MULTI CRITERIA EFFICIENCY ANALYSIS OF FUEL STATIONS

by

Yusuf Şire

B.S., Industrial Engineering, Boğaziçi University, 2003

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Industrial Engineering
Boğaziçi University
2006

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my Thesis Supervisor, Professor İlhan Or for his invaluable guidance, help and patience during preparation of this Thesis. I would like to mention his giving inspiration and hope when I could not found any way out.

I also would like to express my sincere gratitude to my Advisor, Professor Kuban Altınel for his support and patience during the Master's Education Period.

ABSTRACT

MULTI CRITERIA EFFICIENCY ANALYSIS OF FUEL STATIONS

As the oil prices increase, the oil products distribution sector becomes more attractive in Turkey. In recent years, new distributor companies entered the market and the number of gas stations increased significantly, while it is known that rapidly expanding industries experience "overshoot and collapse" behavior in the middle run. In this study, a multi-criteria efficiency analysis of gas stations of a distributor firm is conducted. Two different methods, the Data envelopment Analysis (DEA) and the Stochastic Production Frontiers (SPF) are used for this purpose. The CCR, the BCC models of the DEA, with and without Assurance Region (AR) extensions are applied.

Results of the DEA studies suggest that, after a threshold level of physical resources, scale of station does not provide competitive advantage and better performance depends on management practices. Stations are grouped by cities and a considerable difference in average efficiency scores is observed among different cities. On the other hand, best performing cities coincide in the CCR and the BCC formulations to a great extent, whereas efficiency scores of inefficient units may change dramatically. The AR restrictions have different effects on different stations; some having higher efficiency scores and some having lower, while the statement about scale is still valid and the ranking of cities change considerably. Correlation of efficiency scores to input and output factors are also analyzed.

In the SPF study, different outputs are aggregated into one single output. In this case, results differ considerably from results of the DEA study, especially when the BCC formulation of the DEA (which allows variable returns to scale), is considered.

On the overall, results confirm usefulness of the DEA in multi output problems and provide useful information about performances of gas stations, which is especially valuable when investing in a new station.

ÖZET

AKARYAKIT İSTASYONLARININ ÇOK KRİTERLİ VERİMLİLİK ANALİZİ

Petrol fiyatlarının artması, akaryakıt dağıtımı sektörünü daha çekici bir hale getirmektedir. Son yıllarda yeni dağıtıcı firmaların pazara girmesine ilaveten istasyon sayısında büyük bir artış yaşanmıştır. Hızlı bir genişleme süreci yaşayan sektörlerin, orta vadede daralma eğilimi gösterdikleri bilinmektedir. Bu çalışmada bir dağıtıcı firmaya bağlı akaryakıt istasyonlarının çok kriterli verimlilik analizi gerçekleştirilmiştir. Bu kapsamda, Veri Zarflama ve Stokastik Üretim Fonksiyonları olmak üzere iki farklı yaklaşım kullanılmıştır. Veri Zarflama yönteminin CCR ve BCC modelleri, Güven Bölgesi (AR) varsayımlarını kapsayacak ve kapsamayacak şekilde uygulanmıştır.

Veri zarflama yaklaşımının sonuçları, belirli bir fiziksel büyüklükten sonra ölçeğin verimliliğe katkı sağlamadığını ve yüksek performansın yönetimsel uygulamalara bağlı olduğunu ortaya koymuştur. İstasyonların şehir bazında gruplanarak incelenmesi, verimlilik seviyelerinin şehir bazında önemli oranda değişebildiğini göstermiştir. Ayrıca BCC ve CCR modellerinde, verimli istasyonlar büyük oranda örtüşmesine rağmen verimsiz istasyonların verimlilik seviyelerinde büyük farklılıklar gözlenmiştir. Güven Bölgesi kısıtlarının değişik istasyonlar üzerinde değişik etkileri olmuş; kimi istasyonlar daha yüksek, kimileri daha düşük verimlilik seviyelerine gelmişlerdir. Ancak ölçek ile ilgili yapılan çıkarım geçerliliğini korumuştur. Verimlilik seviyeleri ile girdi ve çıktı değişkenleri arasındaki ilişki de incelenmiştir.

Stokastik Üretim Fonksiyonları çalışması kapsamında çıktı değişkenleri tek bir değişkene toplanmıştır. Bu çalışmada sonuçların Veri Zarflama Metoduna göre büyük farklıklar gösterdiği görülmüştür. Özellikle değişken getiriye izin veren BCC modelleri söz konusu olduğunda farklılık daha belirgindir.

Bu çalışma genel olarak, Veri Zarflama'nın çok çıktılı problemlerdeki etkinliğini doğrulamış ve özellikle yeni bir akaryakıt istasyonuna yatırım aşamasında kullanılacak değerli bilgiler sağlamıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ARI	Assurance Regions Type 1 Model of the DEA
ARII	Assurance Regions Type 2 Model of the DEA
BCC	Banker, Charnes and Cooper Model of the DEA
CCR	Charnes, Cooper and Rhodes Model of the DEA
COLS	Corrected Ordinary Least Squares Method
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
MOLS	Modified Ordinary Least Squares Method
OLS	Ordinary Least Squares
SPF	Stochastic Production Frontiers

1. INTRODUCTION AND PROBLEM DEFINITION

1.1. Introduction

The fuel distribution sector is usually considered to be a “key” sector in Turkey, due to its high and still increasing demand and huge sales volume. As an attractive industry, competition is harsh and profit margins are relatively small. Tax rates are high and almost 75% of fuel sales revenue is directly collected as tax, through the KDV and the OTV surcharges. Actors in the sector include Oil Refining Companies (TUPRAS), distributor firms (POAS, OPET, Shell etc.) and the stations, at the retail level. Distributor firms either buy fuel from TUPRAS or import it themselves, and then sell it to the stations (distributor companies have not been involved in refining up to now). Distributor firms founded PETDER (Petroleum Products Sector Association). In recent years, the number of oil stations has significantly increased. It is not uncommon that rapid growth in an industry is followed by a decline period. In the literature, this is named as “Overshoot and Collapse”, a behavior encountered not only in ecological systems but also in service and manufacturing industries.

There is not a good sense of efficiency in the sector, overall sales volume (in tons) being the main, and in most cases, the only performance indicator. Stations are paid incentives based on their sales volume, and these payments may reach 5% of total sales. Such a performance evaluation and incentive mechanism causes misallocation of resources both for distributors and stations. Therefore it is necessary to develop and apply a systematic performance evaluation methodology to the oil distribution station chains.

1.2. Problem Definition: Market Conditions, Challenges and Definition of Efficiency

In the oil refining and distribution sector, fuel is classified as “white” and “black”, while aviation fuel is considered separately:

1. White Products
   1.1. Benzine
      1.1.1. Unleaded
      1.1.2. Super
   1.2. Kerosene
   1.3. Diesel
2. Black Products
   2.1. Kalyak (Heater Oil)
   2.2. Fuel Number 6
3. Aviation Fuels

White products are usually used in motor vehicles, while black products are used for industrial and heating purposes.

Table 1.1. Sales amount of Oil products in Turkey (Source: PETDER 2005 Sector Report)

Product		2004	2005	Change
White Products	**m^3**	**16.455.111**	**16.413.881**	**-0,25%**
Total Benzine	m^3	3.751.563	3.473.988	-7,40%
Unleaded Benzine	m^3	2.740.099	2.864.966	4,56%
Super Benzine	m^3	1.011.464	609.022	-39,79%
Keresene	m^3	40.554	34.005	-16,15%
Diesel	m^3	12.662.994	12.905.888	1,92%
Black Products	**Ton**	**4.473.043**	**3.919.289**	**-12,38%**
Heater Oil	Ton	714.995	587.746	-17,80%
Fuel Number 6	Ton	3.758.048	3.331.543	-11,35%
Aviation Fuel	m^3	2.017.732	2.283.490	13,17%
Total Fuel	**Ton**	**18.087.255**	**17.525.383**	**-3,11%**

In recent years, sales amount of oil products in Turkey have shown a decreasing trend, as displayed in Table 1.1. This decrease stems from different reasons. In black products for example, the increase in natural gas consumption (for heating purposes) is the main reason for the downwards trend. However in white products, where the market grows steadily due to ever increasing number of motor vehicles in traffic, the main reason is illegally imported fuel, which is not accounted for in the official statistics. Amount of illegal fuel import is estimated to be more than 1 million ton per year (PETDER Illegally traded fuel report). Otogaz (i.e. Liquefied Petroleum Gas or LPG, that can be consumed in motor vehicles if the vehicle is equipped appropriately), also has some credit for the decrease in white products consumption since it is an alternative product for benzene. Also biodiesel (i.e. diesel fuel distilled from vegetal oil) is a candidate to be an important factor, since it is expected that tax rates for this product will be considerably lower in order to support domestic renewable energy consumption. From the distributors and the fuel stations perspective, however, the overall effect of biodiesel won't be so significant, since biodiesel has to go through the same distribution channels as ordinary fuel.

The fuel distribution sector is expected to remain an attractive one (both for distributors and stations), since it is believed that low margins are well compensated by high volume and increasing prices. As it can be seen in Table 1.2., the sector experienced 30% growth last year, despite stationary and/or decreasing sales volumes.

Table 1.2. Sales value of Oil products in Turkey (Source: PETDER 2005 Sector Report)

		2004	**2005**	**Change**
Total Sector	**billion YTL**	**29,6**	**37,4**	**26,35%**
OTV	billion YTL	13,6	16,6	22,06%
KDV	billion YTL	4,5	5,7	26,67%
Total Tax	**billion YTL**	**18,1**	**22,3**	**23,20%**
Net Sector	**billion YTL**	**11,5**	**15,1**	**31,30%**

There are distributor companies that provide oil products to their "dealers" (i.e. the chains of fuel stations affiated to their companies). Dealers are classified as "stations" and "non-stations". White products are almost exclusively sold by stations. In this study only the "station" type of dealers are considered.

Number of distributor companies was 51 in 2005 and although some distributors have only a small number of stations, a steady growth in this number is experienced over the years (and this trend is expected to continue), as displayed in Table 1.3.:

Table 1.3. Number of oil products distributor companies in Turkey.(Source: EPDK, Energy Markets Regulation Council)

Year	Number Of Distributor Companies
2000	14
2001	17
2002	19
2003	26
2004	28
2005	51

Leading distributors that have well established and widespread station chains and significant market shares are listed in Table 1.4.:

Table 1.4. Number of Fuel Stations, (Grouped by distributor companies) : EPDK, Energy Markets Regulation Council)

Distributor	Number of Stations
Aytemiz Akaryakıt Dağıtım ve Pazarlama A.Ş. (Aytemiz)	650
BP Petrolleri Anonim Şirketi (BP)	599
Erk Petrol Yatırımları A.Ş.(Petrol Ofisi)	458
OPET Dış Ticaret ve akaryakıt (OPET)	620
OPET Petrolculuk A.Ş. (OPET)	643
Pet Line Petrolculük (Petline)	227
Petrol Ofisi A.Ş. (Petrol Ofisi)	3.375
The Shell Company of Turkey Ltd (Shell)	645
Total Oil Turkiye (Total)	525
Turcas Petrol A.Ş. (Turcas)	626
Turkuaz Petrol Ürünleri A.Ş. (Turkuaz)	225
Others	2761
Total	**11.354**

Total number of stations was 8.343 in year 2000; having grown 36,1% in 5 years, there were 11.354 active stations in year 2005. Upwards trend accelerated beginning from 2003. (PETDER 2005 Sector Report). The main reason underlying this growth is high oil prices.

Petrol Ofisi, operating with two firms, which has a total of 3.833 stations, is the leading company in the sector. Likewise OPET has two firms which makes 1.263 stations at total.

Investing in a fuel station is not so cheap, although newcomers are supported by distributor companies. Paradoxically, challenges stem from the increasing volume of the sector. Higher unit prices lead to lower profit rates. Pressure from distributor companies for market share and increasing number of rival stations drop the profit margins of the stations. When considering the sector trends and dynamics, it is safe to conclude that a primary goal is to increase market share. Stations are paid incentives based on their sales volume, in order to stimulate sales. Stations in turn, lower their prices to increase sales and to fill their quotas. On the other hand, the “overshoot and collapse” behavior of the sector, mentioned in the previous section, usually leads to the elimination of the poor performers from the market in the middle run. Nevertheless, the notion of “efficiency” is weak in the sector.

One of the methods used to increase sales volume is promotions, usually supported by distributor companies. This is usually at distributor level, and at the station level market sales gets important as distributors force stations to lower their profits margins from fuel sales. Respectively, market sales become one of the critical issues that determine success or failure in fuel distribution sector along with the fuel sales volume.

Therefore it is necessary to develop some method for efficiency evaluation of fuel stations.

2. LITERATURE REVIEW AND METHODOLOGY

2.1. Efficiency Measurement Methods

Efficiency is a generic term that can be used with different meanings. In this study, "comparative efficiency" is considered.

Problem of comparative efficiency stems from just practical managerial reasons. Some common questions addressed are:

- How efficiently are the operating units (Decision Making Units, DMUs) using their resources?
- Is the industry such that there are economies of scale?
- Is the industry such that there are economies of scope?
- What are the efficient marginal costs of outputs?
- Has there been productivity change within the industry over time?
- How has a given operating unit fared on productivity change?

These universal questions are valid both in public and private sectors. The basic requirement is to compare the DMUs on the levels of outputs they "produce" relative to their input levels.

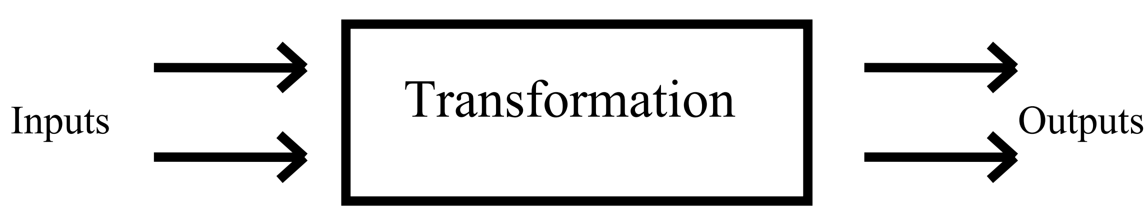


Figure 2.1. Transformation of Inputs into Outputs

In the given operating context, the measure of efficiency is normally one of:

- The distance between observed and maximum possible output for given inputs (output efficiency)

- The distance between observed and minimum possible input for given outputs (input efficiency)

A well accepted classification of comparative efficiency measures is "parametric" versus "nonparametric" methods. The parametric methods hypothesize a functional form and use input and output data to estimate the parameters of that function. The estimated function is then used to arrive at estimates of the efficiencies of the units. Non-parametric methods, on the other hand, does not require any assumption about the production function.

2.1.1. Parametric Comparative Efficiency Measurement

Assume that for each of the DMU's under consideration, there is one single output or multiple outputs can be standardized into one output (The DMU's may correspond to any individual operational unit, process, subsystem or module whose "efficiency" is being evaluated). Suppose there are a total of N DMU's and K inputs. Consider a production function of the form:

$$Y_i = \alpha + \sum_{k=1}^{K} \beta_k x_{ik} + e_i \quad i = 1 \ldots n \tag{2.1}$$

where
Y_i is the (standardized) output of unit i,
x_{ik} are the inputs for unit *i* input *k*, k=1,2,….,K,
e_i is the residual for DMU *i*.

It is the residual term e_i that captures the inefficiency. The residual term however, captures other random effects too, such as measurement errors and omitted variables. Therefore, it is difficult to find out the inefficiency component. If random effects are ignored, then it can be stated that the residual term e_i captures only the inefficiencies involved. Then the model becomes,

$$Y_i = \alpha + \sum_{k=1}^{k=K} \beta_k x_{ik} - u_i \quad i = 1 \ldots n \tag{2.2}$$

where $u_i \geq 0$.

Assume that the Ordinary Least Squares (OLS) method is used for the estimation process (the OLS regression implies zero mean for residuals). Therefore, in order to get the true residuals, the OLS method must somehow be adjusted. Accordingly, the OLS model is modified to:

$$y_i = \alpha + \sum_{k=1}^{k=K} \beta_k x_{ik} - u_i \quad i = 1 \ldots n \tag{2.3}$$

Let a* = [a-E(u_i)] be the intercept of (2.3.) that the OLS regression yields. In order to retrieve the true underlying model in (2.2.), it is necessary to add E(u_i) to a*.

The problem is that E(u_i) is an estimate of the mean inefficiency of DMU's which is not known. In order to get an estimate $\hat{E}(u_i)$ of the mean inefficiency of units, an assumption has to be made about the theoretical probability density function of u_i and then the OLS regression residuals are used to estimate $\hat{E}(u_i)$. Let us define the Residual sum of squares as

$$(\text{RSS}) = \sum (y_i - \hat{y}_i)^2$$

then,

$$\sigma_u^2 = \frac{RSS}{n - (K+1)} \tag{2.4.}$$

Three common estimates of $\hat{E}(u_i)$ are as follows:

i) Corrected OLS (COLS): The intercept a* of the OLS model in (2.3.) is adjusted to $a^* + \max(e_i)$

where e_i is the residual for the ith DMU in the original OLS model (2.3.).

ii) Modified OLS (MOLS): MOLS has two variants, half normal and exponential. Assume that random variable X is distributed normally. If the corresponding distribution is truncated below A and/or above B a *truncated normal* distribution (a sample is displayed in Figure 2.2.) is obtained. If the mean is zero and truncation is done at zero, then the *half normal* distribution is obtained.

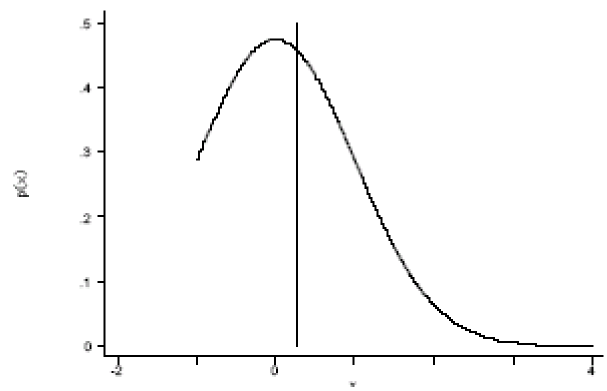


Figure 2.2. Sample probability function of a truncated normally distributed random variable

In half normal MOLS, u_i is assumed to follow a half-normal distribution and the intercept becomes $a^* + \hat{E}(u_i)$. Probability density function of *u* is the same as the normal distribution, except it is truncated at zero. Therefore it can be shown that $E(u) = \left(\frac{\sqrt{2}}{\sqrt{\pi}}\right)\sigma_u$. In exponential MOLS, u_i is assumed to follow an exponential distribution with $E(u) = \sigma_u$. Half normality and exponentiality assumptions imply that larger values of inefficiencies are less likely.

2.1.1.1. The Stochastic Frontier Method In the Stochastic Frontiers Method (SPF) (Aigner et al. 1977, Battese and Corra 1977, Meusen and Van den Broeck 1977), random effects in the residual are not ignored. The approach is motivated by the idea that deviations from the production frontier may not be entirely under the control of the production unit. Therefore, unlike COLS and MOLS, there is a composed error term

$$\varepsilon_i = v_i - u_i \tag{2.5}$$

where
v_i accounts for the random error,
u_i accounts for the inefficiency term.

It can be further assumed that the inefficiency term u_i raises systematically over time. Then the production function becomes:

$$Y_{it} = x_{it}\beta + (v_{it} - u_{it}) \tag{2.6}$$

where
Y_{it} is (the logarithm of) the standardized output for the DMU i at time t,
x_{it} is a k×1 vector of input quantities of the i-th DMU in the t-th time period,
β is the coefficients vector for x (i.e. inputs),
v_{it} are the random variables which are assumed to be iid $N(0,\sigma_V^2)$ and account for random noise,
$u_{it} = (u_i \exp(-\eta(t-T)))$, where the u_i are non-negative random variables, which are assumed to account for technical inefficiency in production and η is a parameter to be estimated. (Distribution of u_i is usually assumed to be exponential, half-normal or truncated normal).

A major step in the SPF procedure is to decompose the source of variability. A well accepted approach is (Battese and Corra (1977) to replace

σ_V^2 and σ_U^2 with $\sigma^2 = \sigma_V^2 + \sigma_U^2$

and

$$\gamma = \sigma_U^2 / (\sigma_V^2 + \sigma_U^2) \quad (2.7)$$

This is done with the calculation of the maximum likelihood estimates. The parameter, γ, must lie between 0 and 1 and thus this range can be searched to provide a good starting value for use in an iterative maximization process, such as the Davidon-Fletcher-Powell (DFP) algorithm (Battese and Coelli (1992)).

Depending on the assumption made about the distribution of the inefficiency u_i, a formula for the conditional value $E(u_i|\varepsilon_i)$ can be obtained. Then the values of ε_i and other values derived from the ε_i are plugged into this formula to arrive at an estimate of the conditional inefficiency u_i of the ith DMU. Consider that u_i is assumed to be distributed half-normally. In such a case, the approach suggested by Jondrow, Lovell, Materov and Schmidt (JLMS) (1982) can be used. It can be shown that Technical efficiency of DMU i can be found as:

$$TE_i = \exp\{-\hat{u}_i\} = \exp\{-E(u_i|\varepsilon_i)\} \quad (2.8)$$

$E(u_i|\varepsilon_i)$ for different distribution assumptions, can be calculated using different methodologies, including Greene (2000) (for the gamma distribution), Kumbhakar and Lovell (2000).

The original specification of model has been vastly used in different applications. Moreover, imposition of one or more restrictions upon the basic model formulation can provide a number of special cases of this particular model, which have appeared in the literature: Setting η to be zero provides the time-invariant model set out in Battese, Coelli and Colby (1989). Furthermore, restricting the formulation to a full (balanced) panel of data gives the production function assumed in Battese and Coelli (1988). The additional restriction of μ

equal to zero reduces the model to the model One in Pitt and Lee (1981). One may add a fourth restriction of T=1 to return to the original cross-sectional, half-normal formulation of Aigner et al (1977). Obviously, a large number of permutations exist. For example, if all these restrictions excepting $\mu=0$ are imposed, the model suggested by Stevenson (1980) results.

2.1.2. The Data Envelopment Analysis (DEA)

The Data Envelopment Analysis, unlike the parametric methods described above, does not assume an underlying production function. Rather, it focuses on the construction of a hypothetical efficient unit, as a weighted average of "efficient" units. In other words, instead of trying to fit a regression plane, DEA is directed to frontiers. Let us assume a multiple input multiple output benchmarking problem. Again, each unit considered for efficiency measurement is called a Decision Making Unit (DMU). What the DEA imposes is:

Efficiency$_i$= Weighted sum of outputs$_i$/ Weighted sum of inputs$_i$ for DMU i.

In this regard, the DEA aims to generalize the single input single output efficiency measurement to multiple inputs and outputs.

While calculating efficiency, a common set of weights must be used to balance all individual input and output factors. The DEA model is formulized and run for one DMU each time. Therefore, a different set for values of weights are deployed for each DMU in order to maximize its efficiency score.

Assume that there are N DMUs (j=1,….,N), each consuming varying amounts of *m* different inputs, x_{ij} i=1,2,…..,m , to produce *s* different outputs, y_{rj} r=1,2,……s. The efficiency concept described above can be mathematically formulated as follows:

$$\text{Efficiency}=\sum_{r=1}^{s} u_r y_{rj} \Big/ \sum_{i=1}^{m} v_i x_{ij} \tag{2.9}$$

Let us further consider the ratio form of the DEA suggested by Charnes et al (1978):

$$\begin{aligned}
&\max \quad \frac{\sum_{r=1}^{s} u_r y_{rj0}}{\sum_{i=1}^{m} v_i x_{ij0}} \\
&s.t. \\
&\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1 \qquad j = 1,2,\ldots, N \\
&\frac{u_r}{\sum_{i=1}^{m} v_i x_{ij0}} \ge \varepsilon \qquad r = 1,2,\ldots, s \\
&\frac{v_i}{\sum_{i=1}^{m} v_i x_{ij0}} \ge \varepsilon \qquad i = 1,2,\ldots, m \\
&\varepsilon > 0
\end{aligned} \tag{2.10}$$

Here ε represents a non-archimedean constant, which is smaller than any positive valued real number. In practice, the non-archimedean term is handled by software packages, but here it is preferred to explicitly state it.

In formulation 2.10, the weighted sum of outputs divided by the weighted sum inputs, for unit j_0, is maximized, where the first constraint limits the values of the weighted sum of outputs divided by the weighted sum inputs by the upper limit 1, for all DMUs. The other constraints imply positive weights for inputs and outputs. Therefore, unit j_0 turns out to be efficient if the objective function value equals to 1.

This formulation can be restated in linear form as follows (Charnes et al 1978):

$$
\begin{aligned}
&\max \sum_{r=1}^{s} u_r y_{rj0} && [1] \\
&s.t. && \\
&\sum_{i=1}^{m} v_i x_{ij0} = C && [2] \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0 && [3] \\
&-u_r \le -\varepsilon && [4] \\
&-v_i \le -\varepsilon && [5]
\end{aligned}
\qquad (2.11)
$$

where $x_{ij} \geq 0$ and $y_{rj} \geq 0$ for all i, r, j.

In this formulation [1] represents the weighted sum of outputs; [2] represents the weighted sum of inputs which are fixed at a constant level *C*. (C is an arbitrary value usually set to 1 or 100.) [3] represents the constraint that the weighted sum of inputs must be greater than or equal to the weighted sum of outputs, for any of the DMUs. Constraints [4] and [5] come from the non-archimedean conditions. If this linear optimization problem is solved, objective function value be either *C* (in case of full efficiency), or less than *C* (in case of inefficiency). Accordingly, the *efficiency score* of a DMU is the ratio of the related function's value to *C*. Also note that this formulation is for one DMU and there should be a similar formulation for each of the DMU's.

According to the Extended Pareto Koopmans Definition, full (100%) efficiency is attained by any DMU, if and only if none of its inputs or outputs can be improved without worsening some of its other inputs or outputs. Note that this definition does not apply to the DEA, since this is an internal benchmarking tool. This is not a disadvantage of DEA: In most managerial and social science applications, "best" performance levels can not be known anyway, unlike in the engineering sense of efficiency. Accordingly, almost all comparative efficiency measures, applied to such situations, are limited likewise. Therefore, the above definition can be restated as: "A DMU is to be rated as fully (100%) efficient on the basis of available evidence, if and only if the performances of other DMUs does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs".

The DEA was introduced in 1978 and has been used in different areas such as, banking, retail, hospitals, non-profit organizations etc extensively. This is mainly due to two major advantages; there is no need of any assumption about the underlying production function and multiple outputs can be well incorporated in the model. The DEA provides new insights into activities or industries, where efficiency has been evaluated by different methods. For example, some firms that are considered as benchmarks with reference to certain measures (e.g. profitability), turn out to be inefficient in DEA studies, pointing to interesting inefficiencies and possible improvement areas (Cooper et al 2000).

Although it is relatively easy to implement the DEA, it has some weaknesses and results of a DEA study may be misleading, unless these weaknesses are taken into consideration. First of all, the random error term is not considered at all. Secondly, results are very dependent on the input/output variables; even a single unnecessary or unrelated variable included in the study may completely distort the results. Thirdly, the "optimal" set of weights generated and assumed by the DEA model solution, may not be feasible nor reasonable nor practical. Also the DEA is very susceptible to outliers.

2.1.3. Comparison of Comparative Performance Evaluation Methods

The methods mentioned in previous sections are all designed for the same purpose; efficiency evaluation by internal benchmarking. All of these methods depend on *identifying a reference boundary relative to which efficiency is assessed*. Consider the simple one input one output case, displayed in Figure 2.3.:

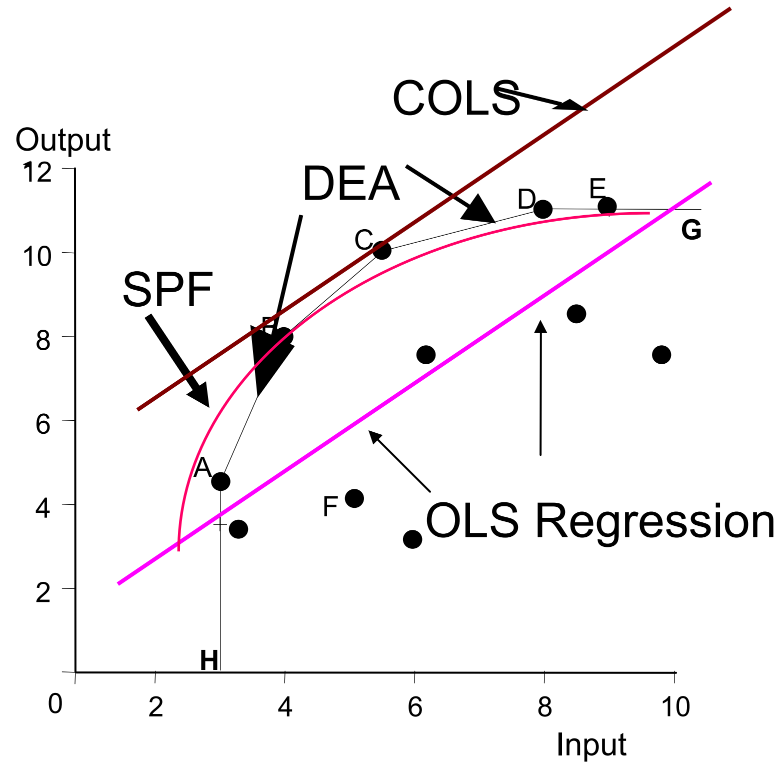


Figure 2.3. Reference boundaries in different performance assessment methods.

As can be seen from Figure 2.3., each method imposes a different reference boundary. The DEA is nonparametric, can handle multiple inputs and outputs, *but* assumes that all distance from the boundary (the efficient frontier) is inefficiency. The SPF, on the other hand, allows for random noise in the distance from the boundary, *but* needs assumptions on the probability distribution of the inefficiency term.

The DEA reveals information about unit-specific peers, returns to scale and productivity change, whereas the SPF, the COLS and the MOLS give industry level information. Additionally, the COLS and the MOLS are very susceptible to outliers and do not allow for random noise.

At many studies in different fields, these methodologies have been used for efficiency measurement. In general terms, it cannot be said that any one of them is better than other.

Rather, the most adequate one must be selected, considering the production process, number of inputs and outputs and possibility of aggregation.

As mentioned in Chapter 1, the Fuel Stations problem has multiple outputs. Different type of sales (e.g. benzene, diesel, market) is considered for maximization. All of the given outputs are in monetary terms (i.e. value of sales in YTL is available), and it can be proposed that aggregating them into one single value of "total sales" is reasonable. However, there are some important problem specifications:

1. In practice, different types of sales have different sales quotas. Both the distributor firm and the station keep track of different types of sales.
2. Profit margin for different types of sales activities are not necessarily the same. Especially market sales has a different (higher) profit margin.

Thus, although all of them are monetary variables, different types of sales values must not be aggregated into one single variable. Accordingly, a methodology that allows dealing with multiple output problems must be chosen. Among the mentioned methodologies, the DEA is the most appropriate one for handling multiple outputs.

Another factor is the existence of random noise. The DEA, the COLS and the MOLS do not take random error and outliers into account.

Another weakness of the DEA is the possibility of unrealistic benchmarks, resulting as a consequence of unrealistically high weights. In the literature there are methodologies developed to cancel out such effects on efficiency evaluation. Examples include assurance region methods and cone ratio method. Therefore, the DEA is chosen to be the main methodology, since it can handle multiple outputs and at the same time consider extreme values.

The Stochastic production function (SPF) approach is dominant to the DEA in distinguishing the random noise and inefficiency terms. One disadvantage of the SPF is that it requires an assumption about the probability distribution of the inefficiency term. This is overcome by applying different functional forms and choosing the most appropriate one. Accordingly, it can not be used effectively when there are multiple outputs (rather when there is a reasonable aggregation of outputs). Therefore, the SPF is used in this study as a complementary approach for the DEA.

2.1.4. The DEA Models

Basic formulation of the DEA was introduced by Charnes, Cooper and Rhodes in 1978. Thus, the initial DEA model is called the CCR. Another version of the DEA was introduced by Banker, Charnes and Cooper in 1984. This version is called the BCC formulation.

2.1.4.1. The CCR Model As mentioned before, *the output oriented CCR Model* can be formulated as:

$$
\begin{aligned}
&\max \sum_{r=1}^{s} u_r y_{rj0} && [1] \\
&s.t. \\
&\sum_{i=1}^{m} v_i x_{ij0} = C && [2] \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0 && [3] \\
&-u_r \leq -\varepsilon && [4] \\
&-v_i \leq -\varepsilon && [5]
\end{aligned}
\qquad (2.12)
$$

where $x_{ij} \geq 0$ and $y_{rj} \geq 0$ for all i, r, j.

It is convenient to call the $\sum_{r=1}^{s} u_r y_{rj0}$ term as *virtual output* and the $\sum_{i=1}^{m} v_i x_{ij0}$ term as *virtual input* for the DMU considered. The formulation above can be interpreted as, “maximization of virtual output, subject to having virtual input fixed at a constant level, while maintaining the condition that virtual output can not exceed virtual input for any DMU”. As noted in Charnes et al. (1985), this implies that the conditions for Pareto-Koopmans optimality are fulfilled, since further increases in this maximal value can be attained only if some of the input values are increased or if some of the output values are decreased.

Like the output oriented model, *there also exists an input oriented CCR model,* where the objective function is to minimize virtual input, while virtual output is fixed:

$$\begin{aligned}
&\min \quad \sum_{i=1}^{m} v_i x_{ij0} \\
&s.t. \\
&\sum_{r=1}^{s} u_r y_{rj0} = C \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0 \\
&-u_r \leq -\varepsilon \\
&-v_i \leq -\varepsilon
\end{aligned} \tag{2.13}$$

where $x_{ij} \geq 0$ and $y_{rj} \geq 0$ for all i, r, j.

In this study, output oriented models are considered, unless explicitly stated otherwise.

<u>2.1.4.2. The BCC Model</u> The BCC version of the DEA model is in common use like the CCR model. The primary difference between the CCR and the BCC formulations is the treatment of *returns to scale*. The CCR model is based on constant returns to scale, whereas the BCC model is more flexible and allows variable returns to scale.

At this point, it is necessary to further explain the term *Returns to Scale.* To start with, consider a single-output, single-input DMU characterized by the production possibility set

$$T = \{(x, y) : y \leq f(x); x \geq a\}$$

where

y = f(x)

is the production function showing the maximum quantity of output "y" producible from input "x", while "a" is the minimum input scale below which the production function is not defined. When there is no minimum scale, a=0.

At some specific point (x, y) on this production function, the average productivity is

$$AP = \frac{f(x)}{x} \tag{2.14.}$$

It is said that *"Locally increasing returns to scale"* holds at this point if a small increase in x results in an increase in AP. Similarly, locally *diminishing returns to scale* exists when AP declines with an increase in x. Under *constant returns*, an increase in x leaves AP unchanged. Thus, $\frac{dAP}{dx}$ is positive under increasing returns, negative under diminishing returns, and zero under constant returns.

Starrett (1977) generalized the above concept of returns to scale in the context of a multi-output, multi-input DMU, by focusing on expansion along a ray. Suppose that the input bundle and the associated output bundle is an efficient pair on the transformation function

$T(x, y) = 0.$

Hence, along the transformation function,

$$\sum_{i=1}^{n} (\tfrac{\partial T}{\partial x_i} x_i) \frac{dx_i}{x_i} + \sum_{j=1}^{m} (\tfrac{\partial T}{\partial y_j} y_j) \frac{dy_j}{y_j} = 0. \tag{2.15}$$

Suppose that all inputs increase at the same proportionate rate and, as a result, all outputs increase at the rate. Then,

$$\frac{\alpha}{\beta} = -\frac{\sum_{i=1}^{n} \frac{\partial T}{\partial x_i} x_i}{\sum_{j=1}^{m} \frac{\partial T}{\partial y_j} y_j} \tag{2.16}$$

is a local measure of returns to scale. Starrett defines,

$$DIR = \frac{\alpha}{\beta} - 1 \tag{2.17}$$

as a measure of the degree of increasing returns. Locally increasing, constant, or diminishing returns hold when DIR exceeds, equals, falls below 0, respectively.

In the BCC formulation a new term, u_0, is introduced, which indicates the returns to scale possibilities. Then, the output oriented BCC formulation can be written as (Banker et al1984):

$$\begin{aligned}
&\max \sum_{r=1}^{s} u_r y_{rj0} - u_0 && [1] \\
&s.t. \\
&\sum_{i=1}^{m} v_i x_{ij0} = C && [2] \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - u_0 \le 0 && [3] \\
&-u_r \le -\varepsilon && [4] \\
&-v_i \le -\varepsilon && [5]
\end{aligned} \tag{2.18}$$

where $x_{ij} \geq 0$ and $y_{rj} \geq 0$ for all i, r, j and u_0 is free.

In the above formulation, $u_0 < 0$ implies locally increasing returns to scale, whereas $u_0 > 0$ implies locally decreasing returns to scale. If $u_0 = 0$, then locally constant returns to scale is encountered. Separate evaluation of returns to scale is more evident in the dual problem and this issue is taken up in section 2.1.4.3.

Note that efficiency score for a given DMU obtained in the BCC formulation is necessarily equal to or greater then the results obtained in the CCR formulation.

2.1.4.3. The Dual DEA Model Like every other linear programming formulation, the DEA formulation has a corresponding dual. Let us consider the output oriented CCR model and reformulate it in the dual form:

Primal	Corresponding Dual Variable
$\max \sum_{r=1}^{s} u_r y_{rj0}$	Z
$s.t.$	
$\sum_{i=1}^{m} v_i x_{ij0} = C$	$\lambda_j \quad j = 1,.....,N$
$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0$	θ
$-u_r \leq -\varepsilon$	s_r^{+}
$-v_i \leq -\varepsilon$	s_i^{-}

Then, the dual is

$$
\begin{aligned}
&\min \quad \theta - \varepsilon[\ \sum_{i=1}^{m} s_i^{-} + \sum_{r=1}^{s} s_r^{+}\ ] \\
&s.t. \\
&\theta x_{ij0} - s_i^{-} - \sum_{j=1}^{n} \lambda_j x_{ij} = 0 \\
&\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^{+} = y_{rj0}
\end{aligned}
\tag{2.19}
$$

where

θ is unrestricted,

$\lambda_j \geq 0$ for j=1,…….,N,

$s_i^- \geq 0$ for i=1,……,m,

$s_r^+ \geq 0$ r=1,……,s.

Actually, it is the dual form that the name Data Envelopment is inspired from. Any admissible choice of λ_j provides an upper limit for the outputs and a lower limit for the inputs of the selected DMU, and against these limits θ is tightened with $\lambda_j, s_i^-, s_r^+ \geq 0$ representing optimization choices associated with the minimization $\theta = \theta^*$. The collection of such solutions then provide an upper bound, which *envelopes* all of the observations.

It can be shown that the dual problem always has at least one solution:

$\theta=1$, $\lambda_0=1$ and all other variables being equal to zero.

It follows that an optimum will be attained where $1 \geq \theta \geq 0$.

Also since the dual problem has a finite optimum solution, using the theory of duality, it can be stated that:

$$\sum_{r=1}^{s} u_r y_{rj0} = \theta - \varepsilon\left[\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right] \quad \text{at optimality,}$$

Note that $\theta=1$ doesn't imply efficiency ($\sum_{r=1}^{s} u_r y_{rj0}=1$), unless s_i^- and s_r^+ are equal to zero. Also, having s_i^- and $s_r^+=0$ is not sufficient for efficiency either. Therefore, in order for the DMU under consideration to be efficient, all of the following conditions must hold:

$\theta=1$ and

s_i^- and $s_r^+=0$.

Accordingly, the dual for the BCC model can be written as:

$$
\begin{aligned}
&\min \quad \theta - \varepsilon\left[\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right] \\
&s.t. \\
&\theta x_{ij0} - s_i^- - \sum_{j=1}^{n} \lambda_j x_{ij} = 0 \\
&\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{rj0} \\
&\sum_{j=1}^{n} \lambda_j = 1
\end{aligned}
\tag{2.20}
$$

The only difference of the above formulation from the dual of the CCR model is the last term, $\sum_{j=1}^{n} \lambda_j = 1$. Accordingly, the difference of interpretation of efficiency between the CCR and the BCC models is more obvious in dual formulations: The additional constraint in the dual of the BCC allows the variable returns to scale condition, in contrast to the constant returns to scale condition in the CCR.

<u>2.1.4.4. The Scale Efficiency</u> A DMU is said to be *total technically efficient*, if it produces on the boundary of the production possibility set. This boundary or the frontier is defined as the best practice observed assuming constant returns to scale. Total technical efficiency can be further decomposed into *pure technical efficiency* and *scale efficiency*. Pure technical efficiency is obtained assuming variable returns to scale. Let us call the efficiency of the Constant Returns to Scale model as TE_{CRS}, the efficiency of the Variable Returns to Scale model as TE_{VRS} and the Scale Efficiency as SE (Coelli et al 1998). By definition,

$$SE = TE_{CRS} / TE_{VRS}. \tag{2.21}$$

Keeping in mind that the CCR model of the DEA assumes Constant returns to scale and the BCC model of the DEA assumes variable returns to scale, the Scale efficiency of a given DMU can be obtained by solving both models. Scale efficiency can be interpreted as follows:

- If SE=1, then the DMU is scale efficient; its combination of inputs and outputs are efficient both under constant and variable returns to scale conditions.
- If SE<1, then the DMU is operating at an inappropriate size for efficiency.

It is not possible for SE to be greater than 1, since result of the VRS problem is an upper bound for the CRS problem efficiency scores.

2.1.4.5. Assurance Region Models Since its introduction, the DEA has been applied to many different types of performance measurement problems. One of the practical problems encountered while using the DEA is dealing and interpreting extreme values for u_r and v_i generated by the model solution. In many cases, some of the DMU's turn out to be efficient by imposing very large weights to one (or more) of the variables. Therefore, misleading conclusions may be drawn, in the sense that such extreme weights may not be feasible, practical or meaningful for the associated input/output factors. In order to avoid such misleading conclusions, methodologies are introduced that impose restrictions on the levels of weights. There are different approaches aiming to limit the allowable levels of these weights. Most prominent of these approaches include (Allen et al1997):

- Restricting the relative ordering or values of inputs/outputs;
- Imposing absolute weight restrictions;
- Restricting virtual inputs/outputs;

These methodologies are systematized into *Assurance Regions* models.

$$
\begin{aligned}
&\max \sum_{r=1}^{s} u_r y_{rj0} && [1] \\
&s.t. \\
&\sum_{i=1}^{m} v_i x_{ij0} = C && [2] \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0 && [3] \\
&-u_r \le -\varepsilon && [4] \\
&-v_i \le -\varepsilon && [5] \\
&k_i v_i + k_{i+1} v_{i+1} \le v_{i+2} && r_1 \\
&\alpha_i \le \frac{v_i}{v_{i+1}} \le \beta_i && r_2 \\
&\gamma_i v_i \ge u_r && r_3 \\
&\delta_i \le v_i \le \tau_i && r_4 \\
&\rho_r \le u_r \le \eta_r && r_5
\end{aligned}
\tag{2.22}
$$

In formulation 2.20, equations up to constraint r_1 define the original, output oriented CCR model. The mathematical relationships r_1 through r_5 describe various implementations of "Assurance Regions" models.

Restrictions r_1 and r_2 are incorporated into the DEA model in case of *Assurance regions of type 1* (ARI). These types of restrictions are used to incorporate relative values of the inputs/outputs. Use of r_2 is more prevalent, reflecting marginal rates of substitution. The upper bound (β) or the lower bound (α) can be omitted in this case.

Note that the bound values for the ARI are dependent on the scaling of the inputs and outputs, that is, they are sensitive to the units of measure of the related factors. The ARI models result in the same relative efficiency scores whether they are input or output oriented and they always turn out to be feasible (Charnes, Cooper, Huang, Sun 1990).

Restriction r_3 depicts *Assurance Regions type 2* (ARII) models (Thomson, Langemeier, Lee, Leee, Thrall 1990). It is sometimes necessary to link input and output weights, as it is the combination rather than the individual values of variables that the efficiency measure should reflect. ARII results in different relative efficiency scores in output oriented and input oriented methods. Moreover, these models may turn out to be infeasible.

R_4 and r_5 represent the *absolute weight restrictions*. These type of restrictions aim to prevent the model from overemphasizing on some of the inputs/outputs and ignoring some others.

3. APPLICATION OF THE DEA TO EFFICIENCY ASSESSMENT OF FUEL STATIONS

3.1. Model Development

The first step of a DEA study is the determination of an appropriate set of Decision Making Units (DMU) for investigation. Since the DEA aims to evaluate performance of the DMUs that are converting inputs to outputs, the very first question to answer is the "definition" of the DMUs.

One further point is that the DMUs must be comparable. In other words, the units must use same inputs seeking to produce same outputs. A typical example for this situation can be observed in performance evaluation of bank branches. In the literature, there are many studies about this topic. In one study, branches are grouped according to their operational classification (e.g. commercial vs. retail branches), then these groups of branches are assessed separately. Another recent example comes from retail stores, where individual stores of a retail chain are classified according to their "levels" that indicate their size and product mix. In the current study, each station is considered as one DMU. As mentioned in section 1.2. , dealers of a fuel distributor are classified as "stations" and "non-stations". Non-station dealers don't sell directly to the end consumers. Therefore, they are omitted and in this study, only the "stations" type of dealers are considered.

The second step is to decide on the inputs that are consumed and the outputs produced. This is usually more difficult than DMU selection. In many studies, it is observed that efficiency scores of the DMUs may highly depend on omitting or adding one factor. Moreover, two common fallacies about this are either having an insignificant variable in the model or a significant variable being omited. A related practical problem is "the degree of significance". By definition, as the number of the inputs and/or outputs increase, number of efficient DMUs increase.

In the fuel stations problem the primary output is sales. Fuel sales include different types of benzene (unleaded, super etc.), diesel and kerosene sales. Benzene Sales and Diesel Sales have very close profit margins. However, distributors impose separate sales quotas for these products. Moreover, dominance of either product over the other one usually reflects the characteristics of the environment that station operates in. In other words, making such a decomposition of fuel sales leads to better peer comparison in the DEA models. Accordingly, this perspective for sales classification is used in this study. Therefore two of the outputs are:

1. Volume (in YTL) of Benzene Sales
2. Volume (in YTL) of Diesel Sales.

Additionally, as mentioned in Chapter 1, the volume of market sales gains importance as a source of profit, as harsh competition limits the profit rates of fuel sales. Market sales are more profitable, furthermore distributors encourage their dealers to increase their market sales, in order to increase the attractiveness and customer base of their stations. Therefore, volume of market sales is included in the model as an output.

Inputs are the resources consumed or required for producing the outputs. In the fuel stations problem, resources can be classified as human resources and physical resources. Number of pumpers is selected as the human resource. As physical resources, number of pumps and size of the station (i.e. total area occupied by the station) are selected. While considering the size of station, it is noted that the size of a station gets higher as the cost of land gets lower. In rural areas for example, it can be observed that the average size of a fuel station is much higher than that of one in metropolitan areas. So, it has been decided that cost of land (which naturally takes into consideration both size and business attractiveness) should replace the station size in the model. This input also better reflects the sales potential of a DMU. The quantitative measure selected to asses this input is unit rent expense. Therefore, size of the station is discarded and instead of it, *total rent expense* of the station is selected as an input. Accordingly, the Data Envelopment Analysis Model for Fuel stations will have 3 inputs and 3 outputs, as displayed in Table 3.1.:

Table 3.1. Inputs and Outputs for the DEA Models

The Inputs:
The Number of fuel pumps in the station *(Input #1)*
The Average number of pumpers in the station *(Input #2)*
The Rent expense of the station *(Input #3)*
The Outputs:
The Volume (in YTL) of Benzene Sales *(Output #1)*
The Volume (in YTL) of Diesel Sales *(Output #2)*
The Volume (in YTL) of Market Sales *(Output #3)*

3.2. Data Collection

Sales volumes of fuel distribution chains are not provided at station level detail publicly. In order to get this information, one of the leading fuel distribution chains is interviewed.

The company has two different types of stations: Type1 and Type2. Type1 is the main brand, whereas Type2 targets extremely price sensitive consumers. Type2 stations are usually located at rural areas. They have limited benzene and market sales. Due to the reasons explained in section 3.1., Type1 and Type2 type of branches are grouped separately. Moreover, Type2 stations provide significant data only for diesel sales and executives prefer to concentrate on Type1 type of stations. Therefore, in this study only Type1 type of stations are considered.

A further classification of stations is based on geographical region. It is said that different policies are valid in different regions and in practice, it is more reasonable to provide peers to inefficient units from the same region. Aiming to reach this, only stations operating in the Marmara Region are considered. Marmara is chosen among others since it has the highest number of station (130 Type1 type stations); however, the model can be applied to other regions as well. As mentioned, in a DEA study having n inputs and m outputs, there is the possibility of nXm efficient units. Therefore, it is advised that the number of DMUs investigated must be substantially greater than nXm. In the current case, there are 3 inputs and outputs, which multiply to 9. Since the number of stations investigated is 130, it is concluded that this number is sufficient for the DEA study.

A problem is encountered about the 3rd output, the volume of market sales. It is noticed that the value of this factor is not reliable for all stations. However, instead of discarding this output, executives of the distributor are asked to mark individual stations regarding reliability or unreliability of their market sales data. Based on this information, stations are flagged as “Market sales reliable” (Market Sales Flag=1) and “Market Sales Unreliable” (Market Sales Flag=0). Also province information for the stations are included in the data set.

Additionally, some of the rent expenses are not actual values, but expertise values since the land is owned by the station. It is reaffirmed that these values are realistic and can safely be used.

Output values (Benzene Sales, Diesel Sales and Market sales) of all stations are totals for first quarter of 2005 and represented in units of 1.000 YTL. Input values are average values for the same period.

It is also necessary to check the correlation between output variables. Although there is not a written rule, correlation levels higher than 0,80 may point to dependent variables, one of which can be eliminated. Correlations of output variables are displayed in Table 3.2.:

Table 3.2. Correlation levels between output levels.

	Market Sales	**Diesel Sales**	**Benzene Sales**
Market Sales	1		
Diesel Sales	0,72	1	
Benzene Sales	0,85	0,63	1

Although strong correlation (0,85) between the outputs market sales and benzene sales is revealed, both variables are kept in the model; these variables reflect results of different business processes that are analyzed separately in practice and usually it’s the deviations from the general trends that reveal the important business dynamics.

Complete set of inputs and outputs are given in Appendix A. In order to solve the resulting DEA formulations, two different software are used: Banxia Frontier Analyst that has advanced visual tools and EMS software that enables advanced modeling options.

3.3. The Application of the CCR DEA Model

In this stage, output maximization type of the CCR formulation of the DEA model is applied to the data set. The complete result set is given in Appendix B.

Efficiency scores distribution of the DMUs are given below. Note that only 9 out of 130 units turned out to be efficient.

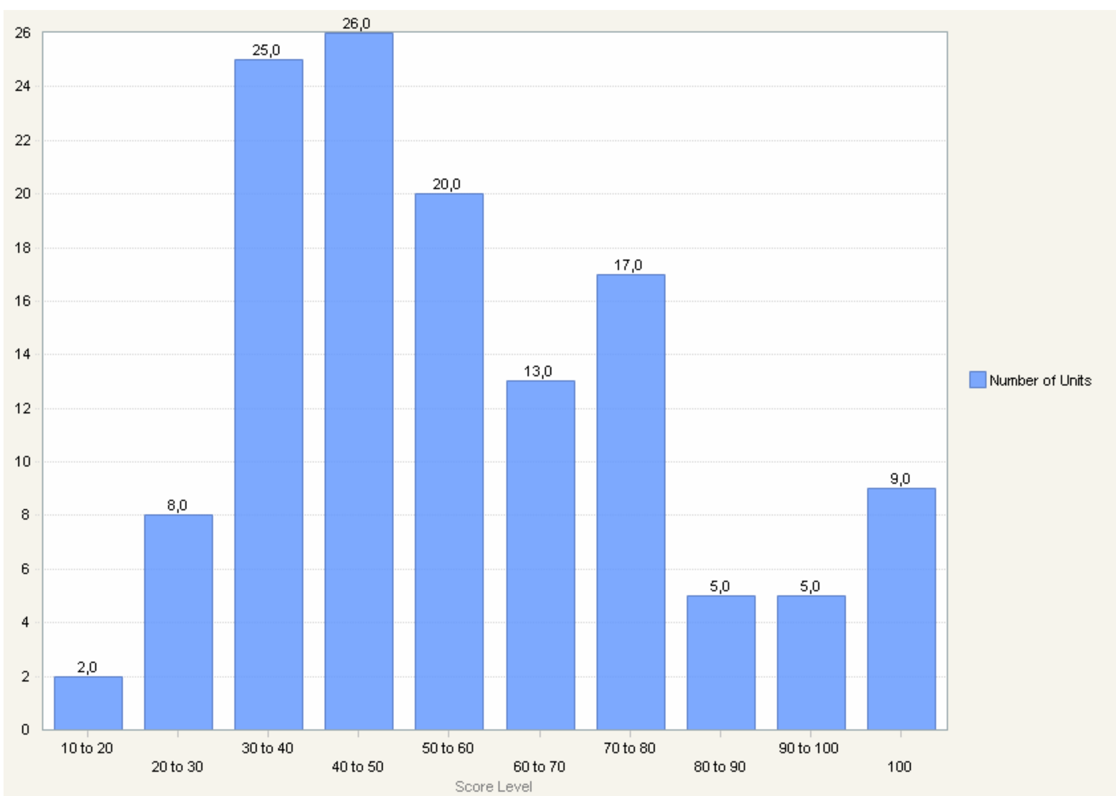


Figure 3.1. Efficiency distribution for the CCR model results.

An important statistic for an efficient DMU is the number of "reference frequency". This statistic indicates the number of inefficient units that the efficient unit is assigned to as a peer, and is displayed in Table 3.3.:

Table 3.3. Reference counts for the efficient units in the CCR model.

Market Sales Flag	City	Station Code	Reference Count
Market Sales are Correct	BURSA	118	86
Market Sales are Correct	İSTANBUL	130	71
Market Sales are Correct	BURSA	116	56
Market Sales are Correct	İSTANBUL	131	51
Market Sales are Correct	KOCAELİ	76	38
Market Sales are Correct	KOCAELİ	111	27
Market Sales are Correct	TEKİRDAĞ	72	18
Market Sales are Correct	İSTANBUL	5	8
Market Sales are not Correct	İSTANBUL	4	2

Correlation graphs (input/output factors vs efficiency scores) are plotted. Figure 3.2. displays a sample correlation graph:

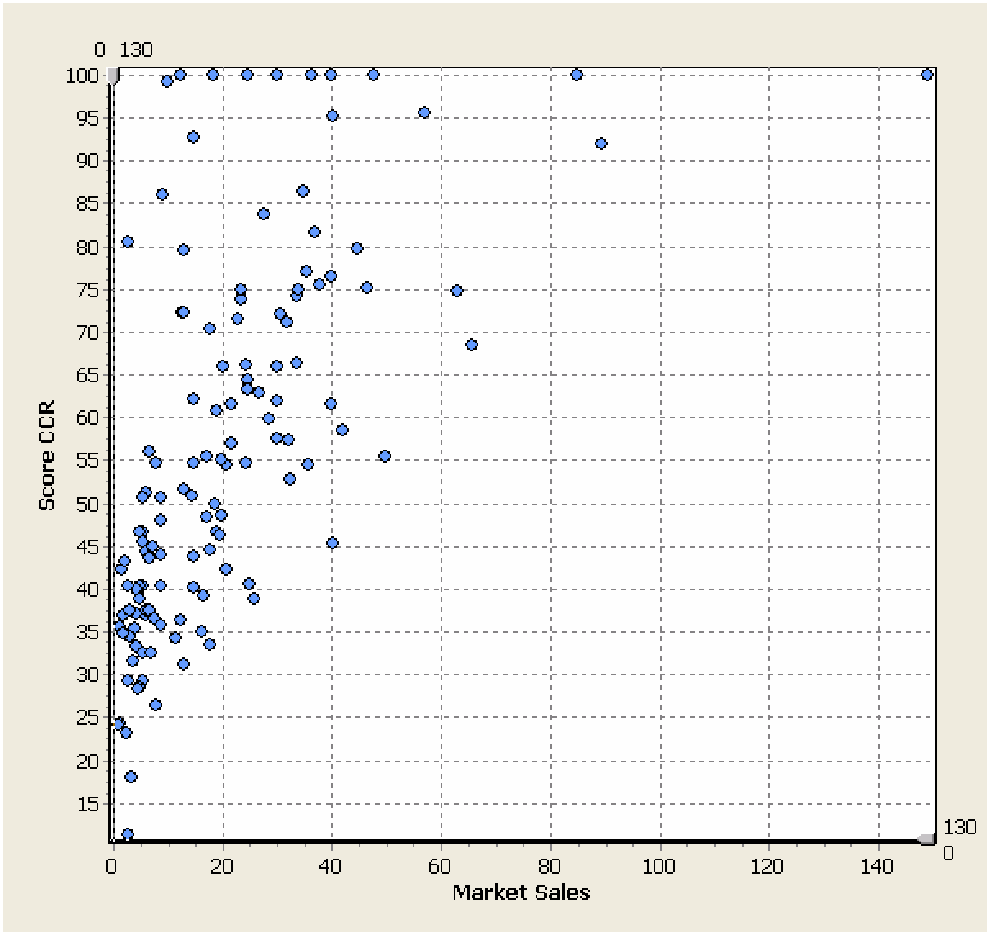


Figure 3.2. Market Sales vs. the CCR Efficiency score; each point represents a DMU

The correlation of efficiency scores to individual inputs and outputs are displayed in Table 3.4.:

Table 3.4. Correlation of efficiency score to input/output variables in the CCR model.

Input/Output	**Correlation**
Number of Pumps	0,33
Average Number of Pumpers	0,34
Rent Expense	0,33
Benzene Sales	0,39
Diesel Sales	0,54
Market Sales	0,63

These results indicate that efficiency scores are most highly correlated to diesel sales and market sales. However, even in these cases the correlation is not a very strong one.

The DMUs are further analyzed according to their cities. Stations are grouped by cities and correlation values calculated separately, as displayed in Table 3.5. (Only cities which have 3 or higher number of stations are considered):

Table 3.5. Correlation of efficiency score to input/output factors in the CCR models, grouped by cities.

	Number of Stations	Number Of Pumps	Average Number Of Pumpers	Rent Expense	Benzene Sales	Diesel Sales	Market Sales
BOLU	5	0,28	-0,81	-0,69	-0,66	-0,55	0,66
BURSA	13	0,35	0,61	0,09	0,22	0,55	0,71
ESKİŞEHİR	6	0,26	0,31	0,48	0,44	0,41	0,92
İSTANBUL	58	0,22	0,21	0,33	0,45	0,48	0,59
KOCAELİ	16	0,22	0,41	0,16	0,39	0,60	0,79
KÜTAHYA	3	0,62	-0,05	0,47	0,65	0,36	0,97
SAKARYA	9	0,49	0,39	0,21	-0,73	0,84	0,81
TEKİRDAĞ	7	0,23	0,17	0,20	0,49	0,72	0,64
ZONGULDAK	3	-0,91	-1,00	-0,99	-0,99	-0,96	-0,93

As it can be seen in Table 3.4, correlations differ between cities. In fact, no common pattern exists at all. When the three cities with highest number of stations (İstanbul, Kocaeli and Bursa) are considered, it is observed that the results are roughly parallel and the market sales has the highest correlation. But for Bursa, Average Number of Pumpers is also highly correlated, whereas Benzene Sales is significantly less correlated when compared to other cities. These results indicate that different patterns exist in different cities, in spite of similar market conditions (dense population, high number of motor vehicles).

The arithmetic mean of efficiency scores of stations of each cities is calculated and displayed in Table 3.6:

Table 3.6. Arithmetic mean of efficiency scores grouped by cities in the CCR model.

	Number of Stations	Aritmetic Mean of Efficiency Scores
BİLECİK	1	81,6
BURSA	13	65,22
İSTANBUL	58	60,24
TEKİRDAĞ	7	57,92
All Cities	**130**	**55,91**
KOCAELİ	16	52,64
DÜZCE	2	52,01
ESKİŞEHİR	6	50,97
KÜTAHYA	3	50,24
SAKARYA	9	48,36
BOLU	5	48,23
EDİRNE	2	47,98
ZONGULDAK	3	39,86
YALOVA	2	38,94
BARTIN	1	24,26
KARABÜK	2	21,05

Except for Bilecik, from where only a single station is considered in this study, cities with higher vehicle inventory have higher average scores. Note that the number of stations for these cities are also higher. It can be concluded that the stations operating in urban areas are working more efficiently.

As mentioned in section 3.2., market sales information is not reliable for some of the stations. It is observed that the stations with reliable market sales information perform better in general:

Table 3.7. Arithmetic mean of efficiency scores grouped by market sales flag.

	Number of Stations	Aritmetic Mean of Efficiency Scores
Market Sales are Reliable	57	65,30
All Stations	**130**	**55,91**
Market Sales are not reliable	73	48,57

As a further study, only the set of stations with reliable market sales are investigated separately and same analyzes are applied. The scores distribution obtained is displayed in Figure 3.3.:

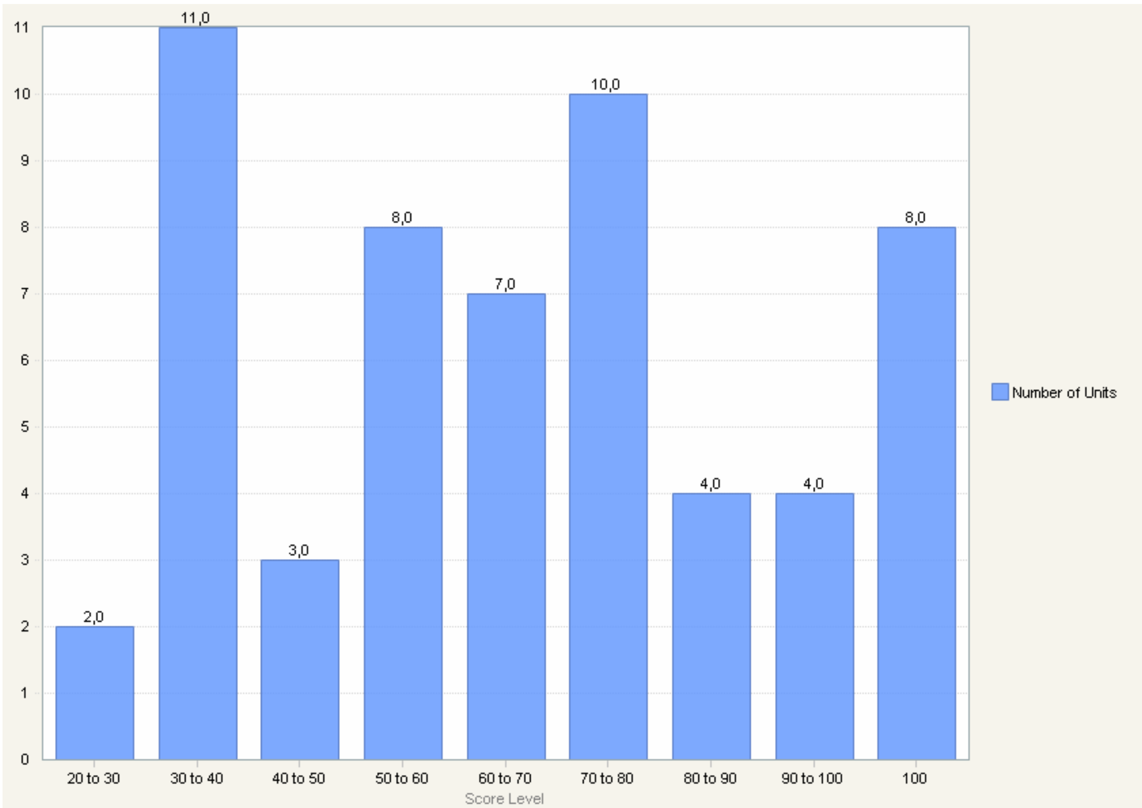


Figure 3.3. Efficiency score distribution of stations with reliable market sales (the CCR model).

The number of efficient units is 8 out of 57 units. Recall that in the original case, efficiency scores of stations are cumulated between 30 and 80. In this case, this accumulation still holds, but is somewhat shifted to the right. In other words, the efficiency scores of stations with reliable Market Sales figures are, in general, higher (maybe this can be regarded as an indication of better organization and procedures contributing to the overall efficiency).

Correlations of efficiency scores to input/output factors are also analyzed. The result are displayed in Table 3.8.:

Table 3.8. The correlation of input/output factors to efficiency scores for the stations with reliable market sales

Input/Output	Correlation
Number of Pumps	0,22
Average Number of Pumpers	0,36
Rent Expense	0,32
Benzene Sales	0,35
Diesel Sales	0,55
Market Sales	0,59

The correlation results are similar with the original model. Diesel Sales and Market Sales still have the highest correlation with efficiency scores. The importance of number of pumps decreased in this instance.

3.4. The Application of the BCC DEA Model

Next, the BCC Model of the DEA is applied. The complete result set is given in Appendix B. Efficiency distribution plot for this model is displayed in Figure 3.4.:

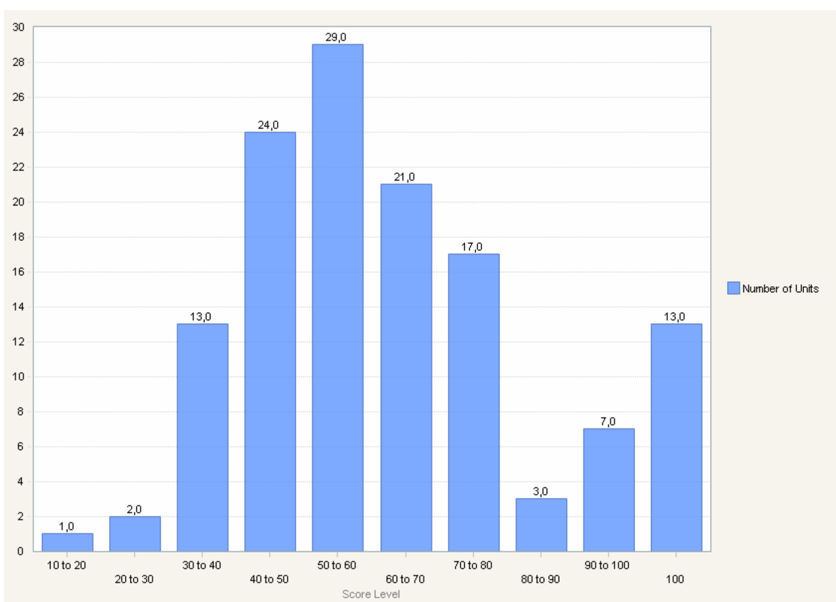


Figure 3.4. Efficiency scores distribution for the BCC model.

Note that, in this case, 13 out of the 130 DMUs turned to be efficient. This is higher than the number of efficient units in the CCR model. Also, it can be the observed that the distribution of efficiency scores has shifted to right, indicating an increase in efficiency scores, in general. This is not surprising, since the CCR model assumes constant returns to scale, whereas the BCC model assumes variable returns to scale (thereby, assigning higher efficiency scores to many smaller DMU's, who are disadvantaged by size alone).

As it can be seen in Table 3.9., the correlations between input/output factors and efficiency scores are investigated. In general, the correlations are lower (then the CCR model case), especially for the inputs:

Table 3.9. Correlation of the efficiency scores to input/output factors in the BCC model

Input/Output	Correlation
Number of Pumps	0,15
Average Number of Pumpers	0,18
Rent Expense	0,17
Benzene Sales	0,31
Diesel Sales	0,41
Market Sales	0,49

Similar to the case of the CCR model, reference frequencies are investigated and displayed in Table 3.10.:

Table 3.10. Reference counts for the efficient units in the BCC model

Market Sales Flag	City	Station Code	Reference Count
Market Sales are not Correct	İSTANBUL	4	86
Market Sales are Correct	BURSA	118	80
Market Sales are Correct	İSTANBUL	130	75
Market Sales are Correct	BURSA	116	63
Market Sales are Correct	BOLU	13	47
Market Sales are Correct	İSTANBUL	131	40
Market Sales are Correct	KOCAELİ	111	23
Market Sales are not Correct	İSTANBUL	117	19
Market Sales are Correct	KOCAELİ	76	18
Market Sales are Correct	TEKİRDAĞ	72	13
Market Sales are Correct	İSTANBUL	5	5

Note that best performers in the BCC model coincide with the ones in the CCR model. This fact is more clearly seen in the Figure 3.5.:

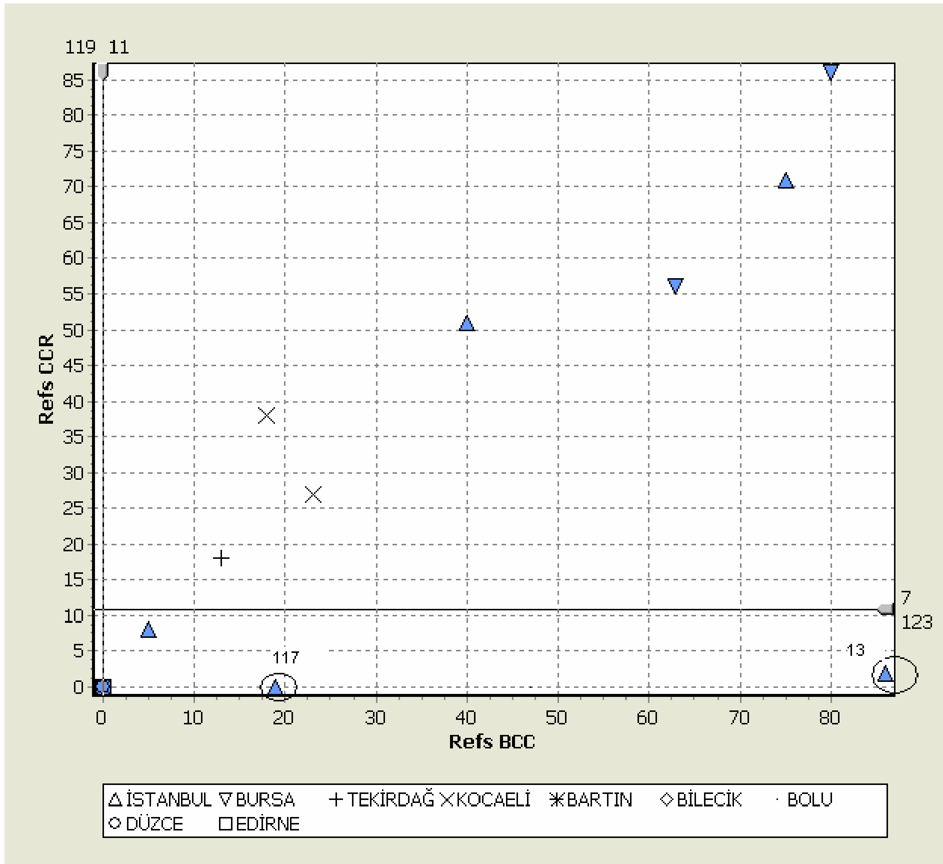


Figure 3.5. The CCR vs the BCC models' reference counts for efficient units.

In the BCC model, there are two additional best performer units: 13 and 117. (These units are placed on the x axis, having zero value on the y axis). This can also be observed in the Table 3.11.:

Table 3.11. The CCR and the BCC models' reference stations and their counts

Station Code	Refs BCC	Refs CCR
4	86	2
5	5	8
13	47	
72	13	18
76	18	38
111	23	27
116	63	56

117	19	
118	80	86
130	75	71
131	40	51

Additionally, efficient units in the CCR and the BCC models are compared:

Table 3.12. Efficient units in the CCR and the BCC models.

Station Code	**Score BCC**	**Score CCR**
1	100	24,26
2	100	51,34
4	100	100
5	100	100
13	100	85,96
72	100	100
76	100	100
111	100	100
116	100	100
117	100	99,17
118	100	100
130	100	100
131	100	100

It can be seen that two bad performing units in the CCR model (stations 1 and 2) turn out to be efficient in the BCC model. This shows that the CCR and the BCC efficiency scores must be further compared. (This study is conducted in section 3.5., regarding the Scale Efficiency)

The arithmetic mean of efficiency scores are analyzed according to cities, and the results are displayed in Table 3.13.:

Table 3.13. The Arithmetic mean of efficiency scores in the BCC model, grouped by cities.

	Number of Stations	**Average Score in the BCC**
BARTIN	1	100
BİLECİK	1	81,72
BURSA	13	69,78
İSTANBUL	58	65,17
TEKİRDAĞ	7	64,44
BOLU	5	63,11
All Cities	**130**	**62,15**
EDİRNE	2	61,65
YALOVA	2	56,73
KOCAELİ	16	56,65
SAKARYA	9	56,12
ESKİŞEHİR	6	54,95
KÜTAHYA	3	53,3
DÜZCE	2	52,94
ZONGULDAK	3	47,71
KARABÜK	2	28,64

All outperforming cities of the CCR model are also outperforming in the BCC model. Bolu, Yalova and Bartın shifted up considerably. It can be seen that the results obtained in the CCR model are in line with the BCC model. This fact can clearly be seen in Figure 3.6.:

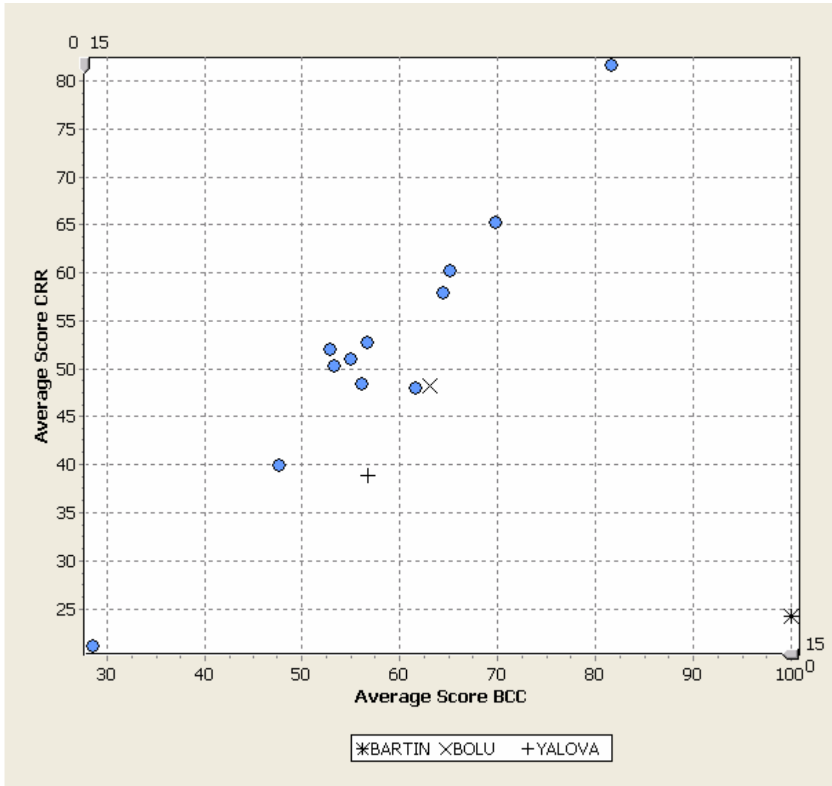


Figure 3.6. The arithmetic mean of efficiencies in the CCR and the BCC models grouped by cities.

Set of stations with reliable market sales are also analyzed separetely. The efficiency score distribution obtained, is displayed in Figure 3.7.:

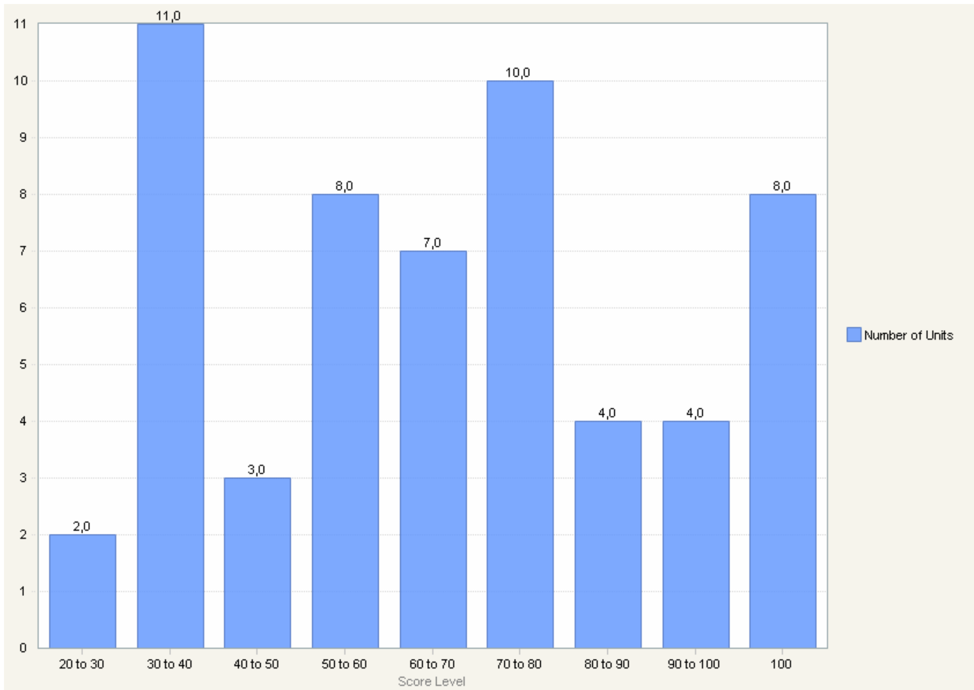


Figure 3.7. Efficiency Score distribution of stations with reliable Market Sales for the BCC model.

The number of efficient units in this case is 12. When the unrestricted (i.e. full data set) and the restricted (i.e. data set for stations having reliable market sales) BCC models are compared, it can be seen that the accumulation between efficiency scores of 30 and 80 is softened and distribution of the efficiency scores is shifted to the right. These results are in parallel with the results obtained in the CCR model case.

Correlations of input/output factors and efficiency scores are also analyzed for this case and the results are displayed in Table 3.14.:

Table 3.14. Correlation of input/output factors to efficiency scores in the BCC model of the stations with reliable market sales

Input/Output	Correlation
Number of Pumps	-0,06
Average Number of Pumpers	0,05
Rent Expense	0,11
Benzene Sales	0,24
Diesel Sales	0,37
Market Sales	0,38

In general, the correlations between the input/output factors and the efficiency score are lower in this case. Number of pumps, average number of pumpers turned to be uncorrelated to the efficiency score. Diesel Sales and Market Sales are again the factors having the highest correlation with efficiency score.

3.5. The Scale Efficiency

Noting that the CCR and the BCC models work under Constant and Variable returns to scale assumptions respectively, a scale efficiency analysis can be conducted. By definition:

$SE= TE_{CRS}/ TE_{VRS}$

In the fuel stations problem, pure technical efficiency (TE_{VRS}) can be interpreted as managerial success, whereas Scale Efficiency reveals whether the stations are operating at appropriate sizes.

Using the definition above, the results obtained in the BBC and the CCR models are further investigated regarding scale efficiency. (Complete set of results for scale efficiencies can be seen in Appendix C). Average scale efficiencies of the stations according to their cities are displayed in Table 3.15.:

Table 3.15. Scale efficiencies of stations grouped by cities

	Number of Stations	Scale Efficiency
BİLECİK	1	1
DÜZCE	2	0,98
KÜTAHYA	3	0,94
BURSA	13	0,93
KOCAELİ	16	0,93
ESKİŞEHİR	6	0,93
İSTANBUL	58	0,92
All Cities	**130**	**0,9**
TEKİRDAĞ	7	0,9
SAKARYA	9	0,86
ZONGULDAK	3	0,84
EDİRNE	2	0,78
BOLU	5	0,76
KARABÜK	2	0,73
YALOVA	2	0,69
BARTIN	1	0,24

The overall score of scale efficiency is equal to 0,9, which is quite high. Thus, it can be concluded that:

1. Inefficiencies mainly result from pure technical efficiency (TE_{VRS}).
2. Greater scale (i.e. physical/human resources), does not provide significant competitive advantage.

Scale efficiency can also be analyzed through Figure 3.8.:

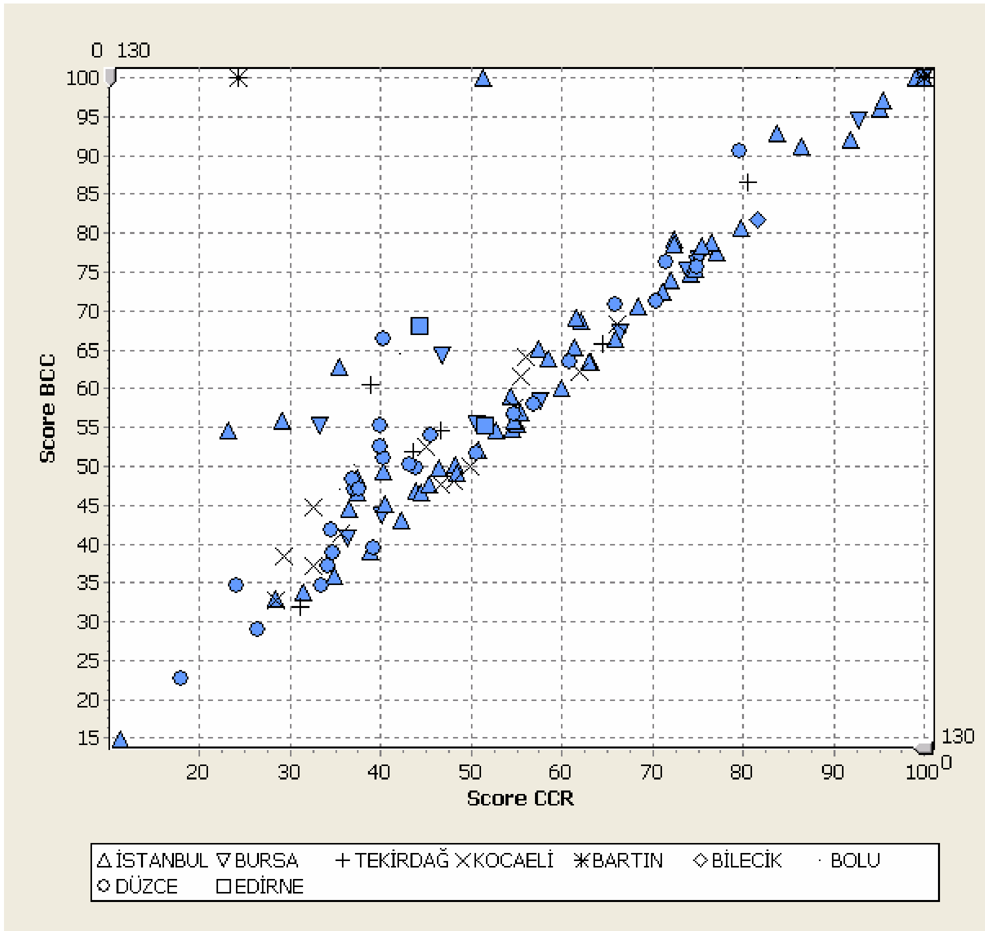


Figure 3.8. The CCR vs the BCC efficiency scores, each point representing a unit.

It can clearly be seen in Figure 3.8. that, in general, the BCC efficiency scores are parallel to the CCR efficiency scores. However, even visually it can be seen that a group of stations turn to be scale inefficient. City is not a distinguishing factor in this situation. It is necessary to further investigate this issue. The most scale inefficient units are displayed in Table 3.16.:

Table 3.16. List of most scale inefficient stations.

City	Station code	Number Of Pumps	Average Number Of Pumpers	Rent Expense	Benzene Sales	Diesel Sales	Market Sales	Scale Efficiency
BARTIN	1	4	4	1.200	61,55	102,35	1,23	0,24
İSTANBUL	24	5	3	4.200	103,32	316,08	2,53	0,43
İSTANBUL	2	6	4	1.100	67,29	98,48	5,97	0,51
İSTANBUL	11	7	4	1.300	70,89	140,56	2,63	0,52
İSTANBUL	17	6	4	1.525	121,54	177,14	3,81	0,56
BURSA	15	6	4	1.529	53,28	198,2	4,21	0,6
YALOVA	18	7	3	1.450	103,12	202,31	4,93	0,61

These units are further analyzed; it is observed that:

- These stations are among the lowest 29 out of 130, when ranked according to number of pumps.
- These stations are among the lowest 33 out of 130, when ranked according to average number of pumpers
- Except for station 24, these stations are among the lowest 12 out of 130, when ranked according to rent expense.

Therefore, the second sentence of the remark above must be rephrased as, “Greater scale does not provide much competitive advantage, after reaching a certain level, which can roughly be defined as 8 pumps, 1.600 rent expense and 5 pumpers ”. This might be a valuable piece of information when investing into a new station. Managerially, these results can be used as minimum levels, so that new stations don’t start the race one step behind.

Table 3.17 displays the correlation between input/output factors and the scale efficiencies. There is no proof of strong correlation between scale efficiency and input/output factors.

Table 3.17. The correlation between the scale efficiency and the input/output factors.

	Correlation to Scale Efficiency
Number Of Pumps	0,534
Average Number Of Pumpers	0,501
Rent Expense	0,474
Benzene Sales	0,286
Diesel Sales	0,458
Market Sales	0,494

3.6. The Assurance Regions Models

In sections 3.3., 3.4. and 3.5., the very basic approaches of Data Envelopment Analysis are used. In these approaches, there have been no restrictions on the decision variables. In a lot of studies, it is observed that some units turn out to be efficient by emphasizing just on a single variable. Note that technically nothing is wrong in this approach. However in practice this is not desired when interpreting the results. In other words, it is not realistic to assign very high (or low) weights to input/output variables. Therefore it may be necessary to apply the assurance region methodology described in section 2.1.4.5. Also assurance region methodology lowers the disturbance caused by false market sales, by restricting weights on variables.

Number of stations that are weighted with one input and one output are displayed in Table 3.18. Note that inputs and outputs are considered separately:

Table 3.18. Number of units that gain non zero weight for only one input and output variable

	CCR	BCC
Average Number of Pumpers	1	1
Number of Pumps	1	6
Rent Expense	37	25
Benzene Sales	0	2
Diesel Sales	11	21
“Market Sales	9	10

When only efficient units are considered, the results displayed in Table 3.19. are obtained:

Table 3.19. Number of the efficient units that gain non zero weight for only one input and output variable

	CCR	BCC
Average Number of Pumpers	0	0
Number of Pumps	0	1
Rent Espense	1	6
Benzene Sales	0	1
Diesel Sales	0	2
Market Sales	3	1
Total Number of Efficient Units	*9*	*13*

Likewise, number of units that ignore input/output variables is displayed in Table 3.20. (i.e. weights of these inputs/outputs being less than 0.05):

Table 3.20. Number of the DMUs that ignore the input/output variable

	CCR	BCC
Number of Pumps	81	57
Number of Pumpers	71	73
Rent Expense	2	7
Benzene Sales	68	73
Diesel Sales	11	19
Market Sales	54	71

These results constitute another proof for unrealistic weight assignments. Thus assurance regions method is applied.

In order to apply the assurance region methodology, a further analysis about ratios of input and output weights is conducted. In this study, weights of outputs are restricted by weights of other outputs and weights of inputs are restricted by weights of other inputs. Therefore, 4 ratios are calculated:

- Number of Pumps/ Average Number of Pumpers
- Rent Expense/ Average Number of Pumpers
- Benzene Sales/ Market Sales
- Diesel Sales/ Market Sales

Then extreme values of these weights are eliminated and the Mean and the Standard Deviation values for these ratios are calculated. Note that the CCR and the BCC models are treated separately:

Table 3.21. Mean and standard deviation values for input and output weights.

		Number of Pumps/ Average Number of Pumpers	Rent Expense/ Average Number of Pumpers	Benzene Sales/ Market Sales	Diesel Sales/ Market Sales
CCR	Mean	0,551	18,943	1,034	0,735
	St. Dev.	0,397	21,501	1,304	0,934
BCC	Mean	2,828	9,485	1,911	0,704
	St. Dev.	4,204	6,610	2,936	0,580

Note that even after eliminating extreme values, standard deviation values are high. Therefore the following approach is applied in order to restrict the weights (u_i, v_i) obtained in the DEA runs:

1. Run1: All ratios given are restricted to a window of $[\mu - \sigma/2, \mu + \sigma/2]$. Size of the window is σ
2. Run2: All ratios given are restricted to a window of $[\mu - \sigma/4, \mu + \sigma/4]$. Size of the window is $\sigma/2$

The CCR and the BCC values are treated separately. When comparing the CCR and the BCC results for scale efficiency, Run1 of the CCR is compared to Run1 of the BCC and Run2 of the CCR is compared to Run2 of the BCC. Unrestricted runs completed in previous sections are named Run0 throughout this section.

Upper and lower bounds are displayed in Table 3.22.:

Table 3.22. Upper and lower bounds for assurance regions.

		Number of Pumps/ Average Number of Pumpers	**Rent Expense/ Average Number of Pumpers**		**Benzene Sales/ Market Sales**	**Diesel Sales/ Market Sales**
CCR Run1	Upper bound	0,749	29,693		1,686	1,201
	Lower Bound	0,352	8,192		0,382	0,268
CCR Run2	Upper bound	0,650	24,318		1,360	0,968
	Lower Bound	0,451	13,567		0,708	0,501
BCC Run1	Upper bound	4,930	12,789		3,379	0,994
	Lower Bound	0,727	6,180		0,443	0,414
BCC Run2	Upper bound	3,879	11,137		2,645	0,849
	Lower Bound	1,778	7,832		1,177	0,559

The data is analyzed using the EMS software, keeping precision up to third digit. Number of efficient units obtained in each run are displayed in Table 3.23. Number of efficient units decreased:

Table 3.23. Number of efficient units in the DEA runs.

	CCR	BCC
Run 0	9	13
Run 1	4	5
Run 2	3	4

Arithmetic mean of efficiency scores also decreased, as displayed in Table 3.24.:

Table 3.24. Arithmetic mean of efficiency scores obtained in the DEA runs.

	CCR	BCC
Run 0	55,91	62,15
Run 1	46	58,6
Run 2	43,9	56,3

Efficiency scores are analyzed according to their cities. It can be shown that there exist different patterns. Consider the CCR problem:

Table 3.25. Arithmetic means of efficiency scores grouped by cities, on the CCR model runs of the DEA.

	Number of Stations	Run 0	Run 1	Run 2
All cities	**130**	**55,91**	**46**	**43,9**
BİLECİK	1	81,6	29,3	28,6
BURSA	13	65,22	55	52,7
İSTANBUL	58	60,24	47,4	46,3
TEKİRDAĞ	7	57,92	43,5	42,6
KOCAELİ	16	52,64	48,2	45,5
DÜZCE	2	52,01	44,9	39,7
ESKİŞEHİR	6	50,97	36,8	35,3
KÜTAHYA	3	50,24	37,7	35,4

SAKARYA	9	48,36	45,3	42,1
BOLU	5	48,23	39,6	37,9
EDİRNE	2	47,98	46,1	43,4
ZONGULDAK	3	39,86	32,1	31,3
YALOVA	2	38,94	37,7	37,7
BARTIN	1	24,26	24,1	24,1
KARABÜK	2	21,05	19,9	19,9

There is a very sharp decrease at Bilecik, which has one station. Scores drop in all of the cities, but not sharp as that. Also note that rate decrease is higher when going from run0 to run1.

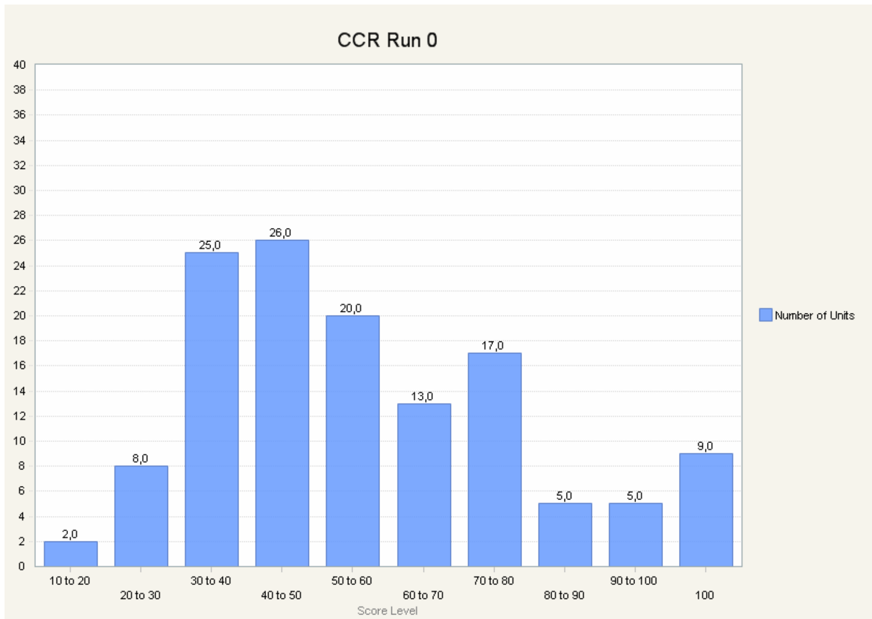
When the BCC problem is considered a different case occurs. Both decreases and increases in efficiency ratios are observed, as displayed in Table 3.26.:

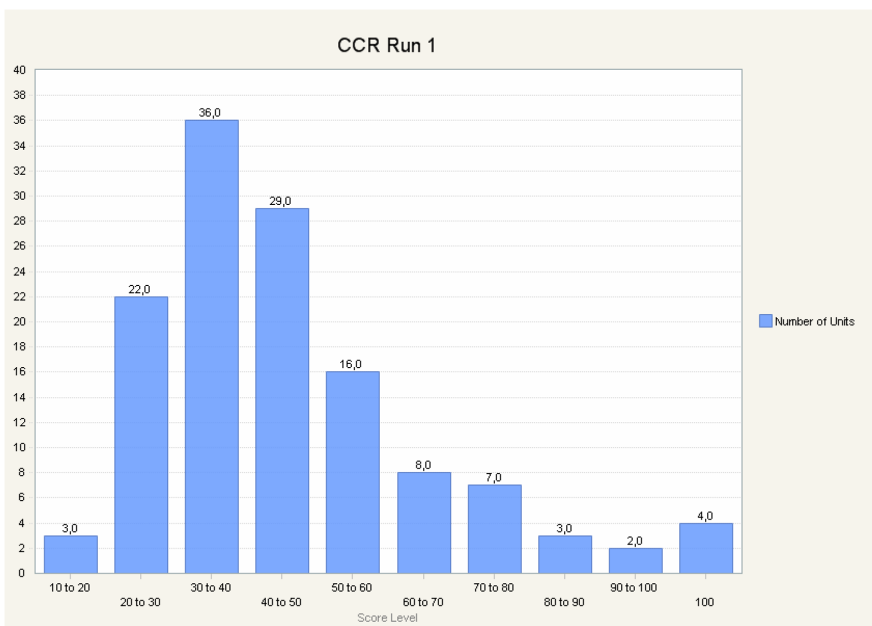
Table 3.26. Arithmetic means of efficiency scores grouped by cities, in the BCC model runs of the DEA.

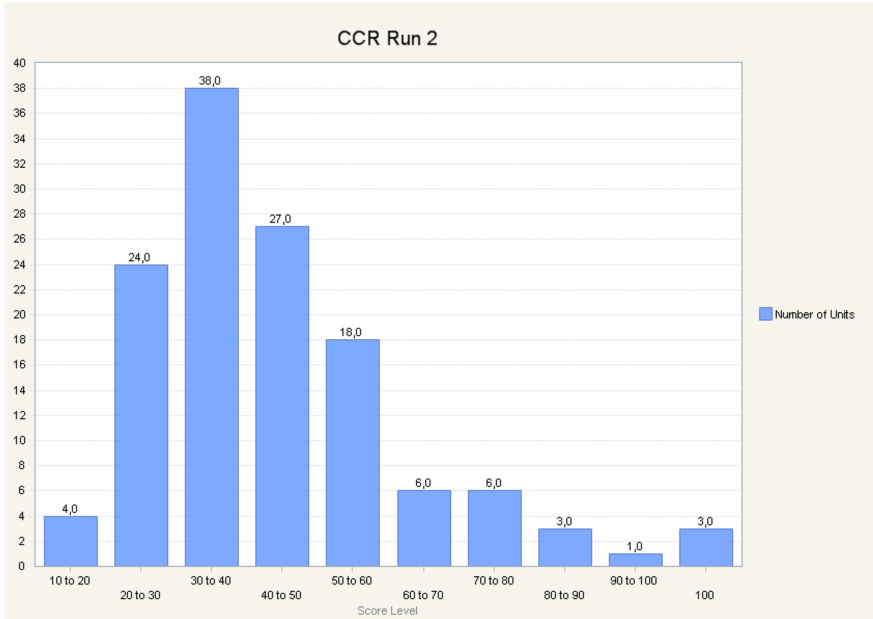
	Number of Stations	Run 0	Run 1	Run 2
All Cities	**130**	**62,15**	**58,6**	**56,3**
BARTIN	1	100	83,6	83,5
BİLECİK	1	81,72	34,9	34,1
BURSA	13	69,78	65,1	61,5
İSTANBUL	58	65,17	56,8	55,2
TEKİRDAĞ	7	64,44	59,4	58,3
BOLU	5	63,11	71,2	70
EDİRNE	2	61,65	65,7	62,8
YALOVA	2	56,73	68,7	68,5
KOCAELİ	16	56,65	60,7	56,8
SAKARYA	9	56,12	61,3	56,7
ESKİŞEHİR	6	54,95	45,9	43,6
KÜTAHYA	3	53,3	53,2	51,2
DÜZCE	2	52,94	46,9	39,3
ZONGULDAK	3	47,71	57,4	56,8
KARABÜK	2	28,64	48,5	48

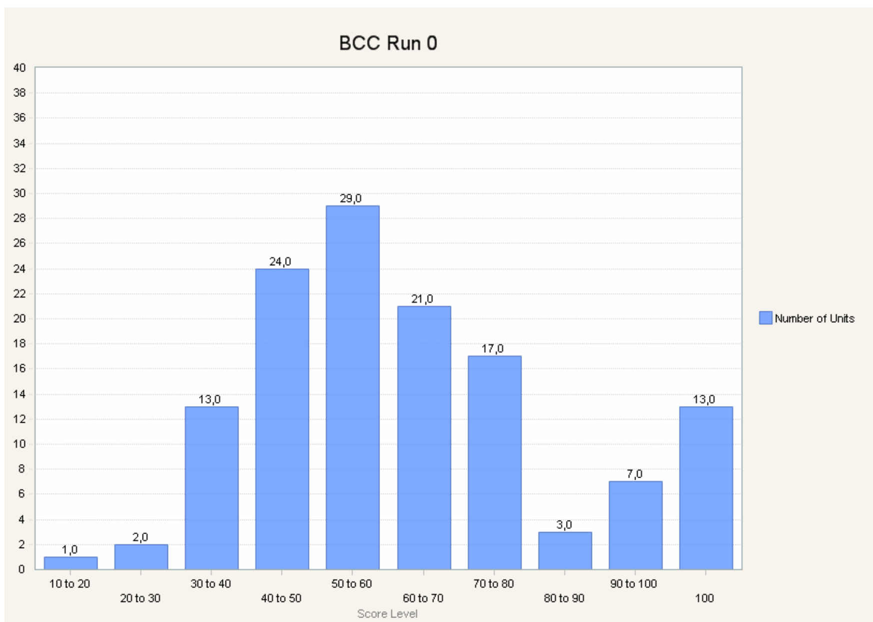
Efficiency scores for Istanbul, for example, drop resulting at a below average score. On the other hand increases are observed for Zonguldak, Karabuk, Bolu and Yalova.

Distribution of efficiency scores is analyzed in Figure 3.9.:






CCR Run 2
40
38
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
4,0
24,0
38,0
27,0
18,0
6,0
6,0
3,0
1,0
3,0
10 to 20
20 to 30
30 to 40
40 to 50
50 to 60
60 to 70
70 to 80
80 to 90
90 to 100
100
Score Level
Number of Units


BCC Run 0
40
38
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
1,0
2,0
13,0
24,0
29,0
21,0
17,0
3,0
7,0
13,0
10 to 20
20 to 30
30 to 40
40 to 50
50 to 60
60 to 70
70 to 80
80 to 90
90 to 100
100
Score Level
Number of Units

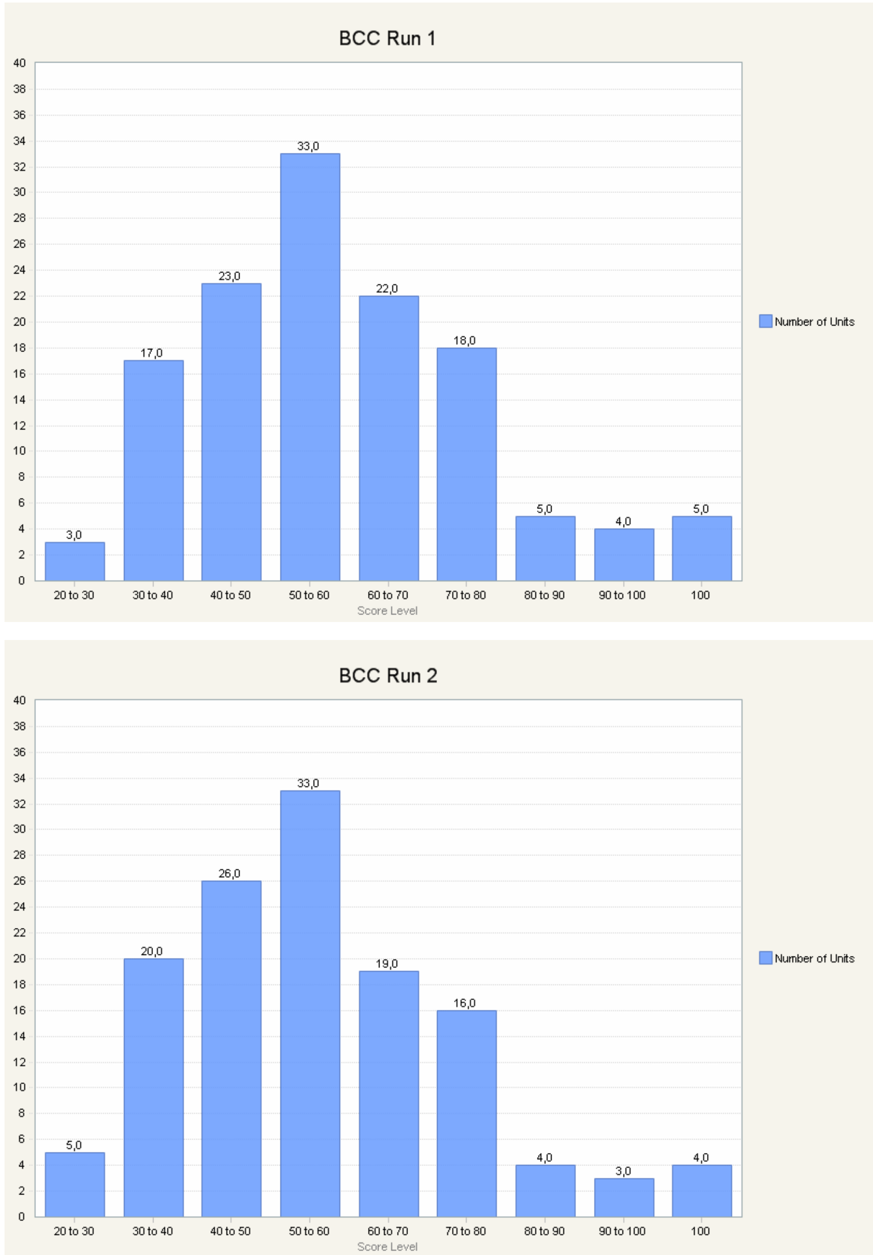


Figure 3.9. Efficiency score distributions for assurance region models.

Figure 3.9 indicates that as weights are restricted, DMUs tend to get lower efficiency scores.

Correlations of efficiency scores to input/ output variables also analyzed and the results displayed in Table 3.27. are obtained:

Table 3.27.Correlation of efficiency scores to input/output variables in the DEA runs

	Score BCC	Score BCC Run1	Score BCC Run2	Score CCR	Score CCR Run1	Score CCR Run2
Average Number Of Pumpers	0,18	-0,02	-0,06	0,34	0,50	0,52
Number Of Pumps	0,15	0,00	-0,04	0,33	0,54	0,55
Rent Expense	0,17	-0,15	-0,16	0,33	0,36	0,38
Benzene Sales	0,31	0,27	0,32	0,39	0,51	0,56
Diesel Sales	0,41	0,21	0,12	0,54	0,68	0,67
Market Sales	0,49	0,15	0,14	0,63	0,53	0,55

A further analysis can be done about scale efficiency. However, when the assurance regions method is applied, the very formulation of scale efficiency (equation 2.18) used in section 3.5 can not be applied. The CCR and the BCC formulations are treated separately when the windows are described. Therefore SE>1 can be obtained. Consider station number 118. At the CCR Run 1, station obtained an efficiency score of 100 whereas at the BCC Run 1 the result was 97,6. Nevertheless, affect of scale can be observed by other methods.

In order to investigate affect of scale, Run1 results for the BCC and the CCR are sketched:

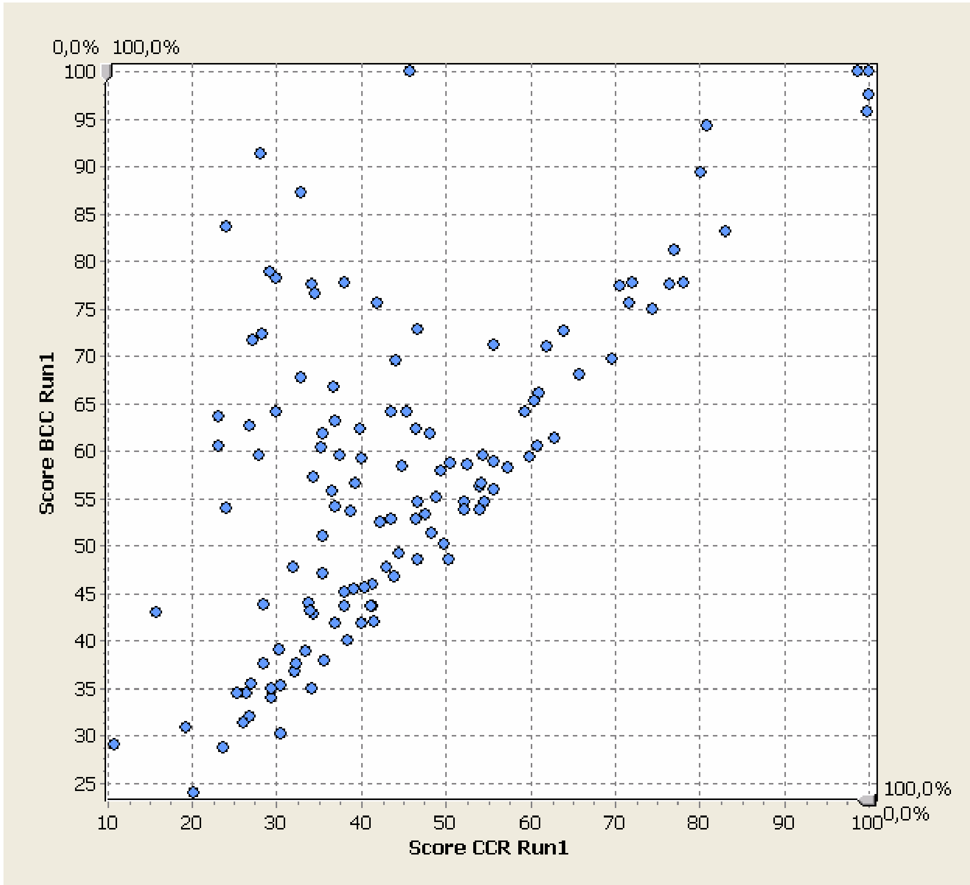


Figure 3.10. The CCR Run1 vs the BCC Run1 efficiency scores, each point represents a unit.

Note that general situation observed in section 3.5. still holds. Affect of scale exits but there are two different patterns: For most of the stations this is low. Bu some of the DMUs deviate significantly from this pattern. By picking the deviating stations, it can be seen that these stations belong to the lowest 25% of the stations. This situation holds for run 2 results. Therefore, although it is not possible directly calculate scale efficiencies, it can be stated that the results obtained in section 3.5. are still valid; there is threshold level of station size that must be reached in order not to be at a disadvantageous scale.

3.7. Regression Analysis

One further analysis that can be conducted on the DEA model results is regression analysis. In this study, dependent variable is the efficiency score, whereas the independent variables are the input and output variables. Consider the functional form assuming linearity:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

where y is the efficiency score x are the variables (both input and output), β_i is the coefficient for variable i and β_0 is the constant term. Corresponding variables are:

i	1	2	3	4	5	6
Corresponding variable	Average Number Of Pumpers	Number Of Pumps	Rent Expense	Benzene Sales	Diesel Sales	Market Sales

Instead of using the input/output values directly, normalized values are used: For each of the input/output variables, maximum levels are fixed to level 100 and all other values are scaled accordingly. Therefore, new levels for input/output variables are obtained:

$X_{in\ normalised}=100*x_{in}/x_{i\ max}$

where $x_{i\ max}= \max\{x_{in}\ n=1..N\}$ for each of the input/output variables.

Therefore, variables for the regression analysis can be rewritten as:

i	1	2	3	4	5	6
Corresponding variable	Normalized (Average Number Of Pumpers)	Normalized (Number Of Pumps)	Normalized (Rent Expense)	Normalized (Benzene Sales)	Normalized (Diesel Sales)	Normalized (Market Sales)

Note that the results of the regression analysis do not change completely, since for any of the variables, new level of values are obtained by multiplying the previous level with a constant. Accordingly, in normalized and non-normalized models, same variables turn out to be significant. However, coefficient values are different and results of the normalized model are valuable for interpreting the effects of changes in the input/output levels.

Let's consider the CCR formulations result and using Least Squares method to obtain the coefficients:

Table 3.28. Regression coefficients and standard errors for the CCR model.

	Normalized (Market Sales)	**Normalized (Diesel Sales)**	**Normalized (Benzene Sales)**	**Normalized (Rent Expense)**	**Normalized (Number Of Pumps)**	**Normalized (Average Number Of Pumpers)**	**Constant Term**
Means	1,97	0,80	-0,60	-1,02	0,17	-0,25	45,49
Standart errors	0,20	0,14	0,19	0,17	0,14	0,16	4,44

An R^2 value of 0,64 is obtained, indicating that regression model is helpful in explaining the y value (i.e. efficiency score)

Also the ANOVA table in Table 3.29 is obtained:

Table 3.29. ANOVA table for regression analysis for the CCR model.

Source	**Sum of Squares**	**Df**	**Mean Square**	**F**
Regression	38.842,23	6	6.473,71	36,66
Residual	21.718,47	123	176,57	
Total	**60.560,7**	**129**		

First it is necessary to check the significance of model. Formally the following hypothesis test is conducted:

$$H_O : \beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$$

$$H_A : \text{ At least one } \quad \beta_k \neq 0$$

F test is conducted. When $F_{6,123}$ is observed, probability of occurrence of 36,66 or more is found out to be 3,38746E-25, which is extremely small.

Therefore the null hypothesis rejected, i.e. at least one of the coefficients is significant.

Accordingly, individual coefficients must be analyzed for significance. Therefore the following hypothesis test is conducted for each of the coefficients:

$H_O \ \beta_j = 0$

$H_A \ \beta_j \neq 0$

The t test is used for this purpose. t- observed values are obtained by dividing the mean values by standard error values, as displayed in Table 3.30.:

Table 3.30. t-observed values for regression coefficients in the CCR model

	Normalized (Market Sales)	**Normalized (Diesel Sales)**	**Normalized (Benzene Sales)**	**Normalized (Rent Expense)**	**Normalized (Number Of Pumps)**	**Normalized (Average Number Of Pumpers)**
T values	10,06	5,57	-3,08	-6,00	1,22	-1,54

Assuming %5 precision and remembering that this is two tailed test,

$t_{0,025,\ 123}$= 2,269172.

Rejection criteria is $\left|t_{observed}\right| \geq t_{0,025,123}$. Therefore, it can be concluded that the number of pumps and the average number of pumpers are not significant and the other four variables are significant. Regression is re run after eliminating the insignificant variables:

Table 3.31. Final results for the coefficients in regression analysis for the CCR model

	Normalized (Market Sales)	Normalized (Diesel Sales)	Normalized (Benzene Sales)	Normalized (Rent Expense)	Constant Term
Means	1,92	0,79	-0,56	-1,09	45,68

The results imply that:

- 1 unit (1% of maximum level) *increase* in market sales results in 1,92 *increase* in efficiency score and 1 unit *decrease* results in 1,92 *decrease*
- 1 unit (1% of maximum level) *increase* in diesel sales results in 0,79 *increase* in efficiency score and 1 unit *decrease* results in 0,79 *decrease*
- 1 unit (1% of maximum level) *increase* in benzene sales results in 0,56 *decrease* in efficiency score and 1 unit *decrease* results in -0,56 *increase*
- 1 unit (1% of maximum level) *increase* in rent expense results in 1,09 decrease in efficiency score and 1 unit *decrease* results in 1,09 *increase*

It is interesting that one of the output variables, benzene sales, has negative effect on efficiency score. Recall that in section 3.3., benzene sales turn to be having lowest correlation with the efficiency score among the output variables. Moreover, market sales turn out to be an important factor.

Same regression analysis is conducted for the BCC formulation. R^2 is 0,55 and the ANOVA table in Table 3.32. is obtained:

Table 3.32. ANOVA table for regression analysis for the BCC model

Source	Sum of Squares	Df	Mean Square	F
Regression	28.512,28	6	4.752,05	24,80
Residual	23.573,04	123	191,65	
Total	**52.085,32**	**129**		

Probability of such an F value is 4,08611E-19. Therefore, at least one of the coefficients is significant. Further hypothesis tests on t values are conducted:

Table 3.33. Regression coefficients and standard errors for the BCC model

	Norm. (Market Sales)	**Norm. (Diesel Sales)**	**Norm. (Benzene Sales)**	**Norm. (Rent Expense)**	**Norm. (Number Of Pumps)**	**Norm. (Average Number Of Pumpers)**	**Constant Term**
Means	1,65	0,83	-0,25	-1,18	-0,08	-0,14	65,11
Standard errors	0,20	0,15	0,20	0,18	0,15	0,17	4,62

Table 3.34. t-observed values for regression coefficients in the BCC model.

	Norm. (Market Sales)	**Norm. (Diesel Sales)**	**Norm. (Benzene Sales)**	**Norm. (Rent Expense)**	**Norm. (Number Of Pumps)**	**Norm. (Average Number Of Pumpers)**	**Constant Term**
t values	8,05	5,58	-1,25	-6,66	-0,51	-0,79	14,09

Recalling that rejection criteria is $|t_{observed}| \geq t_{0,025,123}$ and $t_{0,025,\ 123}$= 2,269172, it can be concluded that the number of pumps, the average number of pumpers and the benzene sales are insignificant whereas the other three variables are significant. Regression is re run after eliminating the insignificant variables:

Table 3.35. Final results for the coefficients in regression analysis for the BCC model.

	Market Sales	**Diesel Sales**	**Rent Expense**	**Constant Term**
Coefficient Estimate	1,49	0,78	-1,35	60,08

Similar coefficients with the CCR case are obtained. However, constant term is higher and market sales coefficient is lower, implying a lower overall effect of variables on regression score, compared to the CCR case.

4. THE STOCHASTIC PRODUCTION FRONTIERS MODEL

In this chapter, the implementation of the Stochastic Production Frontiers (SPF) methodology to the efficiency analysis of the fuel stations is discussed. Let us recall the stochastic production function described in section 2.1.1.1. :

$$Y_{it} = x_{it}\beta + (v_{it} - u_{it}) \tag{4.1.}$$

One requirement of the SPF is that there must be a single output variable, (i.e. multiple outputs are not allowed). In the fuel stations problem, there are three outputs: The Benzene sales, the diesel sales and the Market sales. It is necessary to aggregate this outputs reasonably into one single output.

It can be argued that since all of the three output values are monetary values resulting from different types of sales activities, they can directly be summed. However, 1 YTL of the Benzene Sale is not equivalent to 1 YTL of the Market Sales. Accordingly, it can further be argued that each of these variables must be multiplied with constants before summation, resulting in a "weighted sum of outputs".

In current problem, normalized levels of outputs obtained in section 3.7. are summed and aggregated into one single output called "Normalized Total Sales":

Normalized Total Sales= Normalized Benzene Sales+ Normalized Diesel Sales + Normalized Market Sales (4.2)

Also, there is no time dimension in the model and equation 4.1. can be rewritten as:

$$Y_i = x_i\beta + (v_i - u_i) \tag{4.3}$$

where

Y_i = ln (Normalized Total Sales $_i$),

x_i is the input vector for DMU i ,

vector β representing the set of coefficients.

Additionally, v_i , u_i represent the random noise and the technical inefficiency terms, respectively.

The SPF approach requires an assumption about the probability distribution of the inefficiency term. Since it is less risky not to stick to one distribution, two different distribution assumptions are made, resulting in two different functions: Half normal and truncated normal. Results of these models are then compared.

"Frontier 4.1c" software package is selected for solving the described SPF models. Frontier 4.1c follows a specific approach for fitting (i.e. estimating the parameters of) the production function. As described in Coelli (1996), a three step estimation procedure is applied:

1. Ordinary Least Squares (OLS) estimates of the function are obtained, where all β estimators, with the exception of the intercept are to be unbiased.
2. A two-phase grid search of γ is conducted, with the β parameters (excepting β_0) set to the OLS values and the β_0 and σ^2 parameters adjusted according to the corrected ordinary least squares formula presented in Coelli (1995):

$$\hat{\sigma}^2 = m_2 + \frac{2}{\pi}\left[\sqrt{\frac{\pi}{2}} \cdot \frac{\pi}{(\pi - 4)} m_3\right]^{2/3}$$

$$\hat{\gamma} = \hat{\sigma}^{-2}\left[\sqrt{\frac{\pi}{2}} \cdot \frac{\pi}{(\pi - 4)} m_3\right]^{2/3}$$

$$\hat{\beta}_0 = \hat{\beta}_0(\mathrm{OLS}) + \sqrt{\frac{2\hat{\gamma}\hat{\sigma}^2}{\pi}} \tag{4.4}$$

where

m_2 and m_3 are the second and third sample moments of the OLS residuals, respectively,

Recall that

$\gamma=\sigma_U^2/(\sigma_V^2+\sigma_U^2)$,

and reflects the decomposition of standard deviation among random error and technical inefficiency terms.

Any other parameters are set to zero in this grid search.

3. The values selected in the grid search are then used as starting values in an iterative procedure (using the Davidon-Fletcher-Powell Quasi-Newton method) to obtain the final maximum likelihood estimates.

Note that the resulting estimates are significant if the SPF provides a better likelihood then the OLS results obtained in the first step.

It is also convenient to re run models with different starting points provided to the search. Therefore, in this study different runs with different starting points are made. While interpreting the results, only the most significant ones from each of the assumed distributions are considered.

The results obtained under half normally distributed inefficiency term and truncated normally inefficiency term are displayed in tables 4.1. and 4.2., respectively:

Table 4.1. The SPF estimates under the half normally distributed inefficiency terms assumption (SPF Model1)

	Coefficient	**Standard error**	**t-ratio**
beta 0	1,53	1,11	1,38
beta 1	0,05	0,04	1,32
beta 2	0,09	0,02	3,69
beta 3	0,00011	0,00002	7,18
Sigma-squared	0,18	0,20	0,90
Gamma	0,07	1,66	0,04

Table 4.2. The SPF estimates under the truncated normally distributed inefficiency terms assumption (SPF Model2)

	Coefficient	Standard-error	t-ratio
beta 0	2,41	1,57	1,54
beta 1	0,05	0,04	1,16
beta 2	0,08	0,24	0,35
beta 3	0,00	0,00	1,90
Sigma-squared	0,21	0,81	0,25
Gamma	1,00	0,01	136,93
Mu	0,88	5,29	0,17

Efficiency scores of the stations according to cities are displayed in Table 4.3.As it can be seen from this table, efficiency scores under the models differ considerably:

Table 4.3. The arithmetic mean of efficiency scores for the SPF Model1 and the SPF Model2 (grouped by cities)

	Average Score SPF1	Average Score SPF2
BİLECİK	92,94	72,31
BURSA	92,2	55,72
EDİRNE	92,01	46,88
ESKİŞEHİR	91,84	46,01
KOCAELİ	91,66	45,88
KÜTAHYA	91,65	42,32
İSTANBUL	91,59	44,78
DÜZCE	91,59	41,28
All Cities	**91,56**	**44,13**
TEKİRDAĞ	91,56	44,11
SAKARYA	91,33	39,38
ZONGULDAK	91,18	34,14
YALOVA	90,85	29,64
BOLU	90,71	30,29
KARABÜK	89,56	18,93
BARTIN	89,07	15,93

Correlation of efficiency scores from the SPF1 and the SPF2 models are plotted (as depicted in Figure 4.1.):

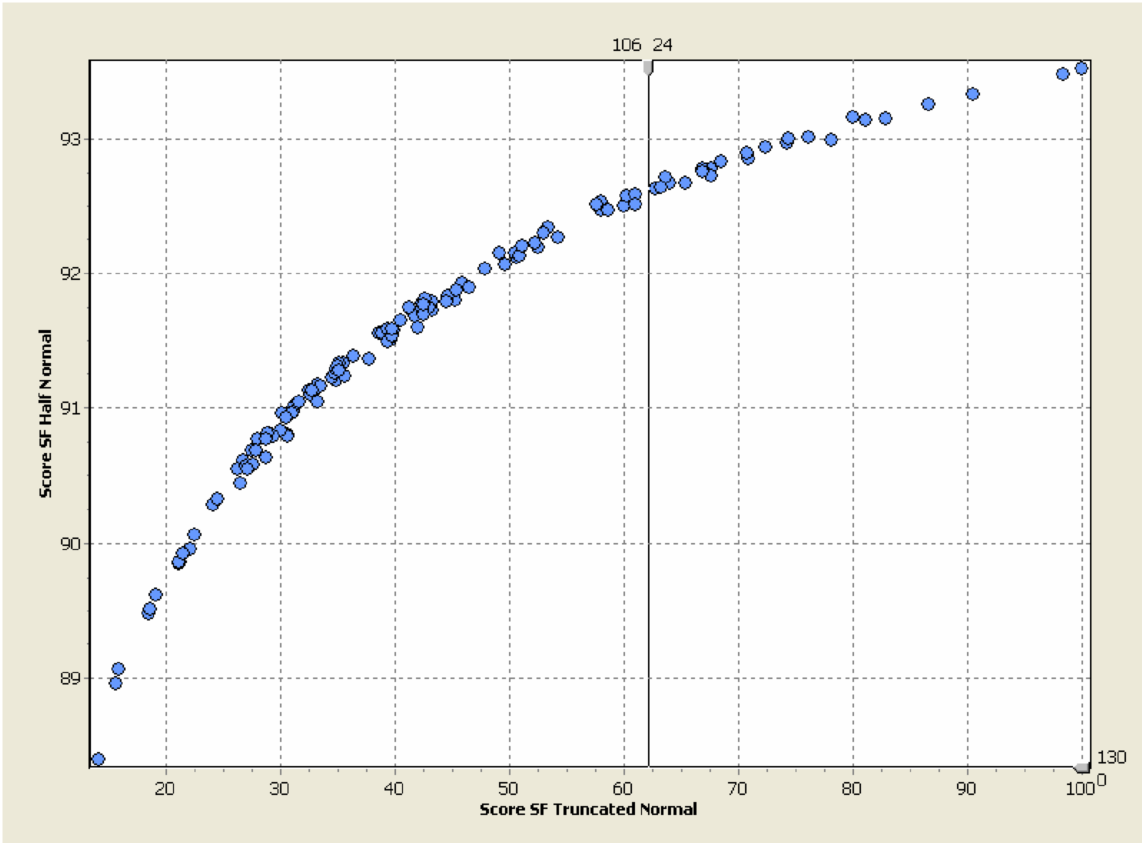


Figure 4.1. The SPF Half Normal vs the SPF Truncated Normal Efficiency Scores.

As the results are analyzed, correlation of efficiency scores is found to be 0,94, pointing to a very strong correlation. It can be concluded that relative efficiency of stations in the two different SPF models are very close. However, efficiency levels are different; under half normally distribution assumption, efficiency levels spread between 87 and 94 whereas in truncated normally distribution assumption levels spread between 15 and 100.

Efficiency scores are also analyzed according to the market sales flag, as displayed in Table 4.4.:

Table 4.4. The arithmetic mean of efficiency scores for the SPF Model1 and the SPF Model2 (grouped by market sales flag).

	SPF Model1	SPF Model2
Market Sales are Correct	91,87	49,95
Market Sales are not Correct	91,33	39,59

It can be seen in Table 4.4. that the Market Sales flag is a distinguishing factor in the SPF models like the DEA models; stations providing correct market sales perform better than the others on average.

At this point it is necessary to compare the results of the SPF and the DEA Models. For this purpose, the efficiency scores are plotted, as displayed in Figure 4.2 and Figure 4.3.:

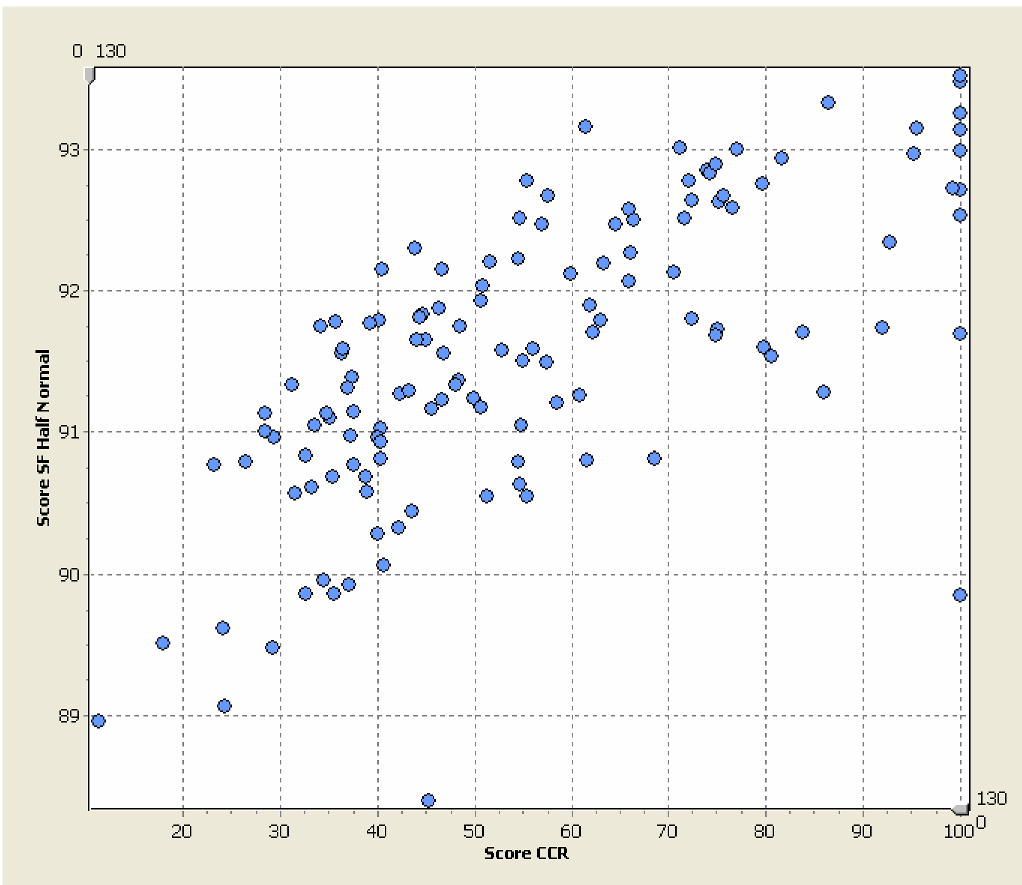


Figure 4.2. The SPF Model1 vs the DEA CCR average efficiency scores

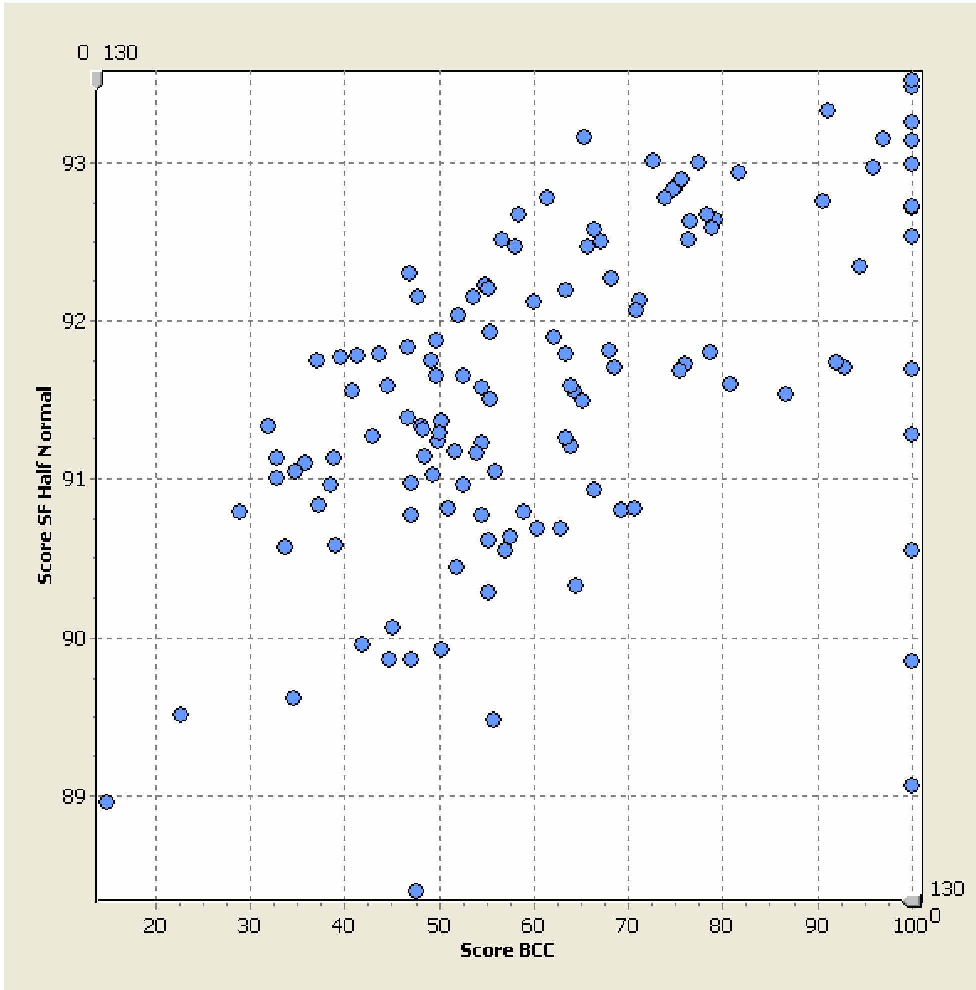


Figure 4.3. The SPF Model1 vs the DEA BCC average efficiency scores.

Likewise graphs result, when the SPF Model2 results are plotted instead of the SPF Model1 results (Recall that correlation between the SPF Models is 0,94). In general, results of the SPF and the DEA approaches are moderately correlated.

Furthermore, the correlations among the efficiency scores from each of the 10 models are investigated. The BCC Run0 and the CCR Run0 represent the unrestricted models; the BCC Run1, the BCC Run2, the CCR Run1, the CCR Run2 represent the assurance regions models, where in Run1 models, all ratios given in section 3.6. are restricted to a window of $[\mu-\sigma/2, \mu+\sigma/2]$ and in Run2 models, all ratios given are restricted to a window of

$[\mu - \sigma/4, \mu + \sigma/4]$; the SPF Run1 represents the SPF model assuming half normal distribution for inefficiency term and SPF Run2 represents the SPF model assuming truncated normal distribution for inefficiency term. The results are displayed in Table 4.5.:

Table 4.5. Correlation of efficiency scores obtained different models.

	BCC Run0	**BCC Run1**	**BCC Run2**	**CCR Run0**	**CCR Run1**	**CCR Run2**	**SPF Run1**	**SPF Run2**
BCC Run0	1,00							
BCC Run1	0,67	1,00						
BCC Run2	0,61	0,97	1,00					
CCR Run0	0,89	0,50	0,43	1,00				
CCR Run1	0,65	0,67	0,59	0,72	1,00			
CCR Run2	0,63	0,68	0,61	0,71	0,99	1,00		
SPF Run1	0,52	0,13	0,07	0,69	0,45	0,43	1,00	
SPF Run2	0,60	0,20	0,13	0,73	0,50	0,48	0,94	1,00

When the DEA and the SPF models are considered it can be stated that:

i) There is a moderate correlation between the CCR results and the SPF results. Lack of a stronger correlation can be explained by the existence of aggregation effect: The SPF and the DEA show stronger correlations when the original DEA model has one output. Moderate correlation still holds when assurance region models of the CCR are considered.

ii) When the BCC unrestricted and the BCC assurance region models' solutions are compared to the SPF models, correlation is weaker or even does not exist, which is also evident in Figure 4.4.. This result stems mainly from the fact that the SPF formulations do not take returns to scale conditions into account, in addition to the aggregation effect.

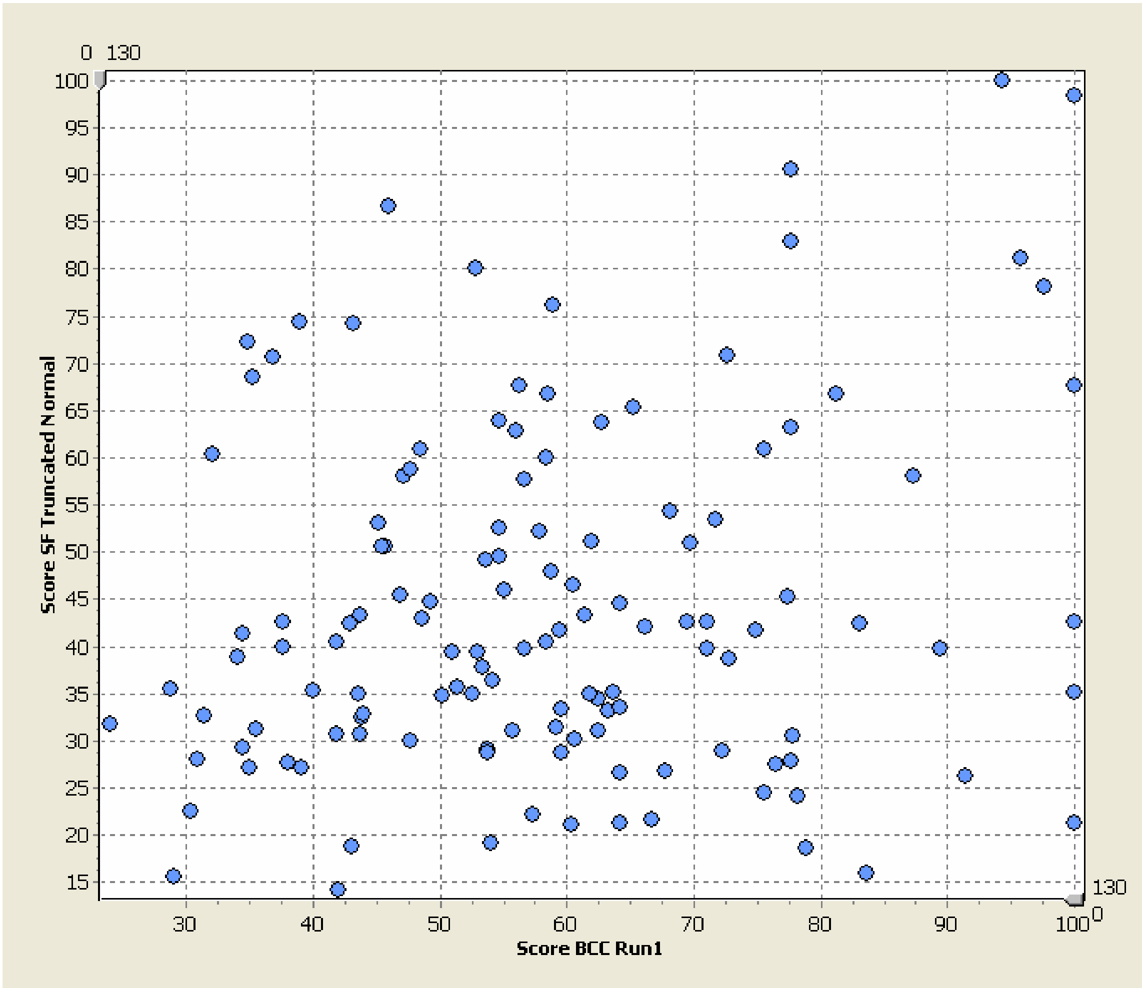


Figure 4.4. The BCC Run1 vs the SPF Half Normal efficiency scores. Each point represents a station.

“Best performing” units in the DEA (which are listed in Table 4.6.) and the SPF models are also compared. The SPF results are not parallel to the DEA models. (This is also observable in Figure 4.3. Units that have 100 efficiency score on the x axis are widespread on the y axis):

Table 4.6. The CCR and the SPF Model 1 efficiency scores for efficient units in the CCR model

Station Code	Score CCR	Score SF Half Normal
All Stations	**55,9**	**91,56**
76	100	93,51
116	100	93,47
72	100	93,25
111	100	93,13
118	100	92,98
5	100	92,71
4	100	92,53
130	100	91,7
131	100	89,85

Note that in the SPF model, one of the best performing units in the CCR model obtained below average efficiency scores.

Likewise, stations that obtained highest efficiency scores in the SPF are not necessarily out performers in the CCR model, as displayed in Table 4.7.:

Table 4.7. The CCR and the SPF Model1 efficiency scores for top performers in the SPF Model1

Station Code	Score CCR	Score SF Half Normal
All stations	**55,9**	**91,56**
76	100	93,51
116	100	93,47
120	86,46	93,33
72	100	93,25
68	61,49	93,15
125	95,57	93,14
111	100	93,13
103	71,2	93
73	77,08	92,99
118	100	92,98
57	95,22	92,97
65	81,6	92,94

Similar results are obtained when the BCC model is considered instead of the CCR. Like in the CCR case, some best performing DMU's in the BCC model performed below average in the SPF Model1, as displayed in Table 4.8.:

Table 4.8. The BCC and the SPF Model 1 efficiency scores for efficient units identified in the BCC model.

Station Code	Score BCC	Score SF Half Normal
All Stations	**62,1**	**91,56**
76	100	93,51
116	100	93,47
72	100	93,25
111	100	93,13
118	100	92,98
117	100	92,72
5	100	92,71
4	100	92,53
130	100	91,7
13	100	91,28
2	100	90,55
131	100	89,85
1	100	89,07

Similar to the CCR case, stations that obtained highest efficiency scores in the SPF are not necessarily out performers in the BCC model:

Table 4.9. The BCC and the SPF Model1 efficiency scores for top performers in the SPF Model1

Station code	Score BCC	Score SPF Half Normal
All Stations	**62,1**	**91,56**
76	100	93,51
116	100	93,47
120	91,16	93,33
72	100	93,25
68	65,27	93,15
125	96,96	93,14
111	100	93,13
103	72,57	93
73	77,47	92,99

118	100	92,98
57	95,91	92,97
65	81,72	92,94

Lack of correlation can be explained by two main reasons: The first one is the aggregations effect that is described, while comparing the CCR and the SPF results. In general, efficiency scores obtained in the DEA and the SPF models tend to have less correlation when the aggregation of outputs (i.e. combining many outputs into a few) increases (aggregation effect). Moreover, efficiency scores obtained in the BCC and the SPF models display even weaker correlation since the SPF models do not consider the returns to scale conditions.

5. CONCLUSIONS AND FUTURE STUDIES

In this study, two different approaches are used for the performance evaluation of fuel stations. The main approach is the Data Envelopment Analysis, whereas the Stochastic Production Frontiers method is applied as a complementary study.

The DEA study reveals that efficiency scores change according to cities, with urban centers, such as Istanbul and Bursa, out performing the others. When the assurance regions models are applied, this dominance remains limited to the CCR formulation. It can be concluded that, under the constant returns to scale assumption, cities with denser urban populations perform better. Note, however that this is a slight dominance; the average efficiency scores for these cities are slightly over the average. On the other hand, in the solution of the CCR and the BCC models, top performing stations coincide to a great extent.

Scale efficiency analysis reveals an average of 0,9 scale efficiency, implying that inefficiencies result mainly from pure technical efficiency (TE_{VRS}). In other words, inefficiencies occur mainly from “bad management practices”. Stations from urban centers (Istanbul, Bursa, Kocaeli) are better than average.

It is necessary to discuss the term bad management practices and interpret results accordingly. Since fuel stations sell commodity products (i.e. standardized, not diversified much, with tight price range) to end customers (eventual users, i.e. car drivers), details in management practices may be quite important, which include both quantifiable and non quantifiable factors; existence and cleanness of toilets, average time a customer spends for fueling and payment, physical attractiveness of station (lights, cleanness, availability of automatic car wash etc.), treatment of customers and conditions in the market (product range, quality, pricing and freshness). These are the management related areas for performance improvement.

In addition to the “internal” factors mentioned above, there are also some “external” factors. Consider the available sales potential of the station. The rent expense variable reflects this variable to a great extent. But other stations in the close neighborhood have significant effect on the performance of the station. It can be argued that this factor can be incorporated into the model. But there are factors that are hard to quantify. First of all it is necessary to define the scope of “being neighbors”. It is not reasonable to set a constant value as a limiting distance, therefore it is necessary to study each of the stations separately for detecting the neighbors. Even this is not enough, since the service levels of neighboring stations are as important as their existence. In other words, if the competitors are good performers, there is harsher competition and if they are bad performers, the station is advantageous. Accordingly, it can be stated that some of the inefficient stations are performing bad due to “external” factors, which they can not control.

When the scale efficiency is analyzed at station level, a different pattern is observed: It remains about same for a broad range of stations. Only stations smaller than a certain size have considerably worse scale efficiencies. Therefore, as discussed in section 3.5., it is concluded that greater scale does not provide a competitive advantage after reaching a certain level of scale, which can be outlined as 8 pumps, 1.600 rent expense and 5 pumpers. The identified threshold level is a valuable piece of information to be used when investing into new stations.

Both the correlation analysis and the regression studies reveal that among the output variables, benzene sales has the lowest impact and market sales has the highest impact on efficiency scores. Among the input variables, most significant one is the rent expense.

In general, the DEA results provide valuable managerial information about the minimum size of stations for efficient operation.

The SPF approach is applied to the problem by aggregating the output values. The results differ considerably from the DEA results. When the SPF and the CCR results are compared, only a moderate correlation is observed, whereas the BCC and the SPF results don't correlate at all. The main reason for this difference is the aggregation effect. It can be concluded that as the number of outputs increase, the DEA and the aggregated SPF results tend to correlate weaker. The lack of any correlation between the BCC and the SPF results can be explained by variable returns to scale conditions of the BCC.

When a comparison between the SPF Models with different assumptions for the inefficiency term is made, it is observed that efficiency scores correlate at a great degree. In other words, relative efficiencies of stations are very close. However when the levels of efficiencies are compared, there is a different picture; in one of the models scores vary from 15 to 100 whereas at the other one they vary from 88 to 100. This difference is quite striking since it reveals that in these models, allocation of inefficiency term between random errors and technical efficiencies differ reasonably.

Accordingly, a study concentrating on Stochastic Production Frontiers, with different distributional assumptions is a reasonable area for future research.

As an immediate future study, models can be expanded to other regions (Only stations in the Marmara region are considered in this study). It is possible that different patterns exist in different regions. Also, a time dimension can be added to the analysis, (There is a single time period in this study) in order to analyze changes in efficiencies over time. In such a case, more complex the SPF models can be developed.