MEASURING THE EFFICIENCY OF COFFEE PRODUCERS IN VIETNAM: DO OUTLIERS MATTER?

A Thesis

Submitted to the Faculty

of

Purdue University

by

Nam Anh Tran

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

May 2007

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

First of all, I would like to express my greatest gratitude to my major professor, Dr. Gerald Shively, for helping me along the way. His invaluable ideas, comments, and constructive criticism play an important role in this thesis. I also want to thank you my other committee members: Dr. Paul Preckel and Dr. Otto Doering for their comments and suggestions.

I would like to thank the Ford Foundation for their financial support during my time in Purdue.

I would like to also thank the faculty and staff at Purdue University who broadened my knowledge in economics as well as aiding me along my program here.

My gratitude goes to my friends who made my two years at Purdue a great life experience.

Last but not least, I want to thank my parents and sister for their continuous support and understanding.

TABLE OF CONTENTS

Page

APPENDICES

LIST OF TABLES

Table Page

Appendix
Table

LIST OF FIGURES

Figure Page

ABSTRACT

Nam Anh Tran, M.S., Purdue University, May, 2007. Measuring Efficiency of Coffee Producers in Vietnam: Do Outliers Matter? Major Professor: Dr. Gerald Shively.

Coffee is one of Vietnam's most valuable exports, earning up to $600 million a year. However, a drop in producer coffee prices in the late 1990s has caused hardship for 4 million Vietnamese people who depend on the production of coffee for income. Also, Vietnam’s coffee production has been associated with negative environmental externalities. Using data from 209 households in Dak Lak province, this thesis focuses on measuring the level of productive efficiency among coffee producers and identifying sources of inefficiency in the sample. Specific attention is focused on the importance of outliers in the dataset. A new, simple method for detecting outliers is introduced.

Initial findings suggest that almost half of the coffee farmers in the study sites are producing at an efficient level. In addition, Kinh farmers are less efficient than non-Kinh farmers. There are strong correlations between efficiency scores and farm size, labor, fertilizer, and pesticide. When interaction variables conditioned by ethnicity are added to the model, an additional positive and significant impact of education is found for ethnic minorities.

With regard to outliers, the thesis develops a new method to detect them based on the weights of observations in the dataset. After outliers are removed, levels of efficiency and sources of inefficiency are re-examined. Results from the new sample suggest that farm size is not significantly correlated with efficiency levels. All the other correlations remain unchanged. Economic inference regarding sources of inefficiency is found to be sensitive to the removal of outliers.

CHAPTER 1. INTRODUCTION

1.1. Motivating Questions

How efficient are Vietnamese coffee producers? What factors help to explain patterns of inefficiency? Are these conclusions sensitive to the presence of outliers in the dataset? These questions motivate this thesis.

A primary objective for many agricultural producers, especially those who grow cash crops, is to produce efficiently. For those that are not efficient, it is important to understand why they are performing inefficiently so that production practices can be improved. To be successful, the process of identifying and improving efficiency must be accurate. This thesis aims to measure the efficiency of a sample of agricultural producers and identify sources of inefficiency in the sample. It also proposes a new method of removing outliers based on the weights of the observations under consideration. Efficiency levels and economic inferences regarding sources of inefficiency are shown to be sensitive to the presence or absence of outliers.

One of the earliest definitions of efficiency comes from Koopmans (1951) who identified a producer as technically efficient if an increase in one of its outputs could be achieved only at the cost of a decrease in another output or an increase in one of its inputs. More recently, Seaver (1995) argues that accurately measuring technical efficiency is a primary concern for economic analysts since efficiency measurements are used to make decisions concerning strategies for improving future performance. Since the data used to measure efficiency are typically not collected expressly for this purpose, it is important to closely scrutinize the approaches to understand the importance of such issues as outliers, collinearity, aggregation, and their impact on economic inference. This thesis examines the impact of outliers on the measurement of efficiency and economic inference regarding sources of inefficiency in a sample of coffee farmers in Vietnam. Outliers are defined as observations that are far away from the rest of the dataset and influential in the construction of the efficiency frontier. A primary concern motivating this thesis is that analysis using a dataset that includes outliers may be misleading. The approach taken here is to try to separate outliers from the rest of the observations before drawing conclusions from the sample. Timmer (1971), among others, highlighted the potential sensitivity of computed technical efficiency measures when there are outliers in the dataset and when linear programming approaches are used to measure efficiency. However, the question of whether the removal of outliers has impacts on economic inference from efficiency measurement has not been extensively studied. The perspective taken in this thesis is both conceptual and empirical: whether the treatment of outliers leads to a significant change in economic inference likely depends on the dataset under investigation and the method used to detect and remove outliers from a dataset constructed to measure productive efficiency.

1.2. Coffee in the World

The crop under investigation in this thesis is coffee. Most coffee is produced in equatorial countries facing severe development challenges. The story of how coffee growing and drinking spread around the world can be traced back to the 9th century in

Africa. The first coffee tree was found in the province of Kaffa, Ethiopia (International Coffee Organization, 2007). Local people used the beans of wild grown coffee to prepare a beverage. However, not until the fifteenth century was the first coffee tree planted by eastern Arabs. Mecca, Yemen was the first place to establish a coffee house. From there, the number of coffee house quickly increased throughout the Arab world and became places where social activities such as playing chess, singing, dancing, and exchanging gossip etc. took place (ICO, 2007). In the period from the 16th to the 18th century, coffee expanded its territory into Asia, Europe, and Americas. A lot of coffee houses were opened in Italy, England, France, German, and America. Nowadays, along with tea and water, coffee is one of the world most popular beverages (Cardenas, 2001). In the USA, the word's largest consumer, 52% of the adult population over 18 year of age drink coffee everyday. This represents 107 million daily drinkers (NCA 2001 cited in Karonen 2003). Despite the fact that coffee is widely consumed and has become one of the world's most important agricultural commodities, it is still mainly produced not on large plantations but on small holdings farmed by peasant households. Some 70% of the world's coffee is grown on farms of less than 10 hectares, many of them family-run operations. An estimated 25 million people are coffee farmers in 50 countries in Africa, Asia, Latin America, and the Caribbean. In total, 6.7 million tons of coffee were produced annually in 1998–2000. The FAO forecasts a rise in production to 7 million tons annually by 2010 (FAO, 2003).

As one of the worlds most traded agricultural products, coffee's importance can not be overstated. In many developing countries, coffee is the second biggest source of foreign exchange after oil. Coffee-related activities such as cultivation, processing, trading, transportation and marketing create job opportunities for millions of people all over the world (ICO, 2007). According to Cardenas (2001, pp.1), coffee plays a very important role in “creation of employment, balance payment, economic growth, public finance, income distribution and regional development” in around 50 producing countries. In some of these countries, coffee even induces economic and social development. As Cardenas (2001, pp. 2) states, with big coffee producing countries like Brazil, Mexico, and Columbia, coffee was the main motivation for “creating urban centers, transport infrastructures, land use, job opportunities and the expansion of both industrial sector and financial system.” In some of the world’s least developed countries, the foreign exchange earning from coffee exportation sometimes accounts for more than 80% of the total value of exports (ICO, 2007).

In developed countries, the coffee industry, in general, is considered to be profitable. However, current weakness in the world coffee price is causing hardship to a lot of developing countries where coffee is one of the main economic activities, and especially to the farmers who produce it. There is a big gap between the earning of coffee producing countries, mostly in developing areas, and the retail sales of coffee in the world, of which industrialized countries have the biggest share. In the early 1990s, this gap was $18 billion with $12 billion for producing countries and $30 billion for consuming countries. By 2002, retail sales of coffee exceed $70 billion but coffee producing countries only received $5.5 billion. One of the main reasons for the big drop in earnings for coffee producing countries is the drop in the world coffee price. Figure 1.1 illustrates the fluctuation of the world coffee price in nominal terms from 1976 to 2006. In 1980, one pound of coffee was worth 120 US cents. However, in 2002, one pound of coffee was only worth 50 US cents. It was the lowest price in real terms in the last 100 years. The drop in coffee earnings is causing severe losses for those countries where coffee plays a major part in export revenues (Osorio, 2002).

The decline in the world coffee price from the mid 1990s to 2002 was caused by a structural oversupply of coffee in the world markets. According to Oxfam (2002) the supply of coffee far exceeds demand. While world coffee demand has grown at the annual rate of 1-1.5 percent, the supply of coffee has grown at an annual rate of around 2 percent. This resulted in a supply of about 115 million bags in 2002, while consumption of coffee stood at 105-106 million bags. The year-on-year excess of supply has created a stock of over 40 million bags.[1] There are many reasons for this oversupply of coffee but one of the main ones is an astonishing increase of coffee production from Brazil and Vietnam. The production in Brazil has been recently boosted by changes in how and where coffee is grown (Oxfam 2002). “Increased mechanization, intense production methods, and a geographical shift away from the traditional, frost-prone growing areas have all increased yields” (Oxfam 2002, pp. 20). While Brazil has long been the world’s number one coffee producer, Vietnam in the 1990s quickly became the world’s second largest coffee producer with 15 million bags (primarily Robusta). Most production in Vietnam occurs on small farms.

---

[1] 1 bag weigh 60kg

1.3. Coffee in Vietnam

1.3.1. Coffee Booming

Coffee was first planted in Vietnam in the 19th century by French missionaries. Several coffee plantations were established in the Central Highlands or nearby areas. Coffee cultivation was encouraged and expanded during the late 1920s and 1930s. However, the coffee sector in Vietnam remained insignificant with almost no export to the world market until recently. Vietnam's high growth rate for coffee was only reached during the *Doi Moi* period beginning in 1986. Over the 13 years from 1990 to 2003 Vietnam's coffee growing area has increased 4.3 times, productivity has increased 1.95 times, and output has increased from 9200 tons in 1990 to 771,200 tons in 2003, an expansion of more than 700% (Thang, 2005). According to Luong (2006), 4 million Vietnamese people depend on the production of coffee for income. It is also one of the country's most valuable exports, with the annual earnings of up to $600 million (VIFOCA, 2005, Luong, 2006). Ninety five percent of total production is sold in the international market.

Figure 1.2 shows production levels for the main coffee producers in the world in 1990 and 2005. The figure shows how rapidly the coffee sector in Vietnam has expanded over the past 15 years. In fact, Vietnam is the only country with a big increase in coffee production. Vietnam went from being anonymous in the world coffee market (with 1.39 million bags in 1990) to being the second biggest producer after Brazil (with 11 million bags in 2005). There are both endogenous and exogenous factors that created the recent boom of coffee production in Vietnam

The country has suitable climate and soil conditions for coffee growing.

Central Highland provinces such as Dak Lak, Gia Lai, Kon Tum, Lam Dong, some Southeastern provinces as Dong Nai, Ba Ria-Vung Tau, Binh Phuoc, and some Coastal places are characterized by a hot and moist climate which are typical for growing Robusta coffee. Also, the Central Highlands, the main coffee production area of Vietnam, has basaltic red soil with high natural fertility and thick soil layers that are suitable for planting coffee. Realizing the favorable conditions for coffee growing, since 1975 the Vietnamese government has been implementing coffee development programs in these areas (Nhan, 2001).

The economic reforms promoting commercialization of agriculture.

Since 1986, after the sixth meeting of the communist party, Vietnam has been involved in a renovation process (*Doi Moi*) in order to promote social and economic development and closer integration with the rest of the world. The Vietnamese government undertook reform policies aimed at moving the country from a centrally-planned economy to a market-based system. New policies such as providing land use rights to farmers, individual access to credit and marketing system, lifting of price controls over inputs and outputs, and liberalization of internal and external trade have boosted the development of the agricultural and forestry sectors. These reforms have led to an increase in the commercialization of agriculture in general and coffee in particular.

Resettlement policies and immigration.

Population pressures and poverty in other areas and government polices have induced migration from the lowlands to the Central Highlands. Through immigration to the Central Highlands, the government of Vietnam wants to make use of natural resources to achieve economic growth and ease poverty in less-developed areas. The first group of Kinh people migrated into the Central Highlands in 1954 (Ahmad, 2000).[2] After 1975, the Vietnamese government set up some new economic zones and state farms in the Central Highlands for a large influx of immigrants from crowed provinces in the northern and coastal areas.

Simultaneously with the planned migration induced by the government, there was also a huge wave of unplanned immigration to the Central Highlands in the same period. This type of immigration is much higher than the official one. The main motivations for this latter one are poverty and demographic pressure in the North (Ahmad, 2000) and economic opportunities created by the coffee boom in the area.

High world coffee prices.

High coffee prices in the 1990s also attracted a lot of immigration in recent years and is one the main reasons for a sharp increase in coffee production area in Vietnam. In the early to mid 1990s, coffee prices increased rapidly due to crop loss and damage in Brazil caused by severe frost. The production of coffee in Brazil fell from 28.2 million

---

[2] Kinh are the ethnic majority in the Vietnamese population.

bags in 1993 to 18 million bags in 1995. Simultaneously, the world coffee price increased from 61.6 cents/lb to 138.4 cents/lb (ICO, 2007).

The increase in the world coffee price together with a flow of immigration into the Central Highlands and liberalization of agricultural commodity trading attracted many newcomers to Vietnam's coffee industry. Production rocketed, quickly making Vietnam the world's number two exporter. However, up until 1999, only state owned enterprises were involved in the export of coffee (Fontenay and Leung, 2002). There are state-licensed assemblers that connect farmers and state processors and exporters. According to Anh (1999, cited in Fontenay and Leung, 2002), the profit margin of state-licensed assemblers from coffee export far surpassed that of the state processors and exporters (40% compared to 0.9%). Since 1999, private coffee enterprises are permitted to export coffee. It is estimated that the export earnings from private firms account for around 30 to 40% of total coffee export revenue (Fontenay and Leung, 2002). The participation of private firms in coffee exporting activities has put an end to the monopoly of state-licensed assemblers. Since private and state-licensed assemblers compete, farm gate prices after 1999 have been a higher percentage of export prices, compared with the period before 1999. As estimated by VICOFA, a total of 149 coffee export companies operate in Vietnam.[3]

---

[3] http://www.vicofa.org.vn/vicofa/vci_s.shtml

1.3.2. Problems with Coffee Production in Vietnam

After the increase in coffee production in the first half of the 1990s, by 2002, the world coffee price had decreased dramatically to 50 cents/lb, its lowest level in the last 100 years. There are many factors associated with this price drop but one of the main reasons, ironically, was because of the rapid development of Vietnam's coffee sector. The coffee price crisis is causing hardship for 25 million coffee farmers world-wide. Vietnam, as the world second largest producer, is suffering from a lot of problems when coffee turns from boom to bust. According to Oxfam (2002), many coffee farms in the Central Highlands of Vietnam sell their beans for much less then they cost to produce- as little as 60 percent of production cost. In the mid-1990s, when coffee prices were high, 1kg of coffee could be exchanged for 5kg of rice. However, in 2002, this ratio dropped to 1 for 1 (ICO, 2003). Associated with economic crashes are social problems. Many children from medium to poor coffee households who largely depend on income from coffee have left school. Also, a survey in March 2002 showed that almost half (45%) of coffee-growing households in Vietnam lack nutritious food, about 66% have bank debts, and 45% have members who have to work for others to earn money (Oxfam-ICARD, 2002; ICO, 2003).

In addition to social and economic problems, coffee production in Vietnam also causes serious environmental externalities, especially in Dak Lak province in the Central Highlands where more than 50% of Vietnam's coffee area is located. Coffee producing families convert upland rice, forest, and other crop land into coffee. The expansion of coffee production in particular and the agricultural sector in general has led to a dramatic loss of forest cover. In the last 20 years, 20,000 hectares of forest a year in Dak Lak has been lost to both public and private coffee plantation and farms. “The forest cover decreased from about 90% in 1960s to 57% in 1995 and less than 50% in the late 1990s” (Oxfam-ICARD, 2002). Ha and Shively (2005) describe how the devastation of forest caused by coffee expansion has led to unsustainable ecological conditions, especially an inability to regulate water resources. Recently, farmers note that floods seem to be increasing in frequency and magnitude. The rapid expansion of coffee into fragile and steeply sloping areas, which is really common in the Central Highlands, also causes soil erosion and loss of soil nutrient (Ahmad, 2000; Ha et al, 2001; D’haeze et al, 2005; Lindskog, 2005). Some farmers report that they can even visibly see the roots of their coffee plants.

The sources of irrigation water for coffee are reservoirs, running water (springs, streams) and groundwater. However, during the dry season in the Central Highlands, the only source of irrigation available for coffee farmers is groundwater. Many farmers have their own wells, pumps, and irrigation tubes. On average, during the dry season, coffee plantations (about 1100 coffee trees/ha) need to be irrigated four times. Each time, every tree needs an average amount of water of 500 liters (Ahmad, 2000). This implies a huge demand for groundwater during the dry season. The continuous use of groundwater for coffee irrigation in recent years has caused a significant decline of groundwater levels. Excessive groundwater extraction has caused the water table to fall significantly, which has led to higher costs of pumping in many places and complete exhaustion of supplied in others (Danida 1998, cited in Ahmad 2000).

Besides the problem of groundwater overuse, coffee plantations in Vietnam also experience contamination of ground and surface water as a result of an inappropriate use of chemical fertilizer and pesticide. Coffee farmers mainly apply fertilizer based on their own perception of the correct proportion or combination of fertilizer, the supply of fertilizer, and affordability (Ahmad, 2000). There is no shortage of fertilizer in the market and because coffee is considered one of the most profitable crops, many farmers tend to use incredibly high amounts of fertilizer for coffee production. This does not always result in higher yields but often results in water contamination.

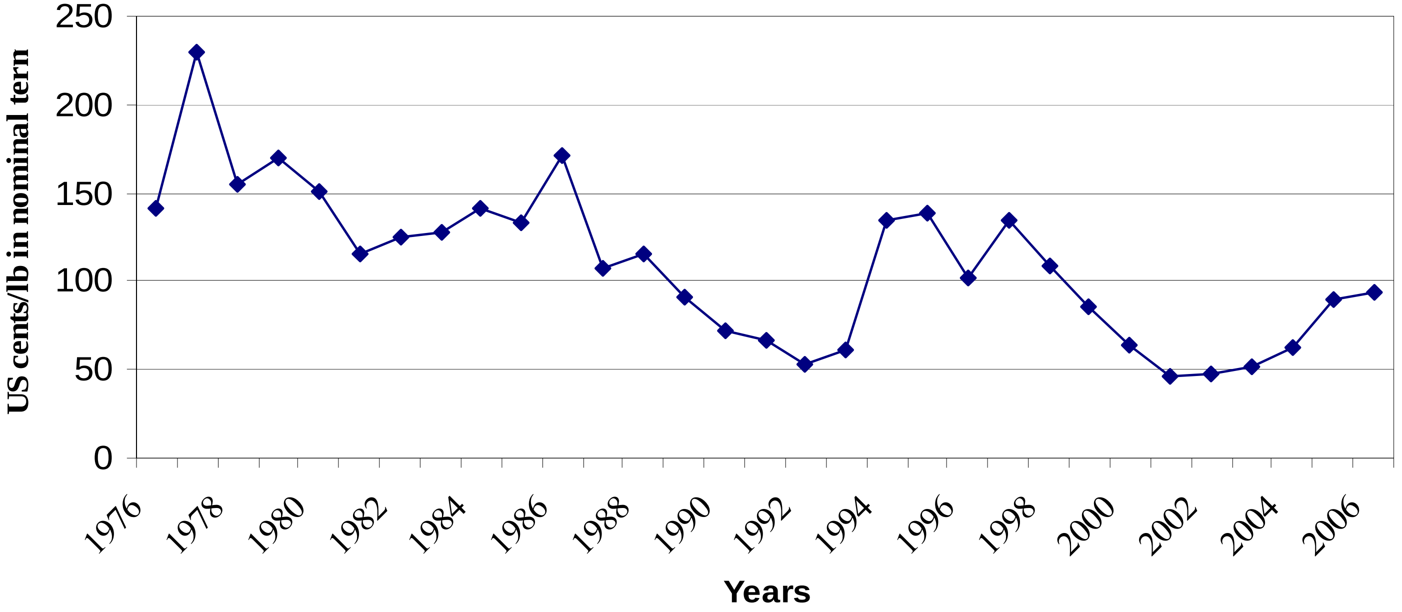
1.4. Objectives of the Thesis

In general, coffee is important to many developing countries including Vietnam, the second-largest coffee exporter in the world. The collapse in the price of coffee is hurting the livelihoods of 25 million coffee producers around the world as well as bringing hardship to coffee producers in Vietnam. In addition, the expansion of Vietnam's coffee production has been associated with inappropriate usage of chemical fertilizers and pesticides, high rates of water use, and environmental stresses, including deforestation, water scarcity, soil degradation, and contamination of surface and ground water. Considering the role of coffee in Vietnam's economy as well as the problems that country has encountered recently, it is of crucial importance to increase the efficiency of coffee operations in Vietnam.

The goal of this thesis is to understand the efficiency of coffee producers in Vietnam. There are two main types of efficiency: technical efficiency and allocative efficiency. Technical efficiency looks at the best combinations of inputs and outputs. Allocative efficiency incorporates input and output prices to study whether the economic benefits of using inputs can be improved relative to the economic costs. Since the prices of coffee fluctuate a lot from year to year and we only have data for one year, the allocative efficiency scores, if calculated for this sample, may be of limited value. For this reason, in this thesis, only technical efficiency is considered. Technical efficiency of coffee producers will be calculated and then based on that, sources of inefficiency will be identified. However, bearing in mind the possible importance of outliers, the thesis proposes a methodology to detect and remove them. A remaining question is whether the removal of outliers will have any impact on our conclusions regarding sources of inefficiency. From all the information gathered with respect to efficiency and sources of inefficiency, the thesis will have policy implications aiming at increasing the efficiency of coffee producers in Vietnam.

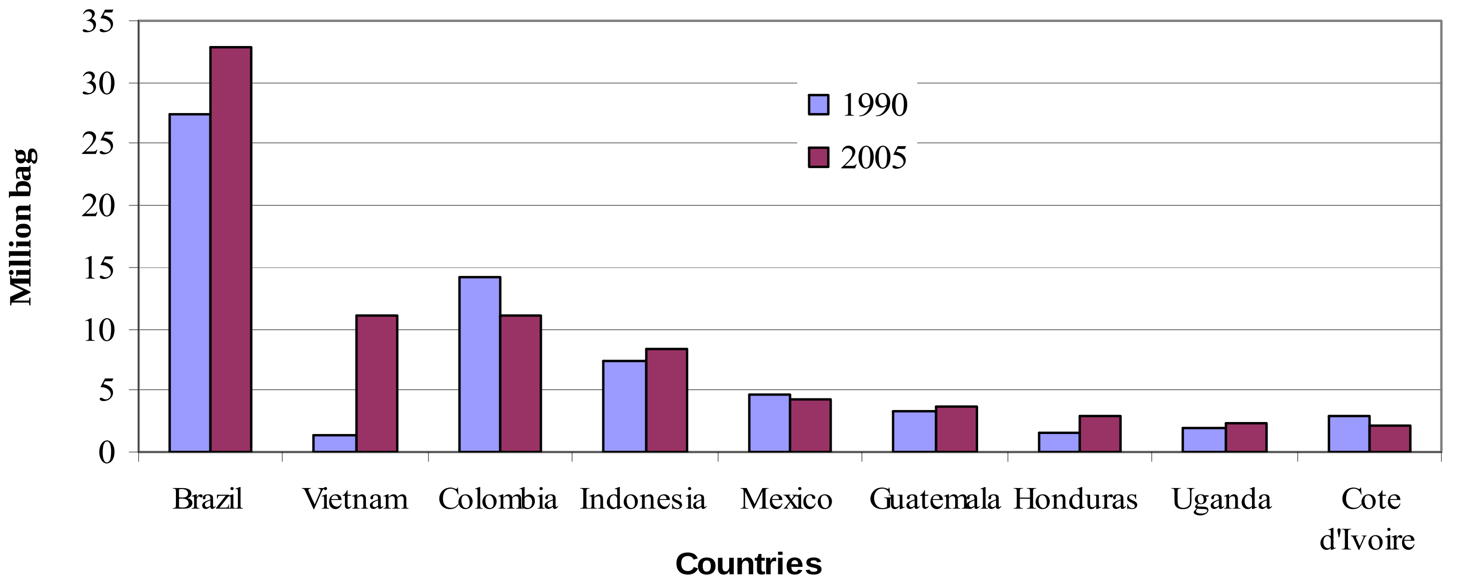
1.5. Hypothesis

One of the main goals of this thesis is to identify the sources of inefficiency in coffee production in Vietnam. The thesis seeks to measure factors correlated with technical efficiency. These can be factors that have nothing to do with farmer choice in the short run, such as land, education, and ethnicity. They can also include inputs, the levels of which farmers choose. In addition to these economic concerns, this thesis also attempts to test a methodological hypothesis about the impact of outliers. We test to see if the removal of outliers has an impact on economic inference regarding sources of inefficiency.



Source: International Coffee Organization, 2007

Figure 1.1 Average Wholesale Coffee Price



Source: International Coffee Organization, 2007

Figure 1.2 Top Coffee Producers

CHAPTER 2. METHODOLOGY AND MODEL

2.1. Efficiency

Efficiency is crucial in production economics as it gives both economic theorists and economic policy makers ideas about how well firms are performing in a relative sense. From productive efficiency, we can expect to know how much a firm/production unit can increase its output without using additional resources, thereby achieving an increase in efficiency (Farrell, 1957). Also, measuring efficiency allows one to seek out the sources of inefficiency and based on that, identify ways to increase efficiency. Koopmans (1951) definition of efficiency is the inability of one firm to increase one of it outputs without decreasing another output or increasing at least one of its inputs. Farrell (1957) proposed that the efficiency of a firm was comprised of two components: technical efficiency, which refers to the ability of one firm to produce as much as possible of one output from a given set of inputs, and price efficiency, which refers to the ability of a firm to use its inputs in the best proportion, in view of prices and technology. The product of these two measurements, a measurement of "perfect" efficiency, is called overall efficiency.[4] In this thesis, we focus solely on technical efficiency as a measure of the performance of coffee producers in Vietnam.

---

[4] Price efficiency and overall efficiency are now called allocative efficiency and economic efficiency respectively.

There are two types of technical efficiency: output-oriented efficiency and input-oriented efficiency. Output-oriented technical efficiency aims to answer the question of how much a firm can increase its output while keeping its inputs unchanged. Input-oriented technical efficiency, on the other hand, addresses the question of how much a firm can reduce its inputs while keeping its outputs fixed.

Input oriented efficiency.

Figure 2.1 illustrates a simple case of input-oriented technical efficiency with two inputs and one output under the assumption of constant return to scale (CRS). The CRS assumption allows the technology to be represented using the unit iso quant. SS' represents the fully efficient firms. Firm Q is efficient as it lays on the efficiency frontier SS'. Firm P and firm Q use the two inputs in the same ratio. However, firm P produces the same amount as Q but uses a fraction OP/OQ as much of each input. It can also be interpreted as firm Q can produce OP/OQ times as much output as firm P from the same inputs. The ratio OQ/OP is defined as the technical efficiency score of firm P. The ratio OQ/OP equals one minus QP/OP thus it takes a value between zero and one. Therefore, a firm is considered technically efficient if it has an efficiency score of one and technically inefficient if it has a score lower than one.

Output oriented efficiency.

Similar to the input-oriented case, Figure 2.2 illustrates a simple case of output-oriented efficiency with two outputs and one input. The arc SS' represents the efficient frontier of the firms. With the same amount of input used, firm Q produces a fraction

OQ/OP as much output as firm P. The ratio OP/OQ measures output-oriented technical efficiency. A value of one indicates that the firm is fully efficient compared with other firms and a value between zero and one means the firm is not producing at its efficient level. This thesis uses output-oriented efficiency as a benchmark.

As can be seen, the production frontier of the efficient firms is not known in practice, and must be estimated from observations. The frontier, and measures of efficiency are, therefore “relative” and valid only in the context of the sample under consideration. After Farrell’s proposal of efficiency measurement, there have been many attempts to estimate production frontiers using either econometric or mathematical programming techniques. One of the most widely used econometric techniques to estimate the efficiency frontier is stochastic frontier analysis (SFA). SFA forms the efficient frontier by setting up a functional relationship between inputs and outputs. It has the advantage over a mathematical programming approach that it actually separates the distance between an inefficient firm and the efficient frontier into statistical error and inefficient effects (Coelli et al, 2005). However, one of the biggest disadvantages of this approach is that it requires one to make assumptions regarding the functional form of the production function. Math programming, on the other hand, is less computationally demanding than econometric methods. It is easier to compute and does not require knowledge or assumptions about the relationships between inputs and outputs. This thesis uses Data Envelopment Analysis (DEA), the most commonly used math programming approach.

2.2. Data Envelopment Analysis

Over the past two decades, DEA has emerged as an important tool for identifying economic efficiency in such settings as the public sector (gas, water, heat, hospitals etc.) and the private sector (banks, mail, insurance, farms etc.). As of 2001, more than 1000 references are listed in this field.[5] This is an impressive number considering that DEA was first proposed 29 years ago by Charnes, Cooper, and Rhodes (1978). One of the main reasons why DEA is so commonly used is its ability to convert multiple inputs into multiple outputs, especially when one lacks a clear functional relationship between inputs and outputs (Murillo-Zamorano, 2004).

The model: Input-oriented DEA model.

To develop the DEA framework, we define the following:

$j = 1, \ldots\ldots, n$	index of firms
$i = 1, \ldots\ldots, m$	index of the inputs of firms
$k = 1, \ldots\ldots, r$	index of the output of firms
$x_j = (x_{1j}, \ldots\ldots, x_{mj})$	(column) vector of inputs of firm j
$y_j = (y_{1j}, \ldots\ldots, y_{rj})$	(column) vector of outputs of firm j
$\lambda = (\lambda_1, \ldots\ldots, \lambda_n)$	(row) vector of non-negative weights
θ	a scalar “shrinking factor”

We define a weighted combination of the input vector, namely

$$x_1\lambda_1 + \ldots\ldots\ldots + x_n\lambda_n$$

---

[5] www.deazone.com

And a weighted combination of the output vector,

$$y_1\lambda_1 + \dots\dots\dots + y_n\lambda_n$$

where all weights are nonnegative.

If we can find a weighting vector λ such that

$$\sum_j \lambda_j y_j \geq y_o \qquad (1)$$
$$\sum_j \lambda_j x_j < x_o$$

where (x^o, y^o) is the actual firm under consideration, then we can say firm (x^o, y^o) is inefficient because we can find a weighted combination of firms that produces equal or more outputs with less inputs. However, if we can not find such a weight vector λ then we can say that firm (x^o, y^o) is efficient.

We can rewrite (1) using the shrinking factor θ as

$$\sum_j \lambda_j y_j \geq y_o \qquad (2)$$
$$\sum_j \lambda_j x_j \leq \theta x_o$$

From (2) we can say that firm (x^o, y^o) is inefficient if $\theta < 1$ and all the conditions stated are satisfied. Firm (x^o, y^o) is efficient if all the conditions are satisfied with $\theta \geq 1$. We now rewrite (2) as the linear programming problem:

$$\text{Minimize } \theta \qquad (3)$$
$$\text{s.t} \quad \sum_j \lambda_j y_j \geq y_o$$
$$\sum_j \lambda_j x_j \leq \theta x_o$$
$$\lambda_j \geq 0$$

From (3) we can see that if $\theta = 1$ then with $\lambda_j = 1$ for j=0 and $\lambda_j = 0$ for $j \neq 0$, (3) will always have feasible solutions; therefore, θ is always less than or equal to 1.

Problem (3) is called the input-oriented DEA model in which a value of $\theta = 1$ indicates technical efficiency. As previously defined by Farrell, the efficiency score ranges in value from 0 to 1; therefore, θ is the efficiency score for the input-oriented DEA model.

As we look at (3), we see that there are no restrictions on λ (the weight vector) except non negativity. This means the efficiency score of one firm is calculated based on all the other firms in the dataset, regardless of their size. Thus, the frontier constructed from the most efficient firms is a linear combination of inputs and outputs. For a firm on the efficient frontier, if inputs increase by *n* times then output also increase *n* times. Therefore (3) is called the Constant Return to Scale (CRS) DEA approach.

The CRS approach, however, is only applicable when firms produce at their optimal level. Nevertheless, in reality, because of government regulation, competitions, and other factors, firm normally do not produce at their optimal level. For an increase in inputs, there is often a smaller increase of outputs. Banker et al (1984) modified work done by Charles (1978) by introducing variable return to scale (VRS) DEA approach. The model (3) is rewritten adding a constraint to λ:

$$
\begin{aligned}
&\text{Minimize } \theta \qquad (4)\\
\text{s.t}\quad &\sum_j \lambda_j y_j \geq y_o\\
&\sum_j \lambda_j x_j \leq \theta x_o\\
&\sum_j \lambda_j = 1\\
&\lambda_j \geq 0
\end{aligned}
$$

By imposing a new constraint on λ_j, the VRS-approach divides the firms under analysis into different classes depending on size. The most efficient firms within each class form the frontier. For that reason, the efficiency frontier in the case of VRS-approach is a *convex combination* of inputs and outputs. Figure 2.3 illustrates a simple case in which firms that have one input and one output lie along two different DEA frontiers, CRS and VRS.

Under CRS, the input-oriented technical inefficiency of firm E is EF. However, under VRS, the input-oriented technical inefficiency of firm E is only EG, which is smaller than EF. The technical efficiency scores of point E under CRS and VRS assumptions are PF/PE and PG/PE respectively.

Output-oriented DEA model.

Following the same steps to build the input-oriented DEA model, we have the CRS output-oriented model written as:

$$\begin{aligned} &\text{Maximize } \theta \qquad (5)\\ \text{s.t} \quad &\sum_j \lambda_j y_j \geq \theta y_o\\ &\sum_j \lambda_j x_j \leq x_o\\ &\lambda_j \geq 0 \end{aligned}$$

As for this case, θ is always greater or equal to 1, therefore, the efficiency score of the output-oriented DEA model is defined as TE = 1/ θ.

Similarly, the VRS output-oriented model is:

$$\begin{aligned} &\text{Maximize } \theta \qquad (6)\\ \text{s.t} \quad &\sum_j \lambda_j y_j \geq \theta y_o\\ &\sum_j \lambda_j x_j \leq x_o\\ &\sum_j \lambda_j = 1\\ &\lambda_j \geq 0 \end{aligned}$$

This thesis uses output-oriented DEA to calculate the efficiency of coffee farmers in Vietnam. As coffee farmers in Vietnam do not produce under perfect conditions and at optimal levels, the assumption used here is CRS. From the DEA results, we can tell whether a farm is efficient or not. Also, the idea of how far can it improve its production without further using other resources can also be identified. However, in order to make such improvement in coffee production, we need to know why the firms are inefficient. The following section presents the approach used here, which is one of the most commonly used techniques for identifying sources of inefficiency in DEA: a two stage DEA Tobit model.

2.3. Two Stage Analysis

The two-stage model involves finding DEA technical efficiency scores in the first stage of the analysis. The scores from the first stage are then regressed on a set of factors that are hypothesized to affect a firm's efficiency score. Because the technical efficiency scores are bounded between 0 and 1, a Tobit model is employed in place of Ordinary Least Square (OLS) regression because of its ability to handle truncated data (Ray, 2004; Coelli et al, 2005). The Tobit model has the following form:

$$\begin{aligned} TE_k &= \beta' X_k + u_k \quad \text{if } TE^*_k < 1 \\ &= 1 \text{ if } TE^*_k \geq 1 \end{aligned} \tag{7}$$

In (7) TE_k is the technical efficiency score for firm *k*, TE^*_k is a true but unobservable efficiency score, β is a row vector of unknown parameters, X_k is a vector of factors that are hypothesized to be correlated with technical efficiency scores, and u_k is an error term that is independently and normally distributed with mean zero and variance δ^2.

Recently, Simar and Wilson (2007) pointed out that the TE scores from DEA estimation are serially correlated. Therefore, economic inference from the two-stage approach may be misleading. However, in this thesis the two-stage approach is used.

The two-stage analysis with DEA in the first stage and Tobit in the second stage has been widely used recently (Simar and Wilson, 2007). In Thailand, Krasachat (2004) applied a two-stage DEA-Tobit model to measure and investigate technical efficiency on rice farms in Thailand. Krasachat finds suggests that the diversity of natural resources has an influence on technical efficiency on Thai farms. Using the same approach, Otsuki et al (2002) examined the effects of the Brazilian government's title granting policies on the efficiency of agricultural and timber production in the Brazilian Amazon. Government expenditures to secure land titles was found to have had a positive effect on the technical efficiency scores of agriculture and timber production. Population density and the number of sawmills per km^2 had negative impacts on joint agriculture and timber production and agriculture-specific production. Audibert et al (2003) used two-stage analysis to identify the social and health determinants of the efficiency of cotton farmers in Northern Ivory Coast. They found that more cotton growers in the village improved efficiency while malaria morbidity had a negative impact on cotton producers' efficiency. However, the intensity of the malaria infection deserves the most attention rather than the presence of the infection. Also in Ivory Coast, Binam et al. (2003) measured the technical efficiency of coffee producers as well as the relationship between technical efficiency and various farms' characteristics in the Central West Region. Their conclusion was that, from a policy point of view, family size, membership in a farmers' club or association and the origin of farmers were the variables most highly correlated with technical efficiency scores. The authors conclude that these factors should be taken into consideration when creating policies to improve the efficiency of coffee production in Ivory Coast.

2.4. Outlier Detection

One of the advantages of DEA is that efficiency is easy to compute and does not require a functional relationship between inputs and outputs. The frontier derived from DEA is a combination of the most efficient firms. In the context of this thesis, those are firms that generate the most output for a given level of inputs. However, because the frontier is constructed using extreme observations, DEA can be very sensitive to extreme points in a dataset. Thus, the technical efficiency scores which are calculated from datasets that include outliers will be misleading. Realizing this problem with DEA, in the last several decades, there have been several efforts to trying to detect and remove outliers. Since the frontier in DEA is not parameterized, the detection of outliers using parameter estimation (e.g. OLS residuals) can not be used (Sexton, 1986 cited in Wilson, 1993). Therefore, all of the methods to identify outliers recently are non-parametric. In 1978, Andrews and Pregibon proposed a geometrical method to deal with outliers. They calculated a proportion of the geometric volume spanned by a subset of the data obtained by removing some observations relative to the volume spanned by the entire dataset. This proportion is then used to detect outliers. The method is only applicable for firms with one output which is a limitation because one of the most appealing advantages of DEA is that this model can handle multiple outputs. Wilson (1993) then developed this model so that it can be used for firms with more than one output. This way of identifying outliers, however, is very computer intensive and does not take into account the frontier aspect of the problem (Simar, 2003). Recently Cazals et al. (2002) proposed a non-parametric estimator that is robust to extreme observations. This method is based on the concept of expected minimum input function (or expected maximum output function). Basically, an expected frontier is formed and then pushed as far away from the data as possible. However, some points will not be enveloped by this expected frontier eventually. Those points are identified as outliers.

In this thesis, we propose a simple alternative method to detect outliers based on ‘weights’ of the observations. In the context of this research, outliers are defined as observations with large impacts on the efficiency frontier. First, let’s look at the output-oriented DEA model to calculate technical efficiency of coffee producers in which λ is a row vector of weights:

$$\text{Maximize } \theta \tag{8}$$

$$\begin{aligned} \text{s.t} \quad & \sum_j \lambda_j y_j \geq \theta y_o \\ & \sum_j \lambda_j x_j \leq x_o \\ & \sum_j \lambda_j = 1 \\ & \lambda_j \geq 0 \end{aligned}$$

Remember that (x^o, y^o) represents the current producer under consideration, thus, the value of 1/ θ calculated from this equation is the technical efficiency score for only producer (x^o, y^o). In order to have the technical efficiency scores for the whole sample of j farms, we must solve (8) j times. The matrix of lambda values then has the form:

$$M_\lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1j} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2j} \\ \dots & \dots & \dots & \dots \\ \lambda_{j1} & \lambda_{j2} & \dots & \lambda_{jj} \end{bmatrix} \tag{9}$$

M_λ represents all the weights that every observation receives during the process of finding all the technical efficiency scores. We can interpret the weights in M_λ in two different ways. First, each non-zero weight represents an occasion when one observation appears during the construction of the DEA hull. The total number of occurrences is indicated by the number of times a non-zero value for λ appears in the corresponding column of lambda values. Second, we can compute the cumulative weights of one observation in all constructed efficient sets. This is the column sum of all the weights for one observation that exists in the matrix of lambda values. Accordingly, we define two new indexes to represent those weights.

Lambda count.

We define C_j as the number of times an observation appears during the construction of the DEA hull. The value of C_j is computed as

$$C_j = \sum_{j \text{ if } \lambda_{ij} > 0} 1 \tag{10}$$

Lambda sum.

We define S_j as the cumulative weight of an observation in all constructed efficient sets. This is computed as

$$S_j = \sum_{j} \lambda_{ji} \tag{11}$$

The DEA model will yield non-zero values for *lambda count* and *lambda sum* for efficient firms. All inefficient firms will have zero values of *lambda count* and *lambda sum*. In identifying outliers, we focus attention only on efficient firms. When we construct the frontier, it is positioned to envelop all observations, including outliers. Note that outliers will always lie on the frontier. So, based on the values of C_j and S_j, we can identify observations in the dataset that exert a great deal of influence in the construction of the efficient frontier. These are considered candidate outliers.

After identifying the observation with the greatest weight, based on either *lambda count* or *lambda sum*, we drop it from the sample. A new dataset with a sample size j-1 results. We then repeat the DEA to get new values for C_j and S_j. The observation yielding the greatest value of either *lambda count* or *lambda sum* is then dropped. We continue to drop observations with the greatest weights after each DEA run. We stop removing of observations once we reach convergence. Convergence is defined as follows. If there is a big difference in the lambda weights between the previously-dropped observation and the current observation, and if the weight difference between the current observation and the most recently dropped observation is small, then we stop the iterations. In other words, we will stop dropping observations when we reach relative convergence of the remaining weight scores. One of the easiest ways to identify convergence is through visual interpretation of the data, using a graph that plots the iteration number on the x-axis and the weights of the dropped observations on the y-axis.

Detecting outliers under VRS and CRS.

This approach of detecting outliers is based on the weights of the outliers during the construction of the DEA hull. However, keep in mind that the frontier of efficient firms is built differently under different assumptions. With CRS, the frontier is a linear combination of inputs and outputs of the most efficient firms. For a firm on the efficient frontier, if inputs increase by *n* times then output also increase *n* times. The VRS-approach divides the firms under analysis into different classes depending on size. The most efficient firm within each class forms the frontier. For that reason, the efficiency frontier in the case of VRS-approach is a *convex combination* of inputs and outputs. Because of the nature of the efficiency frontier under different assumptions, we can see that the weight of outliers differs between CRS and VRS. For CRS, in the one input one output case, because the frontier is a linear combination of observation, the weight of an outlier is put on all the observations in the dataset; therefore, this method of detecting outliers will find the observation in the sample with the greatest weight every time we calculate the weights. Under VRS-approach, this does not necessaryly always happen. As the most efficient observations are identified among different classes depending on size, in some cases, the most efficient observation in the dataset is not the one with the highest weight. In Figure 2.4, for example, we can see that firm D is the most efficient observation and is most likely an outlier, but firm C receives more weight.

We can also see that most of the observations in the dataset lie between the two dotted-lines going through B and C. Therefore, in this case, C but not D is the observation with the most weight. So, in the first iteration, this approach of detecting outliers will drop observation C. However, observation D will be eventually dropped in this case.

This approach of detecting outliers based on dropping observations with the most weight is easy and fast to calculate. Yet, for CRS, in the one input one output case, the most efficient observation is always the one with the most weight.[6] Under VRS, as the efficient frontier is built based on different classes of size, the observations with the most weight is not necessarily the most efficient one. It will take more iterations under the VRS-approach than the CRS-approach to identify outliers in the dataset.

The goal of this thesis is to accurately measure technical efficiency of coffee producers in Vietnam and the sources of inefficiency using a two stage model. In the first stage, a DEA model is run to get the technical efficiency scores. Using Tobit regression

---

[6] In multi-input, multi-output case there may not be a most efficient observation.

in the second stage, sources of inefficiency are identified. Outliers are then removed based on the weight that one observation put on other observations during the construction of the DEA hull. The two-stage model is then re-estimated to find more accurate technical efficiency scores and to correctly attribute sources of inefficiency.

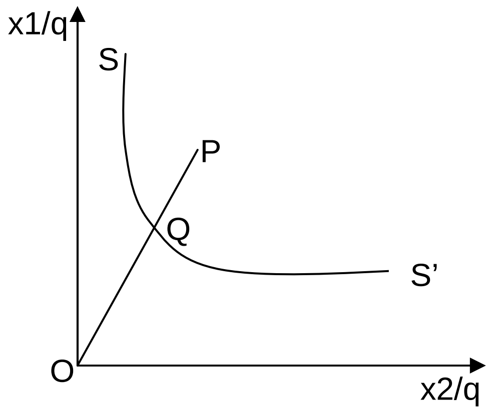


Figure 2.1 Input Oriented Technical Efficiency

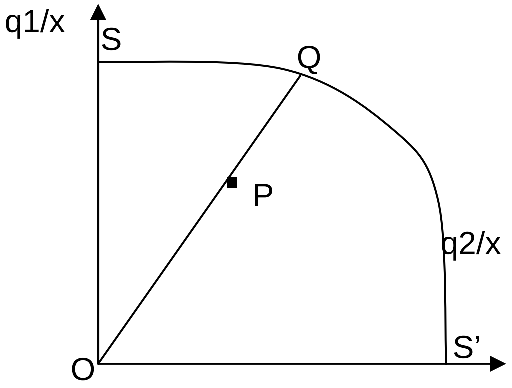


Figure 2.2 Output Oriented Technical Efficiency

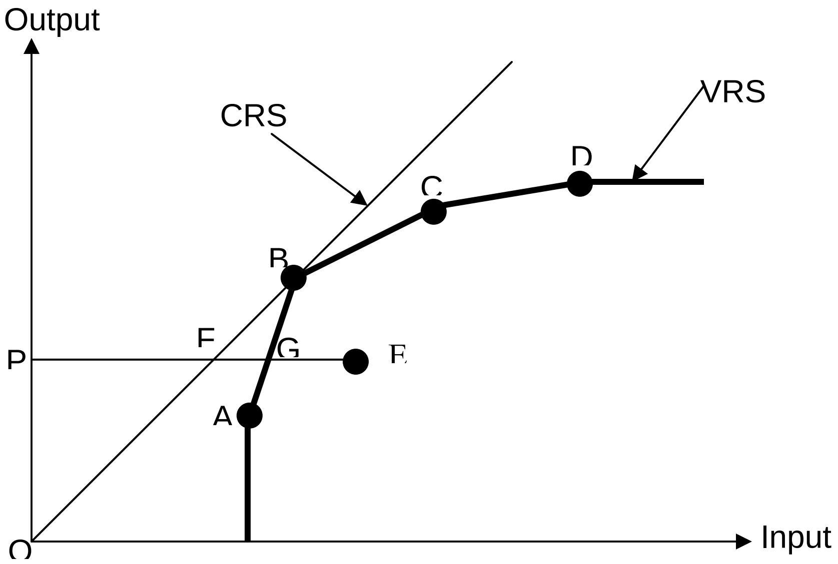


Figure 2.3 CRS and VRS DEA

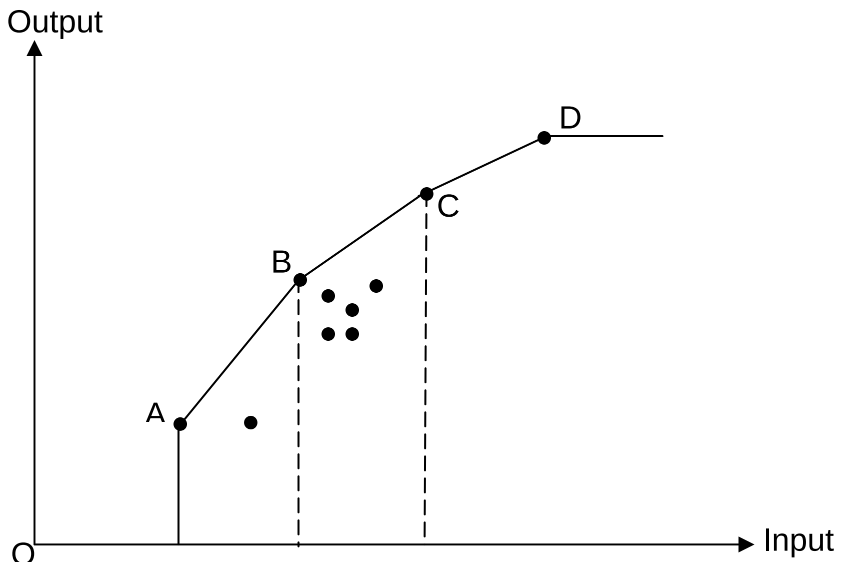


Figure 2.4 Outliers Under VRS

CHAPTER 3. STUDY SITES AND DATA

3.1. Study Sites

Dak Lak province has led the way for coffee production in Vietnam and accounts for more than half of the country's production and area. It is located in the Central Highlands of Vietnam and lies to the East of the Cambodian border (Figure 3.1). It is the largest province in Vietnam covering almost 2 million hectares. Approximately 40% (750,000 hectares) of the province consists of a plateau of rich basaltic red soil with high natural fertility and thick soil layers. These are particularly well-suited for intensive agricultural production, especially the production of coffee. The area is characterized by two distinct seasons: the dry season from December to April and the rainy season from May to November. Most rainfall occurs between July and September. The area's elevation ranges from 400 to 800 meters above sea level which makes its climate substantially milder than the lowlands. Temperatures average 23.7°C annually with 82% humidity. It has an annual average rainfall of 1,650 mm (Cheesman, 2005).

Since the 1960s, there has been a dramatic fall in the forest area of Dak Lak. In 1960, forest represented 90% of the area. By the late 1990s, only 50% of the province was covered with forest (Oxfam-ICARD, 2002). However, in absolute terms, Dak Lak still has the largest area of forest land in Vietnam. The roughly one million hectares of forest land in Dak Lak has many valuable timber and non-timber forest products, many of which are unique to the region (Lindskog, 2005). The decrease of forest cover in Dak Lak in the last 40 years is associated with an increase in agricultural land. As can be seen in Figure 3.2, over the course of 30 years agricultural land increased from less than 100,000 hectares to more than 500,000 hectares. Forest land, on the other hand, decreased from roughly 1.6 million hectares to 1 million hectares.

The agricultural sector in Dak Lak is mainly comprised of coffee, rubber, pepper, cashews and cocoa. Of these, coffee is by far the most important crop. The first coffee tree was introduced in the area in 1857 by the French (Nhan, 2001). However, as in other areas of the country, coffee production in Dak Lak remained insignificant until the mid 1990s. Between 1976 and 2001, coffee area in Dak Lak increased from 20,000 hectares to 264,000 hectares (Babu et al., 2003). Coffee farms in Dak Lak cover approximately half of all agricultural areas of the province and also account for half of all coffee area in Vietnam. By 2001, Dak Lak produced about 465,000 tons of Robusta coffee which represented about 55 percent of Vietnam's annual Robusta output (Cheesman, 2005). However, as in many coffee production areas in Vietnam, expansion of coffee area in Dak Lak has been associated with inappropriate usage of chemical fertilizers and pesticides, high rates of water use, and environmental stresses, including deforestation, water scarcity, soil degradation, and contamination of surface and ground water (Cheesman, 2005; Ha et al., 2001; Koninck, 2000, Ahmad, 2000).

The expansion of the prosperous coffee industry in Dak Lak in the 1990s, the national reform of demographic resettlement, population pressures and poverty in other areas have induced thousands of people to come to Dak Lak. After the war in 1975, Vietnam experienced a sharp increase in population growth with an annual rate of about

2.15%.[7] Of all the provinces, population grew fastest in Dak Lak, with an annual growth rate of 6% between 1975 and 2002 (Lindskog, 2005). In 1976, Dak Lak had a population of 360,000 with a population density of approximately 18 people per square kilometer. At that time, about 300,000 of these people lived in rural areas (Mueller, 2003). After 1975, Dak Lak was designated as a New Economic Zone (NEZ). The Vietnamese government organized mass migration of people from the lowlands to state farms and forest enterprises in the NEZ (Ha et al., 2005). As a result, thousands of people migrated to Dak Lak after 1976, in both spontaneous and state-sponsored patterns. According to official statistics, from 1976-1999, the number of migrants to Dak Lak was about 650,000, almost equally divided between state-controlled and spontaneous migrants (Figure 3.3).

Most of the people who migrated to Dak Lak belong to the Kinh ethnic group, originally from lowland areas with high population and poverty pressure. However, there are also people from other ethnic minority groups from the North of Vietnam (Figure 3.4).

According to Mueller (2003), by 2001, the population of Dak Lak was 1,950,000 with approximately 1,500,000 of these people living in rural areas. The population density has increased to about 100 persons per square kilometer. The massive influx of migrants to Dak Lak not only created more pressures on natural resources but also generated social conflicts between migrants and local ethnic minorities, and between various migrant groups (ADB, 2003). Although there is no doubt about the benefit brought to ethnic minority groups by economic expansion, there are still concerns about

---

[7] http://www.populstat.info/Asia/vietnamc.htm

them being economically disadvantaged compared with their Kinh counterparts (Ha and Shively, 2007).

The data for this study were collected from 209 households in four villages in Ea Tul catchment, one of the main coffee production areas in Dak Lak province. The survey was conducted by researchers at Nong Lam University to assess coffee growing conditions in Dak Lak. The catchment covers two districts: Buon Don with 6 communes and Cu M'gar with 2 towns and 13 communes. Most of the catchment is flat upland area with deep streams flowing through the communes. However, there are also hilly areas with steep slopes. In general, the soil and climate in this area are considered suitable for planting a lot of industrial trees such as rubber, pepper, coffee as well as fruit trees and some annual crops (Ahmad, 2000).

3.2. Descriptive Analysis

Table 3.1 presents basic statistics of household characteristics in the sample. Not surprisingly, all of the farmers interviewed plant coffee as one of their crops. In fact, in the study sites, the average income from coffee production represents more than 70% of total household income. We can see that there is a big gap in term of education between household heads of Kinh families and their counterparts. Household heads of Kinh families, on average, have attended school for 7.9 years, compared with 4.7 years for household heads of other groups. More than 95% of the sample reports land ownership. Similar to the indicator for education of households head, Kinh households also have a higher percentage of families with land use certificates (98.6 per cent vs. 89.2 per cent). However, with respect to inputs for production, households from other ethnicities use more inputs than Kinh do. Non-Kinh have larger farms (1.9 hectares vs. 1.3 hectares), a higher percentage of tractor ownership (83 percent vs. 69.4 percent) as well as more family labor (7.1 people vs. 5.2 people). There is no significant difference in the dependency ratios of the two groups.[8] While ethnic minority households have more inputs, they turn out to have lower income compared with Kinh households (26 million VND and 37 million VND respectively). This phenomenon can be explained, in part, by off-farm activities of Kinh ethnic people. Having higher education, these people find it easier to find off-farm jobs than other ethnicities. Moreover, most of the Kinh people migrated to this area from the lowlands where trading activities are far more developed than in the highlands; therefore, they can more easily get involved in off-farm activities. Regarding rice cultivation, about 25% of the total surveyed households have rice as one of their crops. A higher percentage of Kinh households cultivate rice than minority households (29% vs. 17%)

Table 3.2 presents data on inputs used for coffee in the sample. Out of 14 indicators used to measure inputs, seven of them are significantly different between Kinh groups and non-Kinh households. As non-Kinh have more land for cultivation than Kinh, not surprisingly, they also have larger coffee areas on average (1.74 ha vs. 1.09 ha). However, Kinh families use significantly higher per-hectare amounts of fertilizer and pesticides. For example, Kinh farmers use more than triple the amount of NPK per hectare compared with minority groups (1611 kg/ha vs. 533 kg/ha). Other inputs like organic fertilizer and pesticides also exhibit big differences between the two groups. We

---

[8] (# people>60 + # people <15)/# workers

can also see that since non-Kinh families have more labor in general, they also invest more household labor in planting coffee compared with Kinh people (184 man-days/ha vs. 153 man-days/ha). Nevertheless, Kinh families use more hired labor than other groups (71.5 man-days/ha vs. 36.65 man-days/ha). The difference in hired labor between the two groups compensates for the difference in family labor, so that the two groups devote approximately equal amounts of total labor-days per hectare of coffee. In the year of the survey, Kinh farmers had a higher yield of coffee per hectare than non-Kinh farmers. Higher output can be explained by greater amounts of fertilizers and pesticides.[9]

---

[9] Production function estimates for the sample is reported in Appendix A.

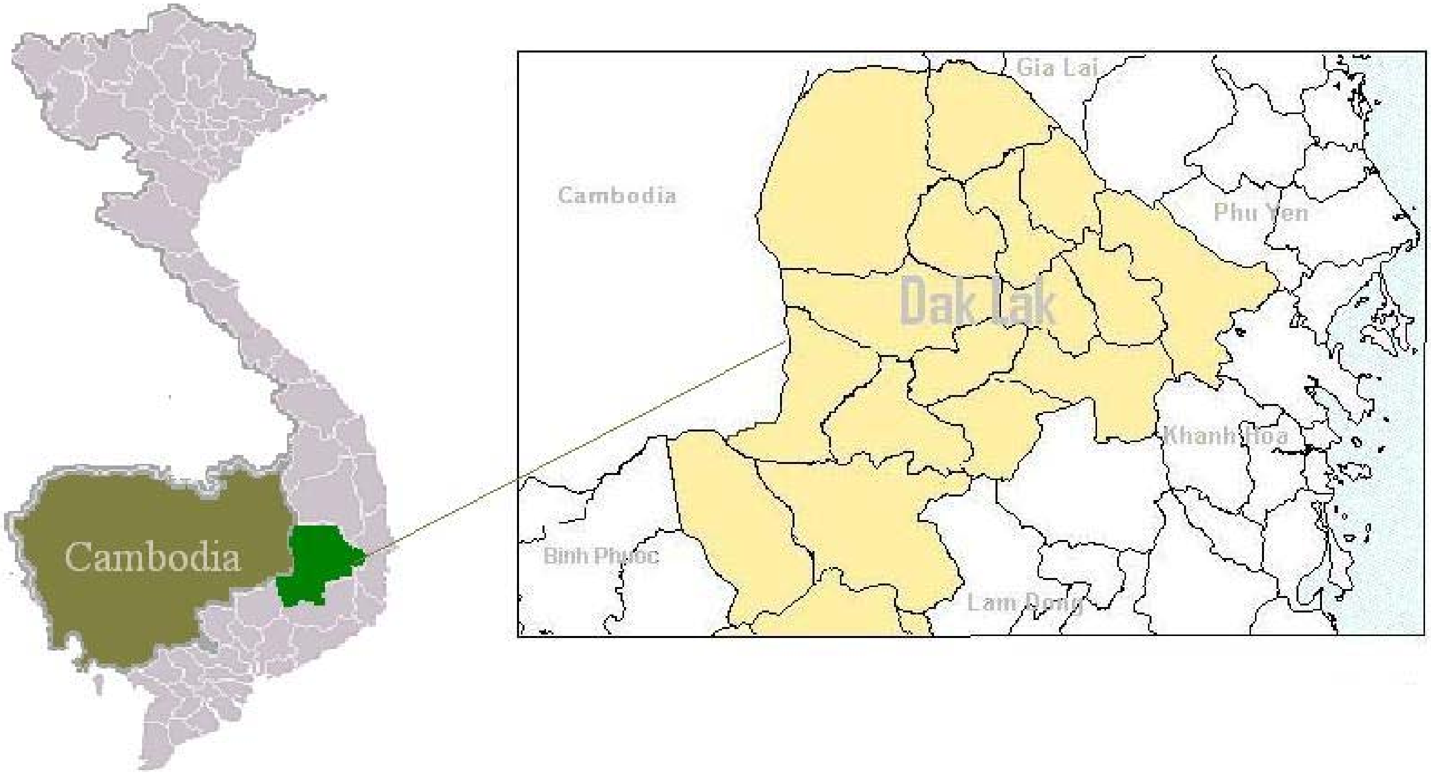
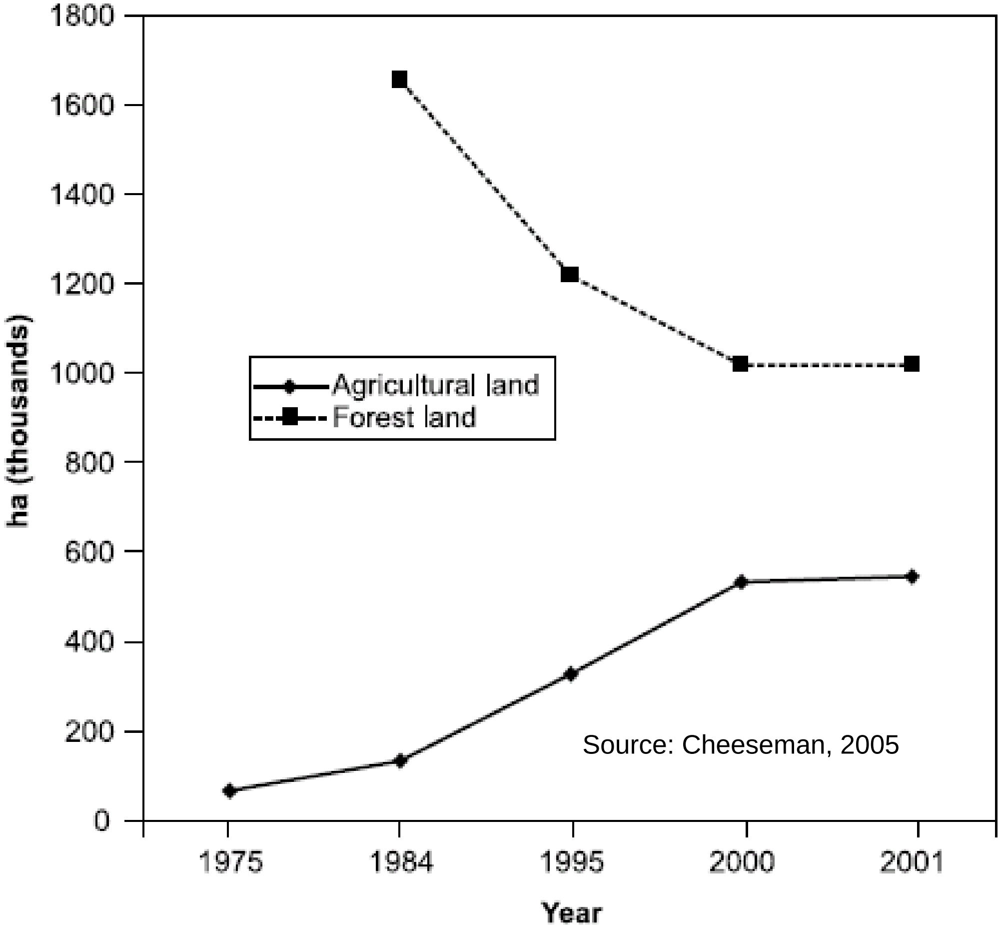
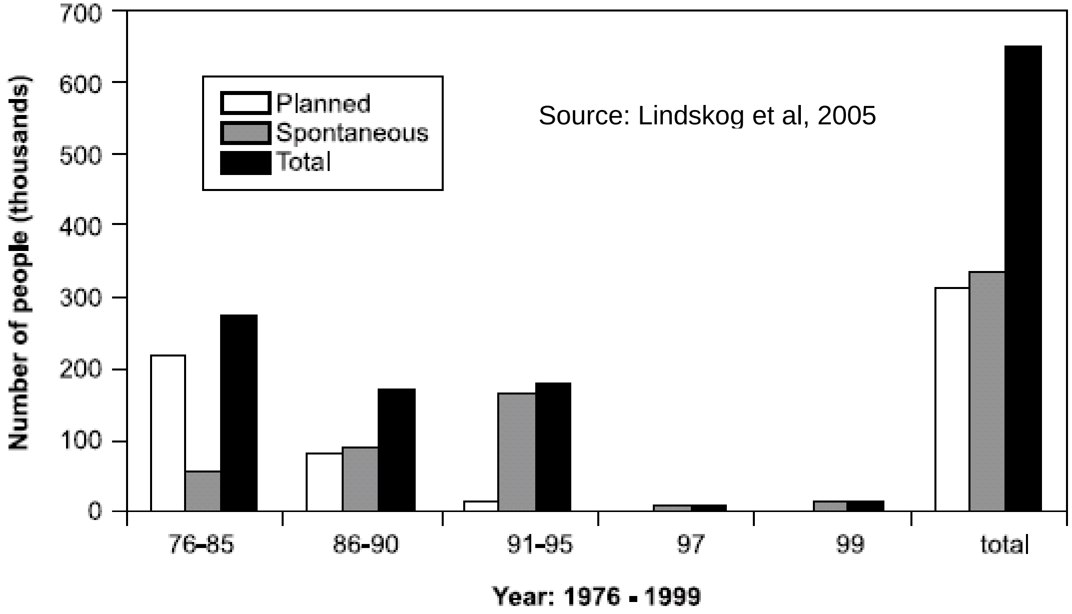


Figure 3.1 Map of Dak Lak



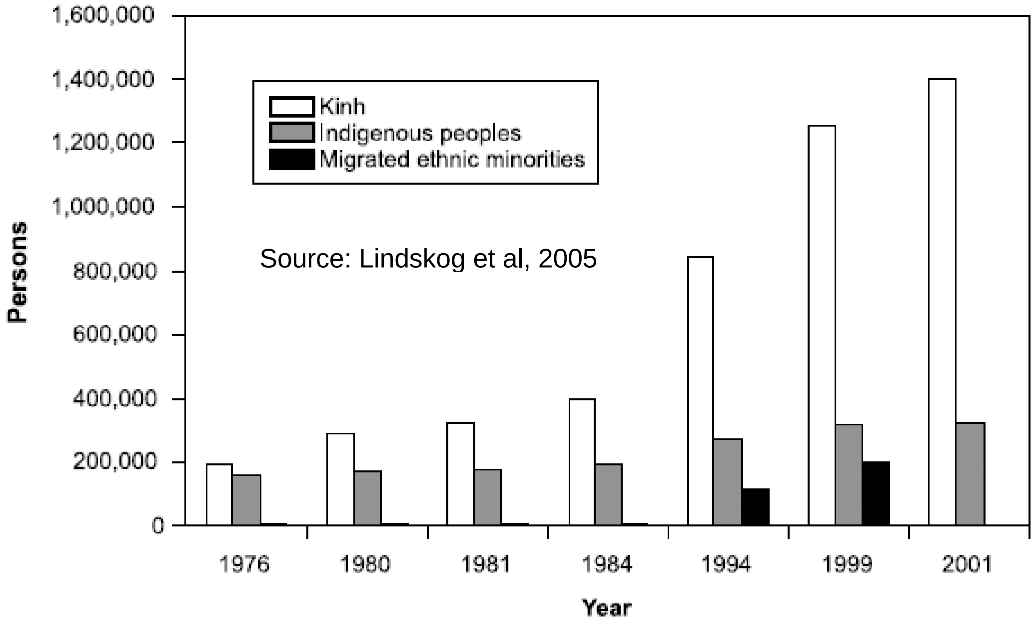
Source: Cheeseman, 2005

Figure 3.2 Changes in Agriculture and Forest Land in Dak Lak (1975- 2001)



Source: Lindskog et al, 2005

Figure 3.3 Migrations to Dak Lak



Source: Lindskog et al, 2005

Figure 3.4 Population and Ethnicity in Dak Lak 1976-2001

Table 3.1 Descriptive Analysis of the Study Site

	Kinh		Non-Kinh		All farms	
Variable	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Age of household head (years)	45	116	45	11	45	11.46
Education of household head (years)	8*	3	4.8*	4.2	7	3.7
Household size (people)	5.3*	1.6	7.1*	2.1	5.9	2
Dependency ratio (#>60 + # <15)/# workers	1	0.6	1.1	0.8	1.01	0.7
Farm area (hectares)	1.3*	0.9	1.9*	1.1	1.50	1
Income (1000 VND)	37,094*	29,649	25,904*	20,729	33,614	27,632
Land ownership (% with land use certification)	98.6*	1	89.2*	39	95.7	20.3
Tractor ownership (% with tractors)	69.4*	46.2	83.1*	37.8	73.7	44.1
Pump ownership (% with pumps)	75.7	43.0	76.9	42.5	76.1	42.8
Livestock ownership (% with livestock)	38.2	48.8	40	49.4	38.8	48.8
Rice cultivation (% with rice)	29.2*	45.6	16.9*	37.8	25.4	43.6
Coffee cultivation (% with coffee)	100	0	100	0	100	0
Irrigation (% with irrigation)	82.6	38	87.7	33.1	84.2	36.6
Number of observations	144		65		209	

Note: An asterisk indicates the difference between means of the two groups is statistically significant at a 95 per cent confidence level in paired t-test.

1 USD= 16,000 VND.

Table 3.2 Descriptive Analysis for Coffee Production

Variables	Kinh		Others		All Farms	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Total area of coffee (ha)	1.09*	0.76	1.74*	0.95	1.29	0.87
Years planting coffee (years)	11.4	4.6	12	5.6	11.6	4.9
Urea (kg/ha)	328	538	260	358	307	489
NPK (kg/ha)	1611*	1,277	533*	617	1,276	1,220
Phosphorus (kg/ha)	212	378	128	230	185	340
Sulphate (kg/ha)	188	357	111	257	164	330
Potasium (kg/ha)	199	454	1794	309	193	414
Other chemical fertilizer (kg/ha)	143	521	23	186	105	447
Organic fertilizer (kg/ha)	252*	656	57*	265	191	570
Herbicides (kg/ha)	0.34	2.57	0.05	0.26	0.25	2.14
Pesticides (kg/ha)	5.40*	5.67	3.05*	2.41	4.67	5.01
Family labor (mandays/ha)	153*	81	185*	81	163	83
Hired labor (mandays/ha)	72*	73	37*	53	61	70
Coffee yield in 2003 (kg/ha)	3045*	1,272	2044*	847	2,734	1,244
Number of observations	144		65		209	

Note: An asterisk indicates the difference between means of the two groups is statistically significant at a 95 per cent confidence level in paired t-test.

1 USD= 16,000 VND.

CHAPTER 4. RESULTS

This chapter reports findings regarding technical efficiency and the sources of inefficiency for 209 coffee producers in Dak Lak, Vietnam. It also provides results for the process of removing of outliers as well as its impact on technical efficiency level and economic inference regarding sources of inefficiency.

4.1. Data Envelopment Analysis Results

This section presents findings from the DEA model for the 209 smallholder Vietnamese coffee farmers in the sample. As mentioned earlier, the assumption made for the DEA model in this thesis is VRS. This is a DEA model consisting of one output and 11 inputs. The only output is coffee production (in kg). The inputs include all factors observed in production. However, there are a number of factors that affect coffee production that are not taken into account in this model. These include exogenous factors like weather conditions, government policies, education etc. that coffee producers have little control over. The 11 inputs used in this model are: urea (kg), NPK(kg), phosphorus (kg), sulfur (kg), potassium (kg), organic fertilizer (kg), other fertilizer (kg), herbicide (kg), pesticide (kg), family labor (man-days), and hired labor (man-days).

The DEA model was implemented in GAMS. The GAMS code used for the model is presented in Appendix B. Table 4.1 summarizes the DEA results for the sample as well as for two sub-samples: Kinh and non-Kinh farmers.

All technical efficiency scores (TE scores) in Table 4.1 measure relative efficiency. This means the farmers who are on the efficient frontier are relatively efficient compared with others in the sample. The TE scores for the two groups range from 0.17 to 1. This means that the farmer with the lowest efficiency score (0.17) could increase his productivity by 83% to reach the highest level of technical efficiency observed in the sample. Almost half of the coffee farmers under consideration are efficient. The average TE score is 0.78. The average TE score implies a 22% (1-0.78) displacement from production frontier, on average. This means nearly one fourth of the coffee production in the study area is lost due to inefficiency. Table 4.1 also provides a comparison of technical efficiency between Kinh farmers and farmers from other ethnic groups. A higher percentage of farmers from other ethnic groups produce efficiently compared with their Kinh counterparts (51% vs. 42%). Also, the mean value of efficiency scores for non-Kinh farmers is significantly higher than for Kinh farners. Farmers from other groups are producing coffee 17% below the production frontier while Kinh farmers are producing 24% lower than the optimum level. This result is somewhat surprising considering the fact that, in Vietnam, Kinh farmers are normally considered to be more advanced than other ethnic groups. As indicated in Table 3.2, Kinh farmers use much higher levels of inputs in coffee production than non-Kinh farmers. They also have higher productivity (3045 kg/ha on average vs. 2044 kg/ha for non-Kinh farmers). Nevertheless, they appear to be producing at a lower efficiency level then farmers of other ethnic groups. However, at this point of analysis, a conclusion regarding the sources of differences in TE scores of the two groups can not be made. Research by Binam et al. (2003) about coffee production in the Ivory Coast had similar findings. The native farmers in the Central West region of Ivory Coast had significantly higher coffee productivity compared with non-native farmers. The reason for such a difference was that native farmers were more homogenous and the farming there was not as intensive as the non-native farming system. They also were more likely to have access to extension personnel, education, technical help, and other information. These reasons, however, are not the answer in the case of Vietnam as Kinh farmers are more homogenous than other groups. Moreover, they have better access to education and information.

4.2. Two Stage Analysis

The TE scores indicate the gap between the optimal production level and the current level of coffee production. In order to gain insight into the causes of inefficiencies, a two stage analysis using DEA and Tobit was carried out (Bravo-Ureta and Evenson, 1993; Coelli et al, 2005). This method uses the TE score from the DEA as a dependent variable in a regression. Independent variables are households' characteristics. The model is:

$$\begin{aligned} TE_k &= \beta' X_k + u_k \quad \text{if } TE^*_k < 1 \\ &= 1 \text{ if } TE^*_k \geq 1 \end{aligned} \tag{11}$$

where TE_k is the TE score for the k^{th} household, TE^*_k is a true but unobservable efficiency score, β is a row vector of unknown parameters, X_k is a vector of factors that are hypothesized to be correlated with TE score, and u_k is an error term that is independently and normally distributed with mean zero and variance δ^2. Because the TE scores are bounded by 0 and 1 by construction, a two-tailed Tobit model is used for the regression.

As reported by Oxfarm-ICARD (2002), even though there has been a sharp increase in the production of coffee in Dak Lak, most of the rise in production reflects expansion of coffee production area. From the period between 1990 and 2000, coffee output growth in Dak Lak was 30.4% per year, of which two-thirds was solely due to an increase in the number of hectares devoted to coffee production. This means that if there is a difference in productivity between two farms, the difference is most likely caused by farm size. For that reason, we suspect a strong correlation between TE score and farm size. A plot of the TE score and farm size is presented in Figure 4.1.

We can see from Figure 4.1 that the TE score and farm size are positively correlated. Most of the coffee farms are small farms with the total area smaller than 2 hectares. Even though there are fewer big farms, there is a higher percentage of big farms that are identified as efficient. As discussed earlier, under the VRS assumption, the efficiency frontier is constructed based on different farm size categories. Therefore, since we calculate the TE score under the assumption of VRS, it is reasonable to see a higher percentage of bigger farms being identified as efficient. In the Tobit model farm size squared is included as a regressor identify a quadratic relationship.

4.2.1. Relationship between Technical Efficiency and Fixed Factors

In farm operation in general and coffee production in particular, there are factors that affect production but have nothing to do with farmer choice in the short term. These factors can be governmental policies, weather, farm land etc. In the search for the sources of efficiency or inefficiency, these factors should always be considered. In this thesis, these factors include farm size (in hectares), ethnicity (minority =1, Kinh=0), education (number of years in school), and tenure security (ratio of land area with land use certification over total land area).

The first column of Table 4.2, labeled Model 1A contains results for a regression of TE scores on fixed factors. The results show there is a positive correlation between ethnicity and technical efficiency but the correlation is not statistically strong. Also, education and tenure are not strongly correlated with the efficiency scores. So among all of these factors, only farmsize has a significant relationship with TE score. This is not a surprising result considering the fact that numerous studies world-wide have pointed to the relationship between farm size and efficiency. Val Zyl et al. (2005) discussed the sources of inefficiency of South Africa's six major grain producing areas and one irrigation area for the period 1975-1990. They found that there is an inverse relationship between farm size and efficiency in the commercial farming areas for the range of farm analyzed. The inverse relationship seems to be stronger as the policy distortions, which tend to favor large farms, are removed. Hardi and Whittaker (1999) examine the relationship between farm size, technical efficiency and the use of agrochemicals with a set of panel data on 35 farms from the South West of England for the years 1987-1991. They find a positive correlation between farm size and technical efficiency. Similarly, Jha et al (2000) also found a positive relationship between farm size and technical efficiency when investigating 300 wheat farms in the Indian Punjab for the period 1981-1982 and 1982-1983. At a larger scale, Gorton and Davidova (2004) drew together results from farm efficiency studies from Central and East European countries that are part of the EU enlargement process. The empirical results showed that as far as size is concerned, in countries in transition where small family farms are well established and managed, they tend to be more efficient than bigger farms. Meanwhile, in countries where small farms are a relatively new phenomenon, small farms appear to be less efficient in compared to big farms.

4.2.2. Relationship between Technical Efficiency and Variable Factors

In addition to farm size, ethnicity, education, and tenure (which the farmers can not change in the short run), there are other factors that also affect farmers' operation. These factors can be changed relatively easily. In order to have a better understanding of the sources of inefficiencies, we add to Model 1A several of these new variables including pump (the number of water pumps used on farm for irrigation in units), pipeline (the total length of irrigation pipeline in meters), family labor (natural logarithm of family labor used in coffee production in one year), hired labor (natural logarithm of hired labor used in coffee production in one year), fertilizer (natural logarithm of the amount of fertilizer used in one year), and pesticide (natural logarithm of the amount of pesticide used in one year). The results of this Tobit regression are presented under the column heading Model 2A in Table 4.2.

Model 2A shows there is a negative relationship between TE score and pipeline and hired labor. However, these correlations are not significant. In addition to the significant relation with farmsize (as found in Model 1A), the TE score also has a significant correlation with pump, family labor, fertilizer, and pesticide. The level of efficiency appears to have positive correlation with the number of irrigation pumps. In contrast, TE score decline with the increase of family labor, fertilizer, and pesticide. That means the more family labor, fertilizer, and pesticide that farmers use as inputs, the less efficient they are. This phenomenon indicates an overuse of those inputs in coffee production. This finding is consistent with previous research, including Cheeseman (2005), Ha et al (2001), Koninck (2000), and Ahmad (2000). In general, all have found that the expansion of coffee production in Vietnam has been associated with inappropriate usage of chemical fertilizer and pesticides which lead to contamination of ground and surface water.

4.2.3. Relationship between Technical Efficiency and Interaction Variables

Efficiency scores in Table 4.1 point toward a higher efficiency level of ethnic minorities compared with Kinh farmers. In order to understand why this is the case, we add interaction variables to Model 2A. The regressions with these interaction variables are labeled Model 3A and Model 4A in Table 4.2.

Model 3A includes interaction variables between ethnicity and factors which farmers can not change in the short run. Those are ethnicity*farm size, ethnicity*education, and ethnicity*tenure. Model 4A also includes interaction variables between ethnicity and factors that can be altered in the short run. The new variables are ethnicity*pump, ethnicity*pipe, ethnicity*family labor, ethnicity*hired labor, ethnicity*pesticide, and ethnicity*fertilizer. Most of the interaction variables do not have a significant correlation with the TE score. Two exceptions are the interaction with education and pump. Pump shows a negative relation with TE score. This means farmers from ethnic minorities will have lower level of efficiency if they have more irrigation pumps (possibly another indicator of over use of inputs in coffee production). The education-ethnicity interaction, on the other hand, is positively correlated with TE score. Ethnic minority farmers with higher levels of education are more likely to have better efficiency performance than their counterparts. In both models with interaction terms, farm size and other inputs remain significantly correlated with the TE score. This means that regardless of ethnicity, these factors still play an important role in the level of efficiency in coffee production.

From the two-stage analysis model, we have some initial findings about the efficiency of coffee production in Vietnam. Ethnic minority farmers, on average, have higher efficiency scores than their Kinh counterparts. Farm size and the level of efficiency are positively and significantly correlated, regardless of ethnicity. Also, there is evidence of the over use of inputs like labor, fertilizer, and pesticide. When adding in interaction variables, we find some additional connections between ethnicity, education and TE score.

The results presented in Table 4.2 are derived from the complete sample of 209 households. It is possible that among these 209 households, there are some which are super efficient, causing distortions in the analysis. In the next section, we examine whether removing these outliers changes the conclusions from the Tobit analysis.

4.3. Outlier Removal

This section discusses the removal of outliers and the impact of this change in the sample on the initial results. It therefore will answer the question of whether or not economic inference with regard to coffee production in Vietnam is sensitive to outliers. The method of identifying outliers is based on the weight that one observation puts on other observations during the construction of the DEA hull. Observations with greatest weights are considered to be outliers and will be removed from the sample. There are two main approaches to identifying weights used here: lambda count and lambda sum. Lambda count is the number of times that one observation appears during construction of the DEA envelope. Lambda sum, on the other hand, is the cumulative weight of one observation in all constructed efficient sets. It is the sum of all the weights for one observation that exist in the matrix of lambda values.

4.3.1. Removing Outliers Using Lambda Count

The value of lambda count (C_j) is

$$C_j = \sum_{j \text{ if } \lambda_{ij} > 0} 1$$

Figure 4.2 shows the greatest values of lambda count and lambda sum when the dataset is trimmed of observations with large weights based on lambda count. The horizontal axis represents the number of observations being dropped and the vertical axis represents the greatest values of lambda count and lambda sum. As we have 209 observations in the sample, we choose to drop 100 observations to observe the change of the highest weights. Looking at Figure 4.2 we see that initially, there is a big drop of both lambda count and lambda sum values. This rapid drop is followed by a more gradual decline and the two values of lambda count and lambda sum tend to be close to each other as we drop more and more observations. Since there is no big change in the greatest weights after we drop 10 observations, we conclude that only the first 10 observation are outliers. The new TE scores calculated with DEA after removing these observations are presented in the middle column of Table 4.1.

After these outliers are removed, more than half of the coffee farms in Dak Lak become efficient. We can see that, compared with the results from the untrimmed dataset, all of the average efficiency scores are higher. This phenomenon is explained by the drop of outliers on the efficiency frontier. In the full sample, the frontier is constructed from all the efficient observations, including the super-efficient ones or outliers. This creates a big gap between inefficient observations and the frontier which results in a large number of low TE scores. When the outliers are removed, the distances between inefficient observations and the efficient frontier become smaller. Some of the observations previously identified as inefficient even become efficient. As a result, the average TE score rises. The average TE score in the new sample is 0.84. Comparing Kinh and non-Kinh, we see a higher percentage of non-Kinh farmers producing efficiently (56.67% vs. 48.2%). Also, the mean value of efficiency scores for non-Kinh farmers is higher than that of Kinh farmers (0.89 vs. 0.81).

Using the dataset trimmed of outliers based on lambda count, the Tobit regressions were repeated. Results are presented in Table 4.3. The models are parallel to those presented in Table 4.2. Models 1B and 2B are short regressions between TE scores and factors that are suspected to have impact on TE scores. The significant correlation between the TE score and family labor and pesticide remains but we no longer see a strong relationship between the TE scores and farm size, pump, and fertilizer. We conclude that the removal of outliers has an important impact on the interpretation of sources of inefficiency. Models 3B and 4B add interaction variables between ethnicity and other inputs. Family labor and the amount of pesticide used remain significantly correlated with TE score which indicates over-use of labor and pesticide in coffee production. Remaining variables, including farm size, show no significant correlation with the TE score.

4.3.2. Removing Outliers Using Lambda Sum

The value of lambda sum (S_j) is computed as:

$$S_j = \sum_j \lambda_{ji} \, .$$

Similar to the lambda count approach, observations with the greatest values of S_j are removed sequentially until the values of the remaining scores converge.

Figure 4.3 shows the greatest values of lambda count and lambda sum when the dataset is trimmed of observation with large weights based on lambda sum. Initially there is a rapid reduction followed by a gradual reduction. As we are searching for the convergence of weights, the observations associated with the quick drop of lambda sum and lambda count values are considered to be outliers. Those observations are removed from the sample. According to Figure 4.3, dropping 10 observations is enough for lambda count and lambda sum to stabilize. The new DEA results based on a sample trimmed using the lambda sum approach (n=199) observations are presented in the final column of Table 4.1

As discussed in the context of lambda count, the TE scores in this new sample are higher than those in the original model. On average, the TE score of 199 coffee households is 0.81. So the coffee farmers in the area only reach 81% of their full efficiency level. Looking at the TE scores of the two different groups, we still see a significantly higher level for ethnic minority farmers. They have, on average, a TE score of 0.86, compared with 0.79 for Kinh counterparts. That means while Kinh farmers, on average, need to increase their output by 21% to reach the efficiency level, farmers from ethnic minorities only have to increase their output by 14%.

The results of a Tobit model based on this trimmed dataset are presented in Table 4.4. Models 1C and 2C show the relationship between TE scores and fixed and variable factors. Models 3C and 4C add to the short regressions the interactions between ethnicity and other variables. Similar to the case when we drop observations base on the values of lambda count, the strong correlation between TE scores and farm size disappears. However, there are still signs of over use of inputs in coffee production. Family labor, the amount of fertilizer, and the amount of pesticide applied have statistically negative correlation with TE scores in Model 2C. That relationship remains robust after adding the interaction variables in models 3C and 4C. In models 3C and 4C, there is a positive and significant relationship between ethnic minorities' farmer's education and the TE score. The more education the head of the household has, the more efficient in coffee production the farm is. The relationship between the number of pumps operated by non-Kinh farmers and the TE score is negative, which possibly indicates an overuse of this input in coffee production.

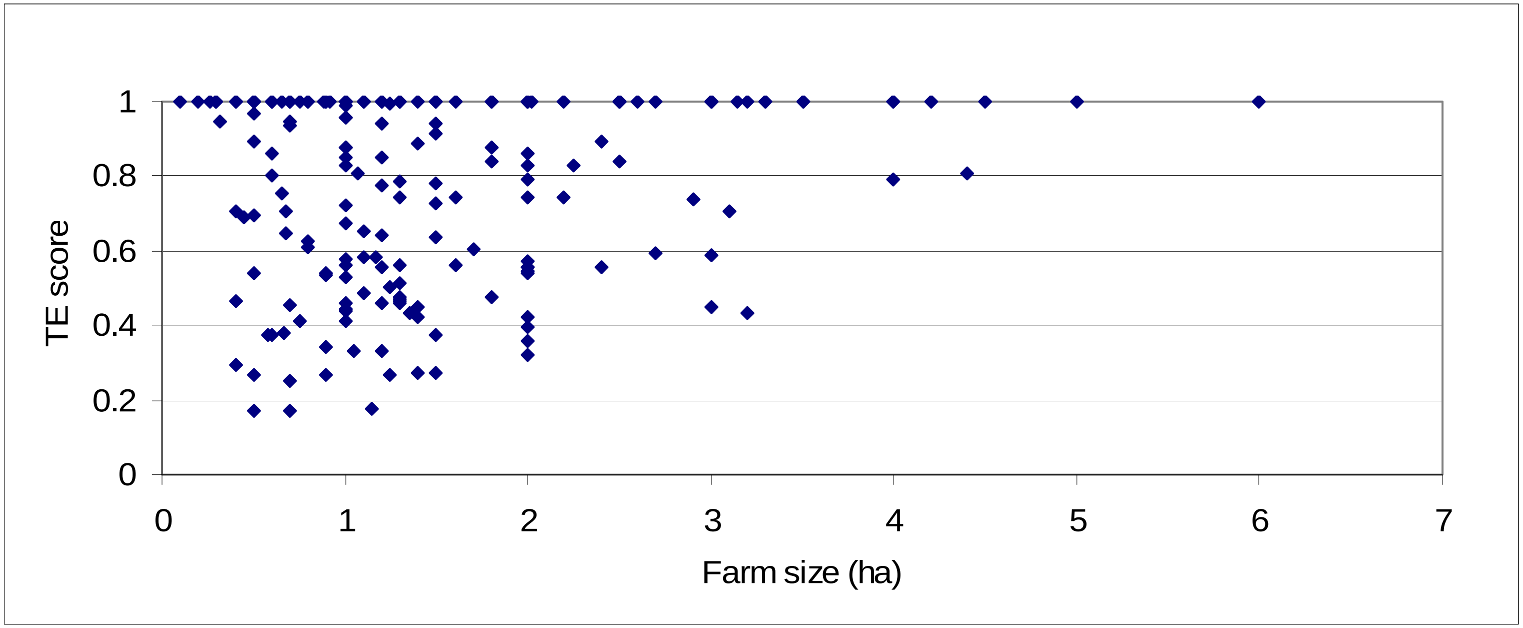


Figure 4.1 Plot of TE Score and Farm Size

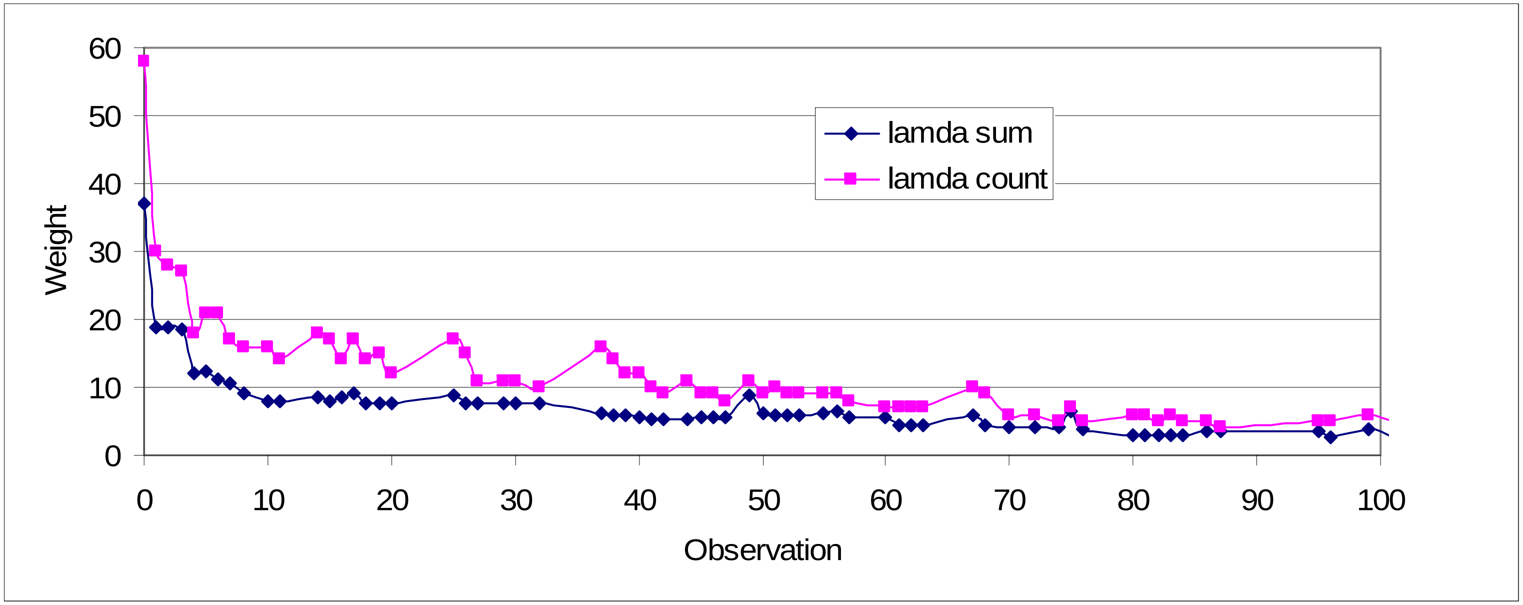


Figure 4.2 Sequence of Weights when Outliers Dropped Based on Lambda Count.

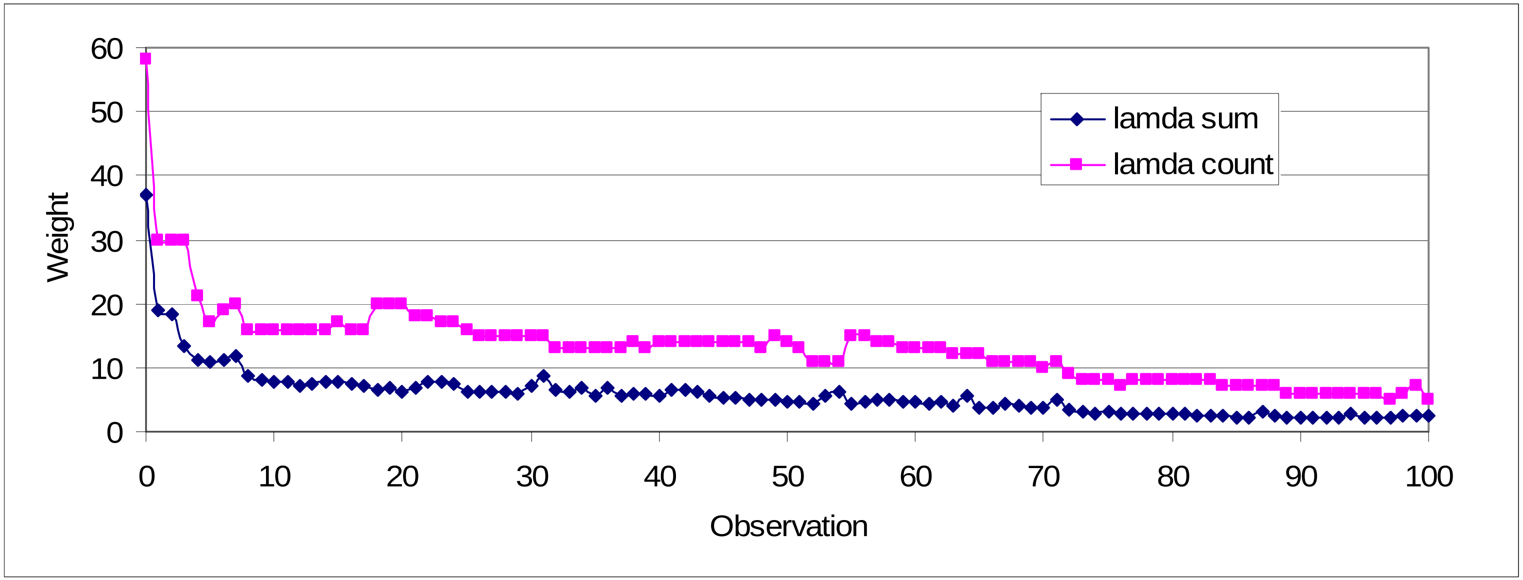


Figure 4.3 Sequence of Weights when Outliers Dropped Based on Lambda Sum.

Table 4.1 Results from DEA models

	Full Sample			Dropping Lambda count			Dropping Lambda sum		
	Kinh	Non-Kinh	All	Kinh	Non-Kinh	All	Kinh	Non-Kinh	All
Efficiency Score	0.76	0.83	0.78	0.81	0.89	0.84	0.79	0.86	0.81
% efficient	42%	51%	45%	48%	57%	51%	45%	53%	47%
Number of observation	144	65	209	139	60	199	137	62	199

Table 4.2 Tobit Model for Technical Efficiency, Full Sample

Independent Variables	Model 1A	Model 2A	Model 3A	Model 4A
Constant	1.0365***	2.4261***	2.4371***	2.2856***
	(-0.1522)	(-0.3693)	(0.3782)	(0.5115)
Farm size (ha)	-0.3101**	-0.2206*	-0.1987*	-0.2372*
	(-0.1251)	(-0.1229)	(0.1197)	(0.1225)
Farm size square	0.0967***	0.0747**	0.0757***	0.0793***
	(-0.0317)	(-0.0296)	(0.0285)	(0.0297)
Ethnicity (0= Kinh, 1= Others)	0.098	-0.0224	0.0312	0.8922
	(-0.076)	(-0.0823)	(0.2403)	(0.9243)
Education (years)	-0.0066	-0.0079	-0.0205*	-0.0114
	(-0.0088)	(-0.0086)	(0.0111)	(0.0087)
Tenure (% area with title)	0.0686	0.0065	0.0792	0.0285
	(-0.11)	(-0.1025)	(0.1363)	(0.1019)
Pump (number)		0.0884*	0.0914*	0.1564**
		(-0.0535)	(0.0529)	(0.0623)
Pipeline (length in meters)		-0.0002	-0.0002	-0.0004
		(-0.0002)	(0.0002)	(0.0002)
Ln(family labor) (days/ha)		-0.1290***	-0.1307***	-0.1271***
		(-0.0306)	(0.0301)	(0.0309)
Ln(hired labor) (days/ha)		-0.0212	-0.0250	-0.0291
		(-0.0164)	(0.0162)	(0.0189)
Ln(fertilizer) (kg/ha)		-0.0772**	-0.0749**	-0.0551
		(-0.0355)	(0.0355)	(0.0577)
Ln(pesticide) (kg/ha)		-0.0892***	-0.0901***	-0.1043***
		(-0.0341)	(0.0336)	(0.0375)
Ethnicity*Farm size			-0.0675	
			(0.0703)	
Ethnicity*Education			0.0301*	
			(0.0164)	
Ethnicity*Tenure			-0.1702	
			(0.1977)	
Ethnicity*Pump				-0.2734**
				(0.1226)
Ethnicity*Pipeline				0.0005
				(0.0005)
Ethnicity*Ln(family labor)				-0.1141
				(0.1282)
Ethnicity*Ln(hired labor)				0.0002
				(0.0367)
Ethnicity*Ln(pesticide)				0.1068
				(0.0878)
Ethnicity*Ln(fertilizer)				-0.0499
				(0.0749)
Pseudo R^2	0.0744	0.2221	0.2393	0.2503
Number of observations	209	209	209	209

Note: *, **, and *** indicate coefficient estimate is significantly different from zero at 90%, 95%, and 99% confidence level respectively. Standard errors in parentheses.

Table 4.3 Tobit Model for Technical Efficiency, Outliers Removed Based on Lambda Count

Independent Variables	Model 1B	Model 2B	Model 3B	Model 4B
Constant	1.0265***	2.2255***	2.2455***	2.0410***
	(0.1550)	(0.3934)	(0.4014)	(0.5188)
Farm size (ha)	-0.1941	-0.0489	-0.0321	-0.0543
	(0.1267)	(0.1247)	(0.1220)	(0.1244)
Farm size squared	0.0686**	0.0379	0.0401	0.0405
	(0.0322)	(0.0299)	(0.0291)	(0.0300)
Ethnicity (0= Kinh, 1= Others)	0.0965	-0.0373	0.0100	0.7224
	(0.0790)	(0.0865)	(0.2563)	(0.9612)
Education (years)	-0.0070	-0.0055	-0.0144	-0.0090
	(0.0091)	(0.0089)	(0.0116)	(0.0091)
Tenure (% area with title)	0.0802	0.0422	0.0870	0.0866
	(0.1137)	(0.1078)	(0.1413)	(0.1078)
Pump (number)		0.0523	0.0547	0.1486**
		(0.0551)	(0.0551)	(0.0643)
Pipeline (length in meters)		-0.0002	-0.0002	-0.0003
		(0.0002)	(0.0002)	(0.0002)
Ln(family labor) (days/ha)		-0.1134***	-0.1149***	-0.1158***
		(0.0308)	(0.0306)	(0.0311)
Ln(hired labor) (days/ha)		-0.0399**	-0.0428	-0.0525
		(0.0174)	(0.0175)	(0.0202)
Ln(fertilizer) (kg/ha)		-0.0601	-0.0595	-0.0375
		(0.0390)	(0.0389)	(0.0586)
Ln(pesticide) (kg/ha)		-0.1061***	-0.1068***	-0.1042***
		(0.0368)	(0.0366)	(0.0399)
Ethnicity*Farm size			-0.0627	
			(0.0743)	
Ethnicity*Education			0.0215	
			(0.0173)	
Ethnicity*Tenure			-0.1078	
			(0.2089)	
Ethnicity*Pump				-0.3755***
				(0.1328)
Ethnicity*Pipeline				0.0003
				(0.0005)
Ethnicity*Ln(family labor)				-0.0638
				(0.1331)
Ethnicity*Ln(hired labor)				0.0156
				(0.0392)
Ethnicity*Ln(pesticide)				0.0290
				(0.0965)
Ethnicity*Ln(fertilizer)				-0.0376
				(0.0796)
Pseudo R^2	0.0556	0.1877	0.1975	0.2295
Number of observations	199	199	199	199

Table 4.4 Tobit Model for Technical Efficiency, Outliers Removed Based on Lambda Sum

Independent Variables	Model 1C	Model 2C	Model 3C	Model 4C
Constant	0.9681***	2.2694***	2.3138***	2.2685***
	(0.1534)	(0.3810)	(0.3919)	(0.5197)
Farm size (ha)	-0.1698	-0.0623	-0.0417	-0.0683
	(0.1251)	(0.1199)	(0.1194)	(0.1193)
Farm size squared	0.0638**	0.0383	0.0397	0.0416
	(0.0316)	(0.0281)	(0.0281)	(0.0282)
Ethnicity (0= Kinh, 1= Others)	0.0894	-0.0362	-0.0465	0.4398
	(0.0774)	(0.0850)	(0.2513)	(0.9275)
Education (years)	-0.0054	-0.0061	-0.0185	-0.0096
	(0.0090)	(0.0088)	(0.0115)	(0.0089)
Tenure (% area with title)	0.0592	0.0060	0.0420	0.0453
	(0.1124)	(0.1065)	(0.1405)	(0.1054)
Pump (number)		0.0832	0.0865	0.1791***
		(0.0552)	(0.0550)	(0.0651)
Pipeline (length in meters)		-0.0001	-0.0001	-0.0003
		(0.0002)	(0.0002)	(0.0002)
Ln(family labor) (days/ha)		-0.1189***	-0.1210***	-0.1233***
		(0.0313)	(0.0310)	(0.0319)
Ln(hired labor) (days/ha)		-0.0221	-0.0262	-0.0317
		(0.0172)	(0.0172)	(0.0199)
Ln(fertilizer) (kg/ha)		-0.0745**	-0.0719*	-0.0712
		(0.0371)	(0.0373)	(0.0584)
Ln(pesticide) (kg/ha)		-0.0969***	-0.0984***	-0.1113***
		(0.0361)	(0.0358)	(0.0395)
Ethnicity*Farm size			-0.0657	
			(0.0722)	
Ethnicity*Education			0.0291*	
			(0.0170)	
Ethnicity*Tenure			-0.0883	
			(0.2057)	
Ethnicity*Pump				-0.3633***
				(0.1263)
Ethnicity*Pipeline				0.0004
				(0.0005)
Ethnicity*Ln(family labor)				-0.0580
				(0.1282)
Ethnicity*Ln(hired labor)				0.0069
				(0.0381)
Ethnicity*Ln(pesticide)				0.1133
				(0.0909)
Ethnicity*Ln(fertilizer)				-0.0218
				(0.0754)
Pseudo R^2	0.0603	0.1926	0.2078	0.2349
Number of observations	199	199	199	199

CHAPTER 5. CONCLUSIONS AND LIMITATIONS

5.1. Conclusions and Policy Implications

Coffee was first planted in Vietnam in the 19th century by French missionaries. At that time several coffee plantations were established in the Central Highlands and nearby areas. In the 1990s, the increase in the world coffee price together with a flow of immigration into the Central Highlands and liberalization of agricultural commodity trading attracted many newcomers to Vietnam's coffee industry. As a result, production rocketed, and Vietnam quickly became the world's number two exporter of coffee. Over the 13 years from 1990 to 2003 Vietnam's coffee growing area increased 4.3 times, productivity increased 1.95 times, and output increased from 9200 tons in 1990 to 771,200 tons in 2003, more than 700%. Coffee is one of the country's most valuable exports, earning up to $600 million a year. Ninety five percent of total production is sold in the international market.

The increase in Vietnam's coffee production along with other factors led to a sharp drop in the producer price of coffee. By 2002, the world coffee price had decreased to 50 cents/lb, its lowest level in the last 100 years. The coffee price crisis is causing hardship for 4 million Vietnamese people who depend on the production of coffee for income. In recent years, many coffee farmers in the Central Highlands of Vietnam have had to sell their beans for much less then they cost to produce. Also, Vietnam's coffee production has been associated with inappropriate usage of chemical fertilizers and pesticides, high rates of water use, and environmental stresses, including deforestation, water scarcity, soil degradation, and contamination of surface and ground water.

Considering the role of coffee in Vietnam's economy as well as the problems that country has encountered recently, it is of crucial importance to increase the efficiency of coffee operations in Vietnam and rationalize the use of inputs. This thesis has focused on identifying the level of efficiency and the sources of inefficiency of coffee producers in Vietnam. The data were collected from 209 households in four villages in Ea Tul catchments, one of the main coffee production areas in Dak Lak province. A main goal of the analysis was to investigate the sensitivity of findings to the presence of outliers in the dataset.

A descriptive analysis of the data shows that all of the households in the dataset plant coffee as one of their main crops. Kinh farmers, on average, have higher education than farmers from ethnic minorities. With respect to agricultural production, ethnic minority farmers use more inputs than Kinh farmers. However, Kinh farmers have more income from agriculture than their counterparts. Ethnic minority farmers also allocate more area for coffee production than do Kinh farmers. Kinh farmers use more pesticides and fertilizers than ethnic minority farmers. They also have significantly higher levels of coffee production.

The results from the first DEA model show that almost half of the farmers in the sample size (44.5%) are on the efficient frontier. On average, coffee farmers in the sample are 22% below the efficient level. Comparing technical efficiency between Kinh farmers and non-Kinh farmers we found that a higher percentage of non-Kinh farmers produce efficiently (51% vs. 42%). Also, the mean value of efficiency scores for non-Kinh farmers is significantly higher than for Kinh farmers (0.83 vs. 0.76).

When outliers are removed from the dataset, the TE scores from both groups rise. This phenomenon can be explained by a shift downward of the estimated efficient frontier when the outliers are removed. The outliers in the dataset appear to have distorted the efficiency interpretation of coffee producers in the area. Nevertheless, non-Kinh farmers still have higher TE scores than Kinh farmers, on average.

In order to have an understanding of the sources of inefficiencies, a two-stage analysis was used. In the first stage the TE score was computed using DEA. The second stage is a Tobit regression using the TE score from the DEA model as a dependent variable. The independent variables were factors that are suspected to be correlated with the TE scores. In addition, some interaction variables were added to have a better understanding the ways in which sources of inefficiency might be conditioned by ethnicity. In the original model (n=209), we found a significant correlation between the TE score and farm size. The most efficient farms were either small farms or big farms. Farms with coffee area of 1.5 hectares were the least efficient. However, after dropping outliers, this relationship ceases to exist. It could be the nature of this particular dataset that makes the difference. If we had additional data from the area, the difference in Tobit results may disappear. Nevertheless, in the context of this thesis, we do not have enough evidence to say if it is the dataset that makes changes in the Tobit outputs. However, we can see that the economic reference from technical efficiency is sensitive to outliers in this case. The removal of outliers causes farm size, one of the main factors correlated with technical efficiency to become no longer statistically correlated with the efficiency level.

In the original model, there is a significant negative relationship between family labor and TE score. This relationship remains robust after adding in interaction terms. The negative relationship means that with more family labor, efficient falls. When outliers are removed, we still see that family labor is negatively correlated with TE score. This is an indication of the over use of labor in coffee production. This result is somewhat surprising considering that farmers actually have to hire labor to work on coffee. This partly may be due to the inefficient allocation of labor. Results from the Tobit regression suggest that coffee farmers in Dak Lak can actually reduce the number of labor hired to work on coffee because they have extra family labor. One of the possible ways to resolve this problem is to create more off-farm activities to attract redundant labor. When more off-farm opportunities are available, farmers will be able to take full advantage of their free time to earn money. On the other hand, it will also release the pressure of over-use of labor on coffee production which will in turn contribute to higher levels of technical efficiency in coffee production.

Similar to family labor, the amounts of pesticide and fertilizer used in coffee production are negatively correlated with efficiency levels. These relationships imply an over use of inputs in coffee production. After outliers are dropped, we still see that fertilizer and pesticide are negatively correlated with efficiency. This result is consistent with the work of Cheeseman (2005), Ha et al (2001), Koninck (2000), and Ahmad (2000). Coffee farmers mainly apply fertilizer and pesticide based on their own perception of the correct proportion or combination of fertilizer/pesticide, the supply, and affordability. As there is no shortage of fertilizer in the market and because coffee is considered as one of the most profitable crop, many farmers tend to apply high amounts of fertilizer and pesticide for coffee trees. In order to deal with this problem, there should be more extension classes and services to teach coffee farmers how to use fertilizer and pesticide judiciously. It is crucial to let them know that by using suitable amounts of fertilizer and pesticide; they not only help to protect the environment but also increase their productive efficiency.

When adding interaction terms between ethnicity and inputs, interestingly, the relationship between being ethnic minorities with education and the TE score stays robust to the removal of outliers. This indicates that the more education ethnic minorities have the more efficient in coffee production they become. In fact, in the study site, the average number of years attending school of household heads from ethnic minorities is very low compared with Kinh household heads (4.7 years vs. 7.9 years). One of the main advantages of education for ethnic minorities is the opportunity to study the Kinh language. In Vietnam, most of the technical requirements and market information are written in Kinh. For that reason, people from other ethnicities who go to school will have greater chances to access this information.

5.2. Limitations and Further Studies

Results from this study can help farmers, researchers and policy makers to understand the level of efficiency in coffee production in Vietnam, as well as the sources of inefficiency and their policy implications. Furthermore, this thesis provides a methodological innovation by introducing a new simple method for detecting outliers. This approach was found to be successful. The removal of outliers was found to have an impact on economic inference with regard to efficiency scores as well as sources of inefficiency. However, there are still some limitations with this research.

First, recall that when we remove outliers, farm size changes from being significantly correlated with TE score to being not statistically related. The question of whether or not it is because of the dataset remains unanswered. Additional data from the same area would shed light on this issue.

A second problem arises when we talk about the negative relationship between labor and the TE score. It is possible that the negative relationship is caused by an inefficient allocation of labor. Nevertheless, as we do not have data on how farmers actually allocate between household labor and hired labor for coffee production, it is impossible to give recommendations on how to allocate labor more effectively.

Third, we use only one year of data in our analysis. It would be advantageous if we had panel data for coffee production in the area. With data from different years, we would be able to take into account factors such as weather or policy that have affects on coffee production. Also, we could compare among different years to see whether our results are sensitive to the time period under consideration.

Fourth, with regards to factors that affect efficiency, some additional explanatory factors could be added to this model. These include land quality, water quality, labor quality etc. These variables are very important to coffee production but were not taken into account in this research due to lack of data.

Fifth, in this study, we only calculated the technical efficiency scores for coffee producers. It would be of interest if in future studies prices could be incorporated. A technically efficient coffee producer is not necessarily economically efficient. Results of a study of allocative efficiency would be interesting because the market value of coffee, the main concern of coffee farmers, would be under consideration.

Sixth, with the new method of detecting outliers based on weights, two sets of outliers were identified. Although 80% of the observations identified using the methods were the same, there are still some differences which could lead to slightly dissimilar results in the two stage DEA-Tobit analysis. Because outliers should always be outliers no matter what method is being used, further study should focus on explicitly identifying which approach (lambda count or lambda sum) should be used, and under what circumstances.

Finally, there are two other methods to identify outliers developed by Wilson (1993), and Simar (2003). The methods used here could be compared to these other methods. Differences among results from the three methods, if any, should be explained.

LIST OF REFERENCES

Ahmad, Alia. (2000). “An Institutional Analysis of Changes in Land Use Pattern and Water Scarcity in Dak Lak Province, Vietnam”. Paper presented at the Nordic Conference on “Institutions, Livelihoods and the Environment: Change and Response in Mainland Southeast Asia.” Copenhagen 27-29 September 2000.

Andrews, D. and Pregibon, D. (1978). “Finding the Outliers that Matter.” *Journal of the Royal Statistical Society*. Series B (Methodological), 40 (1): 85-93.

Audibert, M. et al. (2003) “Social and Health Determinants of the Efficiency of Cotton Farmers in Northern Cote d’Ivoire.” *Social Science & Medicine,* 56:1705-1717.

Babu, K. L. et al. (2003). “Impact of Globalization on Marginal Farmers: A Case Study of Coffee Farmers in India and Vietnam.” WIDER Conference on Sharing Global Prosperity, Helsinki, Finland, 6-7 September.

Banker, R. et al. (1984). “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis.” *Management Science*, 30(9): 1078-1092.

Binam, J. et al. (2003). “Factors Affecting Technical Efficiency among Coffee Farmers in Côte d’Ivoire: Evidence from the Centre West Region.” *R&D Management*, 15(1): 66-76.

Bravo-Ureta, B. E. and Evenson, R. E. (1994). “Efficiency in Agricultural Production: The Case of Peasant Farmers in Eastern Paraguay.” *Agricultural Economics,* 10: 24-37.

Cardenas, J. (2001) “The World Coffee Crisis”. International Coffee Organization.Accessed on April 5, 2007. http://www.ico.org/event_pdfs/cardenas.pdf

Cazals, C. et al. (2002). “Nonparametric Frontier Estimation: A Robust Approach.” *Journal of Econometrics*, 106: 1-25.

Charnes, A. et al. (1978). “Measuring the Efficiency of Decision Making Units.” *European Journal of Operational Research*, 2: 429-444.

Cheesman, J. (2005). “Natural Resources, Institutions and Livelihoods in Dak Lak, Vietnam.” ISSN 1832- 7435 Australian Centre for International Agricultural Research (ACIAR). Project: ADP/2002/015.

Coelli, T. et al. (2005). “An Introduction to Efficiency and Productivity Analysis.” Second Edition- Springer Science + Business Media, Inc.

Farrell, M. (1957). “The measurement of Productive Efficiency.” *Journal of the Royal Statistical Society*. Series A (General), 120(3): 253-290.

Fontenay, P. and Leung, S.(2002). “Managing Commodity Price Fluctuations in Vietnam’s Coffee Industry.” Microeconomics Working Papers No. 574 (2002), East Asian Bureau of Economic Research.

FAO. (2003). “Medium-term Prospects for Agricultural Commodities- Projections to the year 2010.” Accessed on April 5, 2007. http://www.fao.org/docrep/006/y5143e/y5143e00.HTM

Gorton, K. and Davidova, S. (2004). “Farm Productivity and Efficiency in the CEE Applicant Countries: A Synthesis of Results.” *Agricultural Economics*, 30: 1-16.

Ha, D.T et al. (2001) “Impact of Changes in Policy and Market Conditions on Land Use, Land Management and Livelihood among Farmers in Central Highland of Vietnam.” Accessed on April 15th, 2006. http://www.sanrem.uga.edu/sanrem/database/pdf/ImpactsChanges.pdf

Ha, D. T. and Shively, G. (2004). “Coffee Boom, Coffee Bust, and Smallholder Response in Vietnam’s Central Highlands.” *of Development Economics*, forthcoming

Hardi, K. and Whittaker, J. (1999). “Efficiency, Environmental Contaminants and Farm Size: Testing for Links Using Stochastic Production Frontiers.” *Journal of Applied Economics*, II(2): 337-356.

ICARD-Oxfam. (2002). “The Impact of the Global Coffee Trade on Dak Lak Province, Vietnam: Analysis and Policy Recommendations.” ICARD, Oxfam Great Britain, and Oxfam Hong Kong.

ICO. (2007). “The Story of Coffee.” Accessed on April 5, 2007. http://www.ico.org/coffee_story.asp

ICO. (2003). “Impact of the Coffee Crisis on Poverty in Producing Countries.” Accessed on April 5, 2007 http://www.ico.org/documents/icc89-5r1e.pdf

Jha, R. et al. (2000). “Productivity, Technical and Allocative Efficiency and Farm Size in Wheat Farming in India: A DEA Approach.” *Applied Economics Letters*, 7: 1-5.

Karonen, T. (2003). “Stars and Stripes- and Starbucks, Coffee: America’s National Drink.” Accessed on April 5, 2007. http://www.uta.fi/FAST/US7/PAPS/tk-coffe.html

Koopmans, T.C. (1951). “An Analysis of Production as an Efficient Combination of Activities.” in T.C. Koopmans, ed., *Activity Analysis of Production and Allocation, Cowes Commission for Research in Economics* (1951): John Wiley and Sons, Inc., New York.

Krasachat, W. (2004).“Technical Efficiencies of Rice Farms in Thailand: A Non-Parametric Approach.” The Journal of American Academy of Business, Cambridge Volume 4, Number 1&2, March 2004, ISSN 1540 – 1200.

Lindskog, E. et al. (2005). “When Rapid Changes in Environmental, Social and Economic Conditions Converge: Challenges to Sustainable Livelihoods in Dak Lak, Vietnam.” Stockholm Environment Institute. ISBN 91 975238 3 6.

Lovell, C.A.K. (1993). “Production Frontiers and Productive Efficiency.” in Fried, H.O, C.A. Knox Lovell, and S.S. Schmidt, eds. *The Measurement of Productive Efficiency*. New York: Oxford University Press.

Luong, V.Q. and Tauer, W.L. (2006). “A Real Option Analysis of Coffee Planting in Vietnam.” *Agricultural Economics,* 35: 49-57.

Muller, D. (2003). “Land-use Change in the Central Highlands of Vietnam- A Spatial Econometric Model Combining Satellite Imagery and Village Survey Data.” PhD Dissertation, Institute of Rural Development, Georg-August University of Gottingen.

Murillo-Zamorano, L. (2004). “Economic Efficiency and Frontier Techniques.” *Journal of Economic Survey* 18 (1): 33-77.

Nhan, D.T. (2001). “Orientations of Vietnam Coffee Industry.” Speech at International Coffee Conference May 17-19, 2001 in London, UK. Accessed on April 5, 2007. http://www.ico.org/event_pdfs/nhan.pdf

Osorio, N. (2002). “The Global Coffee Crisis: A Threat to Sustainable Development.” Submission to the World Summit on Sustainable Development, Joahnnesburg.

Otsuki, T. et al. (2002). “The Implication of Property Rights for Joint Agriculture-timber Productivity in the Brazilian Amazon.” *Environment and Development Economics,* 7: 299-323.

Oxfam International. (2002). "Mugged: Poverty in Your Coffee Cup." Accessed on April 5, 2007. http://www.oxfamamerica.org/pdfs/mugged_coffee-report.pdf

Ray, S. (2004). *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*. New York: Cambridge University Press.

Rios, A. R. and Shively, G. (2005). "Farm size, irrigation infrastructure, and the efficiency of coffee production in Vietnam." American Agricultural Economics Association Annual Meeting, Providence, Rhode Island. Accessed on April 5, 2007. http://www.agecon.purdue.edu/staff/shively/RS.pdf

Seaver, L. B. and Triantis, P. K.(1995). "The Impact of Outliers and Leverage Points for Technical Efficiency Measurement Using High Breakdown Procedures." *Management Science*, 41(6):.937-956.

Simar, L. and Wilson, P. (2007). "Estimation and Inference in Two-stage, Semi-parametric Models of Production Processes." *Journal of Econometrics*, 136: 31-64.

Simar, L. (2003). "Detecting Outliers in Frontier Models: A Simple Approach." *Journal of Productivity Analysis*, 20: 391-424.

Thang, B.T. (2005). "Trade Liberalization and Marketization of Agro-Products." in Impact of Socio-Economic Changes on The Livelihoods of People Living in Poverty in Vietnam, Institute of Developing Economies, Japan External Trade Organization.

Thompson, G. and Thore, S. (1992). *Computational Economics- Economic Modeling with Optimization Software*. Scientific Press.

Timmer, C.P. (1971). "Using a Probabilistic Frontier Function to Measure Technical Efficiency." *The Journal of Political Economy*, 79(4): 776-794.

VICOFA. (2005). "Development and Prospect for the Vietnamese Coffee Industry." Accessed on March 15th, 2007. http://www.vicofa.org.vn/pls/news/cate$.htnoidung?v_cate_code=1206&v_item_id=60661

Wilson, P. (1993). "Detecting Outliers in Deterministic Nonparametric Frontier Models with Multiple Outputs." *Journal of Business and Economic Statistics*, 11(3): 319-323.

Zyl van, J. et al. (1995). "The Relationship between Farm Size and Efficiency in South African Agriculture." Policy Research Working Paper, Agricultural and Natural Resource Department The World Bank.

Appendix A. Production Function

The production function for the sample is assumed to take the form

$$yield = \exp(\beta_o + \beta_1 fertilizer + \beta_2 pesticide + \beta_3 location + \beta_4 familylabor$$
$$+ \beta_5 hiredlabor + \beta_6 irrigation$$

in which yield (kg/ha) is the yield of coffee in 2003, fertilizer (kg/ha) is the amount of fertilizer applied for coffee production in one year, pesticide (kg/ha) is the amount of pesticide applied for coffee production in one year, family labor (days/ha) is the number of family labor used for coffee production in one year, hired labor (days/ha) is the number of hired labor used for coffee production in one year, location and irrigation are dummy variables.

Results are presented in table A.1

Table A.1 Regression Results

Independent Variable	Coefficients	Standard Error	P>|t|
Constant	7.356	0.118	0.000
Family Labor	0.001	0.000	0.136
Hired Labor	0.002	0.001	0.005
Fertilizer	0.000	0.000	0.000
Pesticide	0.006	0.006	0.330
Location	-0.234	0.075	0.002
Irrigation	0.023	0.012	0.050
Adjusted R squared			0.2467
Observations			209

Except for family labor and pesticide, all other variables have positive correlation with yield. The correlation between fertilizer and yield is significant at the confidence level of 95%. That means in the study sites, the more fertilizer a family apply, the more coffee yield they get.

Appendix B. Gams Code

Gams code for DEA

```
Sets

obs       /1*209/
slo(obs)   selected observation
k          input and output /yield, urea, npk, phos, sulph, potas, other, org, herb, pest, flabor,
hlabor/
ki(k)        input /urea, npk, phos, sulph, potas, other, org, herb, pest, flabor, hlabor/
ko(k)         output /yield/;

option lp=minos;
alias (obs,lobs);
table data(obs,k)

* Data excluded*;

Variables
Theta        efficiency scores
Lambda(obs)        weight;

Positive variables lambda(obs);

Equation

cons1(ko,obs)        DEA constraint for each output
cons2(ki,obs)        DEA constraint for each input
cons3(obs)           Reference tech;

cons1(ko,slo)..
        sum(obs,lambda(obs)*data(obs,ko))- theta*data(slo,ko)=g= 0;
cons2(ki,slo)..
        sum(obs,lambda(obs)*data(obs,ki))-data(slo,ki) =l=0;
cons3(slo)..
        sum(obs,lambda(obs))=e=1;

Model Dea /cons1, cons2, cons3/;
Parameter

techeff (obs);
set oa(obs) set of units to analyze;
oa(obs)= yes;
slo(obs)=no;
```

```
Loop(oa,
slo(oa)=yes;
solve Dea using lp maximizing theta;
techeff(oa) =1/theta.l;
slo(oa)=no);

Parameter
Results(obs,*) Tech Efficiency measures;
Results(obs,'Tech')=techeff(obs);

Display
Results;
```

Dropping Outliers Based on Lambda Count

```
Sets
junk / 0 /
allobs       /1*209/
obs(allobs)
*slo(obs)   selected observation
k         input and output /yield, urea, npk, phos, sulph, potas, other, org, herb, pest, flabor,
hlabor/
ki(k)        input /urea, npk, phos, sulph, potas, other, org, herb, pest, flabor, hlabor/
ko(k)        output /yield/;
obs(allobs) = yes ;
alias (obs,lobs),(allobs,alllobs);
table data(allobs,k)

*Data excluded*

Variables
Theta(allobs)       efficiency scores
Lambda(allobs,allobs)       weight
zoo                   objective;

Positive variables lambda(allobs,allobs);

Equation

cons1(ko,allobs)        DEA constraint for each output
cons2(ki,allobs)        DEA constraint for each input
cons3(allobs)           Reference tech
obj;

cons1(ko,allobs)$lobs(allobs)..
        sum(obs,lambda(obs,allobs)*data(obs,ko))- theta(allobs)*data(allobs,ko)=g= 0;
cons2(ki,allobs)$lobs(allobs)..
        sum(obs,lambda(obs,allobs)*data(obs,ki))-data(allobs,ki) =l=0;
cons3(allobs)$lobs(allobs)..
        sum(obs,lambda(obs,allobs))=e=1;
obj..
        sum(obs,theta(obs)) =e= zoo ;

Model Dea /cons1, cons2, cons3, obj/;

parameter result(*,*);
parameter lamsum(allobs),lamt(allobs,allobs),lamcnt(allobs) ;
option lp=minos,solprint=off,limrow=0, limcol=0;
```

```
solve Dea using lp maximizing zoo;
lamsum(obs) = sum(lobs,lambda.l(obs,lobs)) ;
lamcnt(obs) = sum(lobs$(lambda.l(obs,lobs) gt 0),1) ;
lamt(obs,lobs) = lambda.l(lobs,obs) ;
display lamt,lamcnt ;
display lamsum ;
result('0','lamsum') = smax(obs,lamsum(obs)) ;
result('0','lamcnt') = smax(obs,lamcnt(obs)) ;
result('1','drop')   = sum(allobs$(lamcnt(allobs) eq smax(alllobs,lamcnt(alllobs))),
  ord(allobs)) ;

Set j / 1*1 00/ ;
Loop(j,
obs(allobs)$(lamcnt(allobs) eq smax(lobs,lamcnt(lobs))) = no ;
lambda.l(allobs,lobs)$(lamcnt(allobs) eq smax(alllobs,lamcnt(alllobs))) = 0 ;
lamcnt(allobs)$(lamcnt(allobs) eq smax(alllobs,lamcnt(alllobs))) = 0 ;
lamsum(allobs)$(lamcnt(allobs) eq smax(alllobs,lamcnt(alllobs))) = 0 ;
solve Dea using lp maximizing zoo;
lamsum(obs) = sum(lobs,lambda.l(obs,lobs)) ;
lamcnt(obs) = sum(lobs$(lambda.l(obs,lobs) gt 0),1) ;
display lamt,lamcnt ;
display lamsum;
result(j,'lamsum') = smax(obs,lamsum(obs)) ;
result(j,'lamcnt') = smax(obs,lamcnt(obs)) ;
result(j,'drop') = sum(allobs$(lamcnt(allobs) eq smax(alllobs,lamcnt(alllobs))),
  ord(allobs)) ;
result(j,'remaining') = card(obs);
);
display result;
```

Dropping Outliers Based on Lambda Sum

```
Sets
junk / 0 /
allobs      /1*209/
obs(allobs)
*slo(obs)   selected observation
k         input and output /yield, urea, npk, phos, sulph, potas, other, org, herb, pest, flabor,
hlabor/
ki(k)        input /urea, npk, phos, sulph, potas, other, org, herb, pest, flabor, hlabor/
ko(k)        output /yield/;
obs(allobs) = yes ;
alias (obs,lobs),(allobs,alllobs);
table data(allobs,k)

*Data excluded*

Variables
Theta(allobs)       efficiency scores
Lambda(allobs,allobs)       weight
zoo                  objective;

Positive variables lambda(allobs,allobs);

Equation

cons1(ko,allobs)       DEA constraint for each output
cons2(ki,allobs)       DEA constraint for each input
cons3(allobs)          Reference tech
obj;

cons1(ko,allobs)$lobs(allobs)..
        sum(obs,lambda(obs,allobs)*data(obs,ko))- theta(allobs)*data(allobs,ko)=g= 0;
cons2(ki,allobs)$lobs(allobs)..
        sum(obs,lambda(obs,allobs)*data(obs,ki))-data(allobs,ki) =l=0;
cons3(allobs)$lobs(allobs)..
        sum(obs,lambda(obs,allobs))=e=1;
obj..
        sum(obs,theta(obs)) =e= zoo ;

Model Dea /cons1, cons2, cons3, obj/;

parameter result(*,*);
parameter lamsum(allobs),lamt(allobs,allobs),lamcnt(allobs) ;
option lp=minos,solprint=off,limrow=0, limcol=0;
```

```
solve Dea using lp maximizing zoo;
lamsum(obs) = sum(lobs,lambda.l(obs,lobs)) ;
lamcnt(obs) = sum(lobs$(lambda.l(obs,lobs) gt 0),1) ;
lamt(obs,lobs) = lambda.l(lobs,obs) ;
display lamt,lamcnt ;
display lamsum ;
result('0','lamsum') = smax(obs,lamsum(obs)) ;
result('0','lamcnt') = smax(obs,lamcnt(obs)) ;
result('1','drop')   = sum(allobs$(lamsum(allobs) eq smax(alllobs,lamsum(alllobs))),
  ord(allobs)) ;

Set j / 1*1 00/ ;
Loop(j,
obs(allobs)$(lamsum(allobs) eq smax(lobs,lamsum(lobs))) = no ;
lambda.l(allobs,lobs)$(lamsum(allobs) eq smax(alllobs,lamsum(alllobs))) = 0 ;
lamcnt(allobs)$(lamsum(allobs) eq smax(alllobs,lamsum(alllobs))) = 0 ;
lamsum(allobs)$(lamsum(allobs) eq smax(alllobs,lamsum(alllobs))) = 0 ;
solve Dea using lp maximizing zoo;
lamsum(obs) = sum(lobs,lambda.l(obs,lobs)) ;
lamcnt(obs) = sum(lobs$(lambda.l(obs,lobs) gt 0),1) ;
display lamt,lamcnt ;
display lamsum;
result(j,'lamsum') = smax(obs,lamsum(obs)) ;
result(j,'lamcnt') = smax(obs,lamcnt(obs)) ;
result(j,'drop') = sum(allobs$(lamsum(allobs) eq smax(alllobs,lamsum(alllobs))),
  ord(allobs)) ;
result(j,'remaining') = card(obs);
);
display result;
```