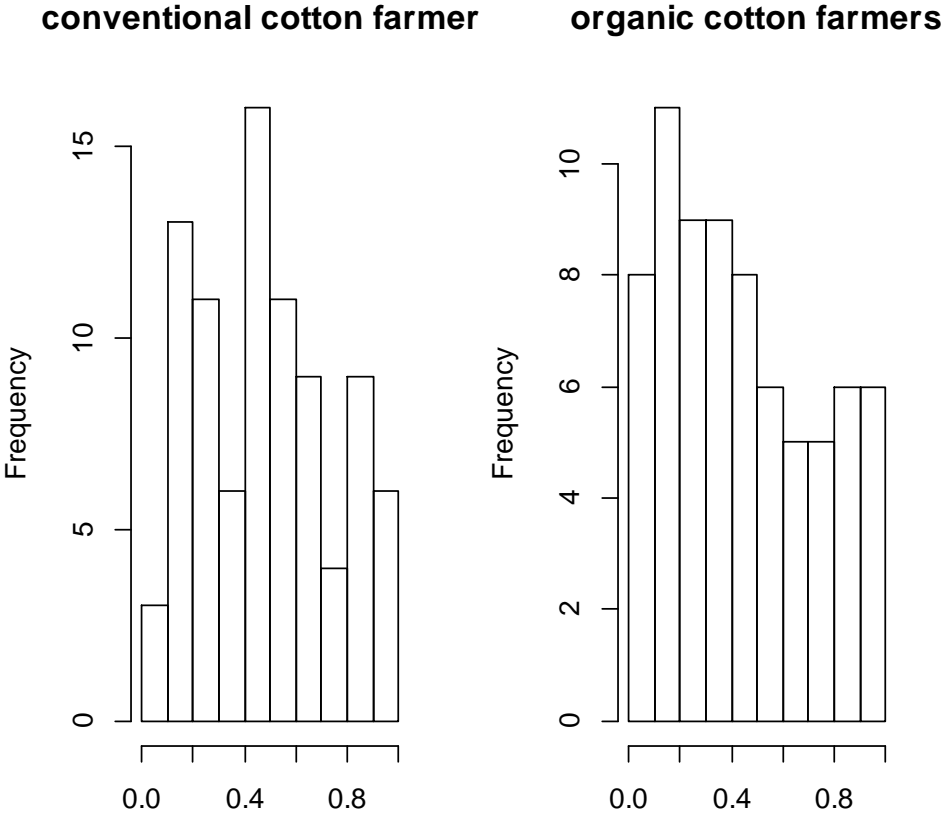
#### 4.1.2. Stochastic Frontier Model

The stochastic frontier models were estimated both with the Cobb-Douglas and the Translog functional form (see Appendix 2 and 3). In all cases (conventional cotton group and organic

cotton group) a likelihood ratio (LR) test revealed that the fit of the Translog functional form was significantly better than the fit of the Cobb-Douglas functional form. Therefore, we use the Translog functional form in our analysis.

The dispersion parameter of the inefficiency term ( $\sigma_u$ ) is much larger for the conventional cotton farmers than for the organic cotton farmers, which indicates that the conventional cotton farmers are more affected by inefficiency than the organic cotton farmers.

The distributions of the technical efficiency estimates of the conventional cotton farmers and the organic cotton farmers are illustrated in Figure 3. While the technical efficiency estimates of the conventional cotton farmers have a very large variation with an average value of 0.517, the technical efficiency estimates of the organic cotton farmers are all close to its mean value of 0.894.



**Figure 2:** Distribution of technical efficiencies estimate with stochastic frontier

**4.2. Stochastic frontier for pooled sample and stochastic metafrontier efficiency**

The differences of farmer’s efficiency across stochastic frontier of pooled data and the corresponding optimum efficiency in the same stochastic frontier are compute. The same thing was done but considering metafrontier instead of pooled data stochastic frontier. This

difference was compute in other to find if stochastic frontier for pooled sample and stochastic metafrontier efficiency well bound all efficiency of farmers. Accordingly to table 4 there is some farmer's efficiency out of pooled data stochastic frontier. Specially, 67.08% of conventional cotton has their efficiencies above pooled data stochastic frontier. But nobody has his efficiency above stochastic metafrontier. Then, based on this empirical result, the stochastic metafrontier well cover stochastic frontiers of all groups of farmers.

**Table 4:** Percentage of farmer's efficiency out of frontier of all groups (%)

	<b>Pooled data SFA</b>	<b>Stochastic Metafrontier</b>
<b>Conventional cotton groups</b>	67.08	0
<b>Organic cotton groups</b>	0	0

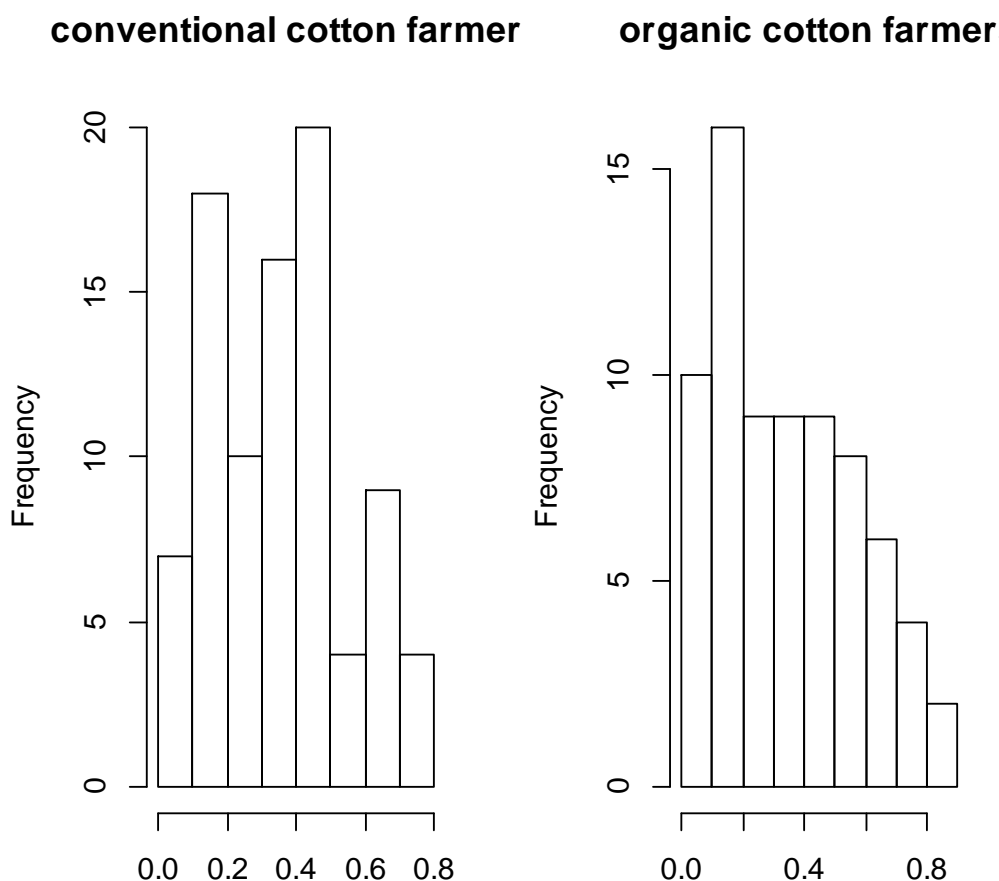
The coefficients of the meta-frontier are presented in Appendix5. In order to facilitate the comparison across the coefficients of the group-specific production frontier, pooled data stochastic frontier and stochastic metafrontier, all of coefficients are presents in Appendix 6. The values of the TE measures for the stochastic frontier and with respect to the stochastic metafrontier are summarized in Table 5. The efficiencies obtain with stochastic metafrontier are statistical different to those obtain with stochastic frontier.

**Table 5:** Descriptive statistics on efficiency using parametric method

	<b>Stochastic frontier</b>	<b>Stochastic metafrontier</b>	<b>Paired t.test</b>
<b>Conventional cotton</b>	0.483423129 (0.255809949)	0.35285599 (0.18578014)	-19.5095 ***
<b>Organic cotton</b>	0.449971022 (0.282542572)	0.34738536 (0.22373983)	-661.6324***

(Standard deviation)

It is one of organic cotton whom obtains the higher efficiencies. All of organic cotton farmer's efficiencies surpass 38.64% of conventional cotton farmer's efficiencies. Using figure 3 and figure 4, the distribution of producer's efficiencies is little different across stochastic frontier and stochastic metafrontier considering each technology.



**Figure 3:** Distribution of Stochastic metafrontier efficiencies

The TGR of conventional group's cotton is statistically different to organic cotton. This result confirms the choice of metafrontier to compare efficiency of those two groups.

**Table 6:** Technologies Gap Ratio with Stochastic Metafrontier

	<b>Mean</b>	<b>Standard deviation</b>
<b>Conventional cotton</b>	0.352856	0.01868448
<b>Organic cotton</b>	0.3473854	0.02252882
<b>t.test</b>		0.1655 **

A higher (lower) TGR value implies a smaller (larger) technology gap between the individual frontier and the MF. A TGR value of 100% is equivalent to a point where a regional frontier coincides with the MF. Under stochastic metafrontier, none of producer can obtain 100% efficiency, even if he produces optimum output under corresponding stochastic frontier (figure 4).

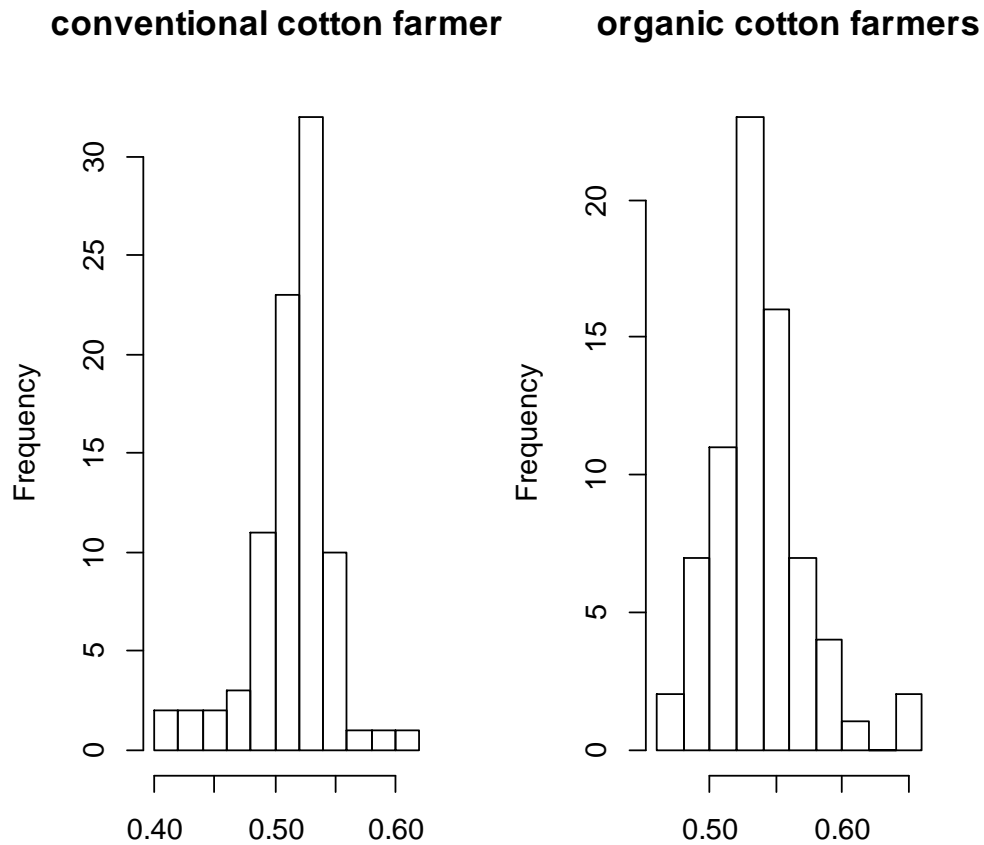


Figure 4: Distribution of Technologies Gap Ratio with Stochastic Metafrontier

### 4.3. DEA metafrontier efficiency

The values of the TE measures for the DEA frontier and with respect to the DEA metafrontier are summarized in Table 7. As we obtain using stochastic method, the efficiencies obtain with DEA metafrontier are statistical different to those obtain with DEA frontier.

Table 7: Descriptive statistics on efficiency using non-parametric method

	DEA frontier	DEA metafrontier	Paired t.test
<b>Conventional cotton</b>	0.4528555 (0.268351)	0.3114306 (0.225066)	-10.1518***
<b>Organic cotton</b>	0.5290699 (0.3691766)	0.452268 (0.3619285)	-6.2937***

Using figure 2 and figure 5, the distribution of producer's efficiencies is different across DEA frontier and DEA metafrontier considering each technology.

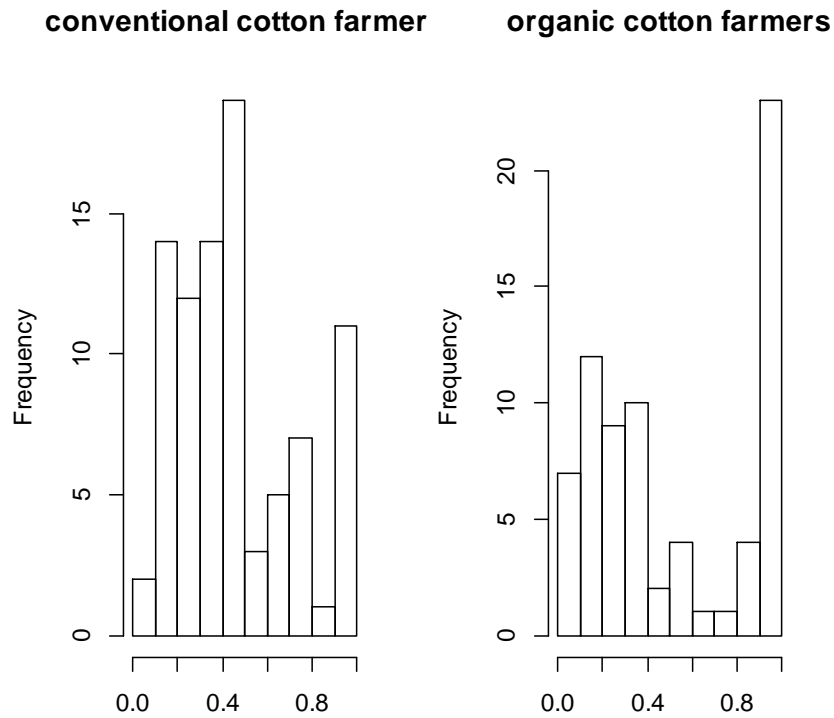


Figure 5: Distribution of stochastic metafrontier efficiency

The TGR of conventional group’s cotton is statistically different to organic cotton. This result confirms the choice of metafrontier to compare efficiency of those two groups.

Table 8: Technologies Gap Ratio with DEA Metafrontier

	Mean	Standard deviation
<b>Conventional cotton</b>	0.681099	0.1896277
<b>Organic cotton</b>	0.8061028	0.1728769
<b>t.test</b>		-4.3706***

Using DEA metafrontier, several producers have opportunity to obtain 100% efficiency if they produce optimum output under their corresponding stochastic frontier (figure 4).



**conventional cotton farmer**

**organic cotton farmer:**

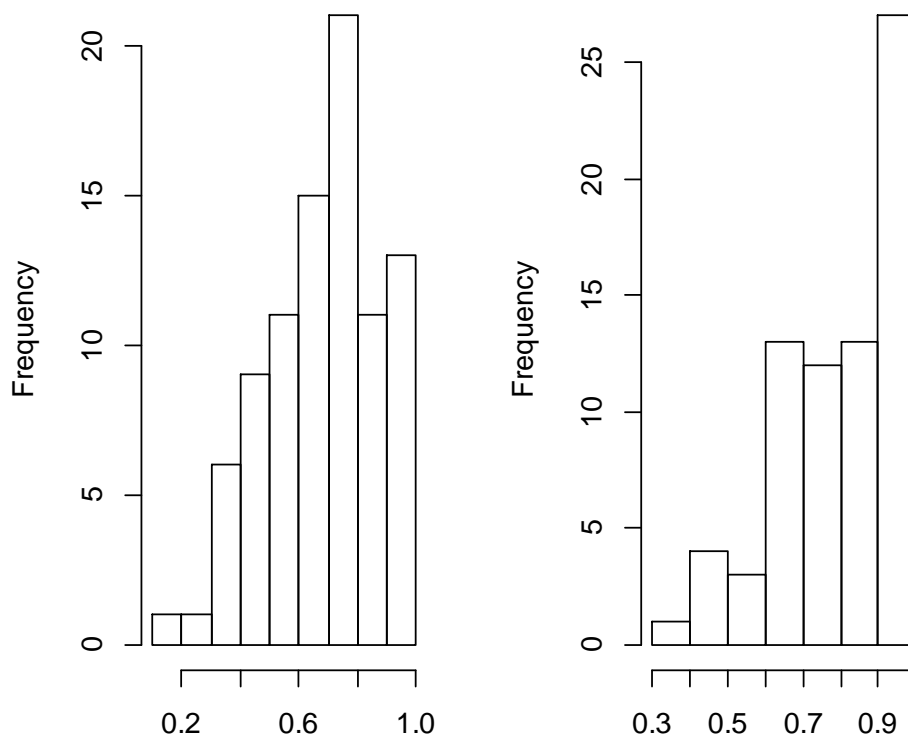


Figure 6: Distribution of Technologies Gap Ratio with DEA metafrontier

#### 4.4. Comparison of DEA metafrontier and Stochastic metafrontier analysis

##### 4.4.1. Result of paired t-test

An paired t-test comparison between the efficiency measures derived from the two approaches reveals that there is no significant difference in the metafrontier technical efficiency derived from stochastic analysis and Data Envelopment Analysis approaches ( $t = -1.089$ ,  $df = 126$ ,  $p\text{-value} = 0.2782$ , which is significant at the 10% level). A comparison between the TGR derived from the two approaches reveals that they are significantly different at the 1% significance level ( $t = -12.5047$ ,  $df = 126$ ,  $p\text{-value} < 2.2e-16$ ). A **Erreur ! Source du renvoi introuvable.** allows confirming that the distributions of efficiencies are similar for both approaches, stochastic and DEA metafrontier. It also allows seeing that more than 75% of TGR values with DEA approaches are greater than more than 75% of TGR values with stochastic approaches.

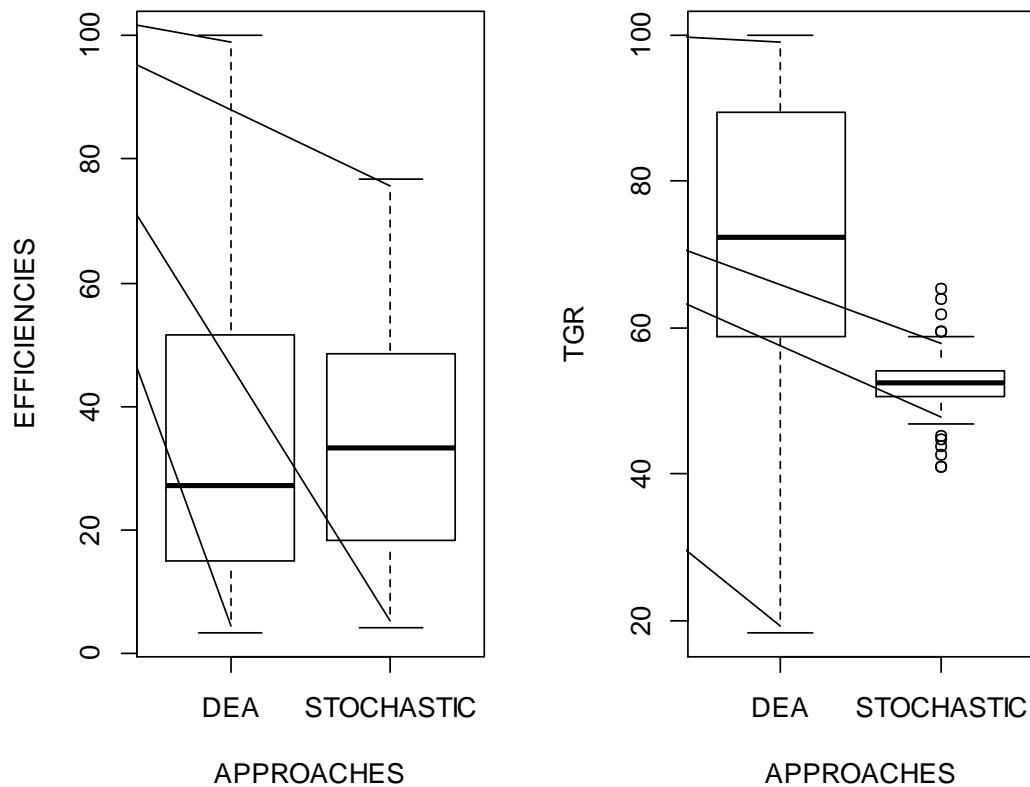


Figure 7: Box plot of stochastic and DEA approaches

The effect of the different methodologies was also further examined using the Spearman's rank order correlation between the different efficiency and TGR measures. The results indicated that there is a strong significant correlation between the efficiency results reported from the different approaches (Spearman's rho = 0.7433 and Prob > |t| = 0.0000). The similar result are obtain with TGR (Spearman's rho = 0.4342 and Prob > |t| = 0.0000).

#### 4.4.2. Matching methods to efficiencies comparison

##### 4.4.2.1. Result of logistic regression

The first step in the use of PSM is to specify the balancing score. For this purpose, the logit model was used to predict the probability to adopt "Organic Cotton", given the socio demographic characteristics. The results of this logistic regression show that there are three main determinants in organic cotton adoption, namely the sex of producers, education level of producer and number of visits between January and October 2012. Indeed, the producer's sex significantly influence the adoption of organic cotton at 1 per 1000 level where education level of producer and number of visits was significant at 5% level.

Table 9: Logistic regression

	Parameters	Standard errors	z	P>z
Age	0.011	0.023	0.470	0.639
Sex of the producer	-2.358	0.590	3.990	0.000
Class of poverty as perceived by producers	0.198	0.262	0.750	0.451
Distance between the producer's house and his farm (km)	-0.023	0.043	0.530	0.596
Formal education of the producer (year)	-0.093	0.077	1.210	0.226
Experience in cotton production in general (year)	-0.001	0.029	0.050	0.961
Membership of producers' organization	1.071	0.738	1.450	0.147
Contact with extension agents	0.301	0.131	2.300	0.022
Constant	-1.260	1.185	1.060	0.288

#### 4.4.2.2. Characteristics of organic cotton producers versus conventional cotton producers after matching

Results in table 10 allowed concluding that both individuals in each paired matching had the same characteristics. Then, “ceteris paribus”, all difference between farmer's efficiencies or Technology Gap Ratio must be attributed to production system (conventional or organic cotton)

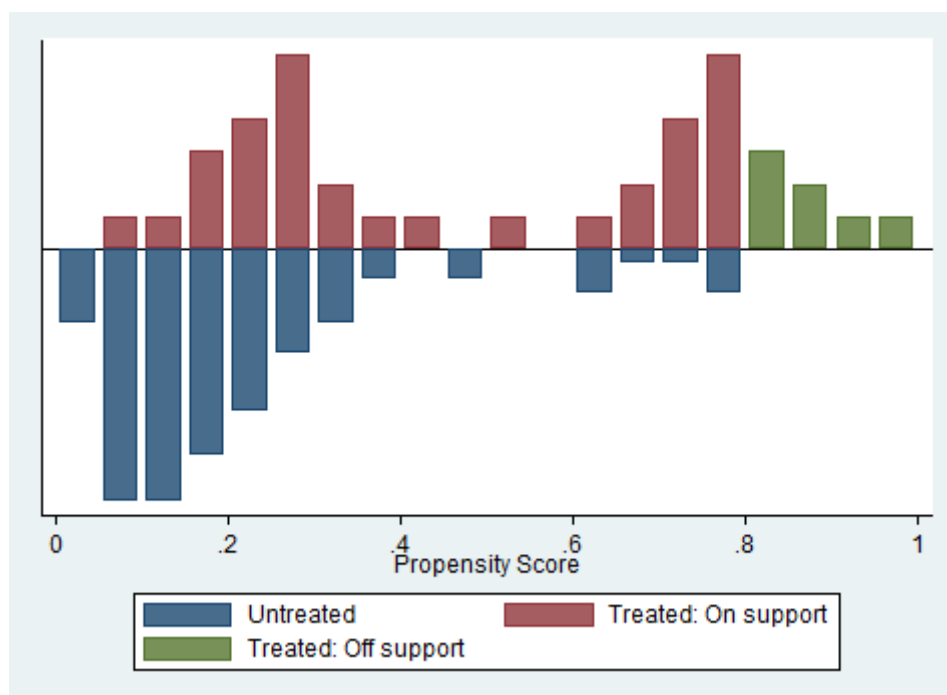
Table 10: Descriptive statistics of socioeconomic characteristics of farmer's after matching:

Variables	Mean		t-test	
	Treated	Control	t	p>t
Age	41.848	41.697	0.050	0.958
Sex of the producer	0.576	0.636	-0.500	0.621
Class of poverty as perceived by producers	2.242	2.273	-0.150	0.883
Distance between the producer's house and his farm (km)	3.568	3.030	0.690	0.495
Formal education of the producer (year)	1.727	1.546	0.260	0.797
Experience in cotton production in general (year)	13.333	10.364	1.260	0.213
Membership of producers' organization	0.848	0.848	0.000	1.000
Contact with extension agents	2.576	2.909	-0.600	0.554

#### 4.4.2.3. Support region

Figure 8 shows the distribution of the estimated propensity score for the organic cotton and the conventional cotton groups. It is clear, from the observation of this figure, that there is

overlap between two groups. All of conventional cottons units have been used in the matching process but there some organic units (in colors green) leaved.



**Figure 8:** The distribution of the propensity score

#### 4.4.2.4. Variance of treatment effects

For both metafrontier methods, stochastic metafrontier and DEA metafrontier, PSM found difference in mean of efficiencies and TGR between a two groups of cotton's farmers.

Table 11: Variance of treatment effects

Metafrontier methods	Difference in mean	Standard error	z	P>z	[95% Interval]	Confidence
Efficiencies						
Stochastic metafrontier	12.462	7.559	1.650	0.099	-2.353	27.277
DEA metafrontier	30.775	11.064	2.780	0.005	9.090	52.460
Technologies Gap Ratio						
Stochastic metafrontier	2.448	1.185	2.070	0.039	0.125	4.771
DEA metafrontier	16.048	6.666	2.410	0.016	2.983	29.112

NB: standard-error estimates include only complete replications;

## 4.5. Discussion

The comparative study of metafrontier methods have been done in the case of two technologies: conventional cotton and organic cotton. Three methods have been compared: stochastic frontier on pooled data, stochastic metafrontier and data envelopment analysis metafrontier. The tests prove the necessity to compute the metafrontier instead of group's frontiers. Estimation of the metafrontier production model is justified on the basis of the paired student t test between efficiencies under group's frontier and efficiencies under metafrontier. The values of the test statistic are significant at the 1 per thousand level. Furthermore, the results of t Test indicate that significant differences exist between the estimated mean TGRs of the two technologies. These results imply that there is shortcoming in the estimation of individual group frontiers in that their efficiency levels cannot be compared and these problems are overcome by estimating the metafrontier model (Battese and al, 2002 and Villano and al, 2010). Recently, O'Donnell et al. (2007) have also argued that SFA or DEA might sometimes lead to inaccurate results when comparing the efficiency of firms that operate under different technologies. This is because frontiers for different groups may not be identical to provide an unbiased comparison.

The choice of methods is a pertinent problem..The first criterion was to see if the boundary of the output set compute by each approach showed the production possibility frontier. Analyses revealed there are some producers' productions set above frontier compute with stochastic frontier with pooled data. Then only metafrontier approach using mathematics programming can well estimate the metafrontier.

Among methods using mathematics programming, stochastic metafrontier and DEA metafrontier have been compared. These methods allow estimating simultaneously efficiencies and Technologies Gap Ratio (TGR).

The efficiencies of both methods were compared using paired-t-tests. For both the stochastic metafrontier production function and the metafrontier data envelopment analysis, there are no differences in the technical efficiencies. This result is in stark contrast with the results of several authors like Matawie K. M. and Assaf A. (2008). In the literature there is some empirical evidence that the mean of efficiencies are sensitive to the method used in estimating the frontier model (Kumbakhar and Lovell, 2004). Then, methods are not the unique source of variation among efficiencies, perhaps interaction between methods and some characteristics of data. It will be interesting for future study to well master the parameters that influence efficiency in order to help in choice of methods.

The choice among the various frontier methods has no important effect on the estimated EE scores. It is now interesting to know whether these approaches produce consistent rankings of banks in terms of efficiency scores. Results from this study indicate that the ranking is invariant with the different approaches, making therefore possible to draw conclusion from the results of this study. Consequently, the two models ranked the sample producers in same orders when evaluating producer' performances.

When examining the efficiency results, both methods have indicated that the first group representing organic cotton producers is operating at a higher efficiency level. The difference in efficiency between the two groups could be attributed to different factors.

The TGR of both methods were also compared using paired-t-tests. For both the stochastic metafrontier production function and the metafrontier data envelopment analysis, there are significant differences in the TGR. The differences in the predicted TGR are attributable to the fact that the two approaches measure technical efficiency relative to different metafrontier (Battese & Coelli, 1995).

## 5. Conclusion

This study has tried to shed light on the sensitivity of empirical results to the selection of the estimation method base on real data collected on conventional cotton and organic cotton. The estimation methods concerned by this study were: stochastic frontier analysis using pooled data, stochastic metafrontier and Data Envelopment Analysis metafrontier. In addition, this research examines ability of Propensity Scores Matching (PSM) to compare efficiencies of two groups of producers.

This thesis is organized as four main chapters. Chapter 1 gives a brief background and the objectives for this study. Chapter 2 provides a literature review on frontier and metafrontier estimation methods, especially stochastic and Data Envelopment (DEA) approaches. The matching methods and algorithms are also presented in this chapter.

The details of the dataset used in the work and methods adopt in this study are given in Chapter 3.

In order to achieve both goals, the research has taken into account the values observed at individual farm level of productivity, labour, depreciation of equipment and the cost of two large groups of agricultural tasks: fertilizer and pesticide. Scores have been calculated using frontier and metafrontier estimation methods within the framework of DEA and stochastic analysis. The inefficiency measures the potential to proportionally improve output without reducing the agricultural tasks.

Chapter 4 gives details on efficiency analysis using the dataset on conventional and organic cotton with both approach stochastic analysis and DEA. Results have been discussed in this chapter. The results revealed no significant difference between metafrontier efficiency results, derived from the two approaches. No significant difference was also found between the efficiency rankings of the different firms. The ranking of the technology gap ratios was also consistent across the two approaches. Then, methods are not the unique source of variation among efficiencies, perhaps interaction between methods and some characteristics of data. The PSM found a significant difference between conventional and organic cotton. The results obtain are robust and consistent.

This study encounter some limitation and provides some indicators which are worthy of further research proof and/or clarification.

First, concerning the statistical software environment “R” (R Core Team) and STATA, it will be interesting to conceive packages because it will be fastidious for non-statistician to write large code.

Second, it will be interesting for future study to well master the parameters of methods and dataset that influence efficiency score in order to help in choice of methods. Simulation methods as Monte Carlo can be way to handle this objective.



## 6. References

- Ajibefun I. A. (2008). An evaluation of parametric and non-parametric methods of technical efficiency measurement: application to small scale food crop production in Nigeria. *Journal of Agriculture & Social Sciences*, 4 (3): 95–100
- Ali K. M. E. and Samad Q. A., 2013, A metafrontier production function for estimation of technical efficiencies of wheat farmers under different technologies, *Journal of Asian Scientific Research*, 3(9):933-944.
- Austin P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine*, 27(12):2037–2049.
- Austin P. C. (2010). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, 33: 1057–1069.
- Austin P. C. (2010). The performance of different propensity score methods for estimating marginal odds ratios. *Statistics in Medicine*, 29: 2137—2148.
- Austin P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, 10:150–161.
- Austin P. C. (2012). The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in Medicine*, 32 2837–2849.
- Austin P. C. and Schuster, T. (2014). The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: A simulation study. *Statistical Methods in Medical Research*, 24p.
- Austin P. C., Grootendorst P. and Anderson G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in Medicine*, 26(4):734–753.
- Ayinde O. E., Adewumi M. O. and Ojehomon V. E. T., 2009, Determinants of technical efficiency and varietal-gap of rice production in Nigeria: a meta-frontier model approach, Contributed Paper prepared for presentation at the International Association of Agricultural Economists Conference, Beijing, China, August 2009, 16-22.
- Banker R. D., Cooper W. W., Seiford L. M., Thrall R. M. and Zhu J., 2004, Returns to scale in different DEA models, 154: 345–362.

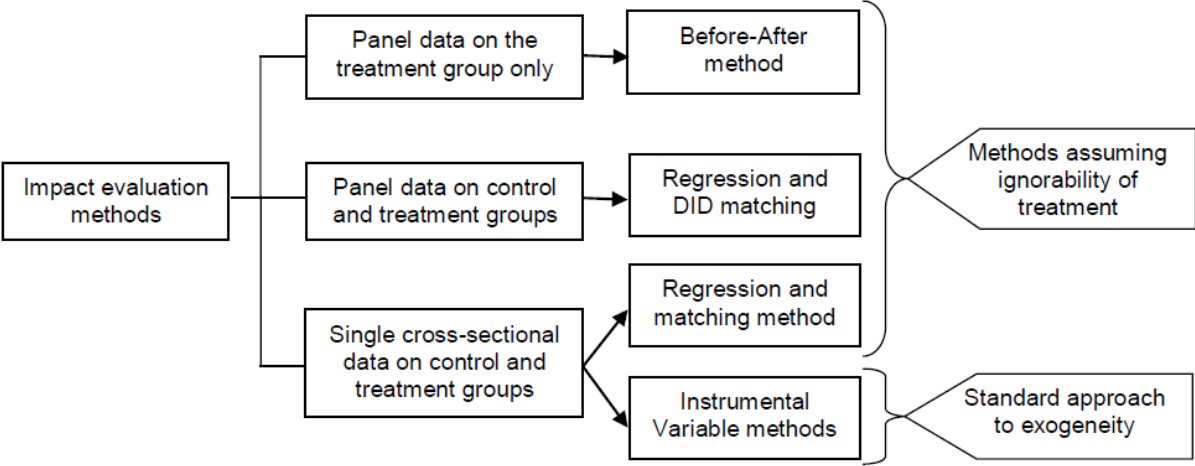
- Basso A. and Funari S., 2012, Constant and variable returns to scale DEA models for socially responsible investment funds, Working Papers Department of Economics Ca'Foscari University of Venice, 20, 26p.
- Battese G. E. and Rao D. S. P. (2002). Technology Gap, Efficiency, and a Stochastic Metafrontier Function. *International Journal of Business and Economics*, 1(2): 87-93.
- Battese G., Rao D. and O'Donnell, C. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *J. Product. Anal*, 21 (1): 91–103.
- Beltrán-Esteve M. and Reig-Martínez E. (2014). Comparing conventional and organic citrus grower efficiency in Spain. *Agricultural Systems*, 129: 115–123.
- Berkelaar M. and others, (2015). lpSolve: Interface to 'Lp\_solve' v. 5.5 to Solve Linear/Integer Programs. R package version 5.6.11.
- Bogetoft P. and Otto L., (2015). Benchmark: Benchmark and Frontier Analysis Using DEA and SFA. R package version 0.26.
- Caliendo M. and Kopeinig S. (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22(1): 31–72.
- Chen Z. and Song S., 2006, Efficiency and Technology Gap in China's Agriculture: A Regional META-Frontier Analysis, UNR Economics Working Paper Series, (06).
- Coelli T. and Henningsen A., (2013). frontier: Stochastic Frontier Analysis. R package version 0.1-0.
- Coelli, T.J., Rao D.S.P., O'DONNELL C. J. and Battese G.E., 2005. An introduction to efficiency and productivity analysis. 2<sup>nd</sup> ed, Springer, 341p.
- De Ponti T., Rijk B. and Van Ittersum M.K. (2012). The crop yield gap between organic and conventional agriculture. *Agricultural Systems*, 108: 1–9.
- Demsetz H. 1995, *The firm of theory: its definition and existence*, the Press Syndicate of the University of Cambridge, (17), 34p.
- Doug T. (2014). Comparing regression, propensity matching and coarsened exact matching in healthcare observational studies using SAS<sup>®</sup>: An example from the Medical Expenditure Panel Survey (MEPS). Cary NC: SAS Institute, 16p.
- Färe R., Grosskopf, S. and Lovell, C.A.K., 1994, *Production frontiers*. United Kingdom, Cambridge University Press.
- Greene W. (2004). Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model. *J. Econometrics*, 126: 269-303.

- Greene, W. (1980), "Maximum Likelihood Estimation of Econometric Frontier Functions", *Journal of Econometrics*, 13, 27-56.
- Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, 15,234-249.
- Hayami Y. and Ruttan V. (1970). Agricultural productivity differences among countries. *Am. Econ. Rev*, 60 (5), 895–911.
- Hayami Y. and Ruttan V.W. (1971), Agricultural development: An international perspective, *Journal of Development Economics*, 26, 197-200.
- Hayami, Y. (1969). Sources of agricultural productivity gap among selected countries. *American Journal of Agricultural Economics*, 51, 564-575.
- Henningsen A., Mpeta D. F., Adem A. S., Kuzilwa J. A. and Czekaj T. G., 2015, A Meta Frontier Approach for Causal Inference in Productivity Analysis: The Effect of Contract Farming on Sunflower Productivity in Tanzania, Selected Paper for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28.
- Henningsen G., Henningsen A. and Jensen U. (2013). A Monte Carlo Study on Multiple Output Stochastic Frontiers: A Comparison of Two Approaches. IFRO Working Paper, 7: 1-31.
- Hothorn T. and others, (2015). *lmtest: Testing Linear Regression Models*. R package version 0.9-34.
- Itang'ata, M. J. J., (2013). "A Comparative Study of Exact Versus Propensity Matching Techniques Using Monte Carlo Simulation" (2013). Doctorate dissertations: Educational Leadership, Research and Technology. Michigan: Western Michigan University.
- Kumbhakar S. and Lovell C. A. K., 2000, *Stochastic frontier analysis*, New York, Cambridge University Press, 2000.
- Lau L. J. and Yotopoulos P. A. (1989). The Meta-Production Function Approach to Technological Change in World Agriculture, *Journal of Development Economics*, 31:241-269.
- Matawie K. M. and Assaf A., 2008, Efficiency Comparison of Heterogeneous Operational Groups, *World Academy of Science, Engineering and Technology* 40.
- O'Donnell C., Rao D. and Battese G., 2007, Metafrontier frameworks for the study of firm-level efficiencies and technology ratios, *Empirical Economics*, 34(2): 231-255.

- Peng X., Jing P. and Feng Z. (2011). Bootstrap Confidence Intervals for the Estimation of Average Treatment Effect on Propensity Score. *Journal of Mathematics Research*, 3 (3). 7p.
- RAO P. D. S., O'DONNELL C. J. and BATTESE G. E., 2003, Metafrontier Functions for the Study of Interregional Productivity Differences, Centre for Efficiency and Productivity Analysis, Working Paper Series (01), 37p.
- Rosenbaum P. R. and Rubin D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1): 41-55.
- Rosenbaum P. R. and Rubin D. B. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39:33–38.
- Sodjinou E. (2012). Evaluation d'impact des interventions par la méthode du propensity score matching (PSM) : théorie et application. PISB/CRA-Agonkanmey/INRAB, 16p.
- Sriboonchitta S. and Wiboonpongse A. (2004). On Estimation of Stochastic Production-Frontiers with Self-Selectivity: Jasmine and Non-Jasmine Rice in Thailand. Paper presented at the 4th Asia-Pacific Productivity Conference, Brisbane, Australia, July 14-16-2004.
- Stone C. A. and Tang Y. (2013). Comparing Propensity Score Methods in Balancing Covariates and Recovering Impact in Small Sample Educational Program Evaluations. *Practical Assessment, Research & Evaluation*, 18 (13).
- Varadhan R., 2014. Numerical Optimization in R: Beyond optim. *Journal of Statistical Software*, September 2014, 60(1).
- Vershelde M., Dumont M., Rayp G. and Merlevede B. (2014). A Semiparametric Stochastic Metafrontier Analysis at the Firm Level. *European Central Bank Working Paper*, (1701): 1-145.
- Viechtbauer W., 2010. Conducting Meta-Analyses in R with the metaphor Package. *Journal of Statistical Software*, August 2010, 36(3).
- Villano R., Mehrabi Boshrabadi H. and Fleming E. (2010). When Is Metafrontier Analysis Appropriate? An Example of Varietal Differences in Pistachio Production in Iran. *J. Agr. Sci. Tech*, 12: 379-389

## **APPENDIX**

Appendix 1: Categorization of the commonly used impact evaluation methods according to the available data (Sodjinou, 2012)



Appendix 2: Conventional cotton SFA

	Cobb-Douglas		Translog	
	Estimate	Standard Error	Estimate	Standar Error
(Intercept)	1.868577	2.084529	-8.73273422	11.24567266
Fertilizer	0.173802	0.151586	-4.81731366 (*)	2.36202978
Pesticide	0.351576 (*)	0.144199	11.41376139 (***)	1.80581750
Labour	0.027544	0.098577	-3.19101126 (*)	1.54381081
Depreciation	-0.025076	0.056977	-2.08107197	1.65756714
0.5*Fertilizer*Fertilizer			-0.22859296	0.17708777
0.5*Fertilizer*Pestide			1.76041482 (**)	0.61050198
0.5*Fertilizer*labour			-0.47997282	0.44392342
0.5*Fertilizer*Depreciation			-0.12220104	0.43774139
0.5*Pestide*Pestide			-2.32771157 (***)	0.45960912
0.5*Pestide*labour			0.83718010 (*)	0.34694774
0.5*Pestide*Depreciation			0.28523535	0.32141539
0.5*labour*labour			0.03919098	0.17040333
0.5*labour*Depreciation			0.25845665	0.18739752
0.5*Depreciation*Depreciation			0.05918410	0.12380291
sigmaSq	0.809375 (***)	0.278671	1.28006171 (***)	0.30926989
gamma	0.729061 (***)	0.225292	0.99999999 (***)	0.00057847
Degree of freedom	7		17	
Log-likelihood	-87.169		-74.073	
Mean efficiency	0,59461387		0,48342313	
Likelihood Ratio Test	26.192 [Pr(>Chisq)= 0.003491] with Df = 10			

Appendix 3: Organic cotton SFA

	Cobb-Douglas		Translog	
	Estimate	Standard Error	Estimate	Standard Error
(Intercept)	5.4575 (***)	0.38786	-0.18021 (***)	1.1417
Fertilizer	0.037347	0.043804	-1.3412	1.1542
Pesticide	0.10228 (*)	0.041103	1.3496	0.88453
Labour	0.063753	0.041996	6.0151 (***)	1.3846
Depreciation	0.036872	0.078591	2.1238 (*)	1.0472
0.5*Fertilizer*Fertilizer			-0.13477	0.15941
0.5*Fertilizer*Pesticide			0.45139	0.28695
0.5*Fertilizer*labour			-0.24733	0.22768
0.5*Fertilizer*Depreciation			0.23921	0.17627
0.5*Pesticide*Pesticide			0.023394	0.17376
0.5*Pesticide*labour			-0.76895 (***)	0.16261
0.5*Pesticide*Depreciation			-0.36866 (*)	0.16384
0.5*labour*labour			-0.023698	0.087692
0.5*labour*Depreciation			-0.20943 (*)	0.09657
0.5*Depreciation*Depreciation			-0.11309 (*)	0.049693
sigmaSq	1.9753 (***)	0.18149	1.7273	0.13007
gamma	1.0000 (***)	6.0079 e-06	1.000000	0.00011742
Degree of freedom	7		17	
Log-likelihood	-86.081		-73.655	
Mean efficiency	0,39071927		0,44997102	
Likelihood Ratio Test	24.851[Pr(>Chisq)= 0.005635] with Df = 10			



Appendix 4: Pooled data SFA

	Cobb-Douglas		Translog	
	Estimate	Standard Error	Estimate	Standard Error
Intercept)	5.300432 (***)	0.978191	13.02017652	17.52216529
Fertilizer	-0.059685	0.080830	-2.24199131	1.84245115
Pesticide	0.216814 (**)	0.081593	-1.48836220	1.65575921
Labour	0.098755 (.)	0.057362	2.91825045	1.18466802 (*)
Depreciation	0.006129	0.045069	0.78597167	1.15359817
0.5*Fertilizer*Fertilizer			-0.00036776	0.12909837
0.5*Fertilizer*Pestide			0.45175557	0.27399391 (.)
0.5*Fertilizer*labour			-0.19425128	0.15455778
0.5*Fertilizer*Depreciation			0.09401695	0.13455244
0.5*Pestide*Pestide			0.16583704	0.11416129
0.5*Pestide*labour			-0.39243630	0.15959296 (*)
0.5*Pestide*Depreciation			-0.22662301	0.12481970 (.)
0.5*labour*labour			0.08100537	0.10408878
0.5*labour*Depreciation			-0.06388461	0.10711269
0.5*Depreciation*Depreciation			0.00168813	0.05723610
sigmaSq	1.559846 (***)	0.322321	1.28668093	0.26645168 (***)
gamma	0.937633 (***)	0.060095	0.91065729	0.06775662 (***)
Degree of freedom	5		23	
Log-likelihood	-181.05		-168.32	
Mean efficiency	0,46284576		0,49815812	
Likelihood Ratio Test	25.459 [Pr(>Chisq)= 0.04411] with Df = 15			

Appendix 5: Stochastic Metafrontier parameters

	Estimate	Bootstrap Estimate	Confidence Interval		Standard Error	t.test	Pr(> t )
(Intercept)	0.000000	15.07016981	3.10593386	27.03440576	431.53568647	2.469372	1.356823e-02
Fertilizer	0.000000	0.03327501	0.02831471	0.03823530	0.17891186	13.151159	7.356558e-39
Pesticide	8.462695	8.49587902	8.4400434	8.55170871	2.01370996	298.329639	0.000000e+00
Labour	0.000000	0.49991790	0.38441169	0.61542411	4.16617086	8.484898	2.810976e-17
Depreciation	1.663077	1.75904273	1.73095907	1.78712638	1.01294376	122.793691	0.000000e+00
0.5*Fertilizer*Fertilizer	0.000000	0.00000000	NaN	NaN	0.00000000	NaN	NaN
0.5*Fertilizer*Pesticide	0.000000	0.05171300	0.04079204	0.06263395	0.39390590	9.283083	2.397179e-20
0.5*Fertilizer*labour	0.000000	0.20970237	0.18197018	0.23743457	1.00026711	14.824237	1.090487e-48
0.5*Fertilizer*Depreciation	0.000000	0.01671937	0.01462982	0.01880892	0.07536765	15.686280	3.608571e-54
0.5*Pesticide*Pesticide	0.000000	0.00000000	NaN	NaN	0.00000000	NaN	NaN
0.5*Pesticide*labour	0.000000	0.26919550	0.20507096	0.33332005	2.31289576	8.229941	2.367011e-16
0.5*Pesticide*Depreciation	0.000000	0.02564025	0.01889035	0.03239015	0.24346076	7.446948	1.118915e-13
0.5*labour*labour	0.000000	0.00000000	NaN	NaN	0.00000000	NaN	NaN
0.5*labour*Depreciation	0.000000	0.14760407	0.11971704	0.17549111	1.00585201	10.376461	5.679864e-25
0.5*Depreciation*Depreciation	0.000000	0.00000000	NaN	NaN	0.00000000	NaN	NaN

Appendix 6: Comparison between SFA and SM

	Conventional cotton SFA	Organic cotton SFA	Pooled data SFA	Metafrontier
(Intercept)	-8.73273422	-0.18021 (***)	17.52216529	15.07016981 (***)
Fertilizer	-4.81731366 (*)	-1.3412	1.84245115	0.03327501 (***)
Pesticide	11.41376139 (***)	1.3496	1.65575921	8.49587902 (***)
Labour	-3.19101126 (*)	6.0151 (***)	1.18466802 (*)	0.49991790 (***)
Depreciation	-2.08107197	2.1238 (*)	1.15359817	1.75904273 (***)
0.5*Fertilizer*Fertilizer	-0.22859296	-0.13477	0.12909837	0.00000000
0.5*Fertilizer*Pestide	1.76041482 (**)	0.45139	0.27399391 (.)	0.05171300 (***)
0.5*Fertilizer*labour	-0.47997282	-0.24733	0.15455778	0.20970237 (***)
0.5*Fertilizer*Depreciation	-0.12220104	0.23921	0.13455244	0.01671937 (***)
0.5*Pestide*Pestide	-2.32771157 (***)	0.023394	0.11416129	0.00000000
0.5*Pestide*labour	0.83718010 (*)	-0.76895 (***)	0.15959296 (*)	0.26919550 (***)
0.5*Pestide*Depreciation	0.28523535	-0.36866 (*)	0.12481970 (.)	0.02564025 (***)
0.5*labour*labour	0.03919098	-0.023698	0.10408878	0.00000000
0.5*labour*Depreciation	0.25845665	-0.20943 (*)	0.10711269	0.14760407 (***)
0.5*Depreciation*Depreciation	0.05918410	-0.11309 (*)	0.05723610	0.00000000

Appendix 7: Summary of T.test result for return to scale

Type of test	Hypothesis		DEA frontier for		DEA Metafrontier
	Null	Alternative	Conventional cotton	Organic cotton	
$\theta^{IRS} > \theta^{DRS}$	CRS	IRS	-2.0589 (0.9795)	1.1039 (0.1357)	-1.2978 (0.9024)
$\theta^{IRS} < \theta^{DRS}$	CRS	DRS	-2.0589 (0.0205)	1.1039 (0.8643)	-1.2978 (0.09765)
Efficiency mean for True production technology			0.4528555	0.5290699	0.3752886
Efficiency standard deviation for True production technology			0.268351	0.3691766	0.3023847

(p.value of t.test)

Appendix 8: Comparison of TGR between DEA and Stochastic Metafrontier

data: DTGR and STGR

$t = 14.9308$ ,  $df = 160$ ,  $p\text{-value} < 2.2e-16$

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

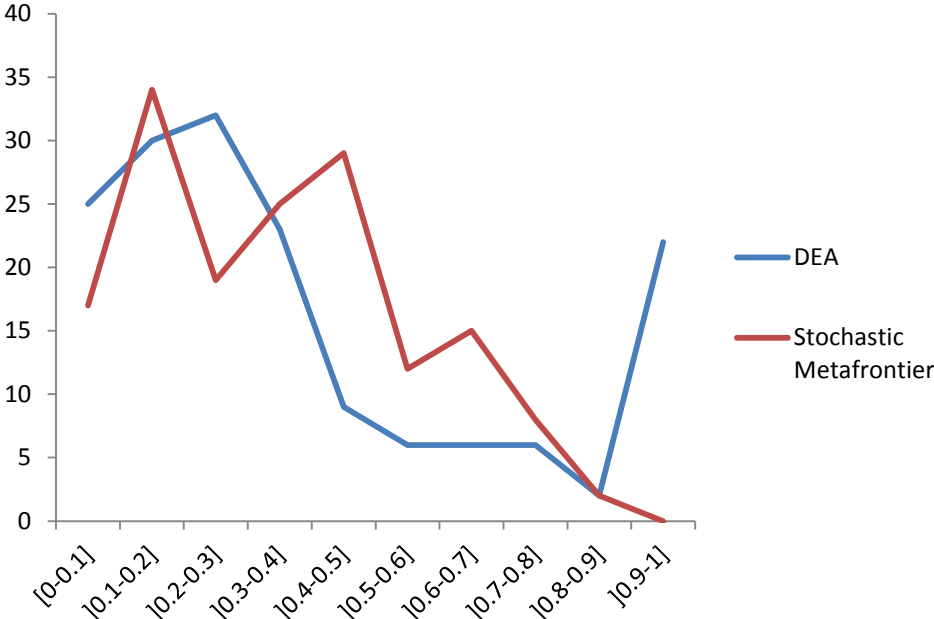
0.1842847 0.2404669

sample estimates:

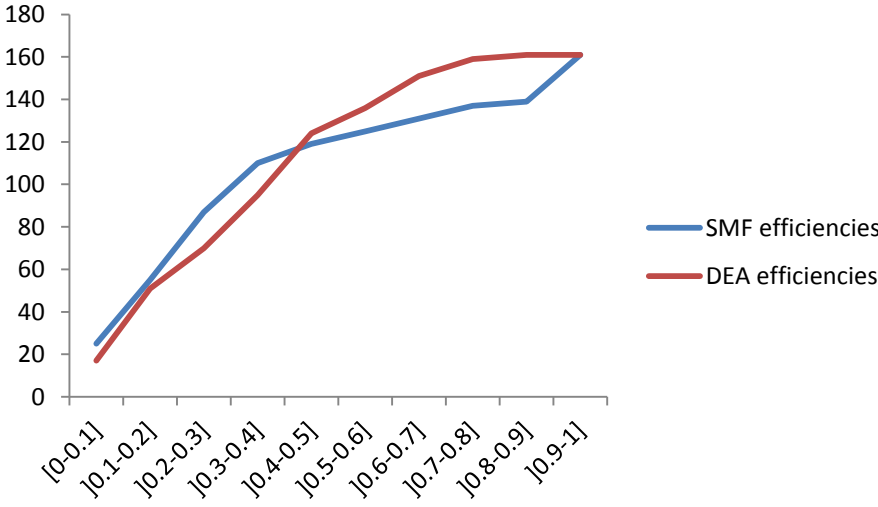
mean of the differences

0.2123758

Appendix 9: Comparison of efficiencies between DEA and Stochastic Metafrontier



A= frequencies polygon of efficiencies



B= cumulative frequencies distribution

## Appendix 10: Sensitivity analysis

Rosenbaum bounds for efficiency (N = 128 matched pairs)

Gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	0	0	2.8325	2.8325	2.59	3.08
1.5	5.6e-16	0	2.59	3.08	2.365	3.315
2	1.9e-12	0	2.43	3.23	2.205	3.49
2.5	2.7e-10	0	2.32	3.365	2.065	3.62
3	7.2e-09	0	2.235	3.46	1.94	3.71
3.5	7.7e-08	0	2.16	3.54	1.835	3.775
4	4.6e-07	0	2.08	3.605	1.745	3.84
4.5	1.8e-06	0	2.02	3.665	1.68	3.895
5	5.7e-06	0	1.95	3.705	1.605	3.935
5.5	.000014	0	1.9	3.735	1.54	3.965
6	.000031	0	1.85	3.77	1.485	3.995
6.5	.000059	0	1.81	3.8	1.44	4.025
7	.000103	0	1.7725	3.825	1.385	4.055
7.5	.000169	0	1.74	3.845	1.34	4.085
8	.000259	0	1.71	3.865	1.3	4.11

gamma - log odds of differential assignment due to unobserved factors

sig+ - upper bound significance level

sig- - lower bound significance level

t-hat+ - upper bound Hodges-Lehmann point estimate

t-hat- - lower bound Hodges-Lehmann point estimate

CI+ - upper bound confidence interval (a= .95)

CI- - lower bound confidence interval (a= .95)

## Appendix 11: R code for Metafrontier analysis

```
##### IMPORTATION OF DATA#####
conbio<-read.table("clipboard",h=T,dec=",")
N1=88
N2=161
NIN=N1+1
N3=N2-N1
conven<-conbio[1:N1,]
organi<-conbio[NIN:N2,]
#attach()
#View(conbio)
#View(conven)
#View(organi)

##### LOG TRANSFORMATION #####
####conventionnal cotton
A<-log(conven$ENGR)
B<-log(conven$PEST)
C<-log(conven$MOT)
E<-log(conven$AMORT)
inter1<-0.5*A*A
inter2<-0.5*A*B
inter3<-0.5*A*C
inter5<-0.5*A*E
inter6<-0.5*B*B
inter7<-0.5*B*C
inter9<-0.5*B*E
inter10<-0.5*C*C
inter12<-0.5*C*E
inter15<-0.5*E*E
P<-log(conven$YIELD)

#### organic cotton
AA<-log(organi$ENGR)
BB<-log(organi$PEST)
CC<-log(organi$MOT)
EE<-log(organi$AMORT)
intero1<-0.5*AA*AA
intero2<-0.5*AA*BB
intero3<-0.5*AA*CC
intero5<-0.5*AA*EE
intero6<-0.5*BB*BB
intero7<-0.5*BB*CC
intero9<-0.5*BB*EE
intero10<-0.5*CC*CC
intero12<-0.5*CC*EE
intero15<-0.5*EE*EE
PPO<-log(organi$YIELD)

#### Pooled data
```



```

PA<-log(conbio$ENGR)
PB<-log(conbio$PEST)
PC<-log(conbio$MOT)
PE<-log(conbio$AMORT)
pinter1<-0.5*PA*PA
pinter2<-0.5*PA*PB
pinter3<-0.5*PA*PC
pinter5<-0.5*PA*PE
pinter6<-0.5*PB*PB
pinter7<-0.5*PB*PC
pinter9<-0.5*PB*PE
pinter10<-0.5*PC*PC
pinter12<-0.5*PC*PE
pinter15<-0.5*PE*PE
PP<-log(conbio$YIELD)

##### STOCHASTIC FRONTIER#####
library(frontier)

####Conventionnal cotton
## Cobb Douglas
sfacobb<-sfa(P~A+B+C+E)

## Translog
sfatrans<-sfa(P~A+B+C+E+inter1+inter2+inter3+inter5+inter6+inter7
+inter9+inter10+inter12+inter15)

##LR test
require(lmtest)
lrcon <- lrtest(sfacobb,sfatrans)

##result
#summary(sfacobb,which="mle")
#summary(sfatrans,which="mle")
#efficiencies(sfacobb)
#efficiencies(sfatrans)
coecon<-coef(sfatrans,which="mle",extraPar = FALSE)
coecon<-as.vector(coecon)
coecon<-coecon[-16:-17]
sumcoecon<-coef( summary( sfatrans ), which = "mle" )
#coecon<-t(coecon)

#### organic cotton
## Cobb Douglas
sfacobbo<-sfa(PPO~AA+BB+CC+EE)

## Translog
sfatranso<-sfa(PPO~AA+BB+CC+EE+intero1+intero2+intero3+intero5+intero6+intero7
+intero9+intero10+intero12+intero15)

```

```

##LR test
require(lmtest)
lrorg <- lrtest(sfacobbo,sfatranso)

##result
#summary(sfacobbo,which="mle")
#summary(sfatranso,which="mle")
#efficiencies(sfacobbo)
#efficiencies(sfatranso)
coeor<-coef(sfatranso,which="mle",extraPar = FALSE)
coeor<-as.vector(coeor)
coeor<-coeor[-16:-17]
sumcoeor<-coef( summary( sfatranso ), which = "mle" )
#coeor<-t(coeor)

#### Pooled data
## Cobb Douglas
psfacobb<-sfa(PP~PA+PB+PC+PE)

## Translog
psfatrans<-sfa(PP~PA+PB+PC+PE+pinter1+pinter2+pinter3+pinter5+pinter6+pinter7
+pinter9+pinter10+pinter12+pinter15)

##LR test
require(lmtest)
lrpool<- lrtest(psfacobb,psfatrans)

##result
#summary(psfacobb,which="mle")
#summary(psfatrans,which="mle")

##### STOCHASTIC METAFRONTIER #####
#### Numeric vector of coefficients of objective function (vectors of variables means)
##computation of mean vector
optef<-conbio[,-1:-2]
#View(optef)
optef$ENEN<-sqrt(optef$ENGR*optef$ENGR)
optef$ENPE<-sqrt(optef$ENGR*optef$PEST)
optef$ENMO<-sqrt(optef$ENGR*optef$MOT)
optef$ENAM<-sqrt(optef$ENGR*optef$AMORT)
optef$PEPE<-sqrt(optef$PEST*optef$PEST)
optef$PEMO<-sqrt(optef$PEST*optef$MOT)
optef$PEAM<-sqrt(optef$PEST*optef$AMORT)
optef$MOMO<-sqrt(optef$MOT*optef$MOT)
optef$MOAM<-sqrt(optef$MOT*optef$AMORT)
optef$AMAM<-sqrt(optef$AMORT*optef$AMORT)
#View(optef)
f.obj <- c(colMeans(optef))
f.obj<-as.vector(f.obj) ##means vector
#View(f.obj)

```

```

##### Matrix of numeric constraint coefficients
f.con<-data.matrix(optef,rownames.force =NA)
#View(f.con)

#####direction of the constraint:"<," "<=," "=", "==" ,">," or ">=".
f.dir <- rep(">=",N2)
#View(f.dir)

#####Vector of numeric values for the right-hand sides of the constraints
f.rhs1<-f.con[1:N1,]*%*%coecon
f.rhs2<-f.con[NIN:N2,]*%*%coeor
f.rhs<-rbind(f.rhs1,f.rhs2)

#####Metafrontier estimators
library(lpSolve)
lp ("min", f.obj, f.con, f.dir, f.rhs)
solu<-lp ("min", f.obj, f.con, f.dir, f.rhs)$solution

##### Efficiency#####
##### Stochastic Frontier efficiencies

TEC<-efficiencies(sfatrans) ###conventional cotton efficiencies
TEO<-efficiencies(sfatranso)### organic cotton efficiencies
TEP<-efficiencies(psfatrans) ### pooled data efficiencies
TE<-rbind(TEC,TEO)
TEC<-as.vector(TEC)
TEO<-as.vector(TEO)
TEP<-as.vector(TEP)
TE<-as.vector(TE)
t.test(TE,TEP)
par(mfrow=c(1,2))
hist(TEC, main="conventional cotton farmers")
hist(TEO,breaks= c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1), main="organic cotton farmers")

#####stochastic Metafrontier efficiencies
##deterministic part of conventional cotton SFA
interac<-rep(1,N1)
logcon<-cbind(interac,A,B,C,E,inter1,inter2,inter3,inter5,inter6,inter7
,inter9,inter10,inter12,inter15)
#View(logcon)
bexpcon<-logcon%*%coecon

##deterministic part of organic cotton SFA
intera0<-rep(1,N3)
logor<-cbind(intera0,AA,BB,CC,EE,intero1,intero2,intero3,intero5,intero6,intero7
,intero9,intero10,intero12,intero15)
bexporg<-logor%*%coeor

```

```

##deterministic part of SFA (fusion of conventional and organic cotton)
bexpx<-rbind(bexpcon,bexporg)

##deterministic part of metafrontier
logcott<-rbind(logcon,logor)
solu
mbexpx<-logcott%*%solu
as.vector(solu)
is.matrix(logcott)
str(solu)
exp<-cbind(bexpx,mbexpx)

##Technology Gap Ratio
TGR<-bexpx/mbexpx

## metafrontier efficiency
MTE<-TGR*TE
hist(MTE[1:88], main="conventional cotton farmers")
hist(MTE[89:161], main="organic cotton farmers")

efficie<-cbind(TEC,TEO,TEP,MTE)
mean(MTE)
mean(MTE[1:88])
mean(MTE[89:161])
sd(MTE[1:88])
sd(MTE[89:161])
t.test(MTE[1:88],MTE[89:161])

diffpol<-TE-TEP
diffsmf<-TE-MTE
diff<-cbind(diffpol,diffsmf)

##### DEA #####
library(Benchmarking)

####Conventionnal cotton
XXC<-cbind(conven$ENGR,conven$PEST,conven$MOT,conven$AMORT)
YYC<-conven$YIELD
XXC<-as.matrix(XXC)
YYC<-as.matrix(YYC)
tedeac1<-dea(XXC,YYC, RTS="irs", ORIENTATION="out")$eff
tedeac1<-1/tedeac1
tedeac2<-dea(XXC,YYC, RTS="drs", ORIENTATION="out")$eff
tedeac2<-1/tedeac2

```

```

tedeac3<-dea(XXC,YYC, RTS="crs", ORIENTATION="out")$eff
tedeac3<-1/tedeac3
t.test(tedeac1,tedeac2,alternative = c("two.sided"))
t.test(tedeac1,tedeac2,alternative = c("less"))
t.test(tedeac1,tedeac2,alternative = c("greater"))

#### Organic cotton
XXO<-cbind(organi$ENGR,organi$PEST,organi$MOT,organi$AMORT)
YYO<-organi$YIELD
XXO<-as.matrix(XXO)
YYO<-as.matrix(YYO)
tedeo1<-dea(XXO,YYO, RTS="irs", ORIENTATION="out")$eff
tedeo1<-1/tedeo1
tedeo2<-dea(XXO,YYO, RTS="drs", ORIENTATION="out")$eff
tedeo2<-1/tedeo2
tedeo3<-dea(XXO,YYO, RTS="crs", ORIENTATION="out")$eff
tedeo3<-1/tedeo3
t.test(tedeo1,tedeo2,alternative = c("two.sided"))
t.test(tedeo1,tedeo2,alternative = c("less"))
t.test(tedeo1,tedeo2,alternative = c("greater"))

f.obj <- c(colMeans(optef))
f.obj<-as.vector(f.obj) ##means vector
#View(f.obj)

#### Matrix of numeric constraint coefficients
f.con<-data.matrix(optef,rownames.force =NA)
#View(f.con)

####direction of the constraint:"<," "<="," "=", " "==," ">," or ">=" .
f.dir <- rep(">=",N2)
#View(f.dir)
R=5000

```

```
cb0<-rnorm(R,sumcoecon[1,1],sumcoecon[1,2])
cb1<-rnorm(R,sumcoecon[2,1],sumcoecon[2,2])
cb2<-rnorm(R,sumcoecon[3,1],sumcoecon[3,2])
cb3<-rnorm(R,sumcoecon[4,1],sumcoecon[4,2])
cb4<-rnorm(R,sumcoecon[5,1],sumcoecon[5,2])
cb5<-rnorm(R,sumcoecon[6,1],sumcoecon[6,2])
cb6<-rnorm(R,sumcoecon[7,1],sumcoecon[7,2])
cb7<-rnorm(R,sumcoecon[8,1],sumcoecon[8,2])
cb8<-rnorm(R,sumcoecon[9,1],sumcoecon[9,2])
cb9<-rnorm(R,sumcoecon[10,1],sumcoecon[10,2])
cb10<-rnorm(R,sumcoecon[11,1],sumcoecon[11,2])
cb11<-rnorm(R,sumcoecon[12,1],sumcoecon[12,2])
cb12<-rnorm(R,sumcoecon[13,1],sumcoecon[13,2])
cb13<-rnorm(R,sumcoecon[14,1],sumcoecon[14,2])
cb14<-rnorm(R,sumcoecon[15,1],sumcoecon[15,2])
```

```
ob0<-rnorm(R,sumcoeor[1,1],sumcoeor[1,2])
ob1<-rnorm(R,sumcoeor[2,1],sumcoeor[2,2])
ob2<-rnorm(R,sumcoeor[3,1],sumcoeor[3,2])
ob3<-rnorm(R,sumcoeor[4,1],sumcoeor[4,2])
ob4<-rnorm(R,sumcoeor[5,1],sumcoeor[5,2])
ob5<-rnorm(R,sumcoeor[6,1],sumcoeor[6,2])
ob6<-rnorm(R,sumcoeor[7,1],sumcoeor[7,2])
ob7<-rnorm(R,sumcoeor[8,1],sumcoeor[8,2])
ob8<-rnorm(R,sumcoeor[9,1],sumcoeor[9,2])
ob9<-rnorm(R,sumcoeor[10,1],sumcoeor[10,2])
ob10<-rnorm(R,sumcoeor[11,1],sumcoeor[11,2])
ob11<-rnorm(R,sumcoeor[12,1],sumcoeor[12,2])
ob12<-rnorm(R,sumcoeor[13,1],sumcoeor[13,2])
ob13<-rnorm(R,sumcoeor[14,1],sumcoeor[14,2])
ob14<-rnorm(R,sumcoeor[15,1],sumcoeor[15,2])
```

```
convcoedata<-
```

```
as.matrix(cbind(cb0,cb1,cb2,cb3,cb4,cb5,cb6,cb7,cb8,cb9,cb10,cb11,cb12,cb13,cb14))
```

```
#View(convcoedata)
```

```

orgcoedata<-
as.matrix(cbind(ob0,ob1,ob2,ob3,ob4,ob5,ob6,ob7,ob8,ob9,ob10,ob11,ob12,ob13,ob14))
#View(orgcoedata)

R1=5000
con<-as.matrix
concoe<-matrix(0,15,R1)
orgcoe<-matrix(0,15,R1)
concoe<-as.vector
orgcoe<-as.vector
#f.rhs1<-matrix(0,N1,R1)
#f.rhs1<-matrix(0,N3,R1)
#f.rhs<-matrix(0,N2,R1)
solu<-matrix(0,R1,15)
for (i in 1:R1)
{
#####Vector of numeric values for the right-hand sides of the constraints
concoe<-convcoedata[i,]
orgcoe<-orgcoedata[i,]
f.rhs1<-f.con[1:N1,]%*%concoe
f.rhs2<-f.con[NIN:N2,]%*%orgcoe
f.rhs<-rbind(f.rhs1,f.rhs2)
solu[i,]<-lp ("min", f.obj, f.con, f.dir, f.rhs)$solution
}
View(solu)
mfparmean<-colMeans(solu)

quantile (solu[,3], probs= c(0.05, 0.95))
#####Metafrontier estimators
library(lpSolve)
lp ("min", f.obj, f.con, f.dir, f.rhs)
solu<-lp ("min", f.obj, f.con, f.dir, f.rhs)$solution

#mbexpxb<-logcott%*%mfparmean

```

```

#TGRB<-bexp/mbexpb
#MTEB<-TGRB*TE
#hist(MTEB[1:88], main="conventional cotton farmers")
#hist(MTEB[89:161], main="organic cotton farmers")

#### DEA Metafrontier
XX<-cbind(conbio$ENGR,conbio$PEST,conbio$MOT,conbio$AMORT)
YY<-conbio$YIELD
XX<-as.matrix(XX)
YY<-as.matrix(YY)
mtedea1<-dea(XX,YY, RTS="irs", ORIENTATION="out")$eff
mtedea1<-1/mtedea1
mtedea2<-dea(XX,YY, RTS="drs", ORIENTATION="out")$eff
mtedea2<-1/mtedea2
mtedea3<-dea(XX,YY, RTS="crs", ORIENTATION="out")$eff
mtedea3<-1/mtedea3

t.test(mtedea1,mtedea2,alternative = c("two.sided"))
t.test(mtedea1,mtedea2,alternative = c("less"))
t.test(mtedea1,mtedea2,alternative = c("greater"))

##### Summary of efficiencies
tedeac2<-as.vector(tedeac2)
tedeac2<-t(tedeac2)
tedeao3<-as.vector(tedeao3)
tedeao3<-t(tedeao3)
mtedea3<-as.vector(mtedea3)
mtedea3<-t(mtedea3)
mean(tedeac2); mean(tedeao3);
mean(mtedea3[1:88]);mean(mtedea3[89:161]);mean(mtedea3)
sd(tedeac2); sd(tedeao3); sd(mtedea3[1:88]);sd(mtedea3[89:161]);sd(mtedea3)

tedea<-cbind(tedeac2,tedeao3)##the choice of tedeac depend on results of t.test
tedea<-t(tedeac)

```



```

mtedea3<-t(mtedea3)
effici<-cbind(tedeac2,mtedea3)##the choice of mtedea depend on results of t.test
View(effici)
View(tedeac2)

###t, test between metafrontier and frontier
t.test(mtedea3[1:88],t(tedeac2),paired =TRUE)
t.test(mtedea3[89:161],t(tedeao3),paired =TRUE)

##graphical analysis
par(mfrow=c(1,2))
hist(tedeac2, main="conventional cotton farmers")

dea.plot(XXC,YYC, RTS="crs", ORIENTATION="out")
hist(tedeao3, main="organic cotton farmers")

##Technologies gap
DTGR<-mtedea3/tedea
mean(DTGR[1:88]);mean(DTGR[89:161]);mean(DTGR)
sd(DTGR[1:88]);sd(DTGR[89:161]);sd(DTGR)
hist(DTGR[1:88], main="conventional cotton farmers")
hist(DTGR[89:161], main="organic cotton farmers")
t.test(DTGR[1:88],DTGR[89:161])
t.test(DTGR,STGR, paired=TRUE)

```

## Appendix 12: Stata code for Matching methods

```
ssc instal psmatch2
```

```
psmatch2 syst age sexe class distex nivins expcot asso nvisit , outcome( sme ) neighbor(1)  
common logit
```

```
pstest age sexe class distex nivins expcot asso nvisit, sum t( syst) support( _support)
```

```
histogram _pscore, by(syst)
```

```
psgraph
```

```
rbounds sme , gamma(1 (1) 60)
```

```
bootstrap r(att), reps(5000) :psmatch2 syst age sexe class distex nivins expcot asso nvisit ,  
outcome( sme ) neighbor(1) common logit
```

```
psmatch2 syst age sexe class distex nivins expcot asso nvisit , outcome( dme ) neighbor(1)  
common logit
```

```
pstest age sexe class distex nivins expcot asso nvisit, sum t( syst) support( _support)
```

```
histogram _pscore, by(syst)
```

```
psgraph
```

```
rbounds dme , gamma(1 (1) 60)
```

```
bootstrap r(att), reps(5000) :psmatch2 syst age sexe class distex nivins expcot asso nvisit ,  
outcome( dme ) neighbor(1) common logit
```