NEW APPROACHES FOR PERFORMANCE EVALUATION USING DATA ENVELOPMENT ANALYSIS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

N. ÖZGÜR ÖZPEYNİRCİ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF INDUSTRIAL ENGINEERING

MAY 2004

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. Canan Özgen
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Çağlar Güven
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Murat Köksalan
Supervisor

Examining Committee Members

Asst. Prof. Dr. Güldal Büyükdamgacı

Assoc. Prof. Dr. Gülser Köksal

Prof. Dr. Murat Köksalan

Assoc. Prof. Dr. Nur Evin Özdemirel

Dr. Özge Uncu

ABSTRACT

NEW APPROACHES FOR PERFORMANCE EVALUATION USING DATA ENVELOPMENT ANALYSIS

ÖZPEYNİRCİ, N. Özgür

M. Sc. Thesis, Department of Industrial Engineering
Supervisor: Prof. Dr. Murat KÖKSALAN

May 2004, 69 pages

Data Envelopment Analysis (DEA) assigns efficiency values to decision making units (DMU) in a given period by comparing the outputs with the inputs. In many applications, inputs and outputs of DMUs are monitored over time. There might be a time lag between the consumption of inputs and production of outputs. We develop approaches that aim to capture the time lag between the outputs and the inputs in assigning the efficiency values to DMUs. We present computational results on randomly generated problems as well as on an application to R&D institutes of the Scientific and Technical Research Council of Turkey (TÜBİTAK).

Keywords: Data Envelopment Analysis, DEA, Performance Evaluation, R&D Institute

ÖZ

PERFORMANS DEĞERLENDİRMEYE VERİ ZARFLAMA ANALİZİ İLE YENİ YAKLAŞIMLAR

ÖZPEYNİRCİ, N. Özgür

Yüksek Lisans Tezi, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Murat KÖKSALAN

Mayıs 2004, 69 sayfa

Veri Zarflama Analizi (DEA) karar verme birimlerinin (DMU) herhangi bir dönemdeki verimlilik değerlerini girdi ve çıktıları kıyaslayarak belirler. Birçok uygulamada, DMU'ların girdi ve çıktıları zamana bağlı olarak izlenir. Girdilerin tüketimi ile çıktıların üretimi arasında bir zaman farkı olabilir. Bu çalışmada, DMU'ların verimlilik değerlerini belirlerken çıktılarla girdiler arasındaki zaman farkını yakalamayı amaçlayan bir yaklaşım geliştirdik. Rassal olarak oluşturulmuş problemlerin sonuçlarının yanısıra Türkiye Bilimsel ve Teknik Araştırma Kurumu'nun (TÜBİTAK) Araştırma-Geliştirme (Ar-Ge) enstitüleri için yaptığımız bir uygulamayı da sunuyoruz.

Anahtar Kelimeler: Veri Zarflama Analizi, DEA, Performans Değerlendirme, Ar-Ge Enstitüsü

To my mom, my dad and my love…

ACKNOWLEDGEMENTS

I would like to express my gratitude to Prof. Dr. Murat Köksalan for his wonderful supervision. I really appreciate his experience and support. Studying with him was a great experience for me.

I would also present my special thanks to two magnificent people, my parents Suzan Özpeynirci and Ali Riza Özpeynirci. I am so lucky that I have their unlimited belief and love.

And a special thanks is due my fiancée Selin Bilgin for everything she added to my life. With her morale support, I could perform such a good work.

I also express my thanks to all my friends, especially my home-mates, for their kind encouragements.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

CHAPTER 1

INTRODUCTION

Both profit-seeking and non-profit organizations have to perform their best in order to be competitive and to survive in the future. Many people such as the owners, partners, investors, and also public bodies are interested in the performances of the organizations in producing goods and/or providing services. These people are usually decision makers for the organizations. They are responsible from making strategic decisions such as the allocation of the resources, target setting for future periods or shutting down the business, etc.

In order to provide information about the performance of the organizations, researchers developed performance evaluation systems. Many people still develop new systems and/or improve the existing ones in order to provide better information, because as stated above, many crucial decisions are made by utilizing the information obtained by the performance evaluation systems.

This thesis study has started with the motivation of demonstrating a performance evaluation system for the research and development (R&D) institutes of the Scientific and Technical Research Council of Turkey, TÜBİTAK. In this study, we develop a performance evaluation system applicable to R&D institutes as well as many other organizations.

The R&D institutes have multiple inputs such as researchers, technicians, R&D laboratories with high-tech equipments, etc. and multiple outputs such as high-tech end-user products, scientific articles, patents, loyalty incomes, etc. Therefore, we can say that the performance evaluation of R&D institutes and many other organizations is a multi-criteria problem. We have examined several systems for the performance evaluation and we utilize an approach called Data Envelopment Analysis in this study.

After making an introduction in Chapter 1, we briefly review literature on both the performance evaluation and the Data Envelopment Analysis in Chapter 2.

In Chapter 3, we review the basic Data Envelopment Analysis models and introduce the new models and ideas developed during this study.

In Chapter 4, we conduct two experiments. We explain the factors of the experiments and report the results.

In Chapter 5, we consider the case of TÜBİTAK and demonstrate our approach on the R&D institutes of TÜBİTAK. This is an exercise application as we only use the inputs and outputs for which data is readily available and ignore other relevant factors.

In Chapter 6, we conclude the study and discuss the further research areas.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we review the literature on performance evaluation and data envelopment analysis (DEA).

2.1 Performance Evaluation

Researchers developed and reported many different approaches for performance evaluation. Most of these approaches propose the use of non-financial indicators in addition to financial ones. Appropriate non-financial indicators are determined according to the organizations being evaluated.

Suwignjo et al. (2000) developed the Quantitative Model for Performance Measurement System (QMPMS). A hierarchical performance system is constructed by QMPMS in three steps:

1) *identification of factors affecting the performance and their relationships,*
2) *structuring the factors hierarchically,*
3) *quantifying the effect of the factors on performance.*

Analytical Hierarchy Process (AHP) is used in the last step. In a further study, Bititci et al. (2001) evaluated the performances of possible strategies by QMPMS. Sarkis

(2003) proposed to use Analytical Network Process (ANP) instead of AHP. He also studied the evaluation of strategies on a planning horizon using QMPMS.

Spronk and Vermeulen (2003) introduced the comparative conditional performance review (CCPR) approach. This approach takes into account the effect of risk on the performance, which is beyond the control of decision maker. CCPR evaluates the performance of firms after removing the effects of risk factors.

Önel and Saatçioğlu (1995) worked on a trend analysis and regression model for the evaluation of university performance. They collected data for the following variables for a period of 21 years:

1) *number of articles published in international journals per professor,*
2) *ratio of full professors to professors(full, associate and assistant),*
3) *ratio of undergraduate students to professors,*
4) *ratio of research assistants to professors,*
5) *number of Ph.D degrees awarded per professor,*
6) *ratio of Ph.D students to professors.*

They tried to find out a relation between the number of published articles and remaining variables. They showed that "*ratio of full professors to professors*" and "*number of Ph.D degrees awarded per professor*" had significant effects on the "*number of articles published in international journals per professor*".

Fandel and Gal (2001) studied fund distribution among universities. They presented the pre-determined criteria set and the final solution accepted by the group of DMs. They analyzed the accepted solution and proposed different solutions using goal programming and distance minimization techniques.

2.2 Data Envelopment Analysis

DEA evaluates the relative performances of comparable units having the authority to make decisions. These units are called decision making units (DMUs). DEA models use the input and output quantities of DMUs in efficiency assignments. Each DMU is considered one by one and a linear program is solved for each DMU. The efficiency of a DMU is defined as the ratio of weighted outputs to weighted inputs. These weights are positive decision variable of the linear program. When a DMU is under consideration, DEA maximizes the efficiency ratio of that DMU by changing the weights of inputs and outputs. There is only one constraint type in standard DEA, none of the DMUs can be more efficient than 100%. Weighted outputs of a DMU can not be more than weighted inputs of that DMU. DEA approach is based on the original work of Farrell (1957) and became popular with the work of Charnes et al. (1978). Standard DEA models only use the input and output quantities of DMUs. However, adding the preferences of decision makers to DEA models is possible and studied in many papers. We discuss some of these below.

An interactive DEA procedure (IDEA) was developed by Post and Spronk (1999). IDEA incorporates the DM preference information to DEA in an iterative manner by setting minimum (maximum) acceptable levels of outputs (inputs). The authors illustrated IDEA model in the assessment of physics departments of UK Universities. They used the following inputs and outputs of the departments:

1) *The amount of general expenditure (input),*
2) *The amount of equipment expenditure (input),*
3) *The amount of research income (output),*
4) *The number of undergraduate students (output),*
5) *The number of post-graduate students on taught courses (output),*
6) *The number of post-graduate students doing research (output),*
7) *University Grant Committee research rating (output).*

The last output is a subjective criterion whereas the other inputs and outputs are objective criteria. They also introduced a combined DEA model which simultaneously solves many DEA models. We will further discuss the combined DEA model in the next chapter.

Halme et al. (1999) introduced the *value efficiency analysis* approach. By this approach, DEA uses the preference information of DM and assigns efficiency scores by using the estimated value function of DM. Korhonen et al. (2001) introduced a systematic approach for the performance analysis of academic research in R&D

institutions and universities using value efficiency analysis. They proposed five criteria:

1) *Quality of research,*
2) *Research activity,*
3) *Impact of research,*
4) *Activity in educating young scientists,*
5) *Activity in scientific community.*

Quantitative values for these criteria are calculated by the weighted sum of indicators (i.e. number of visitors in a research unit, number of citations) whereas AHP is used to generate the weights of indicators. These indicators are introduced as outputs and operation cost as the single input.

Kornbluth (1991) analyzed the policy effectiveness of player teams in a business game. Firstly, the efficient DMUs are determined by DEA. Afterwards, input weights are restricted since the author knows the policies of teams, i.e. the importance of inputs. Restrictions are made with constraints such as the weight of first input is not less than that of second input or the weight of first input is not less than 20% of the sum of all input weights. By introducing weight restrictions in DEA, some efficient DMUs become inefficient. The author classifies inefficiency as *technical inefficiency* and *policy inefficiency*. Technical inefficiency is the inefficiency of DMU in DEA model. Policy inefficiency is the difference of efficiency values of DMU in DEA model and weight restricted DEA model. Since this difference is due to the inefficient policy of the DMU.

Dyson and Thanassoulis (1988) used coefficients of regression analysis as lower bounds for output weights in a single input DEA model. Wong and Beasley (1990) introduced a bound restriction on the ratio of a weighted input to the sum of all weighted inputs. Thompson et al. (1990) described the Assurance Region (AR) concept and defined two types of ARs. In AR I, the restrictions on input and output weights are separable, but AR II type constraints create relations between input and output weights.

All restrictions represented up to now are set for a single DMU which is under observation. The input and/or output weights and/or level for that DMU are constrained. To our best knowledge, all restrictions are within DMU restrictions; however, between DMU restrictions have never been considered before.

A general analysis of DEA models is made by Kleine (2004). In this study DEA models are classified and new DEA approaches are shown with the use of multi criteria decision making approaches. Joro et al. (1998) analyzed and compared data envelopment analysis models and multiple objective linear programming. Both articles have represented different reviews of DEA models and approaches.

DEA is used in efficiency, effectiveness and/or performance analysis of banks and bank branches by Yeh (1996) and Golany and Storbeck (1999), transit systems by Karlaftis (2004) and technology selection problems by Khouja (1995). Multiple period analyses are done in both banking papers. Many inputs (financial ratios, number of employees, ATM's etc.) are collected and the irrelevant ones are determined with a regression analysis. The banks and the bank branches are compared and evaluated with other banks and bank branches in the same period. The efficiencies of banks and bank branches are monitored over a time period. Karlaftis (2004) analyzed the efficiency and effectiveness of urban transit systems of US. He studied the scale economies of these systems using DEA. Khouja (1995), established a two phase procedure for technology selection problems and used DEA in the first phase to find out the non-dominated alternatives. In the second phase, DM selects a technology from the non-dominated alternatives set.

CHAPTER 3

MODELS

In this chapter, we first represent a general overview of data envelopment analysis model and then introduce two new DEA models considering the time lag between consumption of inputs and production of outputs. Finally, we introduce a new type of constraint set that implies some restrictions on the input and/or output weights.

3.1 DEA model

We have n DMU's consuming m inputs and producing s outputs. X_{ij} and Y_{rj} stand for the amount of i^{th} input and r^{th} output of j^{th} DMU, respectively. All inputs and outputs are assumed to be non-negative. v_i and u_r are the weights of i^{th} input and r^{th} output, respectively.

DEA defines the efficiency of a DMU as the ratio of weighted sum of outputs to weighted sum of inputs. The weight set (v and u vectors) is chosen so as to maximize the efficiency of the DMU under consideration, DMU_0. The v and u vectors can be thought of as the prices of inputs and outputs, respectively. The weight set is constrained with the fact that a DMU can not be more efficient than 100%. This model, M1 is given below. Note that it is a non-linear model. The linear version of DEA model, M2 is also given below. The efficiency of DMU_0 is represented by h_0. ε is a very small positive number which ensures that every input and output has a value greater than zero.

(M1)

$$\text{Max } h_0 = \frac{\sum_{r=1}^{s} u_r Y_{r0}}{\sum_{i=1}^{m} v_i X_{i0}}$$

$$\text{Subject to}$$

$$\frac{\sum_{r=1}^{s} u_r Y_{rk}}{\sum_{i=1}^{m} v_i X_{ik}} \leq 1 \quad \forall k=1,...,n$$

$$u_r, v_i \geq \varepsilon \quad \forall r=1,...,s, \quad \forall i=1,...,m$$

(M2)

$$\text{Max } h_0 = \sum_{r=1}^{s} u_r Y_{r0}$$

$$\text{Subject to}$$

$$\sum_{i=1}^{m} v_i X_{i0} = 1$$

$$\sum_{r=1}^{s} u_r Y_{rk} - \sum_{i=1}^{m} v_i X_{ik} \leq 0 \quad \forall k=1,...,n$$

$$u_r, v_i \geq \varepsilon \quad \forall r=1,...,s\,, \quad \forall i=1,...,m$$

In M2, most favorable weight set for DMU_0 is chosen which maximizes the weighted sum of outputs of DMU_0. M2 assigns higher weights to the outputs which DMU_0 is good at producing (ie. produces in large amounts). Also M2 assigns higher weights to the inputs which DMU_0 is good at consuming (ie. consumes small amounts of).

M2 is the original formulation represented in Charnes et al.(1978). Dual of this model is M3. M3 constructs a hypothetical DMU which is the weighted sum of all DMUs. λ_j is the weight of j^{th} DMU in the hypothetical DMU. Hypothetical DMU produces each output at least at the level of DMU_0 and consumes inputs at most θ% of DMU_0. θ value represents the efficiency of DMU_0. M3 is an input oriented model since it points out the inefficiencies in the input consumption of DMU_0.

If the DMU_0 is not 100% efficient, this means that there exists a hypothetical DMU which dominates DMU_0. The input and output levels of the hypothetical DMU can be considered as target levels for DMU_0 since DMU_0 will be 100% efficient if it consumes inputs and produces outputs at the level of hypothetical DMU. Also the DMUs used to form the hypothetical DMU can be used as the benchmarks for DMU_0.

M2 assumes that there is a constant returns to scale (CRS). According to CRS, if the inputs of a DMU is doubled (or halved) also the outputs will be doubled (or halved). In our study, we assume constant returns to scale. See Kleine (2004) and Karlaftis (2004).

Although M2 is the original model, it is called as CCR-Input Dual and M3 is called as CCR-Input Primal in the literature. In addition to input oriented models, output oriented and combined models exist, see Joro et al. (1998). We use the input oriented model.

(M3)

$$\text{Min } z_0 = \theta + \varepsilon\left(\sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^-\right)$$

Subject to

$$\sum_{j=1}^{n} \lambda_j Y_{rj} - s_r^+ = Y_{r0} \quad \forall r=1,\ldots,s$$

$$\sum_{j=1}^{n} \lambda_j X_{ij} + s_i^- = \theta X_{ij} \quad \forall i=1,\ldots,m$$

$$\lambda_j \geq \varepsilon \quad \forall j=1,\ldots,n$$

$$s_i^- \geq \varepsilon \quad \forall i=1,\ldots,m$$

$$s_r^+ \geq \varepsilon \quad \forall r=1,\ldots,s$$

In order to find all DMUs' efficiency values, M2 has to be solved n times, once for each DMU_j (j=1,…,n). Post and Spronk (1999) offered to combine n models in a single model and to find all DMU's efficiencies simultaneously in a single model. Combined model is represented as M4. In this model, v_{ij} and u_{rj} are the weights of i^{th} input and r^{th} output when j^{th} DMU is under consideration.

(M4)

$$\text{Max } \sum_{j=1}^{n} h_j = \sum_{j=1}^{n} \sum_{r=1}^{s} u_{rj} Y_{rj}$$

Subject to

$$\sum_{i=1}^{m} v_{ij} X_{ij} = 1 \qquad \forall j=1,...,n$$

$$\sum_{r=1}^{s} u_{rj} Y_{rk} - \sum_{i=1}^{m} v_{ij} X_{ik} \leq 0 \quad \forall k=1,...,n \;, \forall j=1,...,n$$

$$u_{rj}, v_{ij} \geq \varepsilon \qquad \forall r=1,...,s \;, \; \forall i=1,...,m \;, \forall j=1,...,n$$

If multiple periods of data are available, a period index t is added to all variables. X_{ijt} and Y_{rjt} stand for the i^{th} input and r^{th} output of j^{th} DMU at period t (t=1,…,T). For multiple period instances, combined DEA model is to be solved once for each period. We add period index t to all terms and combined models over periods in model M5. The efficiency values of all DMUs for all periods are found by a single model. The efficiency value of j^{th} DMU in period t is represented by h_{jt} in the model.

(M5)

$$\text{Max} \sum_{t=1}^{T}\sum_{j=1}^{n} h_{jt} = \sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{r=1}^{s} u_{rjt} Y_{rjt}$$

Subject to

$$\sum_{i=1}^{m} v_{ijt} X_{ijt} = 1 \qquad \forall j=1,\ldots,n, \forall t=1,\ldots,T$$

$$\sum_{r=1}^{s} u_{rjt} Y_{rkt} - \sum_{i=1}^{m} v_{ijt} X_{ikt} \le 0 \quad \forall k=1,\ldots,n, \forall j=1,\ldots,n, \forall t=1,\ldots,T$$

$$u_{rjt}, v_{ijt} \ge \varepsilon \qquad \forall r=1,\ldots,s, \forall i=1,\ldots,m, \forall j=1,\ldots,n, \forall t=1,\ldots,T$$

3.2 Multi-period Input (MpI) Model

All DEA models represented above assume that inputs are converted to outputs in the same period. However in some cases, inputs are consumed and outputs are produced after a period of time. Number of published articles is an output example in which there is a time lag between the usage of inputs and obtaining the corresponding output. Also some transactions occurring in a bank may be due to the advertisements done in previous periods. Multi-period input model (MpI) we developed tries to capture the time lag while assigning efficiency values to DMUs.

DEA uses the input and output values of a single period and assigns the efficiency values to DMUs. In addition to these data, MpI uses the input values of previous periods. MpI model is given below. The inputs of previous periods are introduced as if they were new types of inputs for the current period.

(MpI) $$\text{Max} \sum_{t=(P+1)}^{T} \sum_{j=1}^{n} h_{jt} = \sum_{t=(P+1)}^{T} \sum_{j=1}^{n} \sum_{r=1}^{s} u_{rjt} Y_{rjt}$$

Subject to

$$\sum_{p=0}^{P} \sum_{i=1}^{m} v_{ijt}^{p} X_{ij(t-p)} = 1 \qquad \forall j=1,...,n, \quad \forall t=P+1,...,T$$

$$\sum_{r=1}^{s} u_{rjt} Y_{rkt} - \sum_{p=0}^{P} \sum_{i=1}^{m} v_{ijt}^{p} X_{ik(t-p)} \leq 0 \quad \forall k=1,...,n, \quad \forall j=1,...,n, \quad \forall t=P+1,...,T$$

$$u_{rjt}, v_{ijt}^{p} \geq \varepsilon \qquad \forall r=1,...,s, \forall i=1,...,m, \forall j=1,...,n, \quad \forall t=P+1,...,T, \forall p=0,...,P$$

In the MpI model, $X_{ij(t-p)}$ represents the amount of input i consumed by DMU j in period t-p. MpI uses input data of *P+1* periods and output data of the current period. Since there is a time lag of P periods, first output is produced in period *P*+1. v_{ijt}^{p} is the weight of i^th^ input of j^th^ DMU p periods ago while DMU j is under consideration. MpI assigns efficiency scores for periods [P+1, T]. Note that when P = 0, MpI is equivalent to the combined DEA model.

Optimal v_{ijt}^{p} weights give the relative input values of the current and the previous periods. Consider the i^{th} input of j^{th} DMU in period t. Suppose that the $v_{ijt}^{p'}$ value for period $t-p'$ is less than other v_{ijt}^{p} values. This implies that the DMU is willing to use the inputs of period $t-p'$ more than other periods' inputs, since the inputs of period $t-p'$ are cheaper than the inputs of other periods. Using this fact, we developed the Effective Input model.

3.3 Effective Input (EI) Model

EI model introduces the *Effective Input* concept. Effective input is the "real" amount of input used in the current or the previous periods in order to produce outputs in the current period. Effective input is the weighted sum of current and previous periods' inputs. There is an inverse relation with v_{ijt}^{p} values and input consumption. As the v_{ijt}^{p} value increases, DMUs want to use the input i of the other periods where it is cheaper. We assume that all DMUs in all periods have similar patterns of using current and previous periods' inputs. The weight w_{ip} represents the portion of input i used p periods ago. Number of v_{ijt}^{p} values used to calculate each w_{ip} weight is $n(T\text{-}P)$. The weights are calculated and normalized as follows:

$$w'_{ip} = \frac{1}{\sum_{t=P+1}^{T} \sum_{j=1}^{n} v_{ijt}^{p}} \quad \forall i=1,...,m\ ,\ \forall p=0,...,P$$

$$w_{ip} = \frac{w'_{ip}}{\sum_{p=0}^{P} w'_{ip}} \qquad \forall i=1,...,m,\ \forall p=0,...,P$$

E_{ijt} represents the effective input value of i^{th} input of DMU j in period t. Effective input values are calculated using the input values and w_{ip} values as follows:

$$E_{ijt} = \sum_{p=0}^{P} w_{ip} X_{ij(t-p)} \quad \forall i=1,...,m\ ,\ \forall j=1,...,n\ ,\ \forall t=P+1,...,T$$

The effective input (EI) model is given below. EI model assumes that weights of effective input do not change in time and are the same for all DMU's. Note that EI is very similar to the combined DEA model. X_{ijt} values are changed with E_{ijt} and index of t starts at P+1 instead of 1. Also note that, E_{ijt} values are calculated as above and are knownconstant values in the EI model.

(EI) $$\text{Max} \sum_{t=(P+1)}^{T} \sum_{j=1}^{n} h_{jt} = \sum_{t=(P+1)}^{T} \sum_{j=1}^{n} \sum_{r=1}^{s} u_{rjt} Y_{rjt}$$

Subject to

$$\sum_{i=1}^{m} v_{ijt} E_{ijt} = 1 \quad \forall j=1,\ldots,n, \forall t=P+1,\ldots,T$$

$$\sum_{r=1}^{s} u_{rjt} Y_{rkt} - \sum_{i=1}^{m} v_{ijt} E_{ikt} \le 0 \quad \forall k=1,\ldots,n, \forall j=1,\ldots,n, \ \forall t=P+1,\ldots,T$$

$$u_{rjt}, v_{ijt} \ge \varepsilon \quad \forall r=1,\ldots,s, \ \forall i=1,\ldots,m, \ \forall j=1,\ldots,n, \ \forall t=P+1,\ldots,T$$

3.4 Models with Weight Range Constraint (WRC)

All models represented up to now are free to assign any positive weight sets in order to maximize the efficiency values of each DMU at each period separately. These weight sets for different DMUs and periods may take diverse values. However, this may not be the case in the real life, the weights of inputs and outputs may not be much more different than the average weights and it might make more sense to restrict them to lie within a reasonable range. Using this idea, we introduce a new type of weight restriction. This restriction enables an interaction between DMUs which have not been proposed before.

Consider the single period combined DEA model, M4. This model assigns positive values to input weights, v_{ij}'s. We define a bound for each input i. The center of bound for input i is set as the average value of v_{ij}'s over all DMUs. $\overline{v}_i$ value represents the average value for input i. We let v_{ij}'s take values in the interval $(1 \pm \alpha)\overline{v}_i$. A nonnegative parameter α is used to set the allowable range width. Weight range constraints (WRC) are represented below.

(WRC) $$\frac{\sum_{j=1}^{n} v_{ij}}{n} = \bar{v}_i \quad \forall i=1,...,m$$

$$v_{ij} \le (1+\alpha)\bar{v}_i \quad \forall i=1,...,m\ ,\ \forall j=1,...,n$$

$$v_{ij} \ge (1-\alpha)\bar{v}_i \quad \forall i=1,...,m\ ,\ \forall j=1,...,n$$

However simply adding WRC to M4 is not suitable. Because the resulting model will maximize the total efficiency of DMUs by favoring the ones having similar input levels and sacrificing the DMUs having extreme input levels for some or all inputs (i.e. a DMU using first input very much and all other inputs very little). We change the objective function of the resulting model. Instead of maximizing the total efficiency, we minimize the maximum sacrifice due to the addition of WRC. DEA model with new objective and WRC, namely M6, is represented below.

In M6, h_j and q_j represent the estimated efficiency value of DMU j by M4 and M6, respectively. That is h_j is the efficiency value obtained with the original DEA model when there are no weight restrictions. In M6, we try to minimize the maximum deviation from those h_j values. Note that, we allow only positive deviations in the efficiency estimates, since adding additional constraints to weight space should not increase the efficiency estimates of DMUs.

A model may become infeasible with the addition of WRC, especially with a small α value. In order to handle the infeasibility, we use an iterative procedure. If a model is infeasible, α is increased and the model is solved again. This is done until the model with WRCs becomes feasible. Increasing α allows a wider range for weights. The lower bound range constraint of WRC becomes redundant for α>1. In this procedure, an initial α value is to be determined. In M6, WRC are applied only to the input weights. It is possible to add similar constraints for output weights.

(M6) Min dev

Subject to

$$\sum_{i=1}^{m} v_{ij} X_{ij} = 1 \qquad \forall j=1,...,n$$

$$\sum_{r=1}^{s} u_{rj} Y_{rk} - \sum_{i=1}^{m} v_{ij} X_{ik} \le 0 \qquad \forall k=1,...,n \text{ , } \forall j=1,...,n$$

$$\frac{\sum_{j=1}^{n} v_{ij}}{n} = \overline{v}_i \qquad \forall i=1,...,m$$

$$v_{ij} \le (1+\alpha)\overline{v}_i \qquad \forall i=1,...,m \text{ , } \forall j=1,...,n$$

$$v_{ij} \ge (1-\alpha)\overline{v}_i \qquad \forall i=1,...,m \text{ , } \forall j=1,...,n$$

$$\sum_{r=1}^{s} u_{rj} Y_{rj} = q_j \qquad \forall j=1,...,n$$

$$dev_j^+ + q_j = h_j \qquad \forall j=1,...,n$$

$$dev_j^+ \le dev \qquad \forall j=1,...,n$$

$$dev_j^+, q_j \ge 0 \qquad \forall j=1,...,n$$

$$u_{rj}, v_{ij} \ge \varepsilon \qquad \forall r=1,...,s \text{ , } \forall i=1,...,m, \forall j=1,...,n$$

M6 is constructed for one period in the above example. If multiple periods of data exist then the objective is changed as follows in order to minimize the sum of maximum deviations for each period:

$$\text{Min } \sum_{t=1}^{T} dev_t$$

where dev_t represents the maximum deviation of efficiency values in period t.

We represent an example problem to illustrate the effect of WRC with both types of objective functions, maximizing the total efficiency and minimizing the maximum sacrifice. The data of this example is taken from Thompson et al. (1990).

There are two inputs and a single output. The input and output values are given in Table 3.1. Since output values are same for all DMUs, we represent only the input values in Figure 3.1.

Table 3.1. Output and input values of example problem

DMU	**1**	**2**	**3**	**4**	**5**	**6**
Output, y	1	1	1	1	1	1
Input 1, x1	4	2	1	5	4	3
Input 2, x2	1	2	4	1	4	1.5

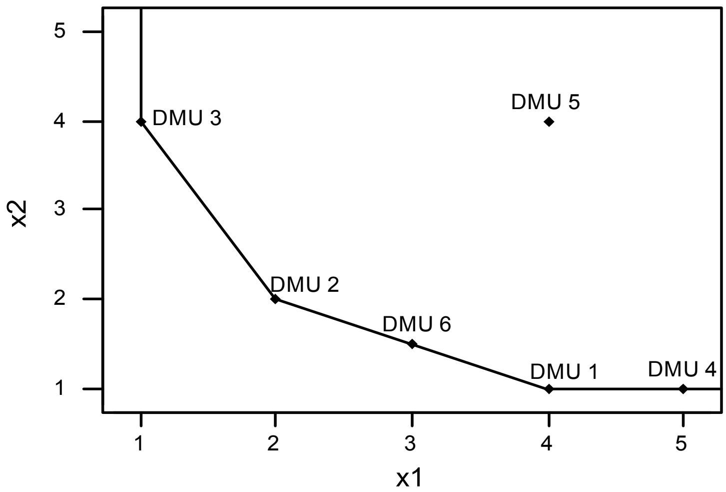

Figure 3.1. Input values of DMUs

We solve this problem using M4, M4 with WRC, and M6 The efficiency values of DMUs are given in Table 3.2 for all models. We use ε =0.00001 for all models and α = 0.5 for the models with WRC. There are two DMUs in the inefficient DMU set for M4; namely, DMUs 4 and 5. DMU 3 is also added to the inefficient DMU set if model M4 with WRC is used. DMU 6 is also added to inefficient DMU set in M6 model. Note that, although the average sacrifice increases, the maximum sacrifice of a DMU decreases.

Table 3.2. Efficiency values

DMU	Efficiency		
	M4	**M4 with WRC**	**M6**
1	100.00	100.00	100.00
2	100.00	100.00	100.00
3	100.00	95.44	88.99
4	99.99	85.71	88.98
5	50.00	50.00	50.00
6	100.00	100.00	99.00
Maximum Sacrifice		14.28	11.01
Average Sacrifice		3.14	3.84

As it can be seen in Table 3.3, optimal weights for DMUs vary between models. The standard deviation of optimal input weights of M4 is much higher than those of other two models. This is due to the WRC, since they enforce the model to take input weights in a range where its center is also determined by the model itself. The weights on the boundaries are written in bold. In M4 with WRC model, the weights of first input for DMUs 2 and 3 are on the upper bound. This can be interpreted as there is a pressure for these DMUs to increase the consumption of the first input while decreasing that of second input. In M6, weights of first input for DMUs 2 and 3, weights of second input for DMUs 1 and 4 are on the respective upper bounds. The second input weight for DMU 5 is also on the lower bound. These results demonstrate that, M6 distributes the pressure between more DMUs while minimizing the maximum sacrifice.

Table 3.3. Optimal input weights

	Weights							
	M4			M4 with WRC			M6	
DMU	v_{1j}	v_{2j}		v_{1j}	v_{2j}		v_{1j}	v_{2j}
1	16.67	33.33		16.67	33.33		13.77	**44.91**
2	33.33	16.67		**30.29**	19.71		**25.99**	24.01
3	100.00	0.00		**30.29**	17.43		**25.99**	18.50
4	0.00	100.00		14.29	28.57		11.02	**44.91**
5	16.67	8.33		12.97	12.34		10.03	**14.97**
6	16.67	33.33		16.67	33.33		17.17	32.34
Average	30.56	31.94		20.20	24.12		17.33	29.94
Std.Dev	35.62	35.91		7.95	8.86		7.15	12.99
Upper Bound				30.29	36.18		25.99	44.91
Lower Bound				10.10	12.06		8.67	14.97

CHAPTER 4

EXPERIMENTS

In order to analyze the effects of the new models and weight range constraints introduced in the previous chapter, we designed two experiments and defined performance indicators. In this chapter, we explain the performance indicators and the design of experiments. Finally, we present and discuss the results.

4.1 Performance Evaluation of Models

In the experiments, we assume that the efficiency values of DMUs exist and we try to estimate these values by the models. We call the assumed efficiency values as “true” and calculated efficiency values as “estimated”. A model performs well if the estimated efficiency values are close to the true efficiency values. We denote the deviation of the estimated value from the true value as the error. In order to see the performances of models and ideas in estimating the efficiency values, we use two performance indicators; root mean squared error (RMSE) and maximum error (ME). RMSE sums up squares of all error terms. ME, on the other hand, looks for the maximum error term in the efficiency estimates. Let H_{jt} and h_{jt} be the true and estimated efficiency values of DMU j in period t, respectively. RMSE and ME are calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n} \sum_{t=P+1}^{T} \left(h_{jt} - H_{jt}\right)^2}{n \times (T-P)}}$$

$$ME = \max_{j,t} \left\{ \left| h_{jt} - H_{jt} \right| \right\}$$

where P represents the time lag between the consumption of inputs and production of outputs and (T-P) represents the number of periods where efficiency estimates are done for n DMUs. We mention this issue in the next section.

Moreover, the correlation between the true and the estimated efficiencies is calculated. However, this information is only used to monitor the results. Correlation is calculated as follows:

$$Correlation = \frac{\sum_{j=1}^{n} \sum_{t=P+1}^{T} \left[(h_{jt} - \overline{h}) \times (H_{jt} - \overline{H}) \right]}{\sqrt{\left[\sum_{j=1}^{n} \sum_{t=P+1}^{T} (h_{jt} - \overline{h})^2 \right] \times \left[\sum_{j=1}^{n} \sum_{t=P+1}^{T} (H_{jt} - \overline{H})^2 \right]}}$$

where $\overline{H}$ and $\overline{h}$ are the mean values of true and estimated efficiencies, respectively.

4.2 Design of Experiments

We designed two experiments to study the factors that may affect the model performances. We generated random inputs, input consumption patterns, and efficiency values. Using assumed production functions related with inputs, we generated outputs and added random error terms. Finally, we used six different models to find the efficiency values of DMUs by using the generated inputs and outputs.

There are seven different factors used in the experiments. These factors are:

1. Model
2. Input Set
3. Input Standard Deviation
4. Weight Set
5. Efficiency Standard Deviation
6. Error Standard Deviation
7. Time Lag

The details of factors and their different levels are described below.

- **Model**

The model used for the efficiency estimation is the first factor of the experiment. There are six levels for this factor:

1. *DEA*: Combined DEA model is used.
2. *DEA*-WRC: Combined DEA model is used with weight restriction constraints.
3. *MpI:* Multi-period input model is used.
4. *MpI-WRC:* Multi-period input model is used with weight restriction constraints.
5. *EI*: Effective input model is used.
6. *EI-WRC*: Effective input model is used with weight restriction constraints.

Note that models DEA and DEA-WRC are represented in Section 3.1 and Section 3.4 as M5 and M6, respectively. We call these models as DEA and DEA-WRC in order to have a consistency in the names of compared models in the experiments.

- **Input Set**

The amount of inputs consumed by a DMU in a period may be related to the amount consumed in previous periods in real life. We use three different patterns for the inputs of consecutive periods. These are independent, positively correlated and negatively correlated patterns. Remember that, X_{ijp} represents the amount of i^{th} input consumed by DMU j in period p.

1. *Independent* : All X_{ijp} values are independently generated from a normal distribution with mean μ_1 and variance σ_1^2

   $X_{ijp} \sim N(\mu_1,\sigma_1^2)$

2. *Positively Correlated*: Input of a DMU is positively correlated with the previous period's input. X_{ijp} and $X_{ij(p+1)}$ are positively correlated.

   $X_{ijp} \sim N(\mu_1,\sigma_1^2)$ for p=1

   $X_{ijp} = 0.7\, X_{ij(p-1)} + 0.3\, ß$ for p=2,…,T

   where $ß \sim N(\mu_1,\sigma_1^2)$

3. *Negatively Correlated*: X_{ijp} and $X_{ij(p+1)}$ are negatively correlated.

   $X_{ijp} \sim N(\mu_1,\sigma_1^2)$ for p=1,3,5,7…

   $X_{ijp} = 0.7\, (2\, \mu_1 - X_{ij(p-1)}) + 0.3\, ß$ for p=2,4,6,8…

- **Input Standard Deviation**

We set μ_1 to 100 and use two levels for the standard deviation σ_1. Two levels of σ_1 value are:

1. *Low*: The standard deviation σ_1 is set to 20,
2. *High*: The standard deviation σ_1 is set to 40.

Using a higher standard deviation value decreases the correlation between the input values of current and previous periods and vice versa. In addition to the factor "input set", we wanted to see the effect of standard deviation of input on the performance of models.

- **Input Weight**

This factor controls the distribution of input consumption weights of current and previous periods. Let w_{ijr}^{p} be the true weight of i[th] input of p periods ago used by DMU j in order to produce the current value of output r. We produce the weight set for P periods. We use four different patterns of weight sets. These are:

1. *Independent*: w_{ijr}^{p} values are generated independently from a uniform

distribution between 0 and 1.

$w_{ijr}^{p} \sim U\,(0, 1)$ for p=0,…,P-1

2. *Non-increasing*: w_{ijr}^{p} values are generated with a non increasing pattern. The weights of periods follow a non-increasing pattern as you go previous periods. Weights are calculated as follows:

   $w_{ijr}^{0} \sim U\,(0,1)$ and

   $w_{ijr}^{p+1} = w_{ijr}^{p}\ \mu$ for p=1,…,P-1

   where μ~U(0,1) is a uniformly distributed random number between 0 and 1.

3. *Special*: In this pattern, the inputs of the current period have no effect on the outputs. All inputs of current period have zero weights. The weights of the inputs of previous periods take values with a non-increasing pattern as follows:

   $w_{ijr}^{0} = 0,$

   $w_{ijr}^{1} \sim U\,(0,1)$ and

   $w_{ijr}^{p+1} = w_{ijr}^{p}\ \mu$ for p=2,…,P-1

4. *No time lag*: In this pattern, the inputs are used without any time lag. Only current period's inputs affect current period's outputs as assumed by standard DEA. As a result, the weights of the previous periods are all zero.

   $w_{ijr}^{0} \sim U\,(0, 1),$

   $w_{ijr}^{p} = 0$ for p=1,…,P-1

After the generation of weights, w_{ijr}^{p} values are normalized to satisfy:

$$\sum_{i,j,p} w_{ijr}^{p} = 1 \quad \forall r=1,...,s$$

Note the difference between v and w vectors. v vector is used in the models and keeps the optimal weights of the inputs. However, w vector is generated in the experiments and keeps two types of information; (i) the contribution of each input to each output, (ii): distribution of these contribution between periods.

- **Efficiency Standard Deviation**

The true efficiency values of DMUs, H'_{jt}'s, are generated from a normal distribution with mean μ_2 and standard deviation σ_2. We set μ_2=100. If we use a high standard deviation then all DMUs will have very distinct efficiency values and it will be easier for models to distinguish the efficient and inefficient DMUs. If a small standard deviation is used, the task of the models will be harder. In order to examine the effects of the standard deviation, we use two levels for this factor:

1. *Low*: The standard deviation σ_2 is set to 10,
2. *High*: The standard deviation σ_2 is set to 20.

We shift the efficiency values as follows in order to obtain at least one DMU in each period to have a true efficiency value of 100:

$MaxH_t = \max_j \{ H'_{jt} \}$ for t=1,…,T

$H_{jt} = H'_{jt} - MaxH_t + 100$ for t=1,…,T and j=1,…,n

- **Error Term**

Different levels of the error term are used in order to see the performance of the models when the relation between inputs and outputs change from a deterministic structure to a stochastic one. Error term for output r of DMU j in period t is represented by e_{rjt}.

$e_{rjt} \sim N(\mu_3, \sigma_3^2)$ where μ_3= 0.

Two levels for the standard deviation of the error term are used:

1. *Deterministic*: The standard deviation σ_3 is set to 0,

2. *Probabilistic*: The standard deviation σ_3 is set to 20.

- **Time Lag**

There are two types of time lags. The first type is the time lag used in the generation of outputs. We may call it as "production time lag". The second one is the time lag used in the models for the estimation of efficiency values. This may be called as "model time lag". In the experiments, we set the production time lag as two periods. This means that the inputs are converted to outputs at most in two periods excluding the current period. This factor is analyzed to examine the under- or over-estimation of the production time lag in the real life application. Three levels for model time lag are used:

1. *Short time lag*: The model time lag is set to 1 (Model uses the inputs of current and the previous period),
2. *Normal time lag*: The model time lag is set to 2,
3. *Long time lag*: The model time lag is set to 3.

After describing the levels of factors, we represent the output function which uses these factors. The output is a function of a weighted combination of inputs from several periods that is adjusted by an efficiency factor plus the error term. That is,

$$Y_{rjt} = \left[H_{jt} \sum_{i=1}^{m} \sum_{p=0}^{P} \left(w_{ijr}^{p} \; X_{ij(t-p)} \right) \right] + e_{rjt} \quad \forall r = 1,...,s \;, \forall j = 1,...,n \;, \forall t = P+1,...,T$$

We set the number of inputs m=3, outputs s=2 and DMUs n=15. The number of periods is set to 10. P value changes as the level of "time lag" factor changes. We increase the T value as P increases in order to keep the number of periods at 10.

$T = P + 10 \;\; \forall \mathrm{P} \in [1,3]$

There is a relation between "weight set" and "time lag" factors. If "no time lag" level of "weight set" factor is used then the levels of "time lag" do not affect the output generation function. For this case, only the inputs of current period will be used for output generation. Due to this reason, we have done two separate experiments. In Experiment I, we do not use the "no time lag" level of "weight set" factor. There are 1296 treatments in this experiment. The factors and factor levels for Experiment I are given in Table 4.1. In Experiment II, we do not use “time lag” factor. There are only 144 treatments in this experiment. The factors and factor levels for Experiment II are given in Table 4.2. 20 replications are done for each factor level combination in both experiments. The inputs, efficiency values, error terms, outputs and the models were generated by C++ code compiled with Borland C++ Builder Version 6.0. The models were solved on GAMS IDE 2.0 with Cplex 7.5 solver. The experiments are done on an Intel Celeron 1.700 GHz CPU and 256 MB Ram PC.

Table 4.1. Factors and Factor Levels for Experiment I

Factor	**Levels**		
Model	DEA	MpI	EI
	DEA-WRC	MpI-WRC	EI-WRC
Input Set	Independent	Positive Correlated	Negative Correlated
Input Standard Deviation	Low	High	
Weight Set	Independent	Non-increasing	Special
Efficiency Standard Deviation	Low	High	
Error Standard Deviation	Deterministic	Probabilistic	
Time Lag	Short time lag	Normal time lag	Long time lag

Table 4.2. Factors and Factor Levels for Experiment II

Factor	**Levels**		
Model	DEA	MpI	EI
	DEA-WRC	MpI-WRC	EI-WRC
Input Set	Independent	Positive Correlated	Negative Correlated
Input Standard Deviation	Low	High	
Efficiency Standard Deviation	Low	High	
Error Standard Deviation	Deterministic	Probabilistic	

Before analyzing the results of above experiments, we consider the effects of assigning efficiency values randomly, without using any input-output data and models. By doing it, we aim to create a base level for both performance indicators.

We made 20 replications for both levels of “Efficiency Standard Deviation” factor since only this factor affects the “true” efficiency values of DMUs. Random efficiency values are generated by a uniform distribution between 0 and 100. Afterwards, as it is done in the “true” efficiency value calculations, all random efficiency values are shifted in order to set the maximum random efficiency value of each period to 100.

4.3 Results

In this section, we represent and discuss the results of experiments and analyze the effects of factors on the performance.

If we have input and output values of DMUs for a time period, we have to choose the model in order to estimate the efficiency values. Based on this, we group 7 factors in two groups, controllable and uncontrollable factors. The only controllable factor is “model” factor. Remaining factors are all uncontrollable factors. We analyze the main effects of controllable and uncontrollable factors on performance indicators for both experiments. Then, we analyze the two way interactions between controllable factor “model” and uncontrollable factors. If there is a significant interaction and the best model changes with the different levels of uncontrollable factor, we report the best model for each level of uncontrollable factor. We run two general linear models (GLM) in order to see whether the mean values of performance indicators differ at different levels. The null hypothesis, H_0 in GLM is:

H_o: different levels of factor i have the same mean

If the p value of factor i is less than 0.05, we reject H_0 and argue that mean value changes with the different levels of factor i. Otherwise we fail to reject H_0.

For Experiment I, we analyze 25920 runs. The results of GLMs for main effects are given in Figure 4.1. All factors have significant effects on RMSE and ME means since all factors have p values those are practically zero.

Analysis of Variance for **RMSE**, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
MODEL	5	99808	99808	19962	1408.20	0.000
inputSet	2	67776	67776	33888	2390.66	0.000
inputStd	1	67442	67442	67442	4757.72	0.000
wSet	2	29923	29923	14962	1055.48	0.000
effStd	1	15134	15134	15134	1067.65	0.000
errStd	1	149629	149629	149629	1.1E+04	0.000
TimeLag	2	5071	5071	2535	178.86	0.000
Error	25905	367208	367208	14		
Total	25919	801991				

Analysis of Variance for **ME**, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
MODEL	5	515506	515506	103101	1064.13	0.000
inputSet	2	308567	308567	154284	1592.39	0.000
inputStd	1	390702	390702	390702	4032.52	0.000
wSet	2	201347	201347	100674	1039.07	0.000
effStd	1	47893	47893	47893	494.31	0.000
errStd	1	1569131	1569131	1569131	1.6E+04	0.000
TimeLag	2	94075	94075	47037	485.48	0.000
Error	25905	2509882	2509882	97		
Total	25919	5637104				

Figure 4.1. General Linear Model: RMSE, ME versus all factors (Experiment I)

We use a coding system for the factor levels. The six models are coded with their names. We represent the other factors, factor levels and corresponding codes in Table 4.3.

Table 4.3. Factor Level Codes

Factor	Factor Level	Code
Input Set	Independent	L-1
	Positive Correlated	L-2
	Negative Correlated	L-3
Input Standard Deviation	Low	L-20
	High	L-40
Weight Set	Independent	L-1
	Non-increasing	L-2
	Special	L-3
Efficiency Standard Deviation	Low	L-10
	High	L-20
Error Standard Deviation	Deterministic	L-0
	Probabilistic	L-20
Time Lag	Short time lag	L-1
	Normal time lag	L-2
	Long time lag	L-3

The overall mean values of factor levels are plotted for RMSE and ME in Figures 4.2 and 4.3, respectively. In these and later figures, dashed lines show the overall means of corresponding performance indicators. Both graphs have similar shapes for all factors. MpI-WRC model performs best for both performance indicators. All models perform better if the input levels are positively correlated with the previous year's input level, input levels do not deviate so much, input weights of previous periods are not increasing, DMUs have similar efficiency values, there is no error term and the time lag is not less than the level used in models.

In Experiment II, there are 2880 runs. The results of GLMs for main effects are given in Figure 4.4. All factors have significant effects on RMSE and ME means since all factors have practically zero p values.

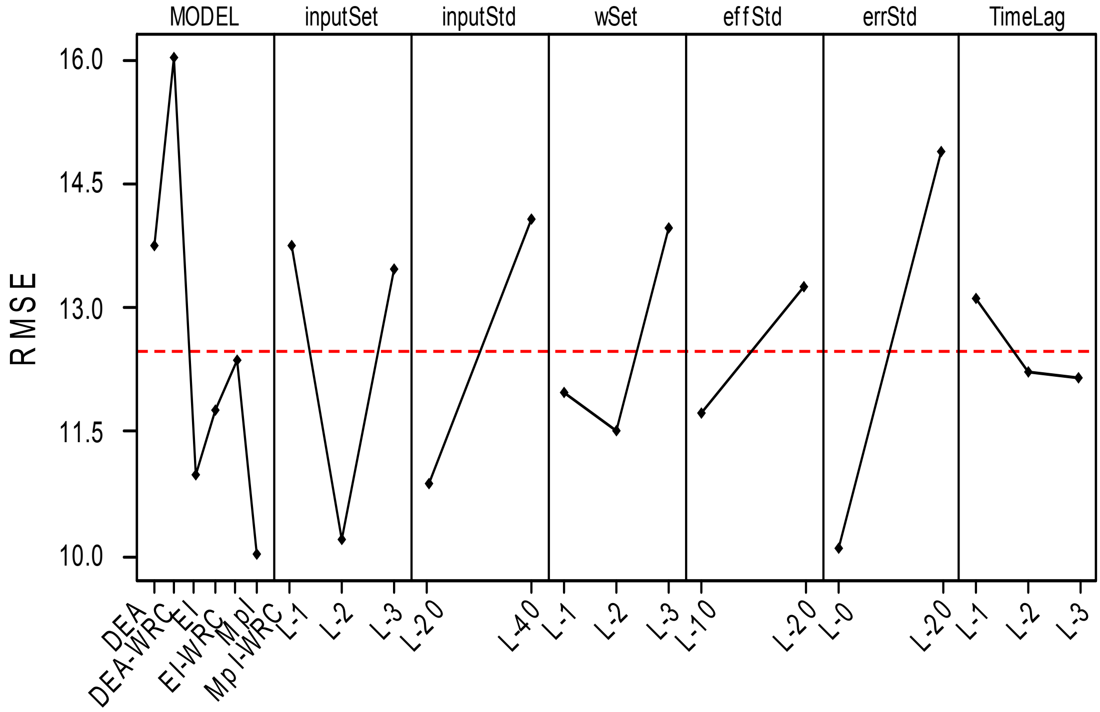

Figure 4.2. Main Effects Plot - Data Means for RMSE (Experiment I)

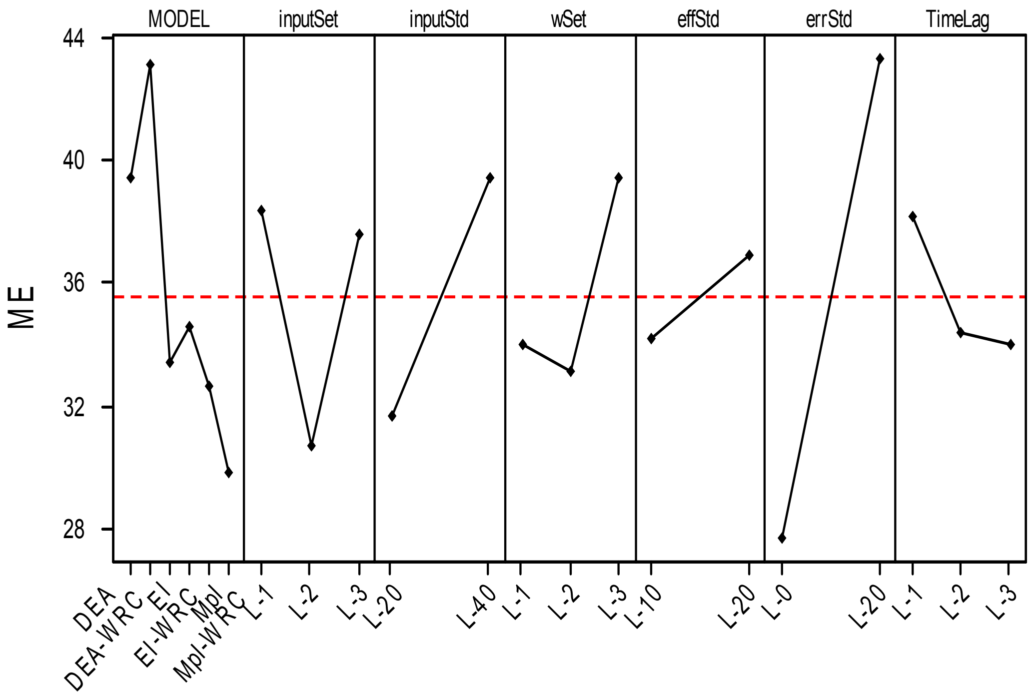

Figure 4.3. Main Effects Plot - Data Means for ME (Experiment I)

Analysis of Variance for **RMSE**, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
MODEL	5	5258.2	5258.2	1051.6	124.17	0.000
inputSet	2	3774.3	3774.3	1887.1	222.83	0.000
inputStd	1	4456.5	4456.5	4456.5	526.20	0.000
effStd	1	2818.3	2818.3	2818.3	332.77	0.000
errStd	1	21453.8	21453.8	21453.8	2533.19	0.000
Error	2869	24297.8	24297.8	8.5		
Total	2879	62058.8				

Analysis of Variance for **ME**, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
MODEL	5	36445	36445	7289	86.49	0.000
inputSet	2	39037	39037	19518	231.61	0.000
inputStd	1	48476	48476	48476	575.23	0.000
effStd	1	9429	9429	9429	111.89	0.000
errStd	1	229935	229935	229935	2728.45	0.000
Error	2869	241779	241779	84		
Total	2879	605101				

Figure 4.4. General Linear Model: RMSE, ME versus all factors (Experiment II)

The overall mean values of factor levels are plotted for RMSE and ME in Figures 4.5 and 4.6, respectively. Both graphs have similar shapes for all factors except the model. MpI performs higher than overall average for RMSE, but performs lower than average for ME. MpI-WRC model performs best for RMSE and DEA for ME. On the average, models perform better if the input levels are positively correlated with the previous year's input level, input levels do not deviate so much, DMUs have similar efficiency values and there is no error term.

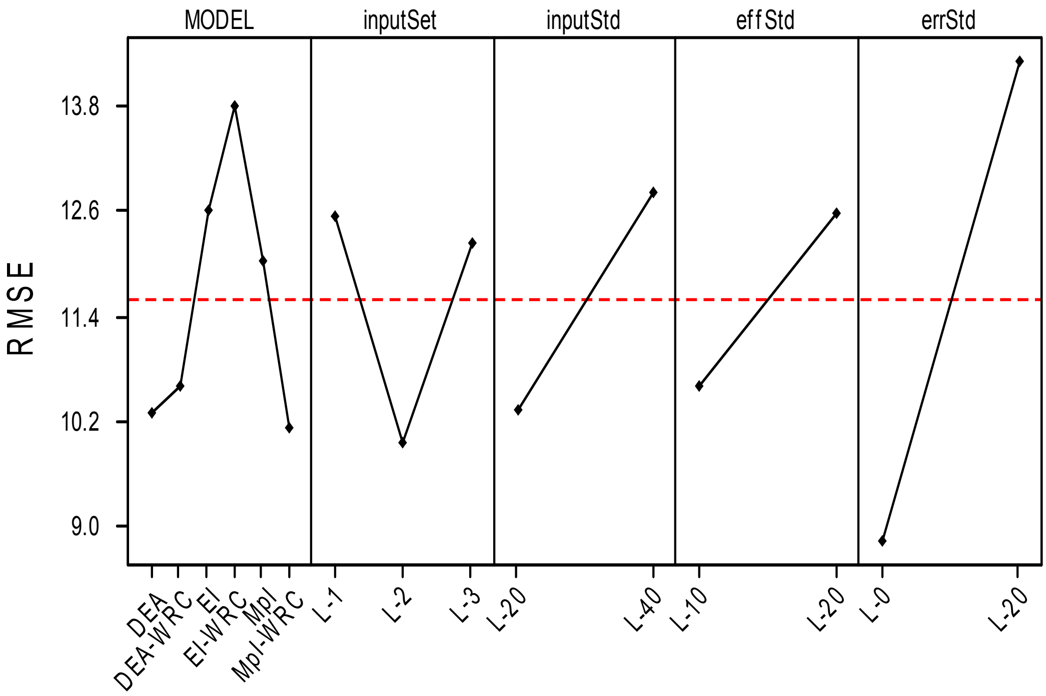

Figure 4.5. Main Effects Plot – Data Means for RMSE (Experiment II)

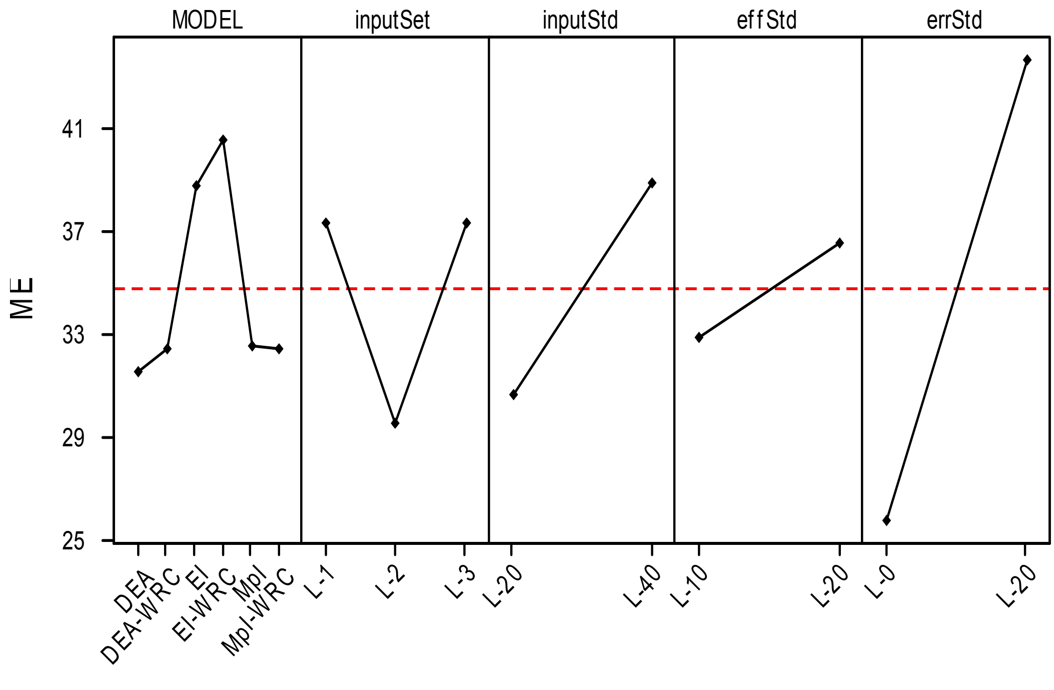

Figure 4.6. Main Effects Plot - Data Means for ME (Experiment II)

For Experiment I, two ANOVA tables for the "Model" factor are represented in Figure 4.7. MpI-WRC is the best model for RMSE and ME. We conduct Tukey, Sidak and Bonferroni tests for the pair-wise comparisons of models. The differences between the means of models are significant for all pair-wise comparisons for RMSE and ME. The only exception is that there is no significant difference between the means of EI and MpI models for ME. MpI-WRC is the best model according to both performance indicators. Using any of the newly introduced models is better than using DEA if the existence of a time lag is for sure.

For Experiment II, two ANOVA tables for the "Model" factor are represented in Figure 4.8. MpI-WRC has the lowest RMSE value; however, the differences between the means of DEA, DEA-WRC and MpI-WRC are not significant. Tukey, Sidak and Bonferroni tests for the pair-wise comparisons of models offer three groups, where DEA, DEA-WRC and MpI-WRC are in the first group, EI and MpI are in the second group and EI-WRC is in the last group. For ME, there are only two groups of models. DEA, DEA-WRC, MpI and MpI-WRC models are in the first group. MpI-WRC is the best model in RMSE and DEA in ME. However, there is no significant difference between these two models when there is no time lag.

The residual graphs and normal probability plots are generated with the ANOVA tables for all main factors. Normal distribution of residuals and equal variance of residuals for all fitted values assumptions are checked. These graphs and plots do not indicate serious violations of the assumptions. . For instance, the residual graphics of model versus ME for Experiment I and model versus RMSE for Experiment II are given in Appendix 1.

All two-way interactions between model and other factors are analyzed. All interactions are significant in both experiments. For all factors, the ranking of models change with different levels of a factor. However, only for some factors, the best model for different levels of a factor changes. The related ANOVA tables are given in Appendix 2. Again, no serious violations of normality and equal variance assumptions are detected in two-way interactions. Two residual graphics for two-way interactions are given in Appendix 3.

One-way ANOVA: RMSE versus MODEL

```
Analysis of Variance for RMSE
Source     DF         SS         MS         F        P
MODEL       5   99807.6   19961.5   736.68    0.000
Error   25914  702183.2      27.1
Total   25919  801990.8
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  -+---------+---------+---------+-----
DEA      4320    13.754     5.359                     (*)
DEA-WRC  4320    16.021     8.186                                (*)
EI       4320    10.981     3.539       (*)
EI-WRC   4320    11.756     5.252           (*)
MpI      4320    12.356     3.741              (*)
MpI-WRC  4320    10.000     3.570  (*)
                                   -+---------+---------+---------+-----
Pooled StDev =    5.205          10.0      12.0      14.0      16.0
```

One-way ANOVA: ME versus MODEL

```
Analysis of Variance for ME
Source     DF         SS         MS         F        P
MODEL       5    515506    103101   521.67    0.000
Error   25914   5121598       198
Total   25919   5637104
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  -------+---------+---------+---------
DEA      4320     39.47     14.29                           (*)
DEA-WRC  4320     43.15     18.96                                    (*)
EI       4320     33.40     11.80           (*-)
EI-WRC   4320     34.57     14.79              (*)
MpI      4320     32.65     10.43          (*)
MpI-WRC  4320     29.87     12.47   (*)
                                   -------+---------+---------+---------
Pooled StDev =    14.06                32.0      36.0      40.0
```

Figure 4.7. One-way ANOVAs: RMSE, ME versus Model (Experiment I)

One-way ANOVA: RMSE versus MODEL

```
Analysis of Variance for RMSE
Source      DF        SS        MS        F        P
MODEL        5    5258.2    1051.6    53.21    0.000
Error     2874   56800.6      19.8
Total     2879   62058.8
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N      Mean     StDev  ------+---------+---------+---------+
DEA        480    10.292     3.147    (--*-)
DEA-WRC    480    10.607     5.220      (--*-)
EI         480    12.617     4.242                   (--*--)
EI-WRC     480    13.809     5.965                           (--*--)
MpI        480    12.022     3.506               (--*--)
MpI-WRC    480    10.128     3.946   (--*-)
                                    ------+---------+---------+---------+
Pooled StDev =    4.446                10.5      12.0      13.5      15.0
```

One-way ANOVA: ME versus MODEL

```
Analysis of Variance for ME
Source      DF        SS        MS        F        P
MODEL        5     36445      7289    36.84    0.000
Error     2874    568656       198
Total     2879    605101
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N      Mean     StDev  ----+---------+---------+---------+--
DEA        480     31.54     10.64   (--*---)
DEA-WRC    480     32.47     15.87     (---*--)
EI         480     38.84     14.55                      (---*---)
EI-WRC     480     40.58     17.05                           (---*---)
MpI        480     32.61     10.04      (--*---)
MpI-WRC    480     32.49     14.80     (---*--)
                                    ----+---------+---------+---------+--
Pooled StDev =    14.07              31.5      35.0      38.5      42.0
```

Figure 4.8. One-way ANOVAs: RMSE, ME versus Model (Experiment II)

In Experiment I, the best model for ME changes for different levels of "Efficiency Standard Deviation" factor. Table 4.4 represents the first and second ranked models for both levels of this factor. According to Tukey's pair-wise comparison with a 5% family error rate, there is no significant difference between the means of EI-WRC and EI when the efficiency standard deviation is 20. Interaction plot is given in Figure 4.9. When the efficiency standard deviation is increased from 10 to 20, only EI-WRC and DEA-WRC models estimate the efficiency values better. Other models estimate better when the efficiency standard deviation is 10.

Table 4.4. Efficiency Standard Deviation Analysis for Experiment I

	Efficiency Standard Deviation			
	Low		**High**	
Rank	**Model**	**ME**	**Model**	**ME**
1	MpI	26.20	MpI-WRC	31.05
2	MpI-WRC	28.69	EI-WRC	33.85
			EI	34.69

Best model for both RMSE and ME changes for different levels of "Error Standard Deviation" factor. First and second ranked models are represented in Table 4.5. Interaction plots for RMSE and ME are given in Figures 4.10 and 4.11, respectively. It can be observed that, MpI is robust to changes in error standard deviation. It becomes the best model for higher error standard deviation. However, according to Tukey's pair-wise comparison, the difference between the means of MpI and MpI-WRC is not significant for RMSE when the error standard deviation is high.

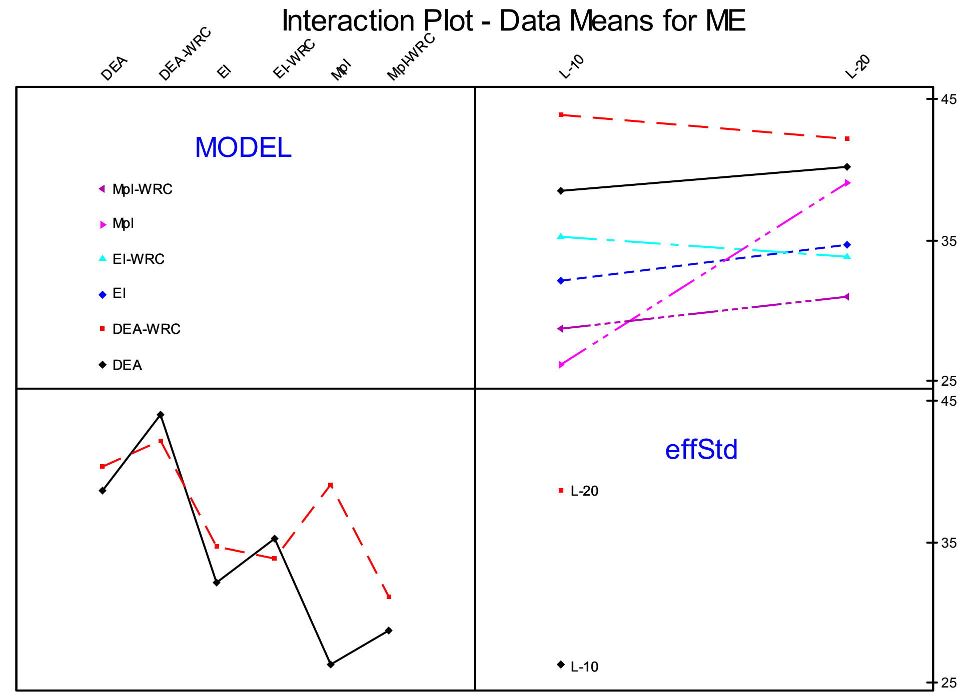

Figure 4.9. Interaction Plot of Model and Efficiency Standard Deviation – Data Means for ME (Experiment I)

Table 4.5. Error Standard Deviation Analysis for Experiment I

	Error Standard Deviation			
	Deterministic		**Probabilistic**	
Rank	**Model**	**RMSE**	**Model**	**RMSE**
1	MpI-WRC	6.91	MpI	12.82
			MpI-WRC	13.09
2	EI-WRC	8.06	EI	13.60
	EI	8.37		
Rank	**Model**	**ME**	**Model**	**ME**
1	MpI-WRC	19.33	MpI	36.14
2	EI-WRC	24.12	MpI-WRC	40.41
	EI	25.25	EI	41.54

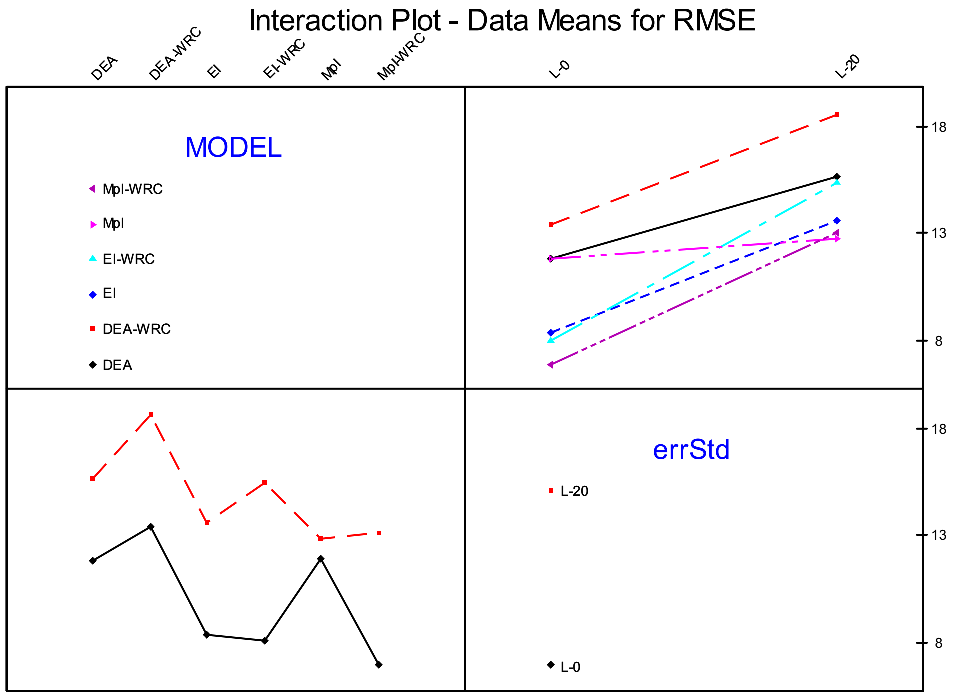

Figure 4.10. Interaction Plot of Model and Error Standard Deviation – Data Means for RMSE (Experiment I)

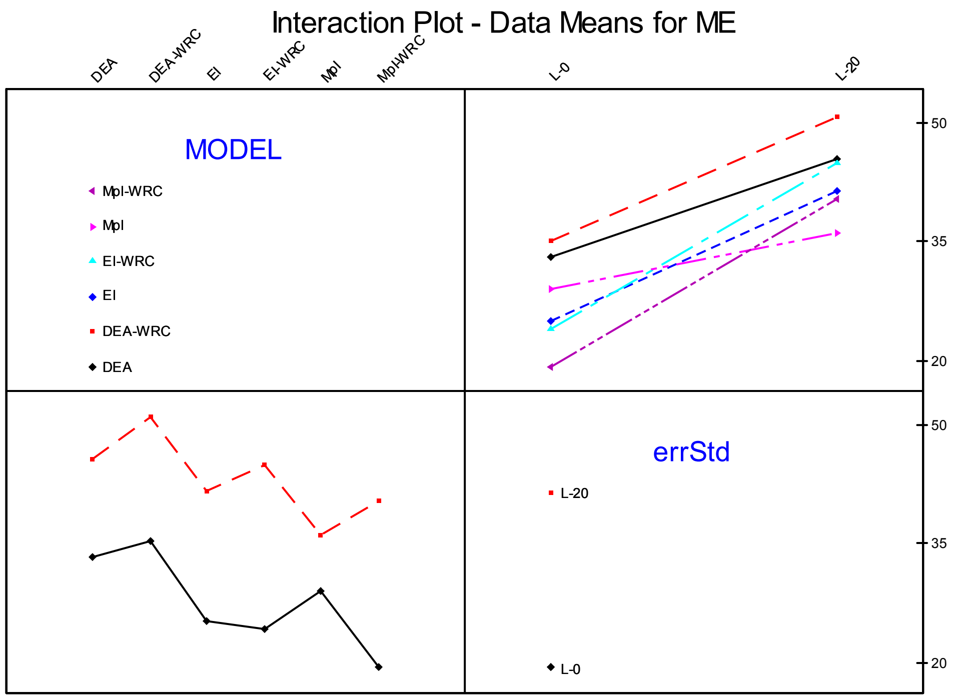

Figure 4.11. Interaction Plot of Model and Error Standard Deviation – Data Means for ME (Experiment I)

For Experiment II, all two-way interactions between model and other factors are significant. Also the best model changes as the levels of factors change. We first observe the “Input Set” factor. Ranking of models is not possible for this factor; we group the models for all levels and both indicators by the information generated by Tukey’s pair-wise comparisons. Groupings are represented in Figures 4.12a and 4.12b. Interaction plots for RMSE and ME are given in Figures 4.13 and 4.14, respectively. All models perform better if the input levels for different periods are positively correlated.

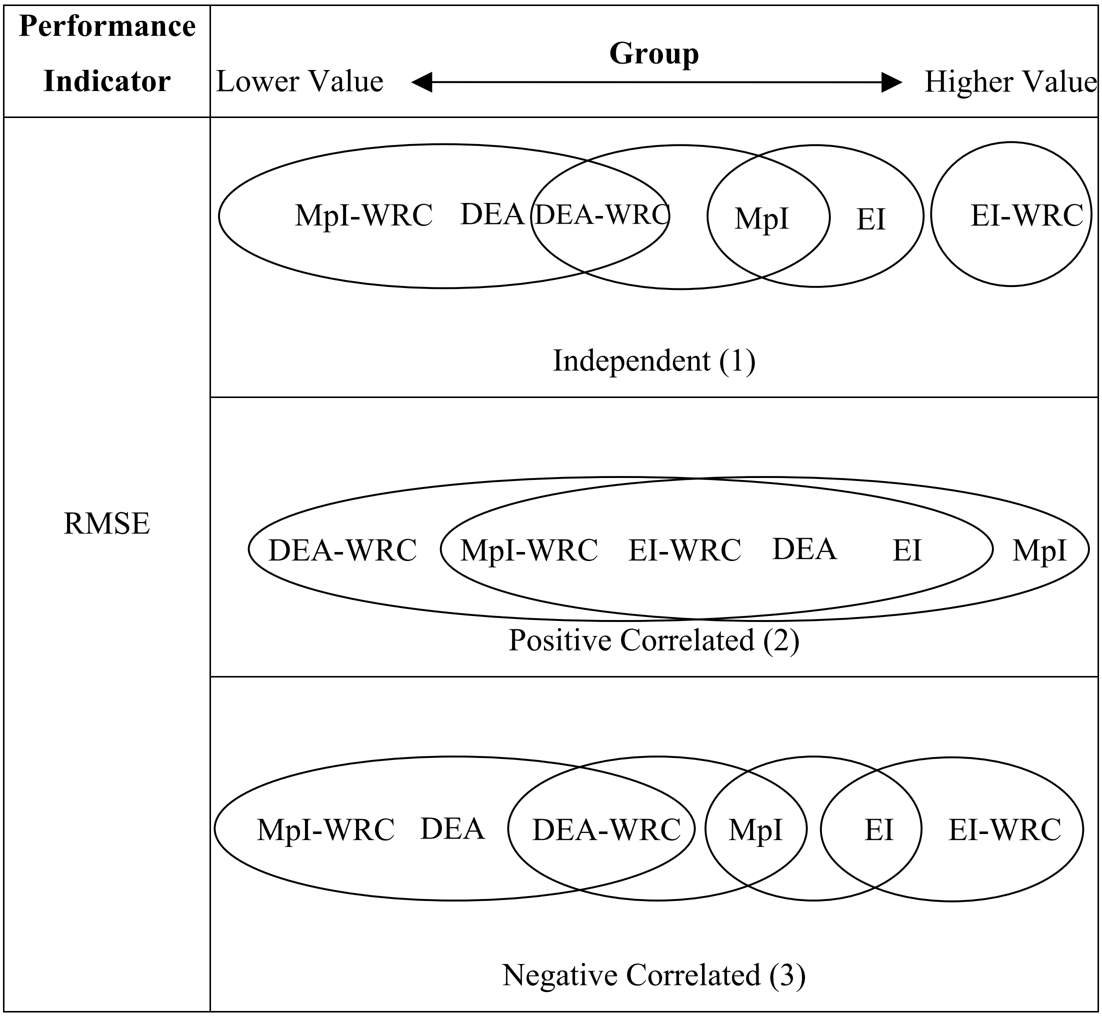

Figure 4.12a. Model Groups for Different Levels of Input Set (Experiment II)

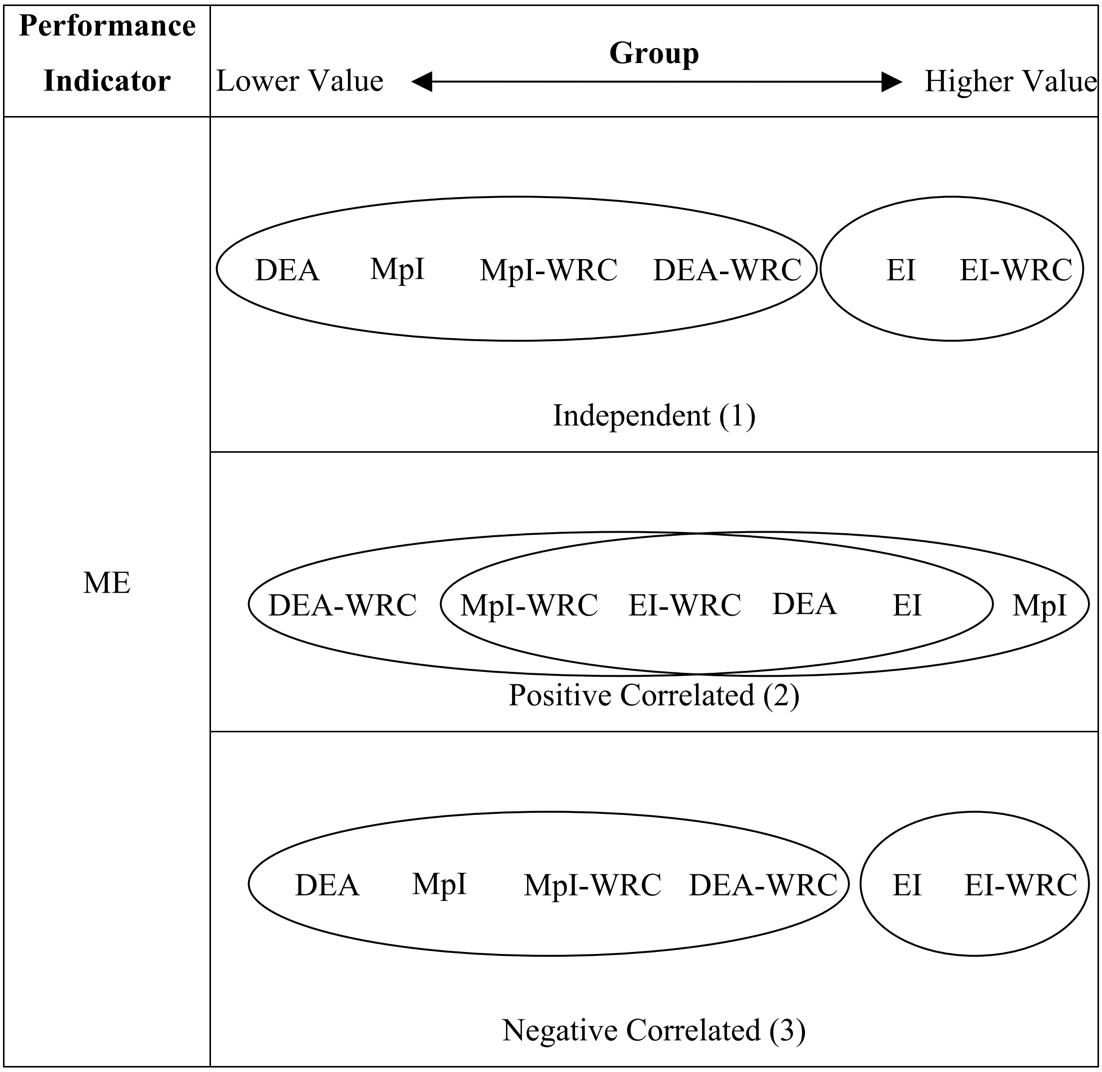

Figure 4.12b. Model Groups for Different Levels of Input Set (Experiment II)

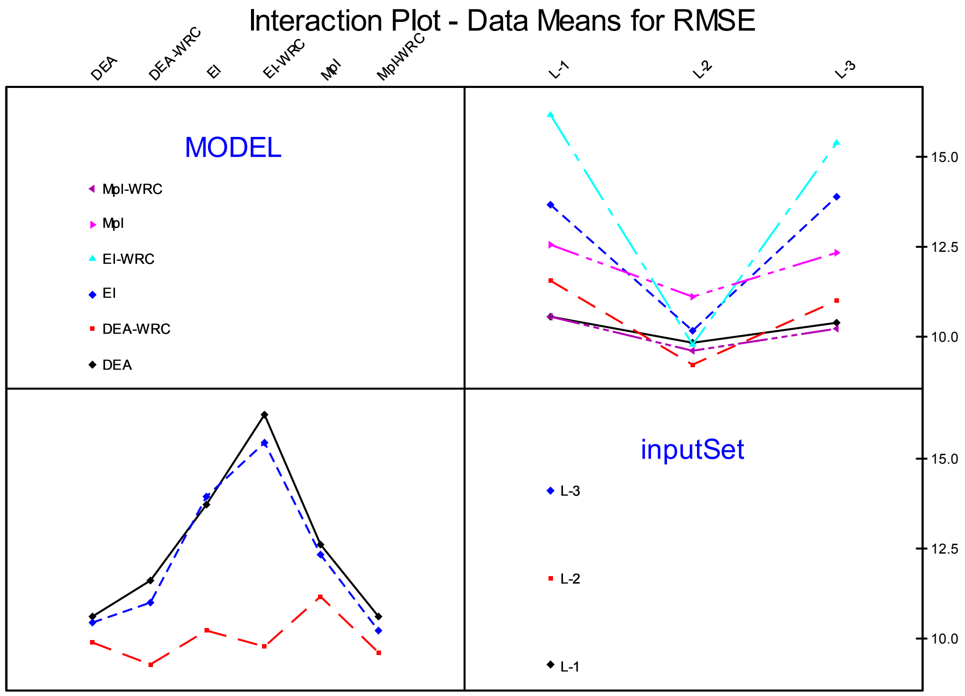

Figure 4.13. Interaction Plot of Model and Input Set – Data Means for RMSE (Experiment II)

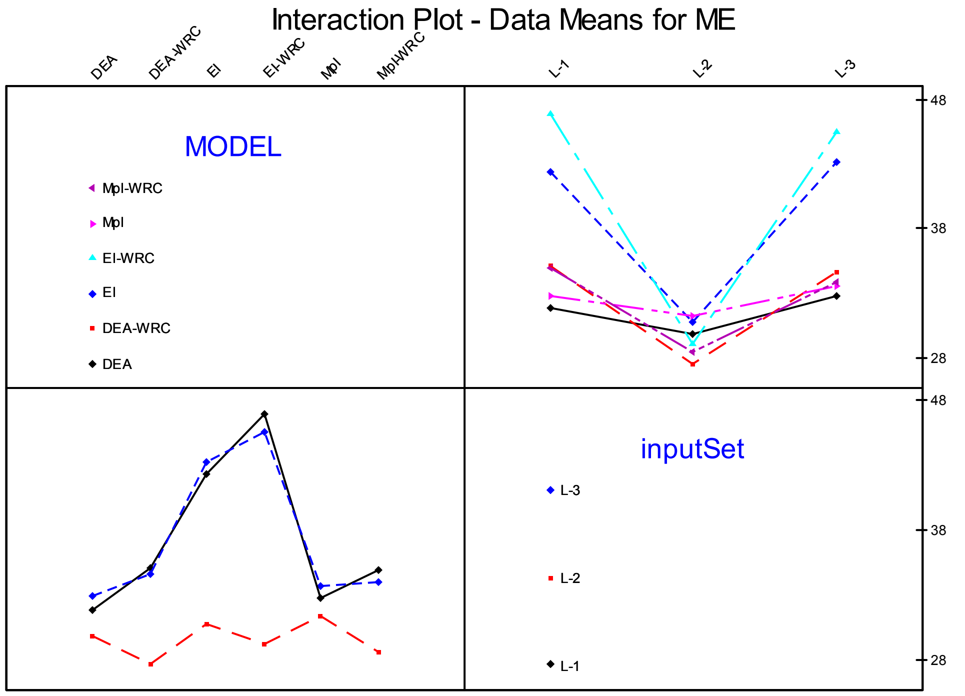

Figure 4.14. Interaction Plot of Model and Input Set– Data Means for ME (Experiment II)

The best model and the rankings of other models change for the different levels of "Input Standard Deviation" factor. It is also impossible for this factor to make a complete ranking for both indicators and levels. Therefore, we group the models as represented in Figure 4.15. Note that, MpI-WRC is always in the best group. The interaction plots are given in Figures 4.16 and 4.17.

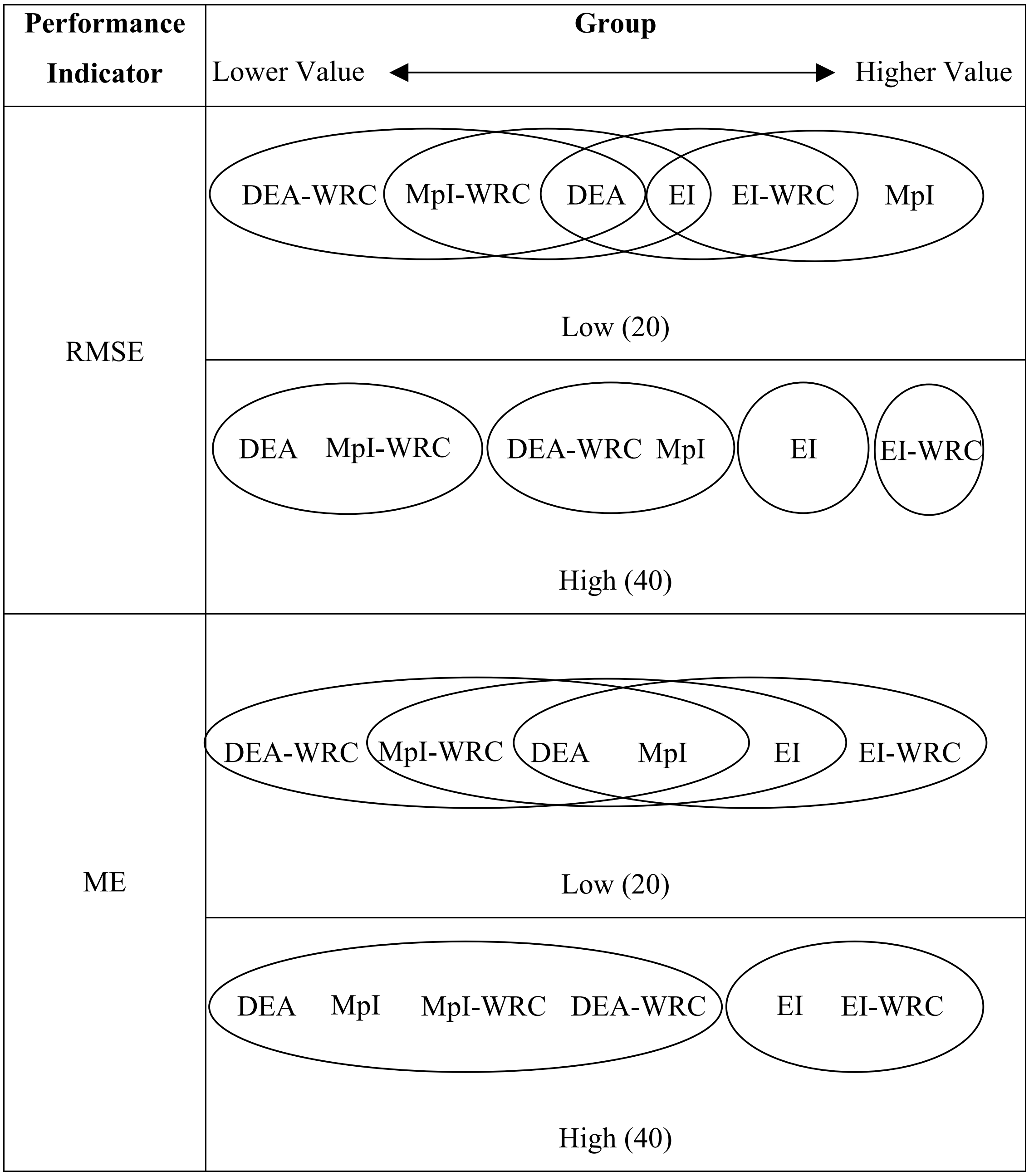

Figure 4.15. Model Groups for Different Levels of Input Standard Deviation (Experiment II)

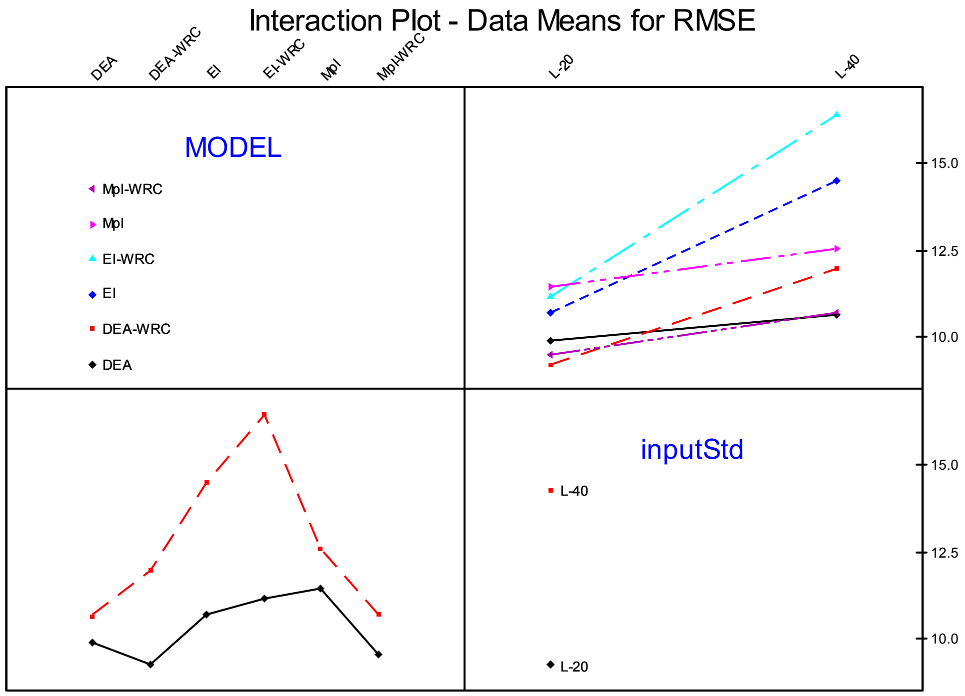

Figure 4.16. Interaction Plot of Model and Input Standard Deviation – Data Means for RMSE (Experiment II)

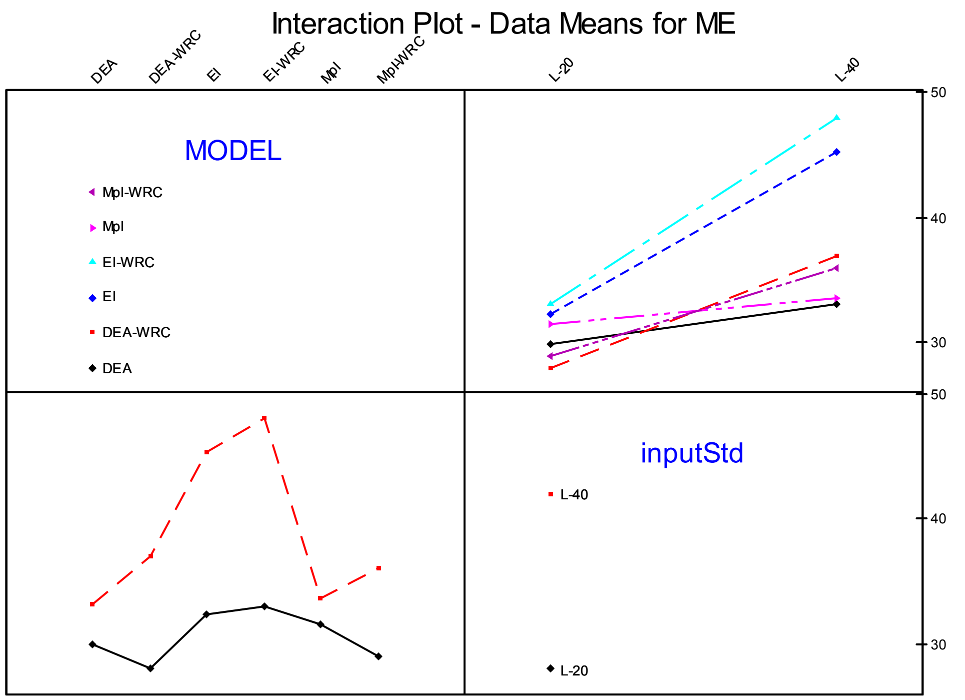

Figure 4.17. Interaction Plot of Model and Input Standard Deviation – Data Means for ME (Experiment II)

The next factor we analyzed is the “Efficiency Standard Deviation”. We represented the groups of models in Figure 4.18. The interaction plots are given in Figures 4.19 and 4.20. Also for this factor, MpI-WRC is in the best group for both factor levels and both performance indicators. MpI-WRC and DEA-WRC are robust to the changes in the efficiency standard deviation; they make good estimates independent of this factor.

Performance Indicator	**Group** Lower Value ⟷ Higher Value
RMSE	DEA MpI MpI-WRC DEA-WRC EI EI-WRC Low (10)
	MpI-WRC DEA-WRC DEA EI EI-WRC MpI High (20)
ME	MpI DEA MpI-WRC DEA-WRC EI EI-WRC Low (10)
	DEA-WRC MpI-WRC DEA MpI EI EI-WRC High (20)

Figure 4.18. Model Groups for Different Levels of Efficiency Standard Deviation (Experiment II)

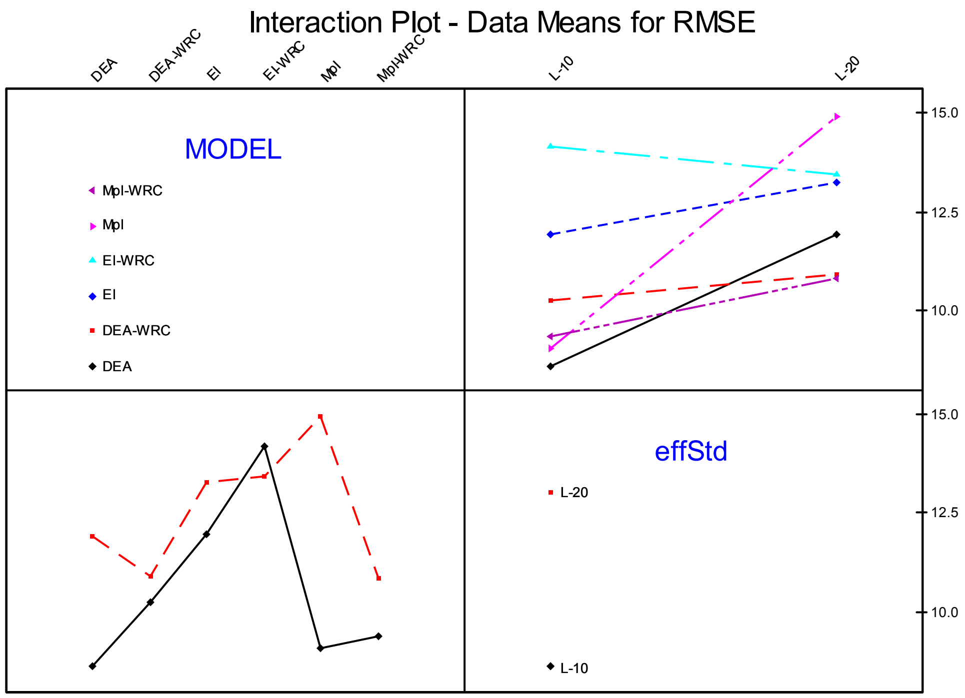

Figure 4.19. Interaction Plot of Model and Efficiency Standard Deviation – Data Means for RMSE (Experiment II)

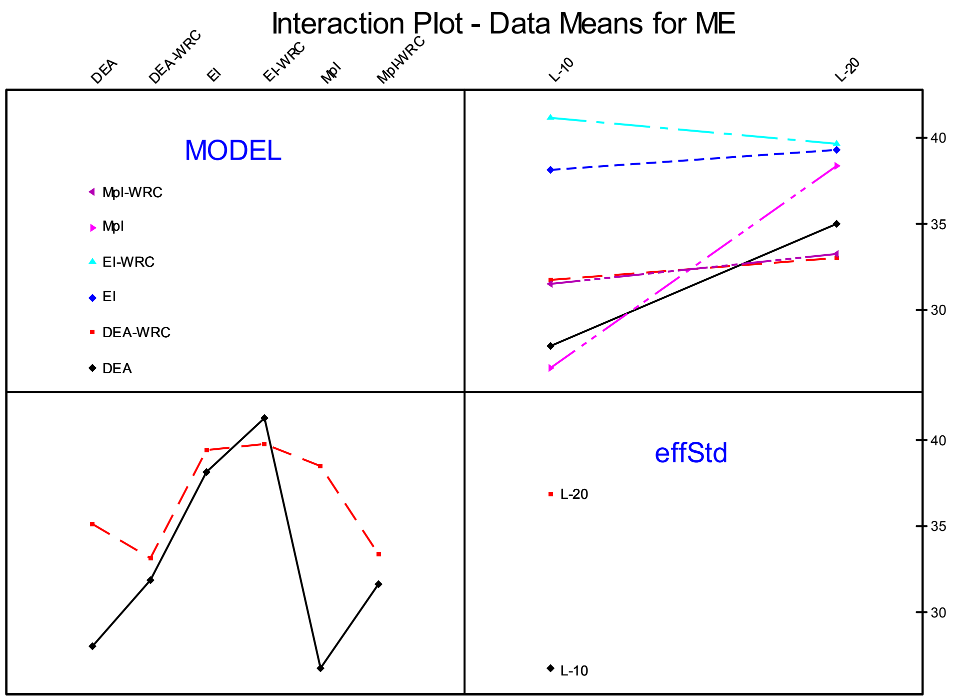

Figure 4.20. Interaction Plot of Model and Efficiency Standard Deviation – Data Means for ME (Experiment II)

The last factor is the "Error Standard Deviation". The best models as well as the rankings of models change with the error standard deviation. The grouping of models and the interaction plots are given in Figures 4.21, 4.22 and 4.23, respectively. MpI is robust to the changes in error standard deviation whereas the other models' performance measure values double.

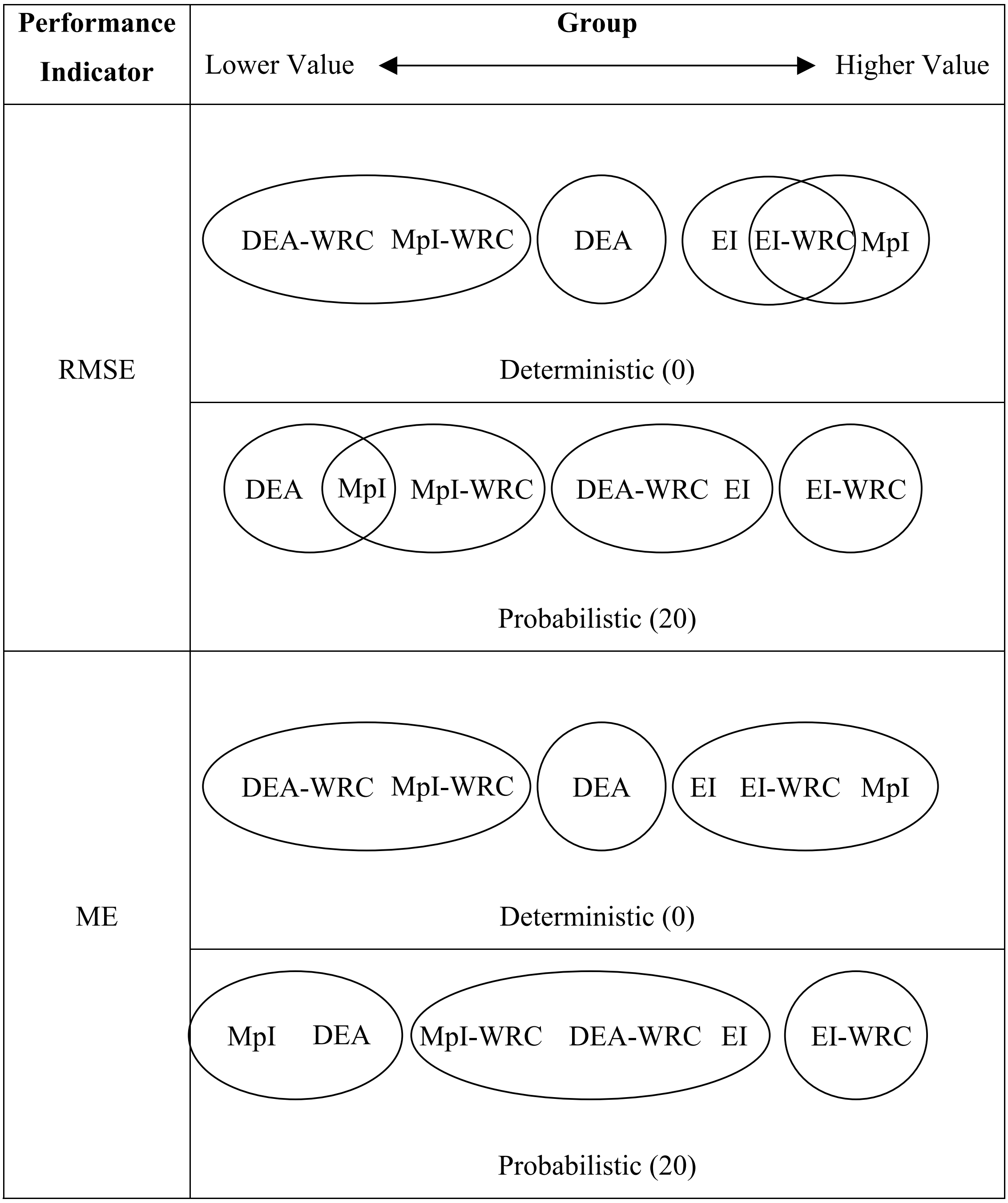

Figure 4.21. Model Groups for Different Levels of Error Standard Deviation (Experiment II)

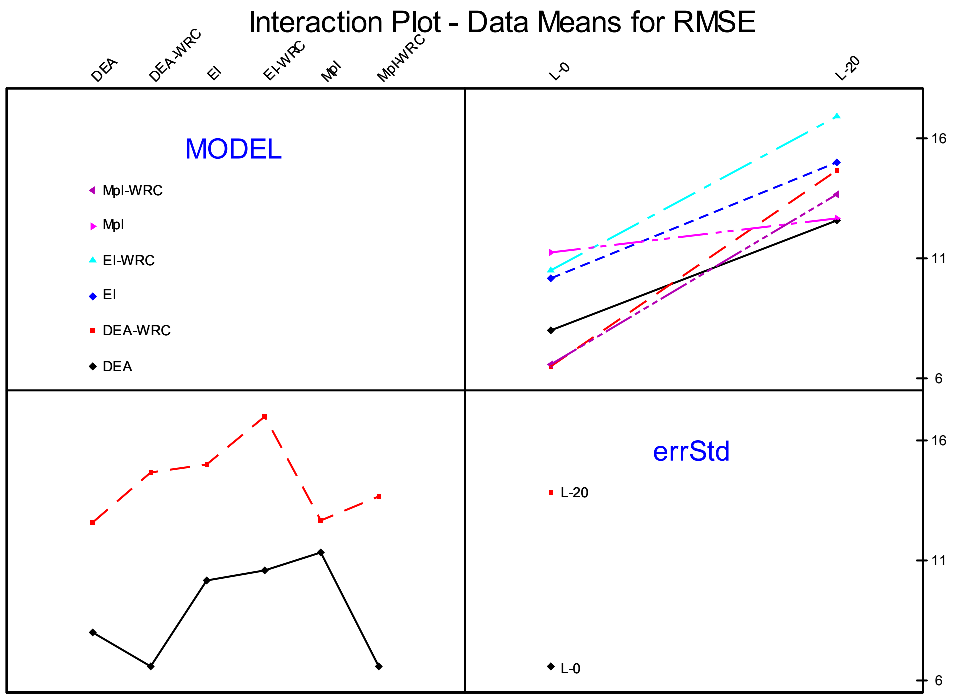

Figure 4.22. Interaction Plot of Model and Error Standard Deviation – Data Means for RMSE (Experiment II)

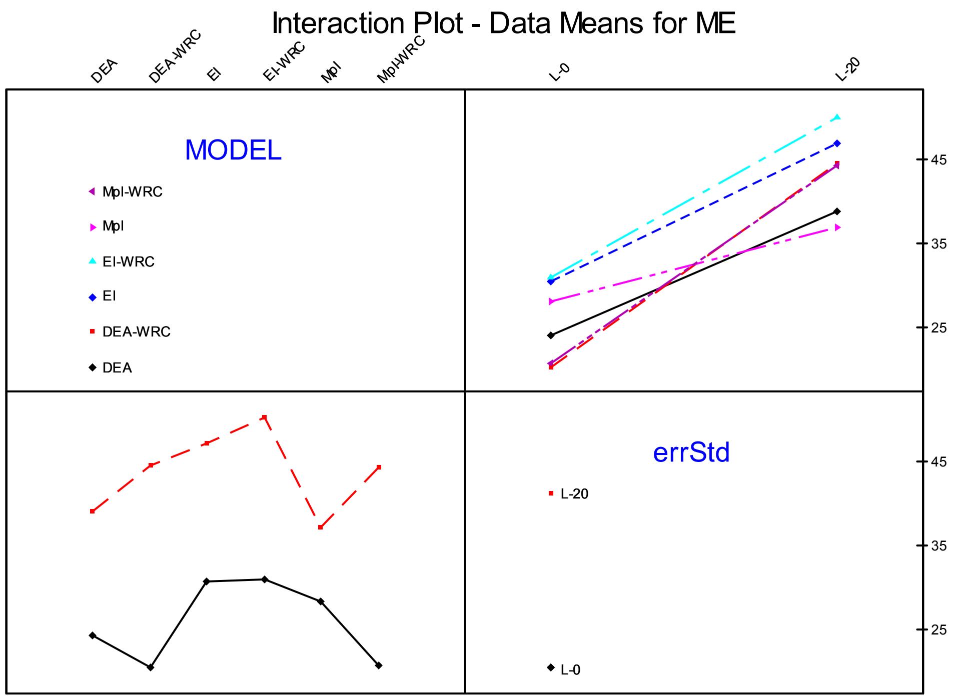

Figure 4.23. Interaction Plot of Model and Error Standard Deviation – Data Means for ME (Experiment II)

We represent the results of the experiments. We analyze the overall effects of uncontrollable factors and the effects of uncontrollable factors levels on the performances of models. It is seen that, MpI-WRC is the best model if the existence of a time lag is known. However, if there is no time lag, no model can be identified as the best model. For different levels of factors, the best model changes. On the other hand, when an overall comparison is made, there is no significant difference between MpI-WRC and DEA models.

Finally, we analyze the experiment where we assign the efficiency values randomly. We intend to see how much the approaches improve over the case where the efficiency values are assigned arbitrarily. On the average, assigning efficiency values randomly will result with a RMSE value of 42.20. RMSE value increases when the efficiency standard deviation decreases. A similar pattern is observed for ME. Results of random assignment are represented in Table 4.6. It is obvious that, using even the worst model is better than assigning the efficiency values randomly.

Table 4.6. Performance of random assignment of efficiency values

	Performance Indicator	
Efficiency Standard Deviation	**RMSE**	**ME**
Low	44.54	88.42
High	39.85	87.46
Average	42.20	87.94

CHAPTER 5

TÜBİTAK APPLICATION

In this chapter, we first introduce the Scientific and Technical Research Council of Turkey (TÜBİTAK) and the research and development (R&D) institutes of it. Afterwards, we propose a method for the performance evaluation for the R&D institutes of TÜBİTAK. Finally, we conduct an assessment with the readily available data and report the results of it.

5.1 About TÜBİTAK

TÜBİTAK is founded in 1963 and responsible for the promotion, development, organization and coordination of research and development in the exact science fields. It functions under the Prime Ministry. Science Board, with 12 members and the President, is the main decision maker of TÜBİTAK. President is also the chairman of the Science Board and responsible for the implementation of the decisions made in the Science Board.

TÜBİTAK performs the following tasks as presented in its web site (2004):

- "determining Turkey's science and technology policies;
- supporting, encouraging and coordinating scientific research;

- establishing and operating special institutes to conduct research and development activities geared to the targets of the five-year economic development plans and the priorities set by the Science Board;
- providing scholarships and other kinds of support to researchers and organizing contests to discover and train future scientists;
- supporting R&D activities and innovations in industry, promoting university-industry collaborations and establishing techno-parks to facilitate their realization;
- implementing tasks undertaken through international scientific and technical cooperation agreements;
- publishing scientific journals, as well as books and monthly popular science magazines that make science accessible to the public;
- supporting scientists and researchers with awards and programs that incent scientific publication;”

TÜBİTAK has two types of institutes. In the first type of institutes, R&D activities are performed. The second type of institutes provides services for R&D activities. TÜBİTAK has 7 first type institutes and 7 second type institutes. The R&D institutes are:

- Marmara Research Center (MAM),
- Information Technologies and Electronics Research Institute (BİLTEN),
- Defense Industry R&D Institute (SAGE),
- National Research Institute of Electronics and Cryptology (UEKAE),
- Feza Gürsey Institute (TBAE),
- Research Institute of Gene Engineering and Biotechnology (GMBAE),
- Çukurova Advanced Agriculture Technologies R&D Institute (ÇİTTAGE).

R&D Support Intuitions are:

- National Metrology Institute (UME)
- Turkish National Academic Network and Information Center (ULAKBİM)
- TÜBİTAK National Observatory (TUG)
- Ankara Test and Analysis Laboratory (ATAL)
- TUBITAK DNA/Cell Bank & Gene Research Laboratory

- Bursa Test and Analysis Laboratory (BUTAL)
- Turkish Institute for Industrial Management (TÜSSİDE)

Institutes send their budget requests to TÜBİTAK. Requested budgets are analyzed and approved by TÜBİTAK. Institutes have the authority to use allocated budgets in the line of their visions without violating the TÜBİTAK law. We treat these institutes as decision making units (DMUs).

After mentioning the types of TÜBİTAK institutions, we briefly state the method offered for the performance evaluation method of these institutes.

5.2 Performance Evaluation and Results

In this section we describe the study in which real TÜBİTAK data is used and report the results. We perform a simple study in order to demonstrate how the MpI-WRC can be used for the performance evaluation of R&D institutes of TÜBİTAK if the data is readily available. In this study, we use the data in the TÜBİTAK's Annual Activity Reports for the period of 1998-2003.

We use two inputs and a single output data. The first input is the actual budget. This is the total amount of money spent during the reporting period. This amount is the summation of personnel costs, provisions, infrastructure expenses and costs of acquired services.

The second input is the number of researchers employed. The researcher is one of the five types of employees in TÜBİTAK. The cost of researchers is reported in personnel costs and actual budget includes personnel costs. Although there seems to be a duplicate use of researcher information, we think using these two inputs provides better information since researchers perform most of the value-added tasks and are one of the main sources of scientific and/or economic outputs.

We use the actual income of the institute during the reporting period as the single output. This amount does not include the money allocated to the institute by

TÜBİTAK. We collect the data for 5 R&D and 3 R&D support institutes of TÜBİTAK as well as the TÜBİTAK Headquarters itself.

Actual budget and actual income values are converted to the 1994's monetary terms by "other sectors" deflator. The number of researchers and deflated data for actual income and actual budget are given in Appendix 4. The actual names of institutes are not used since this part of the study is done not to evaluate the performances of institutes but to represent the use of the model. We should emphasize that the results are not expected to be meaningful since we only use a subset of the meaningful inputs and outputs; those for which data are readily available.

We use the data of 1998-2003 and assume that the time lag is one year. We get the efficiency values for years 1999-2003. We first solve the MpI model and then use the efficiency values in MpI-WRC model. The efficiency values estimated by MpI-WRC are given in Table 5.1. It can be seen that, there is not any stability in the efficiency values of the DMUs. This may be due to lack of inputs and outputs spanning many of the activities done by these institutes. We think that using more inputs and outputs combined under some generic criteria, not only will the efficiency values of DMUs get more stable but also increase to some extend from the extremely low values of 0.1s.

Although we represent an evaluation with a few indicators, TÜBİTAK may use many of them. These performance indicators can be the easily measurable ones such as number of patents or annual project income. Then, these indicators should be combined in higher level indicators such as quality of research or impact of research. The decision makers of TÜBİTAK should decide the measurable indicators, higher level indicators and their relations.

Table 5.1. Efficiency values estimated by MpI-WRC

	Years				
DMU	**1999**	**2000**	**2001**	**2002**	**2003**
1	39.5	100.0	100.0	100.0	70.4
2	46.7	13.0	34.2	80.3	93.6
3	10.9	11.2	37.0	100.0	87.1
4	47.5	82.1	59.2	27.2	25.4
5	9.7	8.2	19.7	14.1	0.1
6	2.7	13.4	0.6	2.2	3.7
7	0.3	0.9	1.2	8.6	8.9
8	1.1	87.6	59.2	100.0	100.0
9	100.0	100.0	13.6	30.9	0.1

We want to call attention to the interpretation of results for the second time. 100% efficient DMUs are the high performers and utilizing their inputs in the best way to produce outputs. The inefficient DMUs have the possibility of increasing their outputs with the same level of inputs or producing same level of outputs by using fewer inputs.

In Table 5.1, the efficiency values for R&D institutes are represented. As it can be observed, some of the efficiency values are very low and also they are not stable over periods. The actual income is a meaningful outcome, but is not the only outcome of these institutes. Scientific outputs such as number of patents and number of articles should also be considered. Also we may add the number of projects completed in a period and the total budget of those projects. A similar comment is possible in input side. We may add some information about the scientific infrastructure such as the number of laboratories and the total cost of scientific equipment. There could be other relevant inputs and outputs. When such relevant inputs and/or outputs are omitted DEA-related approaches would give results that are not very meaningful as is the case in Table 5.1.

CHAPTER 6

CONCLUSIONS AND FURTHER RESEARCH AREAS

In this study, we focus on the performance evaluation of decision making units such as the R&D institutes and the universities. We represent the use of the Data Envelopment Analysis and introduce new models and ideas as extensions to it. We compare these models and ideas by designed experiments. Finally, we use the best model for a sample performance evaluation of the R&D institutes of TÜBİTAK with the readily available data.

Basic DEA models assume that DMUs consume the inputs and produce the outputs in the same reporting period. However, the research activities produce the outputs with a time lag. Knowing this fact, we offered two new models, namely Multi-period Input (MpI) model and Effective Input (EI) model.

We also introduce the weight range constraints (WRC) and a new objective function applicable to DEA, MpI and EI models. WRC are a new type of weight restriction constraints. WRC force the model not to assign very diverse input or output weights for different DMUs. Optimal weights lie in a reasonable range. The center of the range for an input or output is the average of optimal weights of that input or output for all DMUs.

Using WRC with "maximize total efficiency" objective favors similar DMUs and DMUs having extremely high or low levels of inputs or outputs are punished by getting low efficiency values. We offer to use a new objective function for the models with WRC. This objective function minimizes the maximum deviation of efficiency values from the efficiency values assigned by the model without WRC.

In DEA-WRC, MpI-WRC and EI-WRC models, we use both WRC and new objective function. These models not only find optimal weights within reasonable ranges but also minimize the maximum deviation of efficiency values.

We design two experiments. The first one represents a time lag. We do not use any time lag in the Experiment II. We compare six models with the two experiments. If the existence of time lag is for sure, MpI-WRC model is the best model. However, if there is no time lag, MpI-WRC, DEA and DEA-WRC models are all in the best models set and they have no significant difference. In general, MpI-WRC model dominates the DEA model.

Our study can be extended to a number of research areas, some of which are mentioned below:

- In addition to WRC, decision maker's preference information can be incorporated in models.
- The optimal solution of a DEA model includes the information of possible increases in outputs and decreases in inputs in order to make a DMU efficient. A further research area may be to make a similar analysis using the WRC idea and to further study the implications of WRC.
- In TÜBİTAK application, we use two types of institutes (R&D and R&D support) as well as the TÜBİTAK Headquarters as DMUs in the same model. Further research may be done to find a way of taking differences in the types of DMUs into account.

REFERENCES

[1] Suwignjo, P., Bititci, U.S., Carrie, A.S., 2000, ‘Quantitative Models for Performance Measurement Systems’, International Journal of Production Economics 64, 231-241.

[2] Bititci, U.S., Suwignjo, P., Carrie, A.S., 2001, ‘Strategy Management Through Quantitative Modeling of Performance Measurement Systems’, International Journal of Production Economics 69, 15-22.

[3] Sarkis, J., 2003, ‘Quantitative Models for Performance Measurement Systems--Alternate Considerations’, International Journal of Production Economics 86, 81-90

[4] Spronk, J., Vermeulen, E.M., 2003, ‘Comparative Performance Evaluation under Uncertainty’, European Journal of Operational Research 150, 482-495.

[5] Önel, A., Saatçioğlu, Ö., 1995, ‘Trend analysis and regression modeling for the assessment of research performance at a Turkish University’, Transactions on Operational Research 7, 1-12

[6] Fandel, G., Gal, T., 2001, ‘Redistribution of funds for teaching and research among universities: The case of North Rhine-Westphalia’, European Journal of Operational Research 130, 111-120

[7] Farrell, M.J., 1957, ‘The Measurement of Productive Efficiency’, Journal of the Royal Statistical Society A 120, 253-281.

[8] Charnes, A., Cooper, W.W., Rhodes, E., 1978, ‘Measuring the efficiency of decision-making units’, European Journal of Operational Research 2, 429-444

[9] Post, T., Spronk, J., 1999, ‘Performance benchmarking using interactive data envelopment analysis’, European Journal of Operational Research 115, 472-487

[10] Halme, M., Joro, T., Korhonen, P., Salo, S., Wallenius, J., 1999, ‘A value efficiency approach to incorporating preference information in data envelopment analysis’, Management Science 45, 103-115

[11] Korhonen, P., Tainio, R., Wallenius, J., 2001, ‘Value efficiency analysis of academic research’, European Journal of Operational Research 130, 121-132

[12] Kornbluth, J.S.H., 1991, ‘Policy effectiveness using cone restricted data envelopment analysis’, The Journal of the Operational Research Society 42, 1097-1104

[13] Dyson, R.G., Thanassoulis, E., 1988, ‘Reducing weight flexibility in data envelopment analysis’, The Journal of the Operational Research Society 39, 563-576

[14] Wong, Y.-H.B., Beasley, J.E., 1990, ‘Restricting weight flexibility in data envelopment analysis’, The Journal of the Operational Research Society 39, 829-835

[15] Thompson, R.G., Langemeier, L.N., Lee, C.T., Lee, E., Thrall, R.M., 1990, ‘The role of multiplier bounds in efficiency analysis with application to Kansan farming’, Journal of Econometrics 46, 93-108

[16] Kleine, A., 2004, ‘A general model framework for DEA’, Omega 32, 17-23

[17] Joro, T., Korhonen, P., Wallenius, J., 1998, 'Structural comparison data envelopment analysis and multiple objective linear programming', Management Science 44, 962-970

[18] Yeh, Q.-J., 1996, 'The application of data envelopment analysis in conjunction with financial ratios for bank performance evaluation', Journal of the Operational Research Society 47, 980-988

[19] Golany, B., Storbeck, J.E., 1999, 'A data envelopment analysis of the operational efficiency of bank branches', Interfaces, 14-26

[20] Karlaftis, M.G., 2004, 'A DEA approach for evaluating the efficiency and effectiveness of urban transit systems', European Journal of Operational Research 152, 354-364

[21] Khouja, M., 1995, 'The use of data envelopment analysis for technology selection', Computers ind. Engng 28, 123-132

[22] TÜBİTAK – Türkiye Bilimsel ve Teknik Araştırma Kurumu, 2004, www.tubitak.gov.tr

APPENDICES

A. Residual Graphics (Main Effect)

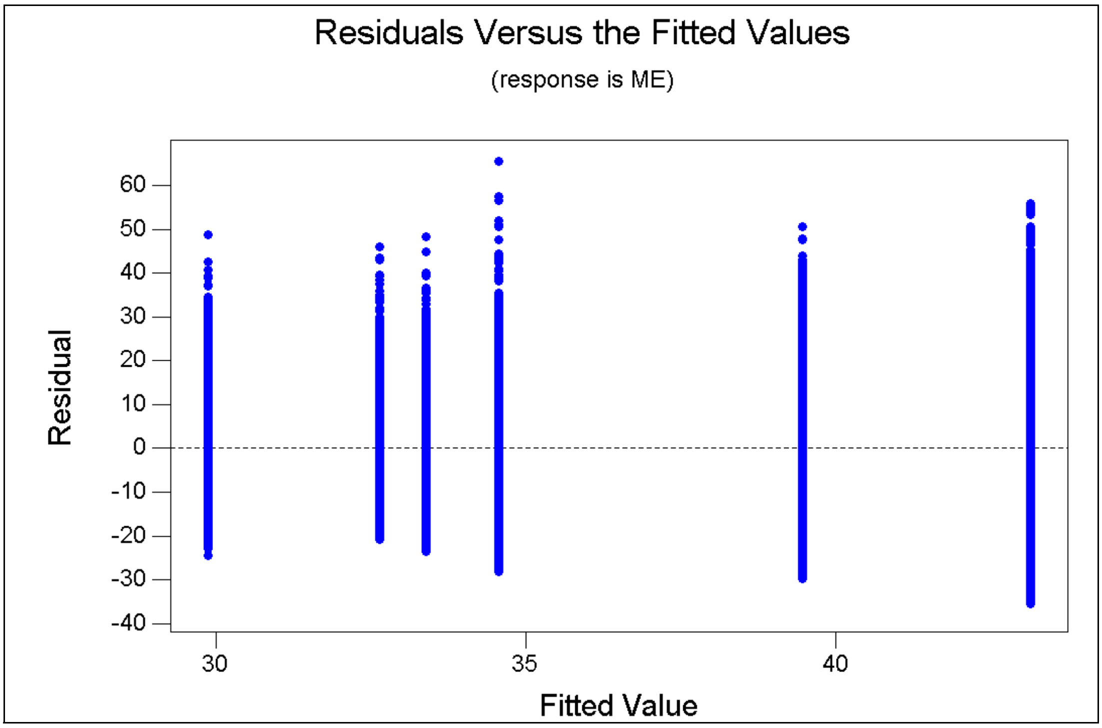

Figure A.1 : Residual versus Fitted Values for ME (Experiment I)

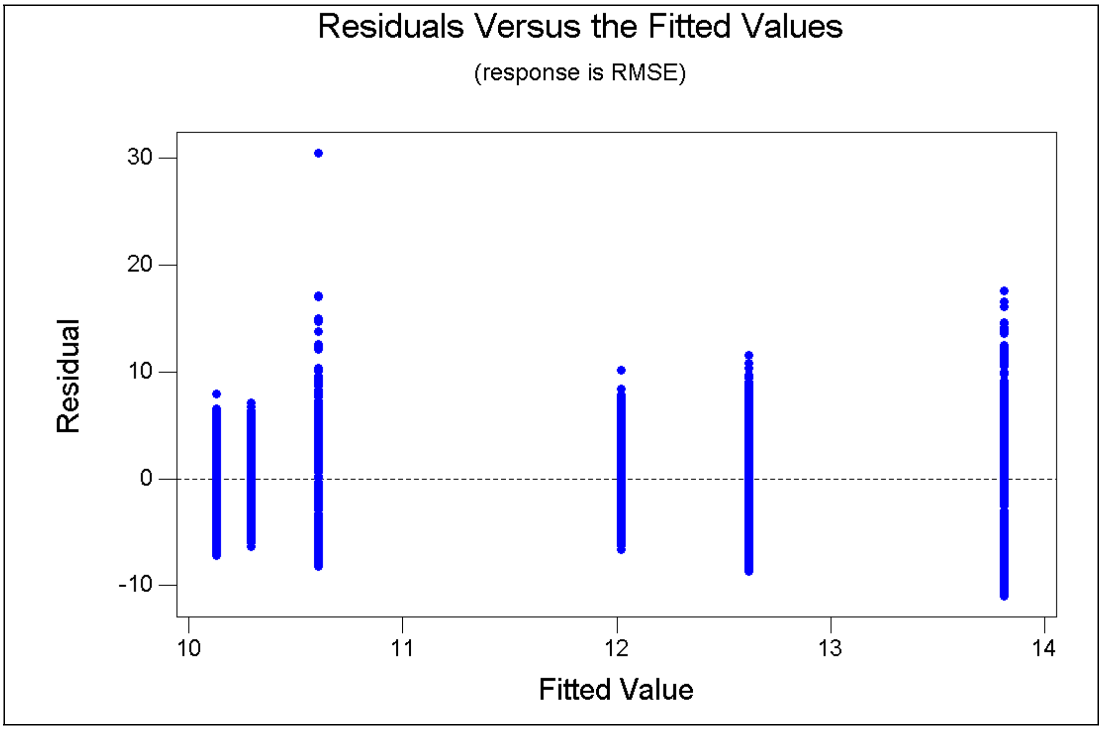

Figure A.2 : Residual versus Fitted Values for RMSE (Experiment II)

B. ANOVA Tables

Two-way ANOVA: RMSE versus MODEL, inputSet

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	99807.6	19961.5	880.45	0.000
inputSet	2	67776.2	33888.1	1494.71	0.000
Interaction	10	47157.7	4715.8	208.00	0.000
Error	25902	587249.3	22.7		
Total	25919	801990.8			

Two-way ANOVA: ME versus MODEL, inputSet

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	515506	103101	586.97	0.000
inputSet	2	308567	154284	878.36	0.000
Interaction	10	263349	26335	149.93	0.000
Error	25902	4549681	176		
Total	25919	5637104			

Two-way ANOVA: RMSE versus MODEL, inputStd

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	99807.6	19961.5	860.61	0.000
inputStd	1	67441.6	67441.6	2907.65	0.000
Interaction	5	33817.1	6763.4	291.59	0.000
Error	25908	600924.5	23.2		
Total	25919	801990.8			

Two-way ANOVA: ME versus MODEL, inputStd

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	515506	103101	585.48	0.000
inputStd	1	390702	390702	2218.67	0.000
Interaction	5	168564	33713	191.44	0.000
Error	25908	4562331	176		
Total	25919	5637104			

Figure B.1. ANOVA Tables of Experiment I

Two-way ANOVA: RMSE versus MODEL, wSet

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	99807.6	19961.5	810.43	0.000
wSet	2	29923.2	14961.6	607.43	0.000
Interaction	10	34274.3	3427.4	139.15	0.000
Error	25902	637985.7	24.6		
Total	25919	801990.8			

Two-way ANOVA: ME versus MODEL, wSet

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	515506	103101	558.02	0.000
wSet	2	201347	100674	544.88	0.000
Interaction	10	134554	13455	72.83	0.000
Error	25902	4785697	185		
Total	25919	5637104			

Two-way ANOVA: RMSE versus MODEL, effStd

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	99807.6	19961.5	792.91	0.000
effStd	1	15134.1	15134.1	601.16	0.000
Interaction	5	34817.9	6963.6	276.61	0.000
Error	25908	652231.3	25.2		
Total	25919	801990.8			

Two-way ANOVA: ME versus MODEL, effStd

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	515506	103101	542.92	0.000
effStd	1	47893	47893	252.20	0.000
Interaction	5	153728	30746	161.90	0.000
Error	25908	4919977	190		
Total	25919	5637104			

Figure B.1. ANOVA Tables of Experiment I (cont)

Two-way ANOVA: RMSE versus MODEL, errStd

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	99807.6	19961.5	983.92	0.000
errStd	1	149629.3	149629.3	7375.32	0.000
Interaction	5	26936.6	5387.3	265.54	0.000
Error	25908	525617.4	20.3		
Total	25919	801990.8			

Two-way ANOVA: ME versus MODEL, errStd

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	515506	103101	785.99	0.000
errStd	1	1569131	1569131	1.2E+04	0.000
Interaction	5	154034	30807	234.86	0.000
Error	25908	3398433	131		
Total	25919	5637104			

Two-way ANOVA: RMSE versus MODEL, TimeLag

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	99807.6	19961.5	745.92	0.000
TimeLag	2	5070.9	2535.4	94.74	0.000
Interaction	10	3948.6	394.9	14.75	0.000
Error	25902	693163.8	26.8		
Total	25919	801990.8			

Two-way ANOVA: ME versus MODEL, TimeLag

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	515506	103101	535.11	0.000
TimeLag	2	94075	47037	244.13	0.000
Interaction	10	36943	3694	19.17	0.000
Error	25902	4990580	193		
Total	25919	5637104			

Figure B.1. ANOVA Tables of Experiment I (cont)

Two-way ANOVA: RMSE versus MODEL, inputSet

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	5258.2	1051.6	59.43	0.000
inputSet	2	3774.3	1887.1	106.64	0.000
Interaction	10	2379.3	237.9	13.45	0.000
Error	2862	50647.0	17.7		
Total	2879	62058.8			

Two-way ANOVA: ME versus MODEL, inputSet

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	36445	7289	40.83	0.000
inputSet	2	39037	19518	109.33	0.000
Interaction	10	18698	1870	10.47	0.000
Error	2862	510921	179		
Total	2879	605101			

Two-way ANOVA: RMSE versus MODEL, inputStd

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	5258.2	1051.6	59.85	0.000
inputStd	1	4456.5	4456.5	253.63	0.000
Interaction	5	1951.4	390.3	22.21	0.000
Error	2868	50392.7	17.6		
Total	2879	62058.8			

Two-way ANOVA: ME versus MODEL, inputStd

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	36445	7289	41.46	0.000
inputStd	1	48476	48476	275.76	0.000
Interaction	5	16003	3201	18.21	0.000
Error	2868	504177	176		
Total	2879	605101			

Figure B.2. ANOVA Tables of Experiment II

Two-way ANOVA: RMSE versus MODEL, effStd

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	5258.2	1051.6	59.34	0.000
effStd	1	2818.3	2818.3	159.03	0.000
Interaction	5	3156.6	631.3	35.62	0.000
Error	2868	50825.8	17.7		
Total	2879	62058.8			

Two-way ANOVA: ME versus MODEL, effStd

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	36445	7289	38.38	0.000
effStd	1	9429	9429	49.65	0.000
Interaction	5	14545	2909	15.32	0.000
Error	2868	544682	190		
Total	2879	605101			

Two-way ANOVA: RMSE versus MODEL, errStd

Analysis of Variance for RMSE

Source	DF	SS	MS	F	P
MODEL	5	5258.2	1051.6	94.65	0.000
errStd	1	21453.8	21453.8	1930.79	0.000
Interaction	5	3479.3	695.9	62.63	0.000
Error	2868	31867.5	11.1		
Total	2879	62058.8			

Two-way ANOVA: ME versus MODEL, errStd

Analysis of Variance for ME

Source	DF	SS	MS	F	P
MODEL	5	36445	7289	65.66	0.000
errStd	1	229935	229935	2071.30	0.000
Interaction	5	20345	4069	36.65	0.000
Error	2868	318376	111		
Total	2879	605101			

Figure B.2. ANOVA Tables of Experiment II (cont)

C. Residual Graphics (Two-way Interactions)

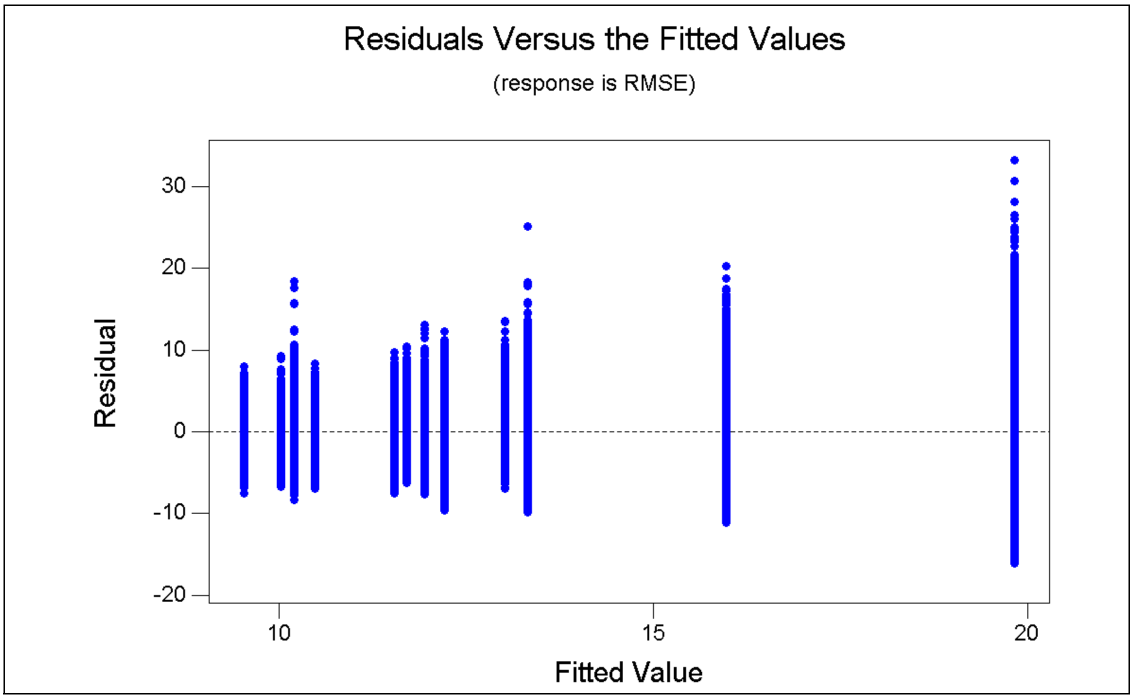

Figure C.1 : Residual versus Fitted Values (ModelinputStd) for RMSE (Experiment I)

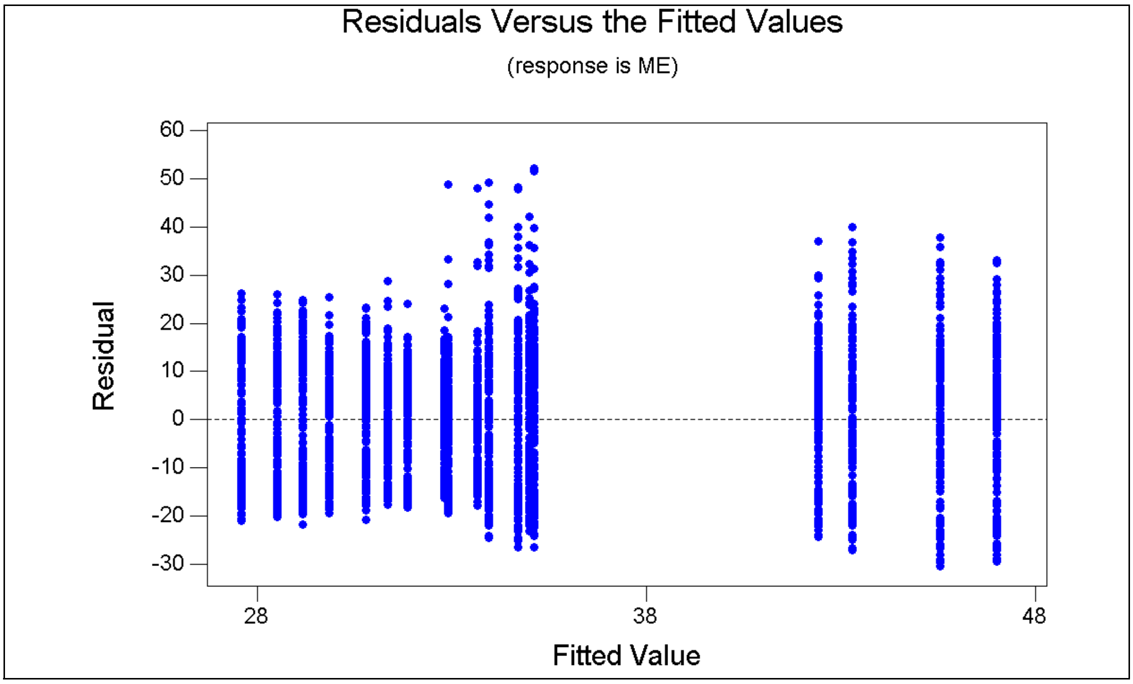

Figure C.2 : Residual versus Fitted Values (ModelInputSet) for ME (Experiment II)

D. TÜBİTAK Data

Table D.1.Actual Budget Values

Actual Budget (Deflated)						
	Years					
DMU	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
1	6762	6079	9713	14533	29815	44597
2	3354	3384	5899	7325	13961	18692
3	4460	8152	23390	15196	16419	15447
4	2441	4706	11341	14134	19466	29382
5	22307	12187	3401	136607	48442	41376
6	13141	13891	35414	17265	29008	80676
7	1601	2459	2520	4295	5803	9432
8	5890	7669	64809	101313	159514	213214
9	3534	9801	3143	13864	17465	30764

Table D.2.Number of Researchers

Number of Researchers						
	Years					
DMU	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
1	4	5	6	6	8	8
2	78	92	105	104	102	90
3	232	263	263	237	274	234
4	73	83	85	89	105	108
5	29	26	27	26	27	26
6	1	1	1	1	2	2
7	94	114	124	125	130	131
8	11	10	10	9	9	10
9	72	74	84	90	103	101

Table D.3.Actual Income Values

Actual Income (Deflated)						
	Years					
DMU	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
1	5093	10100	12096	13059	9191	16170
2	3491	6198	5802	9800	5785	4687
3	11828	13471	7575	15969	20267	25449
4	3811	3841	2813	3235	4559	4760
5	97	376	2148	77	275	419
6	112	215	147	425	423	144
7	12888	22406	44651	62533	44693	34242
8	2133	3865	732	956	869	1262
9	2616	1710	10825	4756	6200	3964