MONTE CARLO DEA AND BUDGET ALLOCATION FOR DATA COLLECTION: AN APPLICATION TO MEASURE SUPPLY CHAIN EFFICIENCY

WONG WAI PENG

(*MBA, Universiti Sains Malaysia*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF INDUSTRIAL & SYSTEMS ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2008

ACKNOWLEDGEMENT

I would like to express my utmost gratitude to Dr. Jaruphongsa Wikrom, my main supervisor and Associate Professor Lee Loo Hay, my co-supervisor for their patience, constant encouragement, invaluable advice and excellent guidance throughout the whole course of my research.

I would like to thank Professor Chen Chun Hung at George Mason University, Professor Xie Min, Associate Professor Chew Ek Peng and Associate Professor Poh Kim Leng at the National University of Singapore, who served on my oral examination committee and provided me many invaluable and helpful comments on my thesis research and writing. Many thanks to Dr. Zhou Peng and Dr. Teng Suyan who helped me a lot during my PhD study. I also wish to thank Ms. Ow Lai Chun and Mr. Victor Chew for their excellent administrative support pertaining to my PhD study. I am also grateful to the members of Computing Laboratory, past and present, for their friendship and help throughout my thesis research.

Last, but not least, I would like to thank my husband Theng Chye, my two lovely kids Zhi Lyn and Zhi Jie, my parents and my mother in law for their constant support and encouragement throughout the whole course of my study.

Wong Wai Peng

SUMMARY

Supply chain efficiency measurement is a very difficult and challenging task. It needs to take into account multiple performance measures related to the supply chain members and it also requires huge and intensive data collection. In addition, the nature of the data which are highly uncertain rendered many existing tools inoperable and unable to provide an accurate efficiency score. Realizing the challenges in measuring supply chain efficiency, this thesis focuses on some key methodological issues related to applying data envelopment analysis (DEA) to measure supply chain efficiency in stochastic environment.

This thesis is divided into three parts. In the first part, we present a relatively comprehensive literature review of DEA and supply chain efficiency measurement, which justifies the significance of the research work presented in this thesis. In the second part, we focus on the development of a tool based on DEA and Monte Carlo to measure supply chain efficiency in the stochastic environment. We develop a tentative DEA supply chain model to address the efficiency measurement of the entire value chain. Then, we enhance the model with Monte Carlo method to cater for efficiency measurement in stochastic environment. The Monte Carlo DEA method is able to find the distributions of the efficiency and tell where the true efficiency lies most of the time. The information obtained is more meaningful and insightful for managers in making decision compared to a discrete value of the efficiency.

In the third part of the thesis, we examine how to get a better estimate of the efficiency score through budget allocation in data collection. The reason of addressing the research problem within the context of the data collection is due to the fact that in

reality the users need to collect data in order to calculate the efficiency score. In order to solve the research problem, we develop two new methods which are the two-phase gradient technique and the GA based technique. The GA and the two-phase gradient techniques are effective and efficient in solving the budget allocation problem. In addition, the second phase of the gradient technique, the GIS (Gradient Improvement Stage) is flexible and can be incorporated with other existing techniques to further improve the solutions.

The contributions of this research are three-folds. First, we provide an alternative way to measure efficiency in stochastic environment, which is Monte Carlo DEA. To show the usefulness of this method, we conduct an application study in supply chain. Second, in the context where data collection is needed and expensive, we provide a way on how to intelligently allocate the resources in data collection in order to get a better estimation of the efficiency score. Third, we develop two new techniques to solve this difficult problem.

This thesis provides the insights that it is important to conduct the data collection intelligently (i.e. by using the two sophisticated techniques) in order to get a better estimate of the efficiency and to achieve greater savings in the budget. Finally, this thesis provides a potential methodological contribution in the operational research field. It incorporates the use of simulation optimization techniques with DEA to obtain a better and more meaningful result in efficiency measurement. Last but not least, the methodology suggested in this research is widely applicable to other fields as well other than supply chain in the area of efficiency measurement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

CCP	Chance Constrained Programming
DEA	Data Envelopment Analysis
GA	Genetic Algorithm
GIS	Gradient Improvement Stage
IPA	Infinitesimal Perturbation Analysis
LP	Linear Programming
MSE	Mean Square Error
OCBA	Optimal Computing Budget Allocation
OR	Operational Research
LP	Linear Programming

Chapter 1

INTRODUCTION

This thesis contributes to some methodologies issues in applying simulation optimization techniques and data envelopment analysis (DEA) to measure supply chain performance, which could be helpful to analysts and decision makers in dealing with stochastic environment. In this introductory chapter, some background information is first provided, which is followed by the scope and objective of our study. Finally, a summary of the contents of this thesis and its structure are presented.

1.1 Background Information

Supply chain management has become one of the most frequently discussed topics in the business literature. According to Simchi-Levi (2003), supply chain management is a set of approaches utilized to efficiently integrate suppliers, manufacturers, warehouses, and stores, so that merchandise is produced and distributed at the right quantities, to the right locations, and at the right time, in order to minimize system wide costs while satisfying service level requirements. Supply chain is defined as a combinatorial system consisting of four processes namely plan, source, make and deliver, whose constituent parts include material suppliers, production facilities, distribution services and customers linked together via the feed forward flow of materials and the feedback flow of information (Stevens, 1989;

Christopher, 1998). Effective management of an organization's supply chains has proven to be a very effective mechanism for providing prompt and reliable delivery of high-quality products and services at the least cost. This is an essential corner stone for the organizations to develop a sustainable competitive advantage and to remain at the fore front of excellence in a level playing market field. To achieve an efficient supply chain, performance evaluation of the entire supply chain is extremely important. This means utilizing the combined resources of the supply chain members in the most efficient way possible to provide competitive and cost-effective products and services. Supply chain performance measurement needs to take into account the multiple performance measures related to the supply chain members, the complex relationship among the measures as well as the integration and coordination of the performances of those members (Simchi-Levi, 2003). In addition, it requires huge and intensive data collection, which is often not trivial. As such, measuring supply chain efficiency is a very difficult and challenging task.

Ross (1998) mentioned that, even within large corporations such as Sears and General Motors which had large supply chain systems, the supply chain performance measurement systems were not in existence. Rao (2006) and Chou et al. (2005) further highlighted that in view of the current level of complexity in performance measurement, it requires more sophisticated tools to measure efficiency. The absence of the performance measurement tool in supply chain is mainly due to the difficulties in measuring the supply chain efficiency.

1.2 Difficulties in measuring supply chain efficiency

Traditionally, the supply chain is usually managed as a series of simple, compartmentalized business functions. The traditional supply chain was normally driven by manufacturers who managed and controlled the pace at which products were developed, manufactured and distributed (Steward, 1997). At such, measuring supply chain efficiency during traditional times could be carried out fairly easily in a simple manner. Generally, the efficiency is measured by taking the ratio of revenue over the total supply chain operational costs. However, in recent years, new trends have emerged in the efficiency measurement, where, customers have forced increasing demands on manufacturers for quick order fulfilment and fast delivery. This has made the supply chain efficiency difficult to be measured (Stewart, 1997). In addition to the usual financial measures used to measure efficiency, the supply chain performance now also needs to take into consideration other specific indicators such as the delivery rate and percentage of order fulfilment. This measurement is further complicated by the influence of manufacturing capacity and other influential operational constraints.

In view of the increasing performance measures in supply chain, not many companies will know how to gauge the performance of their supply chain. The rise of multiple performance measures has rendered the efficiency measurement task difficult and unchallenging. In addition, supply chain efficiency measurement requires knowing the performance of the overall chain rather than simply the performance of the individual supply chain members. Each supply chain member has its own strategy to achieve efficiency. However, what is best for one member may not work in favour of another member. Sometimes, because of the possible conflicts between supply chain members, one member's inefficiency may be caused by another's efficient operations. For example, the supplier may increase its raw material price to enhance its revenue and to achieve an efficient performance. This increased revenue means increased cost to the manufacturer. Consequently, the manufacturer may become inefficient unless it adjusts its current operating policy. Hence, measuring supply chain performance needs to deal with the multiple performance measures related to the supply chain members, and to integrate and coordinate the performance of those members.

The measurement of supply chain efficiency is also greatly hampered by the difficulties in obtaining a full set of accurate data. Supply chain performance measurement requires data collection from the entire value chain which encompasses the suppliers' suppliers until the direct customer. Due to limited resources and time availability, accurate data is difficult to be obtained. Most of the time, the data are either incomplete or not accurate. The natures of these data which are highly uncertain at present in many organizations render many existing tools inoperable and unsuitable to be used for efficiency measurement. The uncertainties in the data could jeopardize the results of the efficiency measurement and hence, the inaccurate efficiency score obtained will not be useful to managers.

Hence, a tool to effectively measure the supply chain efficiency is greatly needed. This is further supported by Yee and Tan (2004) who mentioned that in view of the current level of complexity to address the supply chain problem, it involves more sophisticated tools. Though, the measurement tool only serves as a stepping stone for companies to achieve more strings of successes in the long term, the foundation of measurement has to be laid out robustly by firstly developing a suitable and useful tool for supply chain performance measurement. This tool will not only perform the quantitative reasoning but will also provide insights to manager in the qualitative perspective of strategic decision making.

1.3 Research scope and objectives

There are two main objectives of this thesis. First, it aims to address the problem mentioned in supply chain performance measurement. It will provide an analytical framework to measure the supply chain efficiency by considering the entire value chain in the stochastic environment. This thesis develops a simple and efficient tool, Monte Carlo Data Envelopment Analysis (DEA) to measure the supply chain efficiency. This new tool will be able to find out the distribution as well as the confidence interval for the true efficiency. These information are more meaningful and insightful for managers in making decision compared to a discrete value of the efficiency.

Secondly, this thesis aims to further examine how to get a better estimate of the efficiency score when there are variations in the data. Existing stochastic DEA method, which only provides a single mean value in the stochastic case, will not be able to tell accurately where the true efficiency lies. This study will address this problem within the context of data collection in the supply chain efficiency measurement. The reason of addressing data collection is due to the fact that in real industry, users would have to collect data in order to calculate the efficiency score. Data collection is extremely difficult to be carried out in supply chain as it requires the data from the entire value chain which encompasses from the suppliers until the direct customers. Hence, this greatly suits the purpose to address how to collect the data effectively. The prominent research question that will be addressed in this part of the study is:

'Given that the users have to collect data within a restricted budget to calculate the efficiency score, what is the best way to allocate the budget for data collection so that he/she can get a good estimate of the efficiency score?'

As there is no explicit model to address this question, this thesis will introduce few methods based on the optimization simulation technique to solve the problem.

1.4 Structure and Organization of the Thesis

This thesis focuses on the development of the methodology using DEA to measure the supply chain efficiency when there are uncertainties in the data and to improve the prediction of the efficiency through budget allocation for effective data collection. In consists of eight chapters.

Chapter 2 presents a literature review in the supply chain efficiency measurement, performance measures in supply chain, traditional methods used to measure supply chain efficiency, DEA and its application in supply chain studies, issues in DEA, and a brief review of other concepts or techniques which are applied in this research. Chapter 3 presents the Monte-Carlo DEA based approach to measure the supply chain efficiency. This approach serves as the basis for the second part of the thesis.

Chapter 4 to 7 address the second part of the thesis which is to provide an approach on determining how to collect data effectively so as to have a better prediction of the efficiency score. Chapter 4 discusses some underlying concepts of efficiency measurement in DEA, which path the way for the formulation of the problem statement and the mathematical model of the research problem. Chapter 5-6 discusses the methodology on how to solve the model. The methodology comprised of two main methods, which are the Two-Phase Gradient technique and the GA technique. Chapter 5 discusses on the Two-Phase Gradient Technique, while Chapter 6 discusses on the GA technique and the combinations of the techniques with other existing heuristic algorithms. Chapter 7 presents the results of the numerical experiments. Finally, Chapter 8 summarizes the conclusions of this thesis and provides suggestions for future research.

Chapter 2

LITERATURE REVIEW[1]

2.1 Introduction

This chapter discusses the literature review on the efficiency measurement of supply chain, performance measures in supply chain, traditional methods used to measure supply chain efficiency, DEA and its application in supply chain, issues in DEA and other miscellaneous concepts which will be used in the research.

2.2 Literature survey of supply chain efficiency measurement

From the literature survey of supply chain efficiency measurement, we found that the works can be mainly categorized into two types of studies, which are practical and theoretical. The theoretical category covers the elements of measurement in supply chain, which are namely the performance measures, concept and trends. On the other hand, the practical aspect encompasses the modelling framework and empirical case studies on supply chain. This classification is chosen based on the underlying intention which is to address the distinctiveness between supply chains efficiency measurement from other fields, and to identify potential research focus in this area

[1] The work presented in this chapter has been published as Wong et al. (2008).

through analyzing the imbalances in the past literature. In addition, the classification used in this thesis has not been used in any of the past studies. Past surveys of supply chain efficiency measurement have either focused on one particular attributes or aspects for instance, purely on the performance measures, or emphasized mainly on a particular type of industry.

Earlier efficiency studies in supply chain management covered types of performance measures or practices and comparison of achievable performance levels. Bogan and Callahan (2001) emphasized on internal performance metrics. Boyson et al. (1999), Gilmour (1999) and Stewart (1995) stressed on the qualitative as well as the quantitative performance measures in supply chain. Stewart (1997) and Lapide (2000) addressed the needs to consider internal and external metrics in performance improvement assessments. The concepts and trends in supply chain efficiency study have also been largely explored since the late 19th century. Simatupang (2004) highlighted the needs for an integrated supply chain performance measurement system. Bowersox (1997) and Cox (1997) discussed the requirement of a novel type of efficiency measurement system in supply chain due to the holistic approach of the supply chain management. Gunasekaran (2001) highlighted that a novel type of performance measurement system is needed for supply chain collaboration because the chain members are concerned with both performance drivers and targets.

Mathematical and non mathematical approaches had been analyzed by researchers to model supply chain efficiency, however the numbers are limited. Davis (1993) and Arntzen et al. (1995) called for more research in the area of mathematical modelling of the supply chain efficiency. Seiford (1999) highlighted that mathematical programming and associated statistical techniques to aid decision-making in supply chain benchmarking is still lacking and more work can be carried out in this area. Chopra and Meindl (2001) mentioned that the linkages of the mathematical models to the strategic level of supply chain management is still lacking. Geary and Zonnenberg (2000), Poirier (1999), Polese (2002), Simatupang (2004) addressed the modelling frameworks for supply chain efficiency measurement. Basnet (2003) illustrated a case study of efficiency measurement on supply chain practices in New Zealand companies. Past literature indicates that empirical studies of supply chain efficiency measurement and benchmarking are scarce. Table 2.1 depicts the contribution of various researchers in each respective categories namely theoretical aspects (i.e., performance measure and integration of supply chain) and practical aspects (i.e., model, framework and case study) in supply chain efficiency studies.

Table 2.1: Classification of supply chain efficiency study literature

Period	Authors	Contribution
1995-1997	Boyson, Stewart, Gilmour	Performance measure
Late 90s	Bowersox, Simatupang, Boyson, Kopcak, Stank, Christopher, Ramdan, Mentzer, Poirier	Integration supply chain / interorganizational level
2001~2004	Van Landghen, Geary and Zonnenberg, Poirier, Polese, Simatupang	Model/ framework
2004	Basnet	Case study

Figure 2.1 provides the statistics of the publications in supply chain benchmarking. As can be seen in Figure 2.1, 60% of the publications deal with the theoretical aspects, while 40% explain the practical aspects of supply chain efficiency studies.

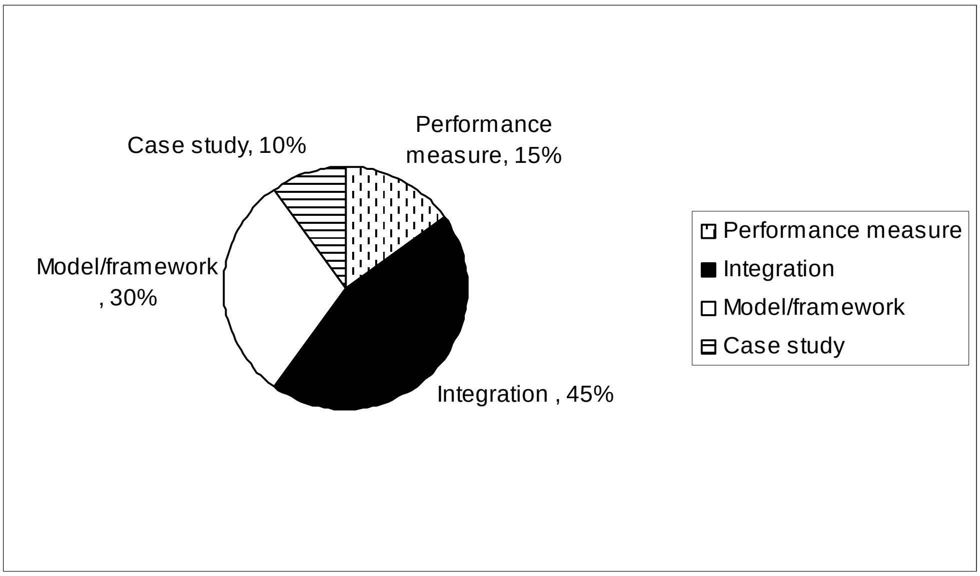

Figure 2.1: Proportion of publications

Appendix A (Table A.1-A.4) shows the summary of the literature on supply chain efficiency studies, with details of the objectives of each study. The tables are categorized according to the classification mentioned previously.

Past publications showed that supply chain efficiency study initiated from the aspects of addressing performance measures and later moved into applying efficiency measurement in an integrated perspective. Hence, this shows the growing trends in supply chain efficiency studies. The present review of literature in this section has identified certain issues which have not been satisfactorily addressed. These issues can be regarded as inadequacies and they offer scope for further research and exploration. Some of the issues identified are as follows:

1. Research in modelling and application of case study is scarce. Past researchers developed theoretical frameworks to address integrated supply chain. Mathematical modelling in supply chain efficiency study can be explored. The

use of mathematical models can help to better gauge the performance and provide clearer representation of the frameworks.

2. Tools used in supply chain efficiency measurement – Literature addressing the sufficiency of the tools are lacking. The suitability of the tools in addressing supply chain efficiency measurement in an integrated perspective needs to be explored.

2.3 Performance measures in Supply Chain

One important issue to address in supply chain efficiency study is to define what are the performances measures because they drive the actions of managers and the correct metrics are critical elements of a company's performance. Performance measures differ from field to field. Hence, this is one of the features that distinguish supply chain efficiency study from general study.

Earlier conceptual developments of performance measurements in supply chain have focused on cost-based performance measures because the cost metric is easily understood and routinely welcomed by management (Ellram, 2002; Ballou et al., 2000). Gradually, more researchers and practitioners seem to understand the shortfalls of having just a unidimesional measure which is rather inflexible and lacks integration with strategic focus. Hence, from the "cost" perspective, researchers began to put in other quantitative as well as qualitative measures in supply chain efficiency measurement. Beamon (1999) identified three types of measure, namely resources, output and flexibility. Extending from this foundation, a framework for measuring the strategic, tactical and operational level of performance in supply chain was developed (Gunasekaran, 2001).

The first universal performance measures that were used in supply chain performance measurement were generated by Pittiglio, Rabin, Todd and McGrath, widely known as the PRTM. It is a comprehensive set of fact-based performance measures that were used to accurately describe a world-class supply chain of planning, sourcing, making and delivering activities (Stewart, 1995). The measurement scheme covers four areas of performance metrics which are identified as the keys to unlocking supply chain excellence. They are delivery performance, flexibility and responsiveness, logistics cost and asset management. This is the first known study that objectively links best practices employed with relative quantitative performance achievements. The PRTM's concept of supply chain efficiency measurement/benchmarking has been extended to become the supply chain operations reference (SCOR) model by the Supply Chain Council (Stewart, 1997). The SCOR then became the first cross-industry framework for evaluating and improving enterprise-wide supply chain performance and management (integrated SCM). SCOR is structured into four levels, based on a plan, source, make and deliver framework. The model integrates the well-known concepts of business process re-engineering, benchmarking and process measurement into a cross-functional framework, which contains:

- Standard descriptions/terminology/definitions of management processes;
- A framework of relationships among the standard processes;
- Standard metrics to measure process performance;

- Management practices that produce best in class performance and
- Standard alignment to software features and functionality.

Having all these features, SCOR provides a standard format to facilitate communication and enable companies to benchmark against others which will then influence future improvement efforts to ensure real progress. The metrics used include key areas such as delivery performance, order fulfilment, production flexibility, and cash-to-cash cycle time. The usefulness of SCOR has been verified. Geary and Zonnenberg (2000) reported that in the benchmarking study conducted by the Performance Measurement Group (PMG), the best-in-class supply chain performers were gaining considerable financial and operating advantages compared to their peers by using the SCOR model.

2.4 Traditional methods to measure supply chain efficiency

Tools used in measuring supply chain efficiency have received numerous attentions. Basically, there are two types of measurements: parametric and non-parametric. The tools use to evaluate these two categories of measurement differ. In the context of parametric analysis, efficiency measurement normally uses gap analysis based techniques for performance measurement. Some of the popular gap analysis based techniques are the "spider" or "radar" diagram and the "Z" chart. These tools are very graphical in nature. Advantages of these tools are the graphical approaches made them easy to be understood and they are capable of showing multiple dimensions simultaneously. However their disadvantage is it causes inconveniences to the analysts since analysts have to integrate all the elements into a complete picture.

Another well known method used is the ratio. It computes the relative efficiencies of the output versus the inputs and is easily computed. However, a problem with comparison via ratios is that different ratios give a different picture and it is difficult to combine the entire set of ratios into a single judgement. Analytic hierarchy process maturity matrix (Eyrich, 1991; Kleinhans et al., 1995) is another alternative technique used in the performance measurement. This technique utilizes a weighted score in the analysis of various benchmarks and provides a single score using perceptual values set forth by decision makers. This is a multi-attribute utility technique. Although this method helps to quantify measure and provide managerial input, it is subjugated to a high degree of subjectivity. In addition, the rank-reversal problem in AHP reduces its usefulness.

Statistical methods (i.e. regression and various descriptive statistics) are also used to analyze data in supply chain efficiency (Blumberg, 1994; Schefczyk, 1993; Moseng, 1995). These are parametric measures. Even though the strong theoretical foundation of statistical tools such as multiple regressions is able to provide meaningful interpretation of the data, a limitation occurs in the number of simultaneous inputs and outputs that needs to be considered. Regression equations can only analyze one single output at a time and one must repeat the regression analysis as often as the number of criteria included. In addition, regression analysis can only determine average values, which probably do not actually occur in any of the units examined. The results therefore can hardly serve as benchmarks because they neither represent "best practice" nor do they exist in the real world. Similarly, regression analysis inherits the assumption that all observed firms combine their input factors in the same way. However in practice, production technology typically varies (Atkinson and Stiglitz, 1969; Freeman, 1994; Imai and Yamazaki, 1992; Vromen, 1995).

Moving to the non-parametric methods, one of the commonly used tools in performance measurement is the Balanced Scorecard (BSC). BSC provides a comprehensive framework that translates a company's strategic objectives into a coherent set of performance measures. Much more than a measurement exercise, the balanced scorecard is a management system that can motivate breakthrough improvements in critical areas such as product, process, customer and market development (Kaplan, 1993). The scorecard basically covers four different perspectives from which to choose performance measures. It complements traditional financial indicators with measures of performance for customers, internal business/processes and innovation and learning activities (Kaplan, 1996). In this way, BSC is distinguished by being able to link the company's strategic objectives to the long-term trend analysis for planning and performance evaluation. However, BSC specifies neither any mathematical-logical relationships among the individual scorecard criteria nor a unitary, objective weighting scheme for them. Hence, it is difficult to make comparisons within and across firms on the basis of BSC. In addition, the inefficient use of resources may go unrecognized and one normally turns to parametric methods in order to arrive at some judgments about the efficiency of resource usage (Rickards, 2003).

2.5 DEA

Data envelopment analysis (DEA) was first introduced by Charnes et al. (1978) as a linear programming (LP)-based methodology for performing analysis of how efficiently a company operates. Its analyzed units are denoted as 'DMU', which stands for decision making units. It is a nonparametric programming approach to frontier estimation (Farrell, 1957). In the sections that follow, we shall first introduce the basic DEA methodology. Next, we present a survey on the publication of DEA studies and the findings from these studies. Lastly, we discuss the application of DEA in supply chain.

2.5.1 Basic DEA methodology

Build upon the earlier work of Farrell (1957), data envelopment analysis (DEA) is a mathematical programming technique that calculates the relative efficiencies of multiple decision-making units (DMUs) based on multiple inputs and outputs. A main advantage of DEA is that is does not require any prior assumptions on the underlying functional relationships between the inputs and outputs (Seiford and Thrall, 1990).

Since the work by Charnes et al. (1978), DEA has rapidly grown into an exciting and fruitful field, in which operations research and management science researchers, economist and experts from various application areas have played their respective roles. For DEA beginners, Ramanathan (2003) and Coelli et al. (2005) provided excellent introductory materials. The more comprehensive DEA expositions can be found in Cooper et al. (2006). In the sections that follow, we shall briefly introduce the basic DEA methodology.

Assume *S* to be the set of inputs and *R* the set of outputs. *J* is the set of DMUs. Further assume that DMU$_j$ consumes $x_{sj} \geq 0$ of input *s* to produce $y_{rj} \geq 0$ of output *r* and each DMU has at least one positive input and one positive output (Fare et al., 1994; Cooper et al., 2004). Based on the efficiency concept in engineering, the efficiency of a DMU, says DMU j_0 ($j_0 \in J$), can be estimated by the ratio of its virtual output (weighted combination of outputs) to its virtual input (weighted combination of inputs).

To avoid the arbitrariness in assigning the weights for inputs and outputs, Charnes et al. (1978) developed an optimization model known as the CCR model in ratio form to determine the optimal weight for DMUj_0 by maximizing its ratio of virtual output to virtual input while keeping the ratios for all the DMUs not more than one. The fractional form of a DEA mathematical programming model is given as follows:

$$\begin{aligned} &\max \frac{\sum_{r \in R} u_r y_{rj_0}}{\sum_{s \in S} v_s x_{sj_0}} \\ &\text{s.t.} \quad \frac{\sum_{r \in R} u_r y_{rj}}{\sum_{s \in S} v_s x_{sj}} \leq 1 \\ &\qquad u_r, v_s > 0 \qquad\qquad s \in S,\ r \in R \end{aligned} \tag{2.1}$$

where u_r and v_s are the weights for the output *r* and input *s* respectively.

The objective function of Model (2.1) seeks to maximize the efficiency score of a DMUj_0 by choosing a set of weights for all inputs and outputs. The first constraint ensures that, under the set of chosen weights, the efficiency score of the observed DMU is not greater than 1. The last constraint ensures that the weights are greater than 0 in order to consider all inputs and outputs in the model. A DMUj_0 is considered efficient if the objective function of the associated Model (2.1) results in efficiency score of 1, otherwise it is considered inefficient.

Using the Charnes-Cooper transformation, this problem can be further transformed into an equivalent "output maximization" linear programming problem as follows:

$$
\begin{aligned}
\max \; & \sum_{r \in R} u_r y_{rj_0} \\
\text{s.t.} \; & \sum_{r \in R} u_r y_{rj} - \sum_{s \in S} v_s x_{sj} \leq 0, \quad j \in J \\
& \sum_{s \in S} v_s x_{sj_0} = 1 \\
& u_r, v_s > 0 \qquad\qquad s \in S,\; r \in R
\end{aligned}
\tag{2.2}
$$

Model (2.2) is known as the CCR model in multiplier form. If the objective function value of (2.2) is equal to 1, it implies that the DMU concerned is relatively efficient since we can find a weight combination to make its efficiency score to be equal to one. Despite the linear form of (2.2), efficiency score is usually calculated based on its dual problem:

$$
\begin{aligned}
\min \quad & \theta \\
\text{s.t.} \quad & \sum_{j \in J} x_{sj} \lambda_j \le \theta x_{sj_o}, \quad s \in S \\
& \sum_{j \in J} y_{rj} \lambda_j \ge y_{rj_o}, \quad r \in R \\
& \lambda_j \ge 0, \qquad\qquad j \in J
\end{aligned}
\tag{2.3}
$$

Model (2.3) is known as the input-oriented CCR in envelopment form or the Farrell model, which attempts to proportionally contract DMUj_0's inputs as much as possible while not decreasing its current level of outputs. The λ_j's are the weights (decision variables) of the inputs/outputs that optimize the efficiency score of DMU j_0. These weights provide measure of the relative contribution of the input/output to the overall value of the efficiency score.˙ The efficiency score will be equal to one if a DMU is efficient and less than one if a DMU is inefficient. The efficiency score also represents the proportion by which all inputs must be reduced in order to become efficient. In a similar way, we can also derive the output-oriented CCR in envelopment form if efficiency is initially specified as the ratio of virtual input to virtual output. A large number of extensions to basic DEA models have appeared in the literature as describe by Ramanathan (2003) and Cooper et al. (2006). We shall limit our discussion to this basic model as this is sufficient to lead us to the formulation of the research model which will presented in the later chapters.

2.5.2 Main features and findings of past studies

A total of 200 studies from the period of the inception of DEA until the year 2007 are reviewed and classified in terms of types of research, application schemes and several other relevant attributes. The list is shown in Table A.5 in Appendix A. These studies have been collected primarily from main OR journals as well as economics and other journals. The classification of journals and the notations used are as follows:

a) Mainline OR Journals (M): Annals of Operations Research, Computers and Operations Research, European Journal of Operational Research, Journal of the Operational Research Society, Management Science, OMEGA, Operations Research and Operations Research Letters.

b) Economics Journal (E): International Journal of Production Economics, Journal of Econometrics, Journal of Productivity Analysis, Socio-Economic Planning Science

c) Other Journals (O): These are the journals which do not fall into category a) or b). For instances, Journal of Banking and Finance, Transportation Research, IEEE Transactions on Engineering Management and etc.

We will discuss the findings in general as well as study the effects of changes over time. Hence, we will separate it into temporal and non-temporal effects.

2.5.2.1 Non-temporal effects

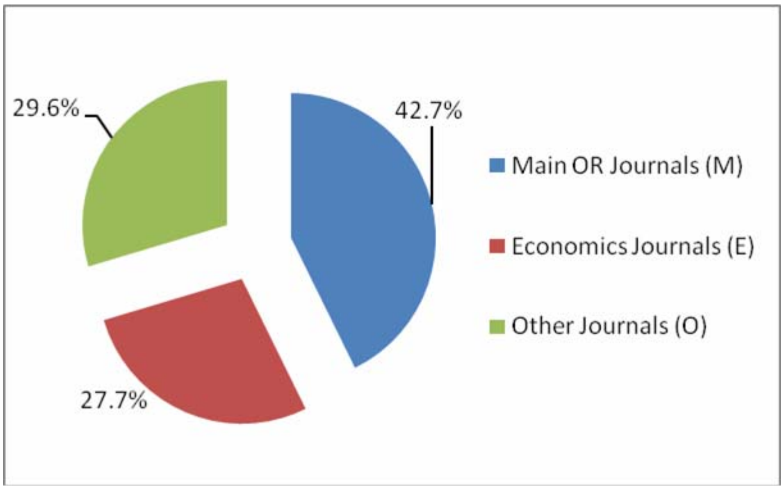

Figure 2.2: Breakdown of publications by types of journal

Figure 2.2 shows that mainline OR journals are the most preferred choice for publication of DEA articles. The reason is clearly that DEA theory and many DEA applications fall within the fields of operations research and management science, exactly the arenas covered by these journals. The economic and other journals have almost equal shares of publications. From the breakdown, one may conclude that the area of DEA is truly multidisciplinary.

In addition, we further classified the studies into 'source of publication', which are journal articles and non-journal publications such as conference papers and book chapters. Our statistics indicate that 89% of the publications are in the form of journal articles, while 11% appearing as book chapters or proceedings, conference papers as well as books themselves.

In the following sections, we categorize the studies in terms of types of research, which refers to the nature of the articles or research strategy. The following categorizations are used.

a. Theoretical developments within DEA

b. Bridging with other theoretic disciplines

c. Real world sectors where an application of DEA can be shown to be useful.

We denote (a) as 'T', (b) as 'B' and (c) and 'A'. Due to fact that the DEA literature has a uniquely high frequency of articles dedicated to theoretical development while simultaneously showing an application of these developments to real-world problems, hence, we also add one additional category which is *theory and application* type paper, which is denoted by 'T/A'.

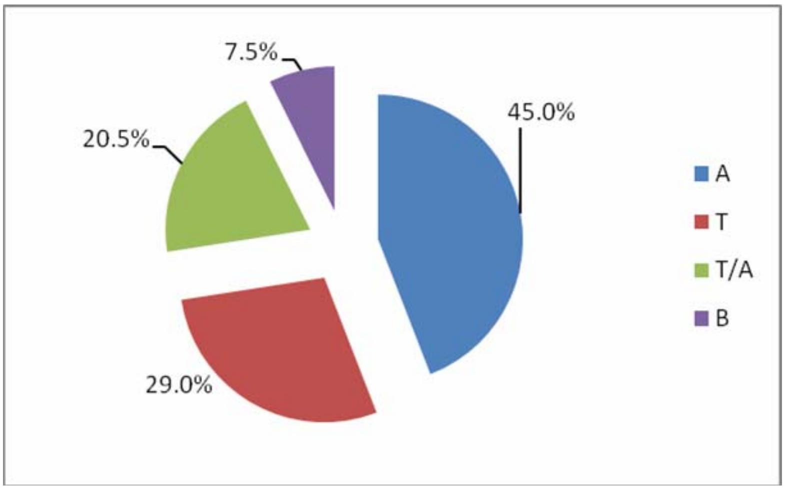

Figure 2.3: Breakdown of publications by research types

Figure 2.3 shows that application types of research comprised the highest percentage of DEA publications. This shows that the application of DEA has been extensive.

The theoretical development types of research in DEA as well as with the real world application have also been largely explored. As can be seen from Figure 2.4, the sum of both types of research accumulated to almost 50% of the total publications.

Some of the significant past works in the theoretical field of DEA are such as Banker et al. (1984), Deprins et al.(1984) and Petersen(1990) who extended and refined the standard DEA model to include variable returns-to-scale properties. Charnes et al. (1994) addressed the non-linear input substitutability and output transformability of the DEA model. Banker and Morey (1986) explored the use of categorical input-output variables, while Cook et al. (1996) addressed how to handle ordinal input-output variables in the DEA model.

Though the research on the bridging of DEA with other theoretic discipline comprised of only 7.5% out of the total publications, it is beginning to become an important research area. Some of distinguished works in this area are such as Kao (2000) who incorporated fuzzy approach in DEA. Yang and Kuo (2003) proposed a hierarchical analytic hierarchy process (AHP) and data envelopment analysis (DEA) approach to solve a plant layout design problem. O'Donnell et al. (2005) adopted the Bayesian approach in finding the frontier in DEA. Van De Meer (2005) incorporated the use of regression analysis with DEA to model the performance of UK coastguard centres.

Due to the large number of DEA publications in application types of research, we further break down the application type of DEA articles into various application schemes. Application scheme refers to the main application studied. The following seven application areas are specified, with the notation given in brackets: Education (E), Public sector(P), Healthcare (H), Banking/finance (B), Industry (I) (i.e. agriculture, manufacturing, airline, telecommunications etc), Utilities (U) (i.e. power, electricity, water etc), and others (O) which cannot be categorized into any of the above six sectors (i.e. computing, R&D, sports, neural network, ERP etc). These schemes are chosen based on the observations from past studies that DEA is mostly applied in these areas.

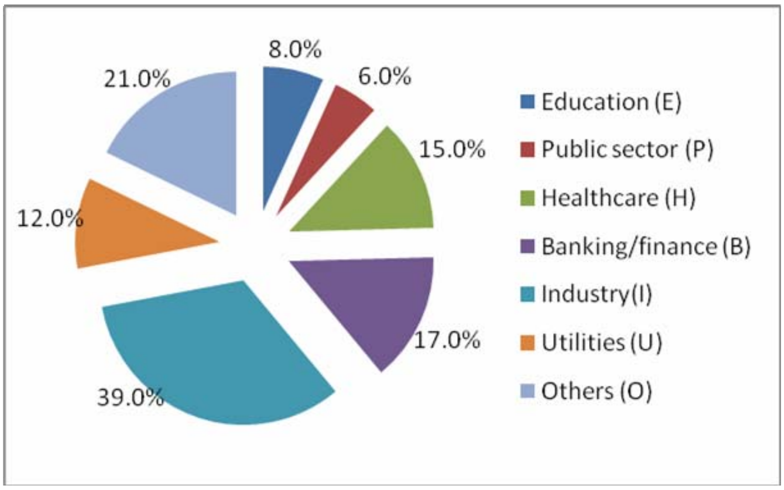

Figure 2.4: Breakdown of publications by application scheme

Banking/finance sector comprised the largest area in the application of DEA. Some examples of the studies in banking are Giokas (1991), Oral et al.(1992), Al-Faraj et al. (1993), Barr et al. (1993), Sherman and Ladino (1995), and Athanassopoulos (1997). In industry sector, DEA has been applied to various assorted activities. For instances, Weber and Desai (1996) employed DEA to construct an index of relative supplier performance. Clarke and Gourdin (1991) applied DEA to the vehicle maintenance activities of 17 separate maintenance shops of large-scale, non-profit logistics systems. Metzger (1993) used DEA to conduct a longitudinal study to measure the effects of appraisal and prevention costs on productivity. Kleinsorge et al. (1991) utilized DEA to conduct a longitudinal monitoring process of one carrier in an effort to assess expected performance improvements over time. Easton et al. (2002)

utilized DEA as a management tool to compare the purchasing efficiency of firms in the petroleum industry.

Other works also includes airline operations (Chan and Sueyoshi, 1991; Schefczyk, 1993); brewing (Day et al., 1995); defense-industrial base (Bowlin, 1995); education (Beasley, 1995); manufacturing (Ray and Kim, 1995; Shafer and Bradford, 1995); retail organizations (Athanassopoulos, 1995); transportation and logistics (Clarke and Gourdin, 1991; Chu and Fielding, 1992) and vehicle maintenance (Clarke, 1992).

Below are some examples of the works for other sectors. Utilities e.g. electricity generation (Charnes et al., 1989; Miliotis, 1992); Health care (Banker et al., 1986; Borden, 1988); non-profit organizations (Charnes et al., 1981; Pina and Torres, 1992); and others e.g. pay equity in professional baseball (Howard and Miller, 1993). For a comprehensive qualitative survey of DEA, please refer to Seiford (1996). As a complement to the qualitative aspect, a quantitative/statistical review of the entire life cycle of DEA is provided by Gattoufi et al. (2004).

2.5.2.2 Temporal effects

To study possible changes over time, we divide the time frame into three 10-year period, 1978-1987, 1988-1997 and 1998-2007. As shown in Figure 2.5, the total number of publications has increased significantly, from 10 in 1978-1987 to 123 in 1998-2007.

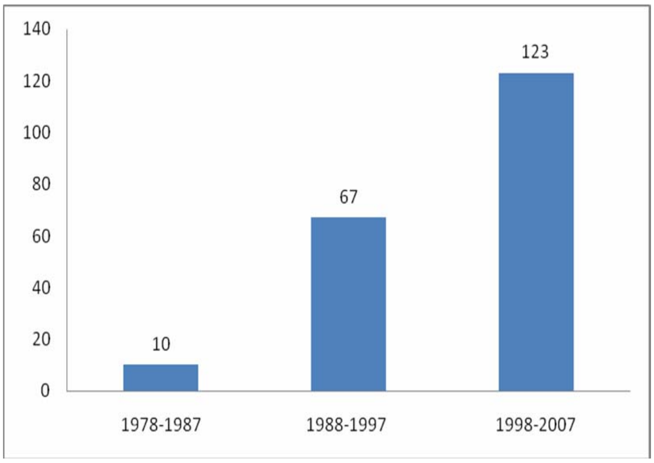

Figure 2.5: Trend of number of studies in DEA.

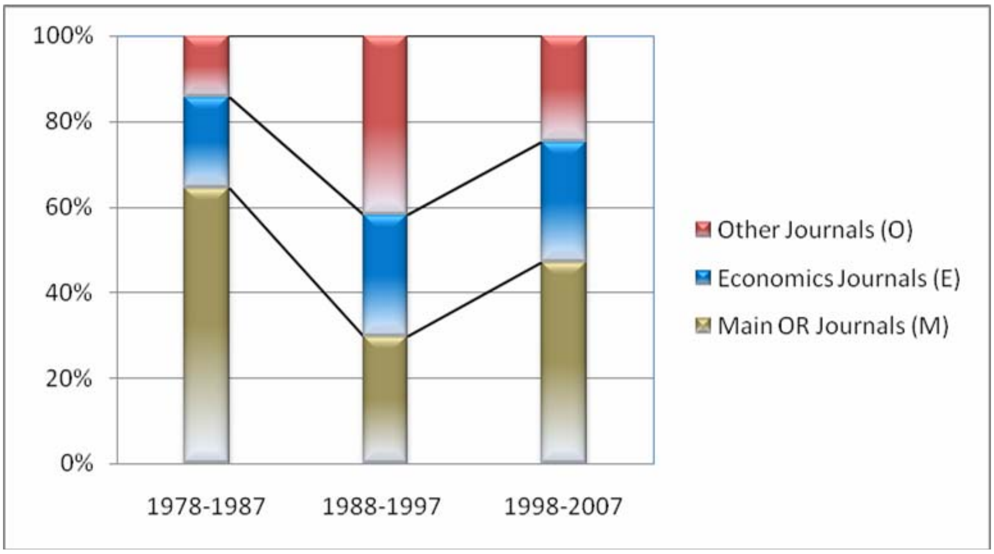

Figure 2.6: Breakdown of publications by source of publication over time

Figure 2.6 shows the breakdown of publications by source of publication over time. It was found that there is a shift in the preferred outlet of publication in the period of 1988-1997. There is a marked increase of publication in other journals from 14.3% in 1978-1987 to 41.8% in 1988-1997. This trend might show the changes in the preferred outlets for researchers that could be influenced by the launch of several new journals in the late 1980s. Also, it could be the result of wider penetration of DEA to different application area.

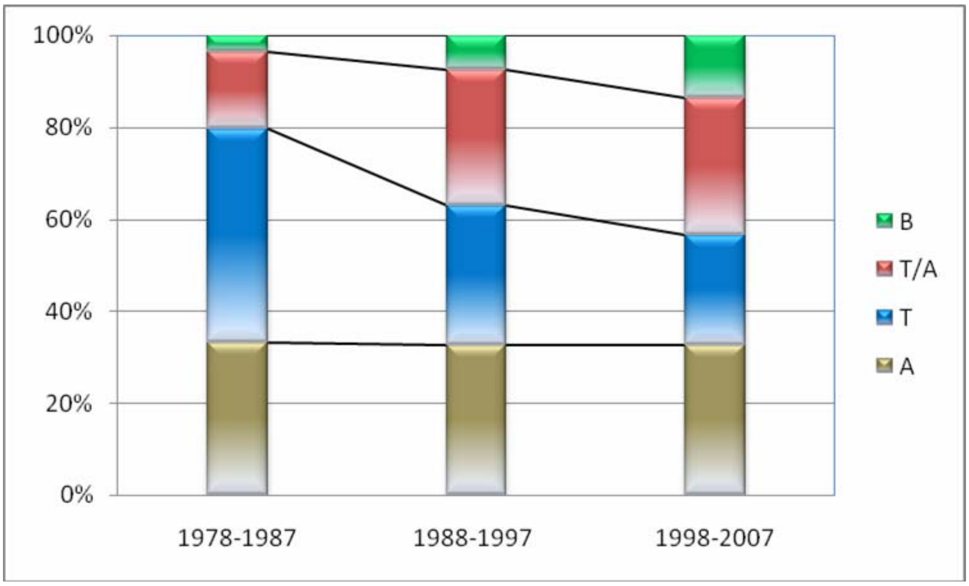

Figure 2.7: Breakdown of publications by type of research over time

Figure 2.7 shows the breakdown of publications by type of research over time. It was found that, the shares taken up by the 'theoretical and application' aspects of DEA increased markedly from 16.7% in 1978-1987 to 20.7% in 1988-1997. This should attribute to the flexibility and ability of DEA in allowing for its application in varying situations. Since various application studies have their individual characteristics, practitioners and researchers may have to present new DEA versions for their use. Another possible reason is that such popular software packages as EXCEL and MATLAB offer researchers huge flexibility to construct and apply their own models. There is also a growing interest in the research area in bridging DEA with other OR disciplines. The shares of publication for this type of research has increased from 3.7% in 1978-1987 to 7.4% in 1988-1997 and finally reaches 13.5% in 1998-2007. Correspondingly, the shares of publication for the theoretical development in DEA reduced from 46.3% in 1978-1987 to 30.4% in 1988-1997 and further decreased to 23.9% in 1998-2007. This marks a saturation level in this type of research. Most researchers focus their works on the theoretical development in DEA during the inception period, hence, the number of publications reached its peak in this period. Since then, researchers gradually started to divert their attention from purely theoretical to other research strategy such as combination of theoretical and application and incorporation of DEA with other OR disciplines. As for the pure application type of research, the breakdown did not change much from 1978-1988 to 1998-2007. This area still remains a popular strategy in research, which proves the vast application of DEA.

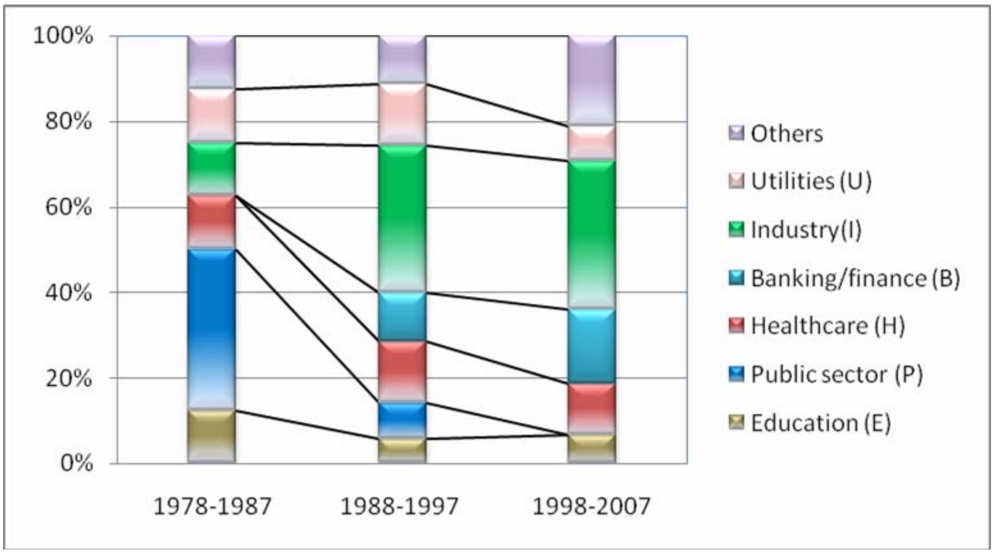

Figure 2.8: Breakdown of publications by application area over time

Figure 2.8 shows the breakdown of the 200 studies by application scheme over the three periods of years. Since the inception of DEA in year 1978, it has gradually become a popular tool for studying the efficiency in various application. Prior to 1988, it was found that the use of DEA in public sector has the highest proportion of publications. The studies are such as Lewin(1984), Miller(1985) and Macmillan (1987). The numbers of studies in this area were exremely huge that it reached saturation level, as later, we can see that there is a significant drop in the number of publications in this field from 37.5% in the period 1978-1987 to 8.6% in 1988-1997 and further reduced to almost 0% in 1998-2007. Other application schemes which exhibit almost similar trends are education and utilities. Temporally, the shares taken up by the studies on industry has increased from 12.5% in 1978-1987 to 34.3% in 1988-1997. This could be explained by the expansion that have occured in the industry since the late 1980s, More types of different indusries have emerged, hence these provided more outlets for the researchers to apply DEA. This proportion does not change much in 1998-2007. Similarly, a growing interest in the application of DEA in banking and financial sectors can be observed by the increment of shares from 0% in 1978-1987 to 17.3% in 1998-2007. This may be largely due to the revolution in the banking industry which provides more opportunities for studies to be conducted in this area. Lately, in the period 1998-2007, there is a marked increase in the study of DEA in other areas. These other areas include rather unique and specialized areas which could not be categorized in any of the above six areas. The reason for this increase is the study scope of DEA has expanded to novel applications. Examples of such studies include bankruptcy (Cielen et al., 2004), neural networks (Vaninsky, 2004). Enterprise Resource Planning (ERP) (Maber et al., 2006) and sports (Lozano et al., 2004). As for the application of DEA in healthcare, there is not much changes in the shares of publication over the three periods. The ever growing development in medical and healthcare studies has simultaneously provided the avenues for DEA application, hence, research interest still remains intact in this field.

2.5.2.3 Other features and findings

We found that a majority of past studies dealt with the input-oriented DEA models. To a large extent, it should be attributed to the characteristic of the industry which widely applies DEA. i.e. public sector. Higher priority has often been given to the goal of meeting demand (Färe et al., 1994a). As a result, input conservation for given outputs seems to be a reasonable logic. Another possible reason is that in many empirical studies, particularly at the macro level, there is only one output such as 'profit' but multiple inputs are often used.

In addition, we also have found that many OR/MS researchers favour DEA models in the multiplier form while economists favour DEA models in the envelopment form. This is likely due to the interdisciplinary nature of DEA and its historical diffusion patterns (F∅rsund and Sarafoglou, 2005).

Lastly, the approach used in our study is different from the past studies. Past surveys of DEA have mainly focused on the compilation of the full bibliographies of DEA (Emrouznejad et al. 2008; Gattoufi, 2004). The analysis carried out in this study is different in terms of attributes used to categorize the studies. In this thesis, we categorize the studies following the applications schemes, publications, and types of research. Past survey focused on the number of publications per authors and keywords used. The next section will discuss the literature of DEA in supply chain.

2.5.3 DEA in supply chain studies

The application of DEA in supply chain studies is still largely unexplored as the numbers of past studies are limited. The reason may be due to the unawareness of the suitability of DEA as a tool to measure supply chain efficiency. In this section, first, we present a brief review on the motivations of using DEA in supply chain, followed by some past studies of DEA in works related to supply chain.

2.5.3.1 Motivations of using DEA in supply chain

DEA is suitable to be used in measuring supply chain efficiency because it can handle multiple inputs and outputs and it does not require prior unrealistic assumptions on the variables which are inherent in typical supply chain optimization models (i.e. known demand rate, lead time etc). These advantages of DEA enable managers to evaluate any measures efficiently as they do not need to find any relationship that relates them. Wong et al. (2008) discussed the motivations of using DEA as a supply chain performance measurement tool, by giving ample evidences, literature supports and reasons on the suitability of DEA as a decision making tool in supply chain management. Some of the distinguished features of DEA that worth mentioned here are as follows:

a) DEA is able to address the complexity arises from the lack of a common scale of measurement. DEA inherits the feature that permits the inclusion of quantitative measures as well as qualitative data in performance analysis. Furthermore, it allows management to analyze simultaneously a relatively large number of inputs and outputs measured on different scales.
b) In DEA, one does not need to assume a priori the existence of a particular production function for weighting and aggregating inputs or outputs. Hence, they are solely dependent on the empirical observations. This fact gives the DEA method a decisive advantage over ordinary optimization procedures.
c) DEA is highly flexible and able to mold with other analytical methods easily to create a more meaningful and efficient way of evaluating performances. Many researchers have studied the extensions of DEA models in evaluating performances, for examples combining with statistical analysis, and other multi criteria decision making techniques (Zhu, 2004; Golany, 1988; Spronk, 1999).

2.5.3.2 Past studies of DEA in supply chain

The application of DEA within the context of supply chain has been scarce. Only a limited number of literature surveys have been reported. These literature are within the context of an individual supply chain member and not the overall supply chain system. For example, Weber and Desai (1996) applied DEA to construct an index of relative supplier performance. Cheung and Hansman (2000) measured the performance of supply chain members based upon single performance measure. Easton et al. (2002) suggested a DEA model to compare the purchasing efficiency of firms in the petroleum industry. Forker et al. (1997) studied the impact of supply chain performance evaluations on management practices. All these models only considered the performance of the individual supply chain members and no attempts have been made to identify best practice in the case of supply chains.

As mentioned earlier, one of the difficulties in measuring supply chain efficiency is the need to consider multiple performance measures related to multiple supply chain members. In addressing the problem of multiple stages/members in the supply chain, several other researchers had developed some methods within the DEA context, which have the potential to be used in supply chain efficiency evaluation. Seiford and Zhu (1999) and Chen and Zhu (2004) provided two approaches in modelling supply chain efficiency as a two-stage process using data envelopment analysis (DEA). Fare and Grosskopf (2000) developed the network DEA approach to model general multi-stage processes with intermediate inputs and outputs. Golany et al. (2006) provided an efficiency measurement framework for systems composed of two subsystems arranged in series that simultaneously compute the efficiency of the aggregate system and each subsystem. Troutt et al. (2001) determined the optimal throughput between the stages in a serial linkage of processes using DEA. Castelli et al. (2004) investigated a two-level hierarchical structure of the DMU composed of consecutive stages of parallel subunits. Chen et al. (2006) developed the DEA-game theory approach to address how to integrate the seller’s and buyer’s efficiency scores and obtain an efficiency score for the supply chain.

The following section will discuss the issues in DEA followed by a brief review on other miscellaneous concept or disciplines which will be used in this research.

2.6 Issues in DEA

DEA is a data-driven approach where it requires the estimation of the inputs/outputs data. An early criticism of DEA is that the data has to be deterministic. In response to this criticism, a number of methods incorporating stochastic variations in data have been proposed. One of the earliest efforts involved the development of chance-constrained formulations of the mathematical programs in DEA in order to accommodate stochastic variations in data (Charnes and Cooper, 1963). The chance constraint approach addresses measurement error by relaxing the constraints so that they are not always binding. Hence, this provides a more conservative estimate of efficiency resulting from a shift in the frontier. In most cases, the efficiency calculated using this way is the minimum efficiency or the worst case efficiency.

Extensive study using Chance Constrained Programming (CCP) has been carried out by Sengupta (1982, 1987, 1988, 1989, 1995). One prominent feature of his studies is to incorporate the stochastic variables into the DEA model and then reformulate the stochastic model into a deterministic equivalent. Similar models based on CCP have also been developed by Desai and Schinnar (1987), Peterson and Olesen (1989), Olesen and Thore (1990), Land et al. (1988), Cooper et al. (1996, 1998) and Sueyoshi (2000). While there exists a broad consensus about the merits of the CCP method which offers a way to breakout from the ‘deterministic’ mold, the severe data requirements such as the necessity to supply information on expected values for all variables, variance-covariance matrices for all variables, and probability levels at which feasibility constraints are to be satisfied, impedes the development of these models to their full potential. In addition, the efficiency measured using the CCP way, which is the minimum efficiency, may not be a fair comparison for the organization's performance as it does not truly reflect the true value of the efficiency score. This may not give a meaningful interpretation to the performance of the organization.

In a parallel strand in the stochastic literature, the treatment of data variations in DEA has also been studied by integrating its nonparametric feature with the parametric approach of the stochastic frontier. This is in line with the effort to bridge the conceptual and philosophical gap between DEA and econometric approaches to frontier estimation. Banker (1993) conceptualized a convex and monotonic nonparametric frontier with a one-sided disturbance term and showed that the DEA estimator converges in distribution to the maximum likelihood estimators. He also specified *F* tests for hypothesis testing. Subsequently, Banker and Maindiratta (1992) introduced an additional two-sided component in the composite error term and proposed an estimation procedure of the nonparametric frontier by DEA. Other different approaches of stochastic DEA has also been studied by Varian (1985), Simar and Wilson (2000), Ferrier and Hirschberg (1997), Gstach (1998), Fried et al. (2002), Triantis and Girod (1998), Park and Simar (1994) and Kniep and Simar (1996). For a selective survey of various stochastic approaches to DEA, see Grosskopf (1996). Past literature indicates that the research on the theoretical development of DEA in the stochastic case has been widely explored. While the earlier researches do offer interesting discussions of DEA in the presence of variations in the data, no study on how to get a good estimate of the efficiency has so far been reported. This area offers scope for further research and exploration.

2.7 Other miscellaneous

This section presents a brief review on other concepts or disciplines which are used in this research. They are the Monte Carlo method, Bayesian framework, OCBA (Optimal Computing Budget Allocation) and IPA (Infinitesimal Perturbation Analysis).

2.7.1 Monte Carlo method

Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to compute the results (Fishman, 1995). Monte Carlo methods are often used when simulating physical and mathematical systems or when it is infeasible or impossible to compute an exact result with a deterministic algorithm. (Rubinstein and Kroese, 2007). The term Monte Carlo was coined in the 1940s by physicists working on nuclear weapon projects in the Los Alamos National Laboratory (Metropolis and Ulam, 1949). Monte Carlo is nonparametric and easily implemented for any systems. In contrast to the simplicity of the approach, the information generated by the Monte Carlo method is very rich. The greater information content and flexibility of the approach are significant advantages in providing statistical information about the precision of the results. Further, the method is more straightforward from a statistical viewpoint, requiring nothing more complicated than a basic ability to generate random numbers from known statistical distribution, a function available in nearly all of statistical and econometrics software packages on the market today (Gentle, 2003). Monte Carlo method has been widely applied in many areas such as finance, risk analysis in investment proposals, reliability engineering, computer science, physical chemistry and in probabilistic design for simulating and understanding the effects of variability (Fishman, 1995). Preliminary analysis of the application of Monte Carlo in DEA had been explored by Zhang and Bartels (1998). They used Monte Carlo to examine the effect of sample size on the mean efficiency in an application study of electricity distribution. Yu (1998) conducted a Monte Carlo study to compare the stochastic frontier method and the data envelopment analysis (DEA) in measuring efficiency in situations where firms are subject to the effects of factors which are beyond managerial control.

2.7.2 Bayesian framework

The Bayesian framework is build on the foundation of Bayesian theory which used the concept of probability to infer or update the degree of belief that a proposition is true in light of new information (Berger, 1999). The central theme in the Bayesian framework involves the need to specify initial uncertainty about unknown parameters by specifying prior distributions for unknown quantities (i.e. unknown outputs or unknown input parameters); followed by the specification of likelihood models to relate unknown parameters to observable data, and finally, the update of the beliefs about unknown quantities as data becomes available using Bayes' rule to obtain posterior distributions for unknown quantities (Winkler, 1972; Carlin and Louis, 2008). Bayesian methods are useful in the simulation context if they are considered to be an analytical tool that informs decisions. They provide a convenient and useful way to represent uncertainty about alternatives (i.e manufacturing system designs, service operations, or other simulation applications) in a way that quantifies uncertainty about the performance of systems, or about inputs parameters of those systems. Bayesian methods for simulation input and output uncertainty have been increasingly applied and developed in recent years (Glynn, 1986; Cooke, 1994; Chen and Schmeiser, 1995; Chen, 1996; Scott, 1996; Andradottir and Bier, 1997; Chick, 1997; Nelson et al., 1997; Chen et al., 1999; Cheng, 1999; Chick and Inoue, 2001a and 2001b). Chick (2001) provided a tutorial on Bayesian methods for simulations. His studies described how Bayesian statistics can help a simulation analyst to deal with issues that arise in the decision-making process, where he discussed the input distribution selection, sensitivity analysis and the selection of the best of several alternative systems. Chick (2006) provide a literature review on the development of theoretical techniques for Bayesian methods in simulation experiments; for applications of those tools (to scheduling, insurance, finance, traffic modelling, public health, water-way safety, supply chain and other areas), the relationship of Bayesian methods to deterministic simulations; and to subjective probability and Bayesian statistics in general. Excellent references to various aspects of Bayesian methods, subjective probability and decision analysis in general can be found in DeGroot (1970), Lindley (1972), Savage (1972), Winkler (1972) , Berger (1999), de Finetti (1990) and Bernardo and Smith (1994).

2.7.3 OCBA (Optimal Computing Budget Allocation)

Simulation, being a popular tool for designing large, complex, stochastic, or any systems where the closed-form analytical solution do not exist, generally requires a huge amount of runs in order to simulate the alternative designs and replicate the stochastic behaviors in the systems. Though the computational power has been dramatically increased with the advancement of new technology, the key issue remains on how to improve the simulation efficiency and to reduce the total computation time. OCBA (Optimal Computing Budget Allocation) is a new control-theoretic simulation technique developed by Chen (1995). The OCBA approach can intelligently determine the most efficient simulation replication numbers or simulation lengths for all simulated alternatives. The basic idea of OCBA is to optimally choose the number of simulation samples for all designs to maximize simulation efficiency with a given computing budget or to attain a desired simulation decision quality using a minimum computing budget. OCBA is ideal for stochastic simulation optimization. Due to the stochastic nature of the objective function, in order to achieve the best computational efficiency, one needs to determine the tradeoff between devoting computational effort for exploration (which refers to searching of the space for new candidate solutions) versus exploitation (which refers to getting more accurate estimates of the objective function at currently promising solutions). In procedure, OCBA sequentially determines which design alternatives need more simulation and how many additional replications are needed. Overall simulation efficiency is improved as less computational effort is spent on simulating non-critical alternatives and more is spent on critical alternatives. Some earlier development of OCBA can be found in Chen

(1996) and Chen et al. (1997). For detail theoretical foundation and derivation of OCBA, readers may refer to Chen et al. (2000) and Chen and Yucesan (2005). Subsequent and related works on OCBA includes Fu et al. (2007), who used OCBA to select the best alternatives when the samples are correlated. Lee et al. (2004) used a sequential procedure called the multi-objective computing budget allocation (MOCBA), which aims to minimize Type I and Type II errors of the solutions within the Pareto sets. Chen et al. (2007) and Shi and Chen (2000) used OCBA for simulation and optimization problems. Literature on the application of OCBA techniques in real industry can be found in Hsieh et al. (2007), Romero et al. (2006), Chen and He (2005), Chen et al. (2003), Hsieh et al. (2001) and Chen et al. (1999).

2.7.4 IPA (Infinitesimal Perturbation Analysis)

Perturbation Analysis (PA) is a technique for estimating the gradient of a system performance measure. Its distinct feature is that derivatives with respect to multiple parameters can be calculated from a single simulation run. IPA (Infinitesimal Perturbation Analysis) is the earliest form of PA and is the well-developed technique. It has been widely adapted in the discrete event dynamic systems (DEDS) such as single-server queues (Suri & Zazanis, 1988) and queuing networks (Ho & Cao, 1983). Suri (1987) and Cao (1985) provided the theoretical foundations for IPA in proving the consistency of the sample gradient estimates for the systems. The assumptions used in IPA are the parameters have to be continuous and the interchangebility conditions[2] in the order of expectation and differentiation (Ho and Cao, 1983). More detailed explanations of the IPA technique and theory can be found in Ho and Li (1988) and Gong and Ho (1987). They showed how to overcome some of the difficulties in IPA. Zazanis (1986b) also provided a comprehensive theoretical work in IPA. As a summary, he proves strong consistency and unbiasedness for the gradient of the system with respect to a parameter. He also demonstrated how strongly consistent the second and higher order derivative estimates can be obtained from a single sample path and also introduced the single-run optimization method utilizing IPA in a preliminary experimental study.

2.8 Concluding comments

In this chapter, we have presented the literature survey on supply chain efficiency measurement which encompasses the performance measures of supply chain and traditional methods used to measure supply chain efficiency. We also presented a literature survey on DEA, its application in supply chain studies, its issues and other miscellaneous techniques which are used in this thesis.

---

[2] The IPA estimator is simply the sample path derivative of the quantity of interest, defined by

$$\frac{dH(\Omega,\omega)}{d\Omega} = \lim_{\Delta\Omega\to 0} \frac{H(\Omega+\Delta\Omega,\omega) - H(\Omega,\omega)}{\Delta\Omega}$$

Thus, for the IPA estimator to be an unbiased gradient estimator, we need

$$\frac{dE[H]}{d\Omega} = E\left[\frac{dH}{d\Omega}\right],$$

$$\lim_{\Delta\Omega\to 0} E[g_{\Delta\Omega}] = E\left[\lim_{\Delta\Omega\to 0} g_{\Delta\Omega}\right], where \quad g_{\Delta\Omega} = \frac{H(\Omega+\Delta\Omega) - H(\Omega)}{\Delta\Omega}$$

The literature endorses the fact that there has been much work done up to present regarding supply chain efficiency measurement. However, the works addressing the supply chain model and the sufficiency of the tools used in measuring the supply chain efficiency is lacking. The analysis from the review of DEA shows that, it has enjoyed a high number and a high incidence of real-world applications. The theoretical development within DEA has also been extensively explored. An area which is increasingly getting the interests from researchers is the bridging of DEA with other theoretical concepts. This area offers much scope for further research and exploration.

Considering the importance of efficiency study and the ability of DEA in handling multiple factors and multistage chain members, it justifies the usefulness of DEA as a tool to measure supply chain efficiency. Therefore, it is reasonable to believe that DEA would play a more important role in supply chain efficiency studies in future. In view of the potential of this area, it is therefore worthwhile to extend our study in the later chapters, whereby, we will address how to measure the supply chain efficiency using DEA as well as address the literature gap in DEA in which we will incorporate other theoretical disciplines with DEA in our objective to get a better estimation for the efficiency.

Chapter 3

MEASURING SUPPLY CHAIN EFFICIENCY IN STOCHASTIC ENVIRONMENT[1]

3.1 Introduction

In this chapter, we introduce the methodology used to measure supply chain efficiency in stochastic environment. The DEA supply chain model will be constructed to measure supply chain efficiency. Then, this model will be enhanced with Monte Carlo technique to cater for efficiency analysis in stochastic environment. The DEA supply chain model is developed based on the conventional DEA CCR model.

3.2 Background

Based on the conventional DEA CCR model (3.1), one way to measure supply chain efficiency is by treating the efficiency of each member or channel separately and then take the average of the efficiencies.

$$\begin{aligned}
\min\ & \theta \\
\text{s.t.}\ & \sum_{j\in J} x_{sj}\lambda_j \le \theta x_{sj_o}, \quad s \in S \\
& \sum_{j\in J} y_{rj}\lambda_j \ge y_{rj_o}, \quad r \in R \\
& \lambda_j \ge 0, \qquad j \in J
\end{aligned} \tag{3.1}$$

[1] The work presented in this chapter has been published as Wong et al. (2008).

Note that all the notations in (3.1) have been previously defined in Section 2.5.1. That is there will be four models, one for each channel (supplier, manufacturer, distributor and retailer) and the supply chain efficiency is assumed to be equivalent to the average efficiency of the four models. The limitation of measuring the supply chain efficiency this way is that it does not capture the efficiency of the entire value chain. What is best for one member may not work in favour of another member. That is, the best practice of one channel does not mean that it fits the other channel. One member's inefficiency may be caused by another's efficient operations. In the following section, we will discuss why the DEA CCR model cannot be directly applied to supply chain and provide a tentative solution on measuring supply chain efficiency.

3.3 Programming model for measuring supply chain efficiency

Consider a simple chain relationship (e.g. supplier – manufacturer) as described in Figure 3.1, where X_A is the input of the supplier, and Y_A is the supplier's output. Y_A is also an input of the manufacturer along with X_B with Y_B being the manufacturer's output. Note that one example of Y_A is ontime delivery; it indicates the performance of the supplier in delivering its products and also as a cost measure to the manufacturer which associated with inventory holding cost.

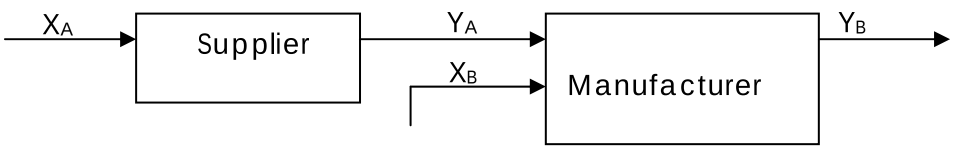

Figure 3.1: A simple chain relationship

Suppose J as the set of supply chain and each chain in the set is such as depicted above. The DEA CCR model (3.1) only considers the inputs and outputs of the supply chain system and ignores measures Y_A associated with supply chain members; hence, it does not characterize the performance of supply chains correctly. If Y_A are treated as both input and output measures in the model, all the supply chains will become efficient. This does not necessarily indicate efficient performance in individual supply chain members. Consequently, improvement to the best-practice can be distorted i.e., the performance improvement of one supply chain member affects the efficiency status of the other, because of the presence of intermediate measures (i.e. Y_A).

Alternatively, we may consider the effect of the intermediates measures. In our propose model, we will separate the measures into two groups, i.e. ‘direct’ and ‘indirect’ (intermediate). We define ‘direct’ measures as associated with a single channel or supply chain member only and intermediate (indirect) measures as associated with two or more members/channels. We will now elaborate how will the supply chain efficiency be characterized if we take into consideration the intermediate (indirect) measures compared to without considering them.

Let’s use a simple scenario; for example, there are two supply chains, i.e. DMU A and DMU B, and each of them is a dual-channel (supplier-manufacturer) system. Let’s say the manufacturer of A and B are the same. Also, let’s assume that supplier A is very efficient while supplier B is less efficient compared to A. Note that the efficiency of the individual supply chain member can be obtained using the DEA CCR model as explained earlier. Recall that the best practice of one channel does not mean that it fits the other channel. In this case, the impact from the performance of the supplier may affect the efficiency status of the manufacturer in such a way that the manufacturer A may seem to be less efficient compared to the manufacturer B; by right, they should be equally good because they are the same manufacturer. This shows that member's inefficiency may be caused by another's efficient operations. Therefore, the efficiency approach (i.e. DEA CCR model) will not characterize the supply chain efficiency correctly.

In order to better characterize the supply chain, we have to 'discount' or remove the impact of the performance improvement of one supply chain member that affects the efficiency status of the other. We will illustrate how this discounting concept can be realized using the intermediate (indirect) measures that we introduce in our model. From the basic DEA model in fractional (ratio) form, let's denote *IS* as the set of intermediate inputs, *DS* as the set of direct inputs, x_{tj} as the t^{th} intermediate input of DMU j and x_{tj_0} as the t^{th} intermediate input for observed DMU j_0. Note that $DS \cup IS = S$.

$$
\begin{aligned}
&\max \frac{\sum_{r\in R} u_r y_{rj0} - \sum_{t\in IS} v_t x_{tj0}}{\sum_{s\in DS} v_s x_{sj0}} \\
&\text{s.t.} \quad \frac{\sum_{r\in R} u_r y_{rj} - \sum_{t\in IS} v_t x_{tj}}{\sum_{s\in DS} v_s x_{sj}} \le 1 \\
&\qquad u_r > 0, \quad r \in R \\
&\qquad v_s > 0, \quad s \in DS \\
&\qquad v_t \ge 0, \quad t \in IS
\end{aligned}
\tag{3.2}
$$

where v_t is the weight for the intermediate variables. All the other notations used have been previously defined in Section 2.5.1. Note that the weight for the intermediate variables may be zero, but for the direct variables, the weights must always be positive. Note also that the difference between (3.2) and (2.1) is the subtraction of the intermediates term. This term represents the performance of one supply chain member (e.g. the upstream channel) that feeds into other supply chain member (e.g. the downstream channel). By subtracting the intermediate term in such a way is analogous to 'discounting' the impact of one's performance that affects the other. From the model (3.2), it is obvious that the impact of the indirect factor is removed; and the efficiency obtained in this model will be the best case efficiency. Though the 'discounting' concept may not have fully addressed all the issues in supply chain, it can serve as a tentative solution to measure the supply chain efficiency.

Model (3.2) can be further transformed into its equivalent linear form as shown in Model (3.3) (the primal model) and Model (3.4) (the dual model) as below.

CCR multiplier model

$$
\begin{aligned}
\max & \sum_{r\in R} u_r y_{rj0} - \sum_{t\in IS} v_t x_{tj0} \\
\text{s.t.} & \sum_{r\in R} u_r y_{rj} - \sum_{t\in IS} v_t x_{tj} - \sum_{s\in DS} v_s x_{sj} \le 0, \qquad j \in J \\
& \sum_{s\in DS} v_s x_{sj0} = 1 \\
& u_r > 0, \quad r \in R \\
& v_s > 0, \quad s \in DS \\
& v_t \ge 0, \quad t \in IS
\end{aligned}
\tag{3.3}
$$

CCR envelopment model

$$
\begin{aligned}
& \min \Omega \\
\text{s.t.} & \sum_{j\in J} \lambda_j x_{sj} \le \Omega\ x_{sj_o}, \qquad s \in DS \\
& \sum_{j\in J} \lambda_j x_{tj} \le x_{tj_o}, \qquad t \in IS \\
& \sum_{j\in J} \lambda_j y_{rj} \ge y_{rj_o}, \qquad r \in R \\
& \lambda_j \ge 0, \qquad j \in J
\end{aligned}
\tag{3.4}
$$

Note that all the notations used have been previously defined in the above section. We will name Model (3.4) as the DEA supply chain model. The model is an input oriented model whereby it aims to reduce the inputs as much as possible while not decrease the level of the output. Note that the third constraint (i.e. for the outputs) can actually be separated into two constraints (i.e. one for direct and another for indirect terms). Since the indirect term for the output will not affect the objective function, therefore, we did not explicitly write it into two separate constraints; for conciseness purpose of the model, we combined them into one constraint.

Given Model (3.4), one way to evaluate the entire value chain efficiency which generally comprised of four channels i.e. supplier, manufacturer, distributor and retailer, is to estimate the efficiency, Ω as the normalized (weighted) efficiency of all the channels. That is,

$$\Omega^* = \frac{w^S\Omega^{S*} + w^M\Omega^{M*} + w^D\Omega^{D*} + w^R\Omega^{R*}}{w^S + w^M + w^D + w^R} \tag{3.5}$$

where Ω^* is the optimal efficiency score of the supply chain or value chain, Ω^{a*}, $a \in \{S, M, D, R\}$, is the optimal efficiency score for a specific supply chain member (channel) and w^a, $a \in \{S,M,D,R\}$ is the weight reflecting the extent of each channel contributing to the evaluation of the entire value chain efficiency. These weights can be estimated using various methods such as AHP (Analytic Hierarchical Process), expert's judgement, pareto analysis and etc. In this research, we consider all channels have equal contribution to the value chain performance. As the indirect effect (i.e. the performance improvement of one channel affecting another channel) has already been removed/discounted from the model (3.4), the weight measures proposed in such way would be reasonable and the 'double counting' effect on the performance of the entire supply chain will not be very significant. Note that in the study we set $w = 1$.

From Model (3.4), a supply chain is efficient if $\Omega^* = 1$. Note that it is possible among all DMUs, the highest value of Ω^* is < 1. In this case, it means that none of the DMUs is efficient. Comparing Model (3.4) to (3.1), as the values of Ω^* and θ^* have to be greater than 0 and less than or equal to 1, and as Model (3.4) has less restriction on the intermediate inputs, the value of Ω^* from Model (3.4) will always be less than or equals to the value of θ^* from Model (3.1) i.e. $\Omega^* \leq \theta^*$.

Proposition 1. *The efficiency score, Ω^* of (3.4) for any DMU j_0 is less than or equal to the corresponding efficiency score from θ^* (3.1).*

To prove this proposition, we note first that $\theta^* \leq 1$ in optimal solution of (3.1) because DMU j_0 is itself one of the $j_0 \in J$ referent observations. By comparing the constraint sets in the two linear programs, we see that any optimal solution to (3.1) is a feasible solution for (3.4); hence, $\Omega^* \leq \theta^*$.

Model (3.4) yields the target values on the performance measures for an inefficient supply chain to reach the best practice by using its slack information. The model assumes that the inputs could be reduced while maintaining all the outputs at the same level. The target values are obtained as follows. We denote $x^*_{sj_o}$ and $x^*_{tj_o}$ as the direct and indirect input targets i.e., $x^*_{sj_o} = \Omega^{a*} x_{sj_o} - s^-_{sj_o}$ and $x^*_{tj_o} = x_{tj_o} - e^-_{tj_o}$

where $s^-_{sj_o}$ and $e^-_{tj_o}$ are the direct and indirect input slacks respectively.

3.4 Efficiency measurement in stochastic environment

In this section, we are going to discuss how to measure the efficiency in stochastic environment. First, we explain what is the common approach that the users used when applying the DEA model in stochastic environment. Then, we discuss the limitation of using this way. Next, we will show how to overcome this problem by introducing an alternative method which is called the Monte Carlo DEA.

3.4.1 Common approach when applying DEA model in stochastic environment

In the deterministic DEA model, the users only use a single value or single data for each input/output and calculate the efficiency score as a discrete value. In other words, the true values of the input/output are known or deterministic. In actual application, that is when the environment is stochastic, the true values maybe unknown. Without loss of generality, if the inputs/outputs values that the users interested in are the true mean of the data, then in order to use the same DEA model to estimate the efficiency, the users have to collect some data and then use the sample mean to replace the true mean. For instances, for cycle time, where its true value is very difficult to be determined precisely, users have to collect a few data and use the sample mean to represent its true mean. By using this way, the efficiency score remains as a single value. The limitation of this method is that the sample mean is the true mean unless we have collected infinite amount of data; if there are only a few data collected, the sample means may be very different from the true mean, hence the efficiency score will not be accurate.

An alternative way to tackle this issue is based on the data collected, we derive the belief (distribution) for the true mean for the stochastic inputs/outputs variables using Bayesian framework. Then, using the distribution of these inputs/outputs, we can estimate the distribution of the efficiency scores. Next, we will explain the Monte Carlo DEA method which is based on the Bayesian approach.

3.4.2 Monte Carlo DEA

The main concept of Monte Carlo DEA is to use Monte Carlo method to sample data from the distributions (belief) that represent the unknown true mean of the input/output variables and then use these samples to estimate the distribution of the efficiency. Without loss of generality, in this study we assume the belief follow a normal distribution with given mean and variance. Note that we want to find a conjugate family where we can derive the posterior distribution, and we are interested in expected value; therefore, normal can be a good approximation. We use Monte Carlo method to randomly generate N sets of data for these unknown variables. Then, for each set of data, we solve the LP model (Model (3.4)) to obtain the efficiency; hence, we can get N efficiency data. We set N to a large number so that it is big enough to get the distribution of the efficiency. Note that in the experiment, we set N to 500. We used linear programming optimization solver to calculate the efficiency scores for each set of data. The efficiency scores are then tabulated and statistical inferences are conducted. To summarize, the Monte Carlo DEA technique with reference to Model (3.4) can be carried out in the general steps below.

a. Generate N sets of input/output data, where the data follow the given distribution.

b. For each set of data, solve the LP Model (3.4) to obtain the efficiency score (Ω^*) and other parameters values i.e. $\Omega^{a*}, \lambda_j^{a*},\ a \in \{S, M, D, R\}$.

c. Estimate the distribution of the efficiency for each DMU.

In order to facilitate the discussion in the later chapters which will be based on the Monte Carlo DEA method, we simplify the representation of the model where the inputs/outputs are grouped together as a single term called ‘data’. To start with, let’s revisit Model (2.3). Let S be the set of inputs and R the set of outputs, where S and R are disjoint sets ($S \cap R = \varnothing$). We denote K as the set of combined inputs/outputs i.e., $K = S \cup R$. J is the set of DMUs. Let $X_D = (x_{kj})_{k \in K;\, j \in J}$, where x_{kj} represents k-th input/output for DMU j. If $k \in S$, then x_{kj} is an input; otherwise if $k \in R$, then x_{kj} is an output. We define $\theta(X_D)$ as efficiency score for DMU j_0. Hence, Model (2.3) can be written as Model (3.6) below:

$$\theta(\mathbf{X_D}) = \min \quad \theta$$

$$\text{s.t.} \quad \sum_{j \in J} \lambda_j x_{sj} \le \theta x_{sj_o} \qquad s \in S$$

$$\sum_{j \in J} \lambda_j x_{rj} \ge x_{rj_o} \qquad r \in R$$

$$\lambda_j \ge 0, \qquad j \in J \tag{3.6}$$

Note that θ^* is the optimal solution obtained by the model. The explanation of the variables and the details of the model are similar to Section 2.5.1. Similarly, Model (3.4) (DEA supply chain model) can be simplified and written as Model (3.7) below. Let S be the set of inputs and R the set of outputs, where S and R are disjoint sets ($S \cap$

$R = \varnothing$). K is the set of combined inputs/outputs i.e., $K = S \cup R$. Let DS be the set of direct inputs, IS the set of indirect inputs and both DS and IS are disjoint sets. Note that $DS \cup IS = S$. J is the set of DMUs. Let $X_D = (x_{kj})_{k \in K; j \in J}$, where x_{kj} represents k-th input/output for DMU j. If $k \in DS$, then x_{kj} is a direct input; if $k \in IS$, then x_{kj} is an indirect input; otherwise if $k \in R$, then x_{kj} is an output. Note that we do not particularly segregate the output into direct or indirect because the indirect term for output does not enter into the objective function; hence it does not affect the model. We define $\Omega(X_D)$ as the supply chain efficiency score for DMU j_0. The optimal solution obtained by the model is given by Ω^*.

$$
\begin{aligned}
&\Omega(\mathbf{X_D}) = \min \Omega \\
&\text{s.t.} \quad \sum_{j \in J} \lambda_j x_{sj} \le \Omega \; x_{sj_o}, && s \in DS \\
&\qquad \sum_{j \in J} \lambda_j x_{tj} \le x_{tj_o}, && t \in IS \\
&\qquad \sum_{j \in J} \lambda_j x_{rj} \ge x_{rj_o}, && r \in R \\
&\qquad \lambda_j \ge 0, && j \in J
\end{aligned}
\tag{3.7}
$$

The explanation of the variables and the details of the model are similar to Model (3.4). Model (3.6) and (3.7) will be used in the second part of the thesis. Next, we explain the structure of the supply chain that we plan to model and the variables which we use to measure supply chain efficiency. Then, it will be followed by a description on how the numerical run is setup,

3.5 An application study

In this section, we discuss an application study on supply chain efficiency measurement. First, we explain the overall conceptual model for measuring the supply chain efficiency, the variables and data used for the study. Then, it will be followed by setup of the experiments and finally results discussion.

3.5.1The overall conceptual model for measuring supply chain efficiency

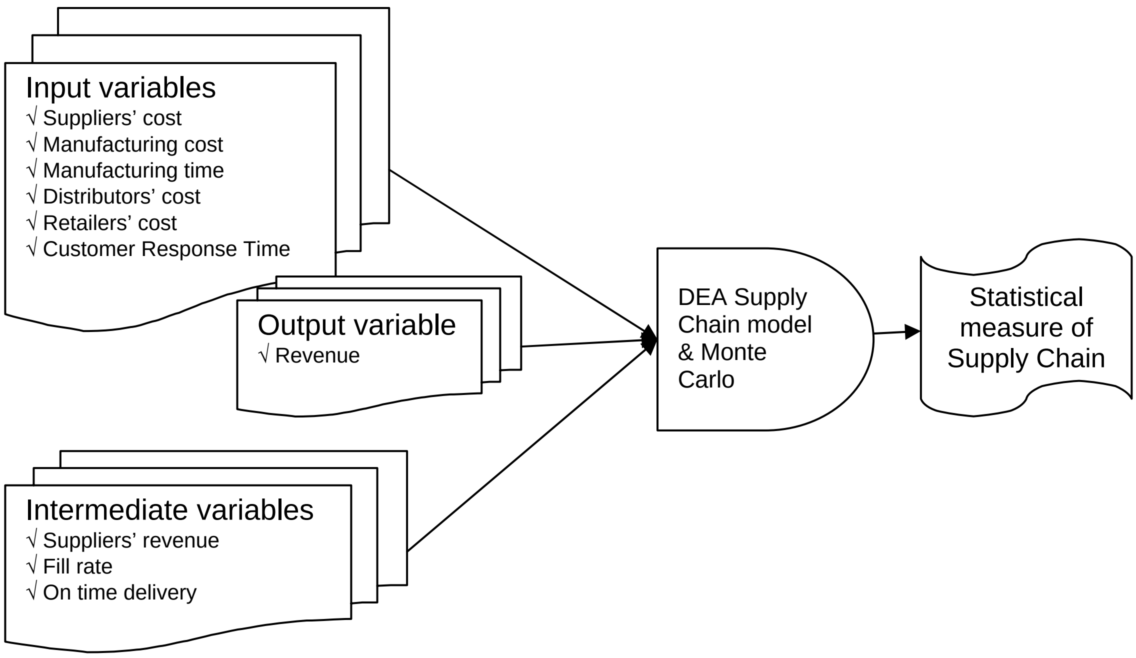

Figure 3.1: Conceptual model for measuring supply chain efficiency in stochastic environment

Figure 3.1 shows the conceptual model for measuring supply chain efficiency. The input, output and intermediate variables used are categorized according to the performance metrics listed in the SCOR (Supply Chain Operations Reference). SCOR is chosen because it is the first cross-industry framework for evaluating and improving enterprise-wide supply chain performance and management (Stewart, 1997) and it is the most common standard used by industry to measure supply chain performance today. The metrics used in SCOR include key areas such as financial measures and operational measures. The operational measures can be further broken down into specific measures which are delivery performance, order fulfilment and production flexibility.

The DEA supply chain model is used as a tool to analyze these variables. The evaluation of the supply chain efficiency needs to consider some "intermediate" variables. The categorization of these intermediate measures is determined through the coordination among related supply chain members (Parlar and Weng, 1997; Thomas and Griffin, 1996). Table 3.3 illustrates the input, output and intermediate variables. The input and output variables are defined following the standard definition used by analysts in supply chain management.

Table 3.1: Variables used in the DEA supply chain model

Measures	Output variables	Intermediate variables	Input variables
Financial measures	Revenue	Supplier's revenue	Each supply chain member's cost
Supply Chain Operational measures	-	Fill rate, On-time delivery	Manufacturing time, Customer response time

Table 3.2: Breakdown of the variables according to supply chain member.

Echelons	Supplier	Manufacturer	Distributor	Retailer
Inputs	Supplier's cost	**Supplier's revenue* Manufacturing cost	Distributor's cost Customer response time **Fill rate*	Retailer's cost **On-time delivery*

		Manufacturing time		
Outputs	**Supplier's revenue*	**Fill rate* **On-time delivery*	**On-time delivery*	Retailer's revenue

***Note**: Supplier's revenue, fill rate and on-time delivery are also the intermediate variables

The definitions for each measure are given below:

1. Financial measures:
    a. Revenue - This is a common measure of efficiency in various profit-oriented organizations.
    b. Cost - This is the performance attribute for supply chain costs, i.e. the costs associated with operating the supply chain.
2. Operational measures:
    a. Fill rate – This is a performance attribute for supply chain reliability. In the broadest sense, fill rate refers to the service level between two parties. It is usually a measure of shipping performance expressed as percentage. In this paper, fill rate is treated as a cost measure to the distributor, which is associated with inventory holding cost and the amount of products required from the manufacturer.
    b. On-time delivery rate - This is a common performance attribute for 'supply chain delivery reliability'. It refers to the performance of the supply chain in delivering the correct product, to the correct place, at the correct time, in the correct condition and packaging, in the correct quantity, and with the correct documentation to the correct customer.
    c. Customer response time - It is the performance attribute for 'supply chain responsiveness'. It refers to the velocity at which a supply chain provides products to the customers.

d. Manufacturing time - This is the performance attribute for ‘production flexibility’. It refers to the agility of a supply chain in responding to marketplace changes to gain or maintain competitive advantage.

In this study, the subject measure for fill rate will be referred from the manufacturer to the distributor (not from manufacturer to retailer, or from distributor directly to retailer). We assume that fill rate is associated with the amount of products required from the manufacturer. The distributor will always try to meet the needs of its customer while setting an appropriate level of fill rate. A high fill rate incurs additional storage and holding cost to the distributor, while a low fill rate may not be able to satisfy customers demand. An optimal level of fill rate is usually determined from the tradeoff between rate of customer order fulfilment and inventory level. As such, we assume that the fill rate between manufacturer to distributor has more significant impact on the supply chain efficiency compared to the fill rate between the manufacturer to retailer and from distributor to retailer.

Table 3.2 shows the breakdown of the inputs, outputs and intermediate variables according to each supply chain members. For the supplier, we use operating cost as direct inputs and revenue as the output. This revenue becomes an intermediate input to the manufacturer. The revenue from the supplier can affect the manufacturer performance in such way e.g. assume that the purchasing cost of the manufacturer can be increased or reduced; when the supplier increases its selling price to enhance its revenue, this increased revenue means increased cost to the manufacturer and consequently, the manufacturer may become inefficient. Alternatively, if the supplier reduces its selling price as part of revenue sharing contract with the manufacturer, this in turn will reduce the purchasing cost of the manufacturer and the manufacturer will subsequently become efficient. For the manufacturer, we use manufacturing cost and manufacturing lead time as two direct inputs, in addition to the intermediate input i.e. supplier's revenue. For the distributor, we use distribution cost and customer response time as two direct inputs in addition to the intermediate input (fill rate) linked with the manufacturer. For the retailer, in addition to the intermediate input from the distributor which is on-time delivery, we have one direct input of number of backorders and one output of profit. Backorders are retailer's cost while profit is equivalent to revenue.

3.5.2 Data used for the study

To make matters more concrete in the use of the proposed supply chain efficiency model, a survey was designed to collect inputs and outputs variables data from various companies. The companies from the semiconductor sector were selected. The sampling source for the companies was obtained from the Penang Development Corporation (PDC), Malaysia. These companies have their manufacturing plants located in the Penang Free Trade Zone. There are about 50 semiconductor companies listed in the PDC database and all these companies are selected for this study. These companies have similar logistic distribution network and operating in the similar businesses. As we are using DEA to measure the efficiency, that is the relative performance of decision making units (DMUs) are measured on the basis of the observed operating practice in a sample of comparable DMUs (i.e., homogenous units), therefore it is a fair comparison.

Data collection of the input and output variables was done via different methods. First of all, revenue and supply chain cost were obtained from the companies' financial reports. Note that, the revenue figures may include revenue generated from other businesses; however, due to the fact the companies which we selected operate in the same business, the effect of revenue generated from other businesses would be minimal. Secondly, fill rate, cycle time and on time delivery rate were collected from the questionnaires which were mailed to the supply chain managers. Thirdly, site interviews and telephone calls were made to follow up on the questionnaires and to validate their answers. We received responses from 30 companies resulting in a response rate of 60 percent. Of these responses, 10 had all items completed and were usable for this study. Since the data are used to compute the rankings of relative efficiency, the low response rate does not affect the accuracy of the DEA outcomes. These data were then used in the DEA supply chain model and the solutions were obtained using Excel and its linear optimization solver. A total number of 10 DMUs was analyzed in this study.

3.5.3 Setup of the experiments

The first part of the study addresses the model from the deterministic perspective. Table 3.5 shows the data of the 10 DMUs.

Table 3.3: Supply chain data

DMU	**Unit**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Supplier-cost	Million USD	130	150	165	170	200	185	135	190	185	190
Supplier-revenue	Million USD	20	21	23	24	27	25	24	30	28	25
Manufacturing cost	Million USD	125	120	110	150	146	115	105	100	135	120
Manufacturing time	Days	3	2	3	4	2	3	2	2	4	3
Distributor cost	Million USD	90	100	80	70	85	77	78	90	78	68
Customer response time	Days	3	3	2	4	2	2	1	3	2	1
Fill rate	%	70	90	78	88	73	95	89	87	95	90
On-time delivery	%	96	95	97	89	99	89	93	88	99	83
Retailer cost	Million USD	100	110	130	125	140	135	125	155	135	130
Retailer revenue	Million USD	310	220	300	230	320	240	350	370	325	355

In the second part of the study, we will address the stochastic case. We generate some stochasticity in the data, following the steps mentioned in Section 3.4. We choose retailer's revenue and manufacturing time (cycle time) as the random variables because they have the most significant impact on the efficiency in Model (3.4) compared to other variables. Furthermore, in reality, revenue and cycle time data are difficult to be obtained precisely. Hence, by choosing these variables as random variables, it can validate the accuracy of the proposed model. We assume that the revenue and cycle time data follow a normal distribution. The mean and variance of the random variables used in the study for each DMU is listed in Table 3. below. Note that the values of the mean and variance for the stochastic variables are determined through an analysis carried out on the data collected from the survey. In addition, we also seek some advice from the users on the appropriate values to be used. Next section will discuss on the results obtained.

Table 3.4: Distribution of the random variables

DMU	Cycle time, Unit : Days			Retailer's revenue, Unit: Million (USD)		
	Mean	Variance	Standard deviation	Mean	Variance	Standard deviation
1	3.08	0.9293	0.96	318	1406.25	37.5
2	2.09	0.8391	0.92	220	625.00	25.0
3	2.81	0.8118	0.90	301	1135.69	33.7
4	3.90	0.8874	0.94	226	761.76	27.6
5	1.91	1.1664	1.08	318	1346.89	36.7
6	3.08	1.0404	1.02	232	761.76	27.6
7	2.44	0.1444	0.38	350	1840.41	42.9
8	1.98	0.8464	0.92	366	2070.25	45.5
9	3.99	0.9409	0.97	324	1797.76	42.4
10	2.85	1.0404	1.02	359	1632.16	40.4

3.5.4 Results and discussions

Table 3. shows the efficiency score for each individual member as well as the overall supply chain.

Table 3.5: Deterministic efficiency score

DMU	Member Efficiency (Model 3.1)					Supply Chain Efficiency (Model 3.4)				
	Supplier	Manufacturer	Distributor	Retailer	Average	Supply Chain	Supplier	Manufacturer	Distributor	Retailer
	θ^{S*}	θ^{M*}	θ^{D*}	θ^{R*}	θ^{*}	Ω^{*}	Ω^{S*}	Ω^{M*}	Ω^{D*}	Ω^{R*}
1	0.865	1	1	1	0.966	0.933	0.918	0.970	0.843	1
2	0.881	1	0.880	0.673	0.859	0.601	0.600	0.625	0.465	0.714
3	0.964	1	1	0.810	0.944	0.791	0.720	0.747	0.747	0.948
4	0.870	0.856	1	0.754	0.870	0.576	0.503	0.447	0.690	0.663
5	0.895	1	1	0.820	0.929	0.795	0.688	0.768	0.768	0.954
6	0.937	0.999	1	0.673	0.902	0.613	0.529	0.579	0.625	0.717
7	1	1	1	1	1.000	1	1	1	1	1
8	1	1	0.811	1	0.953	0.943	1	1	0.770	1
9	0.994	0.904	1	0.856	0.938	0.907	0.994	0.849	1	0.784
10	1	0.986	1	1	0.997	0.992	1	0.968	1	1

To compare whether Model (3.1) or Model (3.4) is better, we evaluate the supply efficiency using both models. The values of θ are obtained by solving Model (3.1) separately for each member with respective to every DMU. The average value for each DMU (column 6) is then calculated by averaging out the entire member's efficiency. Meanwhile, the values of Ω^* and Ω^{a*} are obtained by using Model (3.4). From the analysis, only one supply chain which is DMU 7, is efficient ($\Omega^* = 1$). This means that DMU 7 represents the best practice of the supply chain system and in its case, all its supply chain members are efficient (Ω^{S*}=Ω^{M*}=Ω^{D*}=Ω^{R*}=1) as well as (θ^{S*}=θ^{M*}=θ^{D*}=θ^{R*}=1). Recall that θ^{a*} is the individual member's efficiency score for supply chain member *a* obtained from Model (3.1). The results show that the average supply chain member efficiency score (column 6) which is obtained from Model (3.1) is always greater than or equals to the supply chain efficiency score (column 7) which is obtained from Model (3.4).

Table 3.6: Target values for inputs, outputs and intermediate variables for DMU 1.

DMU 1	**Original value**	**Target value**	**% Change**
Supplier-cost	130	119.37	-8.18
Supplier-revenue	20	21.22	6.11
Manufacturing cost	125	121.26	-2.99
Manufacturing time	3	2.91	-2.99
Distributor cost	90	75.91	-15.65
Customer response time	3	2.53	-15.65
Fill rate	0.7	0.91	29.92
On time delivery	0.96	0.96	0.00
Retailer cost	100	100	0

For example, for DMU 1 its average supply chain member efficiency score θ^* is 0.966 and the supply chain efficiency score Ω^* is 0.933. Note that the reduction of the supply chain efficiency score is due to the removing of the indirect measures from Model (3.4). The value of Ω^{R*}= 1 for DMU 1 (from Table 3.5) indicates that the retailer is efficient; hence no adjustments for measures related to the retailer are required. However, in order to reach the best practice, the supplier, the manufacturer and the distributor could reduce their inputs while maintaining the same level of outputs (based upon Ω^{S*}, Ω^{M*} and Ω^{D*}, which are less than 1). In the case of DMU 1, all its direct input slacks have zero values. Thus, the supplier could reduce its cost to 119.4 (based on Ω^{S*} = 0.918); the manufacturer could reduce its cost to 121.26 and manufacturing time to 2.91 and the distributor could reduce its cost to 75.9 and customer response time to 2.53 while maintaining all the other outputs at the same level. All these target values are listed in Table 3.6. In addition, the supplier and the manufacturer could reach an agreement on the selling price of raw materials to increase the supplier's revenue by 6.11% [(21.2-20)/20]. The distributor's fill rate could be increased to 91% from the current rate of 70%. This solution indicates that based upon the best practice, the distributor could be able to maintain the fill rate of 91% while cutting down costs and cycle time. All these are the potential input savings that the supply chain could achieve. Similarly, the adjustment for other DMUs and their input savings could be interpreted using the same way. Note that if we use Model (3.1) to measure the efficiency, all these savings will not be significant. Thus, Model (3.4) is better than Model (3.1) for supply chain efficiency measurement. Appendix B (Table B.1) lists the target measure adjustments for each DMU.

Table 3.7: Target benchmark for each DMU

DMU	Supplier (λ^S)	Manufacturer (λ^M)	Distributor (λ^D)	Retailer (λ^R)
1	7	2	1	3, 9
2	8, 9	2, 5, 7	7	3, 9
3	8, 10	3	7, 8	3
4	7, 8	2, 3	7, 10	3, 9
5	8, 10	5	8, 10	5
6	8, 10	3, 7	7, 8	3, 9
7	7	7	7	7
8	8	8	8	4, 9
9	8, 10	3	7, 10	3
10	10	2, 3, 7	10	10

Addition managerial information could be obtained from Model (3.4), whereby the non-zero values of λ_j^{S*} ,λ_j^{M*} ,λ_j^{D*} and λ_j^{R*} will indicate on which DMUs are to be used as benchmarks. For example, when DMU 8 is under evaluation using Model (3.4), in the retailer column, we have non zero values for λ_4^{R*} and λ_9^{R*} , hence this indicates that DMU 4 and 9 are used as benchmarks. Similarly for DMU 1, its benchmark for supplier is DMU 7, benchmark for manufacturer is DMU 2 and benchmarks for retailer are DMU 3 and 9. The targets for the other inefficient DMUs can be interpreted using the same way.

The DEA supply chain model provides firstly an approach for characterizing and measuring the efficiency of supply chain as well as supply chain members, and secondly, makes it clear that two supply chains may have different input-output mix yet both may be efficient. This model enables supply chain members to collectively improve the supply chain performance. At the same time, it also provides information on which supply chain members are used as benchmarks in order to achieve best-practice performance and to gain a competitive edge.

Next, we will move on to the results discussion for stochastic case. Figure 3.2 contains box plot of the Monte Carlo efficiency scores by observation number. As can be seen in Figure 3.2, DMU 7 (which has an efficiency score of 1 in deterministic case) was not consistent on the frontier during the Monte Carlo application. The size of the boxes is determined by the span of values from the 25th to the 75th percentiles; as can be seen, they vary quite a bit. This indicates how sensitive a particular DMU's efficiency score is to variations in the efficiency of the other DMUs in the data set. For example, DMU 7 and 10 were both originally found to be efficient from Model (3.4), their respective efficiency scores are 1 and 0.992. In addition, the means of their Monte Carlo efficiency scores are very similar (0.954 and 0.952, respectively), but DMU 7 has a tighter distribution of Monte Carlo efficiency scores than DMU 10, indicating less precision in DMU 10's scores. Appendix B (Table B.2) showed the distribution statistics of the efficiency scores for each DMU.

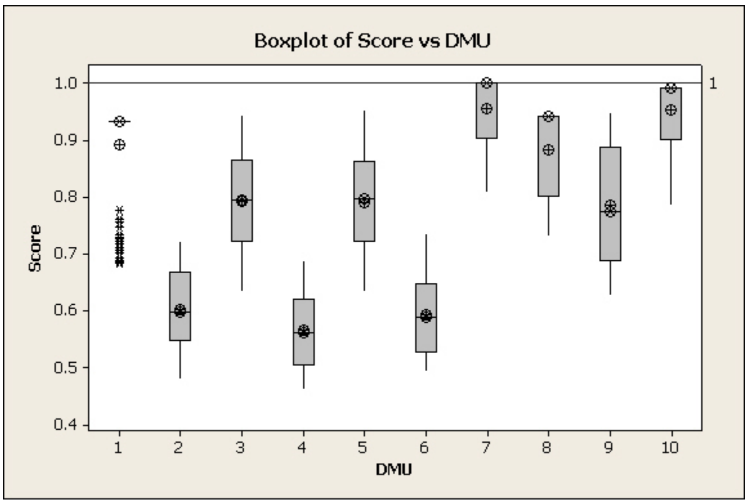

Figure 3.2: Boxplot of the Monte Carlo efficiency score

Compared to the point estimates from traditional DEA model, the distribution of the efficiency is able to provide more useful information to the managers. With it, managers could know where the efficiency normally lies; hence, they would be able to gauge the reliability of the results.

Based on the mean efficiency values from Figure 3.2, we can group the observations into three groups which are high efficiency DMUs, medium efficiency DMUs and low efficiency DMUs. The high efficiency group consists of DMU 1, 8, 10 and 7 (with efficiency scores ranging from 0.8~1.00); the medium efficiency group consists of DMU 3, 5 and 9 (with efficiency scores ranging from (0.7~0.8); and the low efficiency group consists of DMU 2, 4 and 6 (with efficiency scores lower than 0.7). Figure 3.3 to 3.6 showed the excess distribution function for the three categories of DMUs. From Figure 3.4, it is evident that DMU 8 is first order stochastically dominated by DMU 1, 7 and 10. From Figure 3.5, all the DMUs 3, 5, and 9 do not stochastically dominate each other. From Figure 3.6, DMU 2 and 6 first order stochastically dominate DMU 4. From these graphs, managers could be able to find out whether are there any possibilities for a particular DMU to be more efficient than the other. An interesting finding is that, it is possible for DMU 10 to be more efficient than DMU 7 although the average score of DMU 10 is lower than DMU 7. This is not evident from the deterministic model. By using this information, managers will be able to make better decisions and appropriately strategize to improve their supply chain performance.

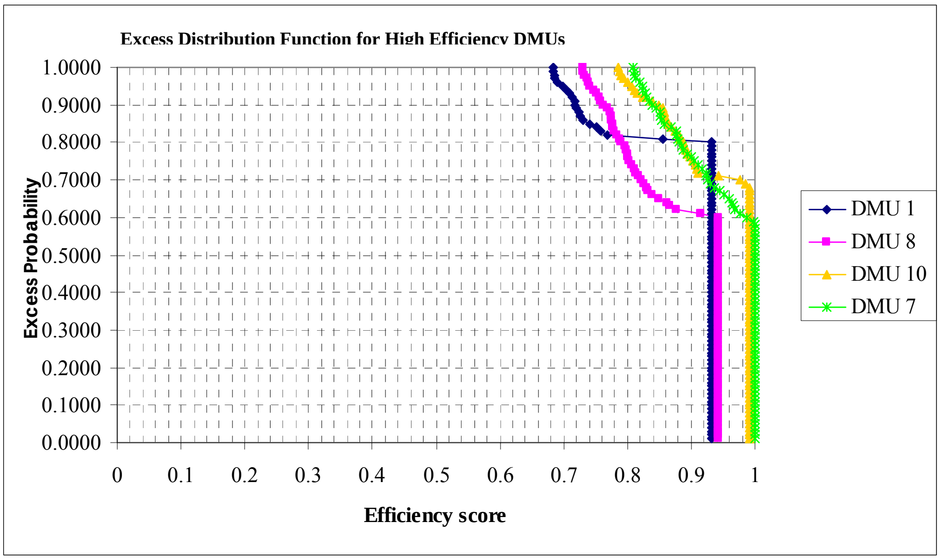

Figure 3.3: Excess Distribution Function for ‘High Efficiency’ DMUs

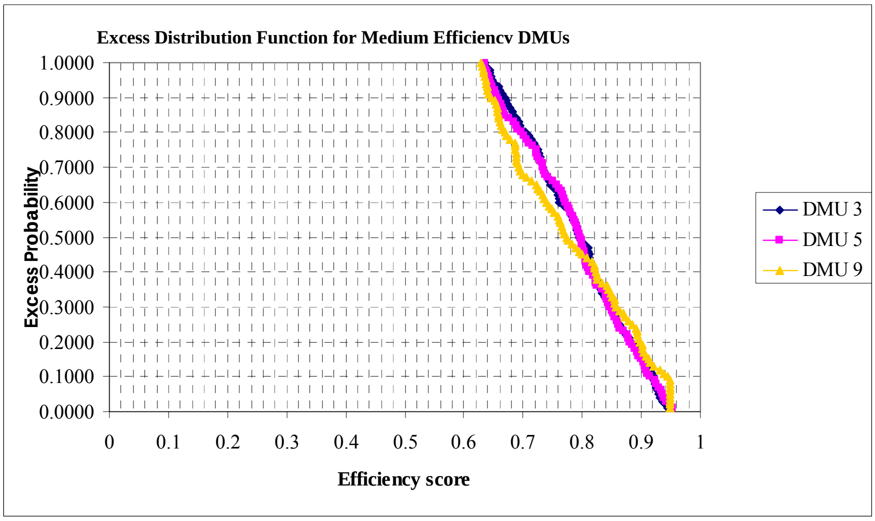

Figure 3.4: Excess Distribution Function for ‘Medium Efficiency’ DMUs

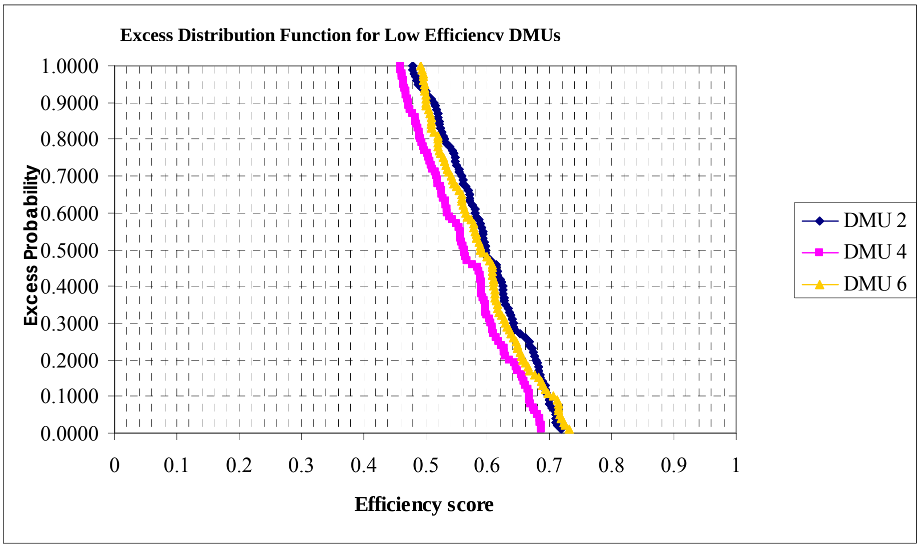

Figure 3.5: Excess Distribution Function for 'Low Efficiency' DMUs

Note that Figures 3.5 and 3.6 do not appear to be approximate "uniform distribution". There is no specific reason on why the distributions behave this way. This may be due to the data used in the evaluation.

Table 3.8: Ranking of DMUs

Rank	Deterministic case		Stochastic case			
	Efficiency score	DMU	Mean efficiency score	DMU	Median efficiency score	DMU
1	1	7	0.954	7	1	7
2	0.992	10	0.952	10	0.992	10
3	0.943	8	0.892	1	0.942	8
4	0.933	1	0.884	8	0.933	1
5	0.795	5	0.793	3	0.798	5
6	0.791	3	0.791	5	0.795	3
7	0.783	9	0.785	9	0.774	9
8	0.613	6	0.602	2	0.597	2
9	0.601	2	0.593	6	0.589	6
10	0.576	4	0.566	4	0.562	4

Another implication of our findings is that care should be taken when ranking DMUs in terms of their efficiency scores. Table 3.8 shows the ranking comparison between stochastic case and deterministic case. The stochastic case is divided into mean-based ranking and median-based ranking. The results showed that all the three methods of ranking differ. It is often difficult for managers to distinguish between mean and median of the efficiency score on which one is a better estimate. The decision maker would have to decide based on his/her own discretion on which estimate to use for performance ranking. Alternatively, the excess distribution functions would be able to shed some lights in handling the discrepancies incurred. Figure 3.3 showed that DMU 10 has higher chances of achieving efficiency score of 1 (about 70% chances) compared to DMU 7, which only has about 60% chances. (Note: these values are read from the y axis in correspondence to the efficiency score of 1 at the x-axis). DMU 1 is also slightly dominant over DMU 8 (as can be seen from Figure 3.3 that the accumulated area under the cumulative frequency curve for DMU 1 is slightly greater than that of DMU 8. Similarly, for the medium efficiency DMUs (refer Figure 3.4), DMU 9 is always dominated by DMU 3 and 5. For the low efficiency DMUs (Figure 3.5), DMU 6 and 2 are always more efficient than DMU 4. Hence, the ranking results provided by the stochastic model are able to highlight some discrepancies and provide important insights to managers which are not evident if we use the deterministic model.

Target benchmarks in stochastic model

Table 3.9: Target peers and percentage of time for target benchmark for each DMU

DMU	Target Supplier	Target Manufacturer	Target Distributor	Target Retailer
1	7 (94.5%), 1(5.5%)	2 (99%), 1(1%)	1 (98%) 7 (2%)	3,9 (97%), 1(3%)
2	8, 9 (98.5%), 2(1.5%)	2 , 5, 7 (100%)	7 (99%), 2(1%)	2(2.5%), 3,9 (97.5%)
3	3(2%), 8,10(98%)	3 (99%), 2(1%)	3(1.5%), 7, 8 (98.5%)	3(99%), 9(1%)
4	7, 8 (96.5%), 4(3.5%)	2, 3 (98.5%), 4(1.5%)	7, 10 (99%), 4(1%)	3, 9(99%), 4(1%)
5	8, 10 (97%),5(3%)	5 (99%),2(1%)	8, 10 (95%), 5(5%)	5 (99%), 3(1%)

6	8, 10 (98%), 6(2%)	3, 7 (98.5%), 6(1.5)%	7, 8 (95%), 6(5%)	3, 9 (98%), 6(2%)
7	7 (99%), 8(1%)	3, 5 (1%), 7(99%)	1, 8 (44.5%), 7(55.5%)	7 (99%), 9(1%)
8	8 (99%), 10(1%)	3,7 (36.2%), 8(63.8%)	8 (75.4%), 7,10(24.6%)	4, 9 (98%), 8(2%)
9	8, 10 (97%), 9(3%)	3 (99.5%), 9(0.5%)	7, 10 (99%), 9(1%)	3 (95%), 9(5%)
10	10 (99%), 7,8(1%)	2, 3, 7 (99.5%), 10(0.5%)	10 (99%), 7,8(1%)	10 (98%), 3,9(2%)

Table 3.9 shows the target benchmarks and percentage of time for each target to become benchmark for each DMU when the model is stochastic. The non bracketed integer in Table 3.9 denotes the target DMU to refer to as benchmark while the numerical value (in bracket) indicates the percentage of time (frequency) for the DMU in becoming the target benchmark. The results obtained in the stochastic case are different from the deterministic case. For instances, for DMU 7, its peer for distributor is itself only in the deterministic model. However, in the stochastic model, it has two additional targets which are DMU 1 and 8, where 44.5% of chance the targets will be these two DMUs. Similarly, for DMU 8, it has two additional targets for manufacturer (DMU 3 and 7) where 36.2% of chance the targets are DMU 3 and 7, and 63.8% of chance the target is DMU 8 and its target distributors are DMUs 7, 8 and 10, where 75.4% of chance the target is DMU 8 and 24.6% of chance the targets are DMU 7 and DMU 10.

Additional targets are expected to occur in stochastic model due to the following reason. When uncertainties occur, additional precaution measures would have to be taken in most of the time. This is often carried out by making more comparisons with other DMUs and setting more targets to improve the existing performances. Hence, this is apparently depicted in the target benchmark results, where more additional targets will be identified in the stochastic case compared to the deterministic case. In addition, if target benchmark is only carried out using the deterministic way, wrong target might be identified and this could jeopardize the overall effort in performance benchmarking.

The Monte Carlo study conducted here manages to point out that in the stochastic environment there may be additional or different target peers for all the DMUs. This piece of important information would be missing if analysis is only based on conventional-LP based DEA model. Especially, for the efficient DMUs, (which have obtained a score of 1), they would be contented and thought that their processes are extremely efficient. Hence, they will not carry out any further improvements in their processes. But, this is not true. In actual case, their performances are not robust and they still need to improve further by fine tuning and comparing with other better (or equivalent) target peers (benchmarks) like some of the examples mentioned above.

In actual industry practices, upon obtaining the value of the efficiency score, managers will then use it to adjust their input or output measures. Hence, in the stochastic case, we could use the distribution of the efficiency to get some additional insights towards the distribution of the measure adjustments. Table 3.10 shows the measure adjustments for DMU 7 for stochastic case.

Table 3.10: Measure adjustments for DMU 7

DMU 7	**Measure adjustments**			
	10%< ≤20%	5%< ≤10%	0%< ≤5%	0%
Supplier-cost	0.0%	0.0%	1.0%	99.0%
Supplier-revenue	0.0%	0.0%	1.0%	99.0%
Manufacturing cost	0.0%	0.2%	0.8%	99.0%
Manufacturing time	0.0%	0.5%	0.5%	99.0%
Distributor cost	14.0%	17.0%	13.0%	56.0%
Customer response time	16.5%	12.5%	15.5%	55.5%
Fill rate	11.0%	14.0%	19.5%	55.5%
On time delivery	18.0%	10.0%	15.0%	57.0%
Retailer cost	0.0%	0.0%	1.0%	99.0%

The adjustments are categorized into 4 groups which are ‘0%’, ‘0%< ≤ 5%’, ‘5%< ≤ 10%’ and ‘10% < ≤ 20%’. These represent the percentage of adjustments: i.e. ‘0%’ means no adjustment is needed, ‘0%< ≤ 5%’ means that adjustment greater than 0% but less than or equals to 5% is needed for the respective measure and etc. This categorization is based on the distribution obtained from the numerical runs. The results showed that there are differences in the measure adjustments between the stochastic and deterministic model. For instances, in the stochastic model, 14% of the time, DMU 7 would need to adjust its distributor cost between 10% to 20%, 17% of the time, it needs to adjust between 5% to 10%, 13% of the time it needs to adjust between 0% to 5% and 56% of the time, no adjustment is needed. In contrast to the deterministic model, DMU 7 does not need to adjust any of its measures because it is fully efficient. These are obviously vast different conclusions. The results show that the choice of DEA model and the implication of that choice are very serious in terms of managerial decision making.

3.6 Conclusion and Managerial Implications

This chapter developed the DEA supply chain model as well as a simple tool which is called the Monte-Carlo DEA to measure supply chain efficiency in stochastic environment. Though the DEA supply chain model which we develop may not have fully addressed all the concerns in supply chain efficiency measurement, it is still better than the conventional method and can be a tentative solution. The Monte Carlo DEA method has given a more meaningful interpretation to the efficiency. In contrast to the point estimate of the efficiency score given by the conventional DEA model, it is able to make statistical inferences on the efficiency. The additional information provided such as the distribution, target benchmarks, measure adjustments and other statistical measures are invaluable to managers. They provide additional useful insights to managerial decision makings. For instances, decision maker could use the confidence intervals to gauge the reliability of the calculated efficiency scores. The stochastic model approach also offers a new avenue to managers in analyzing the supply chain performances rankings. The target benchmarks and adjustments provide additional reference sets for the inefficient DMUs which are not evident from the conventional method. In order to demonstrate the usefulness of this method, an application study on supply chain was conducted. The results obtained from the analysis support the validity of the model. The results were depicted in graphical forms to enhance the understanding of the analysts and vast significant differences were found between the conventional (deterministic) and the proposed methodology. The contribution of this study provides useful insights into the use of Monte Carlo technique combined with DEA as a mathematical modelling tool to aid managerial decision making in measuring supply chain efficiency. Given the ever increasing availability of cheap computing power, the Monte-Carlo DEA based approach appears to be a valuable tool for decision makers. It provides them with a technique for attaching statistical precision and greater confidence to the efficiency analysis that may form the basis of important decisions.

CHAPTER 4

BUDGET ALLOCATION FOR EFFECTIVE DATA COLLECTION IN PREDICTION OF AN ACCURATE EFFICIENCY SCORE

In chapter 3, we presented a Monte Carlo DEA based approach to measure the supply chain efficiency in stochastic environment. By using the Monte Carlo DEA way, we are able to get the distribution of the efficiency and know where the efficiency lies most of the time. Starting from this chapter onwards, we will address the second part of the thesis, where we will provide an approach on how to get a good estimate of the efficiency score. We will focus on given that a user can collect additional data, how he/she collects the data effectively so as to obtain a better estimate of the efficiency score.

4.1 Introduction

In the previous chapter, we have used Monte Carlo method to estimate the distributions for the efficiency scores based on the distributions for the inputs/outputs. The attractiveness of this Monte Carlo DEA approach is its computational simplicity and its ability to give statistical inferences of the efficiency. From the numerical study conducted, we obtained some useful insights which form the basis for the second part of our research. We found that the changes in the distributions of the inputs/outputs have impact on the distribution of the efficiency. The following are some of the observations obtained from the study: a) when the distributions of the inputs/outputs variables are narrower, the distribution of the efficiency will be narrower. In other words, when there is less variation in the inputs/outputs variables, the efficiency score will be more accurate. b) Each input and output has different impact on the efficiency. Some may cause the efficiency to vary more, while some do not have any significant impact at all. Based on these observations, we knew that distributions of the inputs/outputs will affect the efficiency score. As the distributions of the inputs/outputs are determined by how the data are collected, building on the first work, we now proceed to address in the context of data collection, what is the better way to collect data so that the efficiency score will be accurate.

In real application, when the true values of the inputs/outputs are unknown, we need to collect data and use the sample means to estimate these true values or true means. Then the inputs/outputs will be a distribution which depends on the amount of data collected for that input/output. Intuitively, the more data we collect, the lesser the spread of the belief of the true mean. In other words, we know where the true mean locates more precisely when we have collected more data. At such, we would want to collect as much data as possible. However, due to the exorbitant cost in conducting data collection in reality, it is usually a common practice in any organizations that the data collection is not infinite and will be limited to a certain budget. In this case, any attempt to collect data often raises one question. *How should we allocate the budget - how many data should we collect for each input/output?* If we naively allocate the data collection effort fairly, the efficiency estimated might not be accurate. Hence, it is important to know how to allocate our budget for data collection in order to get a better estimate for the efficiency. Different allocation or data collection schemes will affect the accuracy of the efficiency. Intuitively, to ensure high accuracy of the efficiency, more efforts or budget should be spent on collecting those data that are critical in the process of predicting the efficiency. By conducting the data collection intelligently, the users can make better use of their resources. Ideally, one would like to find out what is the best data collection scheme which yields the best accuracy for the efficiency within a given budget. To do this, we first need to define what we meant by accuracy of efficiency.

4.2 Definition of accurate efficiency

Accuracy of an estimator refers to the degree of conformity of the measurement to its actual (true) value. In other words, the degree of variation / deviation of the measurement to its actual value; the lower the variation, the better is the accuracy. The literature on efficiency study had mostly focused exclusively on the variance as the measure for accuracy (Gong, 1995; Grosskopf, 1996). However, there is other indicator, such as bias, which is of no less importance and should also be subsumed into the performance metric for the accuracy of the efficiency. The methodical approach to obtain the efficiency with minimum variance and bias is still missing at large (Simar and Wilson, 2000). In this research, we will use mean square error or MSE as the measure for the accuracy of the efficiency. In statistics, MSE of an estimator is defined as the amount by which an estimator differs from the true value of the quantity being estimated i.e. MSE(b) = E[(b - B)^2] where b is the estimator and B is the true value of the quantity being estimated (DeGroot, 1970).

We choose MSE because it is a more appropriate measure than variance due to it assesses the quality of an estimator in terms of its variation and unbiasedness. In order to suit our case, we define MSE of the efficiency as below.

$$MSE = E(\tilde{\theta} - \theta_D)^2 \qquad (4.1)$$

where $\tilde{\theta}$ is the belief towards where the true efficiency lies and θ_D is the efficiency calculated from the initial data collected. Note that we use sample mean to estimate θ_D. Recall $\theta(\mathbf{X_D})$ from previous section (Model 3.7), we will use this to represent θ_D. We can compute θ_D this way because this is a common approach adopted by users as explained in Section 3.4.1. Without loss of generality, we start off by focusing on how to improve the efficiency score for one particular DMU. The methodology proposed in this paper can be easily extended for future work to include multiple DMUs.

4.3 Problem Statement

The problem statement is, if we can only collect limited additional data, how should we distribute our efforts in collecting data so that we can get a better prediction of the efficiency. Note that we assume the effort in collecting data for different inputs/outputs is the same.

Assume we have collected some initial data for all the inputs/outputs and let **X** be the matrix of their sample averages, we will determine the data collection scheme (or allocation design) so as to minimize the MSE of efficiency score subject to a limited budget for additional samples. In our research, we refer budget as the total amount of additional data to be collected, denoted by *N*. The allocation design is denoted by $\mathbf{n} = [n_k]_{k \in K}$, where n_k represents the number of additional data collected for input/output *k*. Next, we construct the mathematical model for our problem.

4.4 Mathematical Programming Model

The model that represents our problem statement is as shown below:

$$\min F(\mathbf{n}) = E\left[(\tilde{\theta}(\mathbf{X}') - \theta(\mathbf{X}))^2 \right] \tag{4.2}$$
$$\text{s.t. } \sum_{k\in K} n_k = N$$

The objective function $F(\mathbf{n})$ is defined as the MSE of the efficiency score for allocation design **n** where $\mathbf{X}'$ is the belief of the inputs/outputs after additional data are collected following the allocation design **n**. Note that θ(X) is the efficiency score computed using Model 3.7. $\tilde{\theta}(\mathbf{X}')$ represents the belief for the true efficiency. The above model cannot be solved directly because the distribution for $\tilde{\theta}(\mathbf{X}')$ is unknown. In order to estimate $F(\mathbf{n})$, first we need to quantify $\mathbf{X}'$. We will now discuss how to derive $\mathbf{X}'$ under the Bayesian framework. Let $\mathbf{X}' = [x_k']_{k\in K}$ and x_k' is a random variable.

The rationale for the adoption of the Bayesian model in determining how data collection determines the distributions of the inputs/outputs is the ease of derivation of the solution approach. Under the Bayesian model, the belief for the unknown true value/mean of the input/output k denoted by μ_k is treated as a random variable and has a prior distribution. This prior distribution describes the knowledge or the subjective belief about μ_k before any sampling. The posterior distribution is updated after we observe the samples $\{\hat{x}_k(t), t = 1,...,n_{ok}\}$. The posterior distribution $P(\mu_k|\{\hat{x}_k(t), t = 1,...,n_{ok}\})$ summarizes the statistical properties of μ_k given the prior knowledge and sampling information. Note that n_{ok} = total number of samples, $\hat{x}_k(t)$

= the t-th sample of the performance measure. Similar to Chen et al. (2000), we assume that the μ_k has a conjugate normal prior distribution and consider non-informative prior distribution which implies that no prior knowledge is available about the performance of any design before conducting simulation. In that case, DeGroot (1970) shows that the posterior distribution of μ_k is $x'_k \sim N\left(\bar{x}_k, \frac{\sigma_k^2}{n_{ok}}\right)$ where $\bar{x}_k = \frac{1}{n_{ok}} \sum_{t=1}^{n_{ok}} \hat{x}_k(t)$ is the sample mean of the observations and σ_k^2 is the true variance which can be approximated by the sample variance $\sigma_k^2 \approx \frac{1}{n_{ok}-1} \sum_{t=1}^{n_{ok}} \left(\hat{x}_k(t) - \bar{x}_k\right)^2$. In addition to using the Bayesian framework to develop a posterior distribution for the unknown true value of the input/output after collecting the data, we also use it to approximate the belief of the true mean if additional samples are collected. Based on the approximations made in Lee et al. (2008) and Chen et al. (1996), when sample size increases, $\bar{x}_k$ and σ_k^2 do not change and if additional n_k samples are collected, the predicted posterior distribution for μ_k can be approximated by $x'_k \sim N\left(\bar{x}_k, \frac{\sigma_k^2}{n_{ok}+n_k}\right)$ where $\bar{x}_k$ and σ_k^2 are the sample mean and variance of the original n_{ok} independent samples. Hence, this explains how data collection will affect the distributions of the inputs/outputs which will ultimately determine the distribution of the efficiency score.

After using the Bayesian framework to quantify $\mathbf{X}'$, we can then estimate the distribution of the efficiency score through Monte Carlo method. Hence, we are able to estimate F given a value of $\mathbf{n}$. An estimation of $F(\mathbf{n})$ is given by

$$F(n) \approx \frac{1}{M}\sum_{i=1}^{M}(\theta(\hat{\mathbf{X}}_{[i]}) - \theta(\mathbf{X}))^2 \qquad (4.3)$$

where $\hat{\mathbf{X}}_{[i]}$ is the realization of the inputs/outputs $\mathbf{X}'$ in the replication *i* of the Monte Carlo run for allocation design **n** and *M* is the number of random data set.

Remarks: Recall that we only focus on one DMU. Hence, in our model, θ is only for the DMU that we are interested in. However, it can also be generalized to all DMUs as well.

4.5 Summary

In this chapter, we discuss the formulation of the problem statement and the mathematical model. The research problem is to find out what is the best way to allocate the budget for data collection within a restricted budget in order to get a good estimate of the efficiency score. The mean square error (MSE) is used as the metric for the accuracy of the efficiency. In order to solve the mathematical model, we used Bayesian framework to quantify how the data affects the efficiency and then applied the Monte Carlo method to estimate the MSE. Solving the model is a non-trivial task since it has no close-form formulation for the computation of MSE.

In the following chapter, we will present two methods to solve the model. We will first introduce a gradient search method, followed by a method based on Genetic Algorithm (GA). We will demonstrate how to calculate the gradient of the performance by using IPA (Infinitesimal Perturbation Analysis) and then how to use the gradient information to determine which n_k to increase, or in other words which data should we collect. Then, for the GA based method, we will use GA to search for the optimal solutions and OCBA (Optimal Computing Budget Allocation) to efficiently allocate the simulation budget.

CHAPTER 5

TWO-PHASE GRADIENT TECHNIQUE

As there is no close-form formulation to compute MSE for a given allocation design **n** in our model (refer to the model in Chapter 4), in order to find an optimal solution (or allocation), we therefore need to use a search-based method. In this method, first we generate some designs (i.e. different allocations **n)**, then we estimate the MSE. After that we repeat this entire process until we find the best MSE and the associated **n**. We adopt the simulation optimization technique which comprised of the Monte Carlo simulation to estimate the MSE given a certain allocation scheme and optimization techniques to find the better allocation scheme. The estimation of MSE using Monte Carlo method is carried out exactly the same way as in the Monte Carlo DEA. Instead of the end result which is efficiency, in this case, the end result is MSE which is calculated using the formula as previously explained in Chapter 4. The optimization techniques include the two-phase gradient technique and the hybrid GA technique. In this chapter, we discuss the two-phase gradient technique. We will address all the issues that arise.

5.1 Background Information

The two-phase gradient technique is based on gradient search. The approach that we use is that given an allocation design, we estimate the MSE and the gradient of

MSE. The gradient information provides a direction for finding a new allocation design that may have a lower MSE value. There will be two stages for this gradient technique. In the first phase, we find the gradient of MSE using infinitesimal Perturbation Analysis (IPA) approach. Then, using the gradient, we implement the hill-climbing algorithm to locate a solution at the boundary. The second phase is a gradient-based improvement approach to fine tune the solutions from the first phase. Note that the searching method only seeks for a local optimum. This is due to our research problem where the solution space can be very huge and unstructured[1].

The discussion on the two-phase gradient technique is divided into three parts. First, we explain how to find the gradient using IPA. Then we explain how we implement the hill climbing algorithm by using the gradient derived from phase 1. The hill climbing algorithm will guide the search to reach a solution at the boundary. Lastly, we explain the improvement stage on how to fine tune the solutions.

5.2. Finding the gradient using IPA

When we estimate the MSE using Monte Carlo method, at the same time, we can also estimate the gradient without rerunning simulation by using IPA. The idea of IPA is to consider how perturbation in a parameter affects the changes of the random variables generated and eventually how it changes the performance of the system.

[1] When the solution space is unstructured (e.g., the decision variables are not real numbers), it will deter the practicality of our method (i.e. using perturbation analysis (PA) to estimate the gradient for determining the local search direction) in finding the optimal solution. Though, in such case, if the solution space is not large, brute force evaluation for all the design alternatives will be able to find the optimal design; this is impractical in our research problem due to the huge solution space. In order to find the optimal solution for our research problem, we may need to use some AI (artificial optimization tools) e.g. Genetic Algorithm (GA) which could help us to locate the near-optimal designs.

Relating this idea to our research problem, we want to look at when the number of allocation changes or in other words, when the number of allocation is perturbed by a small amount, how it affects the data that is generated (i.e., the realization of the inputs/outputs). After knowing how the perturbation has generated in the data, we want to see how it affects the efficiency value. Finally after observing how perturbation has generated in the efficiency, we want to find out how it affects the overall performance which is the MSE of the efficiency. The gradient of the performance (MSE) with respect to **n**, is denoted by $\nabla \mathbf{f_n} = \frac{\partial F(\mathbf{n})}{\partial \mathbf{n}}$. We relax the problem by assuming that **n** is a real number vector.

Remark: In our case, although all elements in **n** must be integers, we relax it by assuming **n** to be real numbers to suit the conditions in IPA so that we are able to find the gradient of the performance with respect to **n**. This approximation is fine when **n** is large. The **three** main steps in IPA are ***perturbation generation***, ***perturbation propagation*** and ***perturbation in performance***. We will first explain what we want to find in each stage of the IPA and then show how to use chain rule to link all these together. To ease the understanding of the readers and for illustration simplification, we discuss the lemmas and mathematical proofs from the point of view of an element of the vector **n**.

5.2.1 1st stage (Perturbation generation)

In the perturbation generation stage, we want to find when we perturb the parameter by a small quantity, how will it affect the random number that is generated

(Ho and Cao, 1991). Suppose that the number of initial data collection is given by $\mathbf{n_o}$ = $[n_{ok}]_{k\in K}$. Recall that $\hat{\mathbf{X}}_{[i]}$ is the realization after $\mathbf{n}$ allocation obtained in replication i of the Monte Carlo runs. If we perturb n_k value by a small quantity Δn_k, we are interested to see how it will affect the $\hat{\mathbf{X}}_{[i]}$. To simplify the notation for $\hat{\mathbf{X}}_{[i]}$, we will remove the index [i] from it. Let $\hat{\mathbf{X}} = [\hat{x}_k]_{k\in K}$.

Lemma 1: Suppose $x'_k \sim N\left(\overline{x}_k, \dfrac{\sigma_k^2}{n_{ok}+n_k}\right)$, $Z = \dfrac{\hat{x}_k - \overline{x}_k}{\sigma_k \big/ \sqrt{n_{ok}+n_k}}$ *is the normalized value of the realization* $\hat{x}_k$. *Using the common random number to generate the perturbation in* $\hat{x}_k$, *if* n_k *is changed by* Δn_k, *then the change in* $\hat{x}_k$ *can be approximated by* $\Delta\hat{x}_k = -\dfrac{Z\sigma_k}{2(n_{ok}+n_k)^{3/2}}\Delta n_k + \mathrm{O}\left((\Delta n)^2\right)$. *Hence,* $\dfrac{\partial \hat{x}_k}{\partial n_k} = \lim\limits_{\Delta n_k \to 0}\dfrac{\Delta \hat{x}_k}{\Delta n_k} \approx -\dfrac{Z\sigma_k}{2(n_{ok}+n_k)^{3/2}}$.

PROOF: Recall Bayesian framework that $x'_k \sim N\left(\overline{x}_k, \dfrac{\sigma_k^2}{n_{ok}+n_k}\right)$. The normalized value of the realization $\hat{x}_k$ is given by $Z = \dfrac{\hat{x}_k - \overline{x}_k}{\sigma_k \big/ \sqrt{n_{ok}+n_k}}$. If Δn_k additional samples are collected, then the belief of input/output k is given by $x''_k \sim N\left(\overline{x}_k, \dfrac{\sigma_k^2}{n_{ok}+n_k+\Delta n_k}\right)$.

Using the same Z statistic, the realization of x''_k is given by $\dfrac{Z\sigma_k}{\sqrt{n_{ok}+n_k+\Delta n_k}}$. Thus, the change in $\hat{x}_k$ will be $\Delta\hat{x}_k = Z\sigma_k\left(\dfrac{1}{\sqrt{n_{ok}+n_k+\Delta n_k}} - \dfrac{1}{\sqrt{n_{ok}+n_k}}\right)$.

Using Taylor expansion: $f(h+\Delta h) - f(h) = f'(h)\cdot\Delta h + \mathrm{O}\left((\Delta h)^2\right)$, we have

$$\Delta\hat{x}_k = -\frac{Z\sigma_k}{2(n_{ok}+n_k)^{3/2}}\Delta n_k + \mathrm{O}\left((\Delta n)^2\right).$$

Hence, $\dfrac{\partial\hat{x}_k}{\partial n_k} = \lim\limits_{\Delta n_k\to 0}\dfrac{\Delta\hat{x}_k}{\Delta n_k} \approx -\dfrac{Z\sigma_k}{2(n_{ok}+n_k)^{3/2}}$. □

5.2.2 2nd stage (Perturbation propagation)

To solve Model 4.2, we need to generate $\theta(\hat{\mathbf{X}}_{[i]})$ which we need to solve an LP (i.e. Model 3.7) with given $\hat{\mathbf{X}}_{[i]}$. As we know how the perturbation has generated in $\hat{x}_k$, hence, we will see how it will affect the $\theta(\hat{\mathbf{X}}_{[i]})$ value. We use the approximation of the LP problem and dual model to find the perturbation in $\theta(\hat{\mathbf{X}}_{[i]})$.

Lemma 2: Let θ' be the efficiency score obtained from solving Model (3.7), λ' the solution for the particular DMU that we are interested in, and π_k the corresponding dual variable for the constraint related to input/output $k \in K$. If $\hat{x}_k$ changes by $\Delta\hat{x}_k$, then the change in efficiency score is estimated by $\Delta\theta' = (\theta' - \lambda')\pi_k\Delta\hat{x}_k$ if $k \in S$;

otherwise, $\Delta\theta' = (1-\lambda')\pi_k \Delta\hat{x}_k$ *if* $k \in R$. *Hence,* $\dfrac{\partial\theta(\hat{\mathbf{X}})}{\partial\hat{x}_k} = \lim\limits_{\Delta\hat{x}_k \to 0} \dfrac{\Delta\theta(\hat{\mathbf{X}})}{\Delta\hat{x}_k} \approx \left(\theta(\hat{\mathbf{X}}) - \lambda'\right)\pi_k$

if $k \in S$ *and* $\dfrac{\partial\theta(\hat{\mathbf{X}})}{\partial\hat{x}_k} = \lim\limits_{\Delta\hat{x}_k \to 0} \dfrac{\Delta\theta(\hat{\mathbf{X}})}{\Delta\hat{x}_k} \approx \left(1 - \lambda'\right)\pi_k$ *if* $k \in R$.

PROOF: Note that we have dropped the index j_o from $\hat{x}_k$, λ' and π which refers to the particular DMU that we are interested in. Here we need to reintroduce the index j_o in this proof. Also note that we have dropped the index [i] from the θ', λ', π_k and $\hat{x}_k$ because they can be written in general due to the reason that the same principle applies to all replications of the Monte Carlo runs.

First, we discuss how the efficiency score changes when the input $\hat{x}_{kj_o}$ ($k \in S$) changes. When $\hat{x}_{kj_o}$ ($k \in S$) changes by $\Delta\hat{x}_{kj_o}$, the corresponding 1st constraint of Model (3.7) (DEA Model) will change as follows:

$$\sum_{j \in J, j \neq j_o} \lambda_j x_{kj} + \left(\hat{x}_{kj_o} + \Delta\hat{x}_{kj_o}\right)\lambda_{j_o} + \left(-\hat{x}_{kj_o} - \Delta\hat{x}_{kj_o}\right)\theta \le 0$$
$$\sum_{j \in J, j \neq j_o} \lambda_j x_{kj} + \hat{x}_{kj_o}\lambda_{j_o} - \hat{x}_{kj_o}\theta \le \left(\theta - \lambda_{j_o}\right)\Delta\hat{x}_{kj_o}$$

Assuming that the changes in the solutions are very small, we replace $\left(\theta - \lambda_{j_o}\right)\Delta\hat{x}_{kj_o}$ with $\left(\theta' - \lambda'_{j_o}\right)\Delta\hat{x}_{kj_o}$. That is, we approximate that the right-hand side of constraint corresponding to the input k changes from 0 to $\left(\theta' - \lambda'_{j_o}\right)\Delta\hat{x}_{kj_o}$.

Thus, based on the sensitivity analysis[2] of linear programming, we can estimate the change in efficiency score by

$$\Delta\theta' \approx \left(\theta' - \lambda'_{j_o}\right)\Delta\hat{x}_{kj_o}\pi_k$$

$$\frac{\Delta\theta'}{\Delta\hat{x}_{kj_o}} \approx \left(\theta' - \lambda'_{j_o}\right)\pi_k$$

Hence, $\dfrac{\partial\theta(\hat{\mathbf{X}})}{\partial\hat{x}_{kj_o}} = \lim\limits_{\Delta\hat{x}_{kj_o}\to 0}\dfrac{\Delta\theta(\hat{\mathbf{X}})}{\Delta\hat{x}_{kj_o}} \approx \left(\theta(\hat{\mathbf{X}}) - \lambda'_{j_o}\right)\pi_k$

Next, we discuss how the efficiency score changes when the output $\hat{x}_{kj_o}$ ($k \in R$) changes. When $\hat{x}_{kj_o}$ ($k \in R$) changes by $\Delta\hat{x}_{kj_o}$, the corresponding constraint of Model (3.7) will change as follows:

---

[2] For an LP problem $\max\{z = \mathbf{cx} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, suppose, **b** changes by $\Delta\mathbf{b}$, the corresponding increment in the objective function is $\Delta z = \boldsymbol{\pi}\Delta\mathbf{b}$ where $\boldsymbol{\pi}$ is the dual variable for the corresponding primal constraint, which directly reflects the change in z owing to a change in the **b** (see e.g., Winston, 2003). In economic terms, $\boldsymbol{\pi}$ is referred as the shadow price, which is equivalent as the marginal price of **b**). Thus, if there are changes in both **b** and **A**: $\mathbf{A} \to \mathbf{A} + \Delta\mathbf{A}, \quad \mathbf{b} \to \mathbf{b} + \Delta\mathbf{b}$, the corresponding increment in the objective function can be approximated as follows:

$$\begin{aligned}(\mathbf{A} + \Delta\mathbf{A})\mathbf{x} &= \mathbf{b} + \Delta\mathbf{b} \\ \mathbf{Ax} &= \mathbf{b} - \Delta\mathbf{Ax}\end{aligned}$$

Assuming that the changes in the solutions are very small, hence the changes in the right hand side of the constraint can be approximated by

$$\begin{aligned}\mathbf{Ax} &\approx \mathbf{b} - \Delta\mathbf{Ax}' \\ \Delta\mathbf{b} &\approx -\Delta\mathbf{Ax}'\end{aligned}$$

Therefore, the change in the objective function is given by $\Delta z \approx \boldsymbol{\pi}(-\Delta\mathbf{Ax}')$

$$\sum_{j \in J, j \neq j_o} \lambda_j x_{kj} + \left(\hat{x}_{kj_o} + \Delta\hat{x}_{kj_o}\right)\lambda_{j_o} \geq \left(\hat{x}_{kj_o} + \hat{x}_{kj_o}\right)$$
$$\sum_{j \in J, j \neq j_o} \lambda_j x_{kj} + \hat{x}_{kj_o}\lambda_{j_o} \geq \hat{x}_{kj_o} + \left(1 - \lambda_{j_o}\right)\Delta\hat{x}_{kj_o}$$

Assuming that the changes in the solutions are very small, we again replace $\left(1 - \lambda_{j_o}\right)\Delta\hat{x}_{kj_o}$ with $\left(1 - \lambda'_{j_o}\right)\Delta\hat{x}_{kj_o}$. That is, the right-hand side of the constraint corresponding to output k changes approximately by $\left(1 - \lambda'_{j_o}\right)\Delta\hat{x}_{kj_o}$. Thus, based on the sensitivity analysis of linear programming, we can estimate the change in efficiency score by

$$\Delta\theta' \approx \left(1 - \lambda'_{j_o}\right)\Delta\hat{x}_{kj_o}\pi_k$$

$$\frac{\Delta\theta'}{\Delta\hat{x}_{kj_o}} \approx \left(1 - \lambda'_{j_o}\right)\pi_k \ .$$

Hence, $\dfrac{\partial\theta(\hat{\mathbf{X}})}{\partial\hat{x}_{kj_o}} = \lim\limits_{\Delta\hat{x}_{kj_o} \to 0} \dfrac{\Delta\theta(\hat{\mathbf{X}})}{\Delta\hat{x}_{kj_o}} \approx \left(1 - \lambda'_{j_o}\right)\pi_k$. □

5.2.3 3rd stage (Perturbation in performance)

Given the perturbation in $\theta(\hat{\mathbf{X}}_{[i]})$, we want to see how it affects the overall performance, which is to estimate the change in MSE ($\Delta F(\mathbf{n})$) when efficiency changes by $\Delta\theta(\hat{\mathbf{X}}_{[i]})$. However, instead of expressing in terms of perturbation, we will express in terms of derivative directly to suit our aim which is to find the gradient, $\nabla\mathbf{f_n} = \dfrac{\partial F(\mathbf{n})}{\partial\mathbf{n}}$. Taking derivative of $F(\mathbf{n})$ with respect to $\theta(\hat{\mathbf{X}}_{[i]})$, we obtain

$$\frac{\partial F(\mathbf{n})}{\partial \theta(\hat{\mathbf{X}}_{[i]})} = \frac{2}{M}(\theta(\hat{\mathbf{X}}_{[i]}) - \theta(\mathbf{X})) \tag{5.1}$$

Eventually, by using chain rule[3] to link all these together, we are able to find the rate of change in $F(\mathbf{n})$ with respect to $\mathbf{n}$.

Lemma 3: Suppose Lemma 1 and 2 hold, $\frac{\partial F(\mathbf{n})}{\partial n_k}$ *can be approximated by*

$$\frac{2}{M}\sum_{i=1}^{M}\left[\left(\theta(\hat{\mathbf{X}}_{[i]}) - \theta(\mathbf{X})\right) * \left(\left(\theta'(\hat{\mathbf{X}}_{[i]}) - \lambda'_{[i]}\right)\left(-\frac{Z\sigma_k}{2(n_{ok}+n_k)^{3/2}}\pi_{k[i]}\right)\right)\right] \quad \textit{if } k \in S\textit{, otherwise}$$

$$\frac{2}{M}\sum_{i=1}^{M}\left[\left(\theta(\hat{\mathbf{X}}_{[i]}) - \theta(\mathbf{X})\right) * \left(\left(1 - \lambda'_{[i]}\right)\left(-\frac{Z\sigma_k}{2(n_{ok}+n_k)^{3/2}}\pi_{k[i]}\right)\right)\right] \quad \textit{if } k \in R$$

PROOF: Here, we need to reintroduce the index i in this proof because it requires the values of $\theta(\hat{\mathbf{X}})$, λ, π and $\hat{x}_k$ in each replication i of the LP.

From (5.1), $\frac{\partial F(\mathbf{n})}{\partial \theta(\hat{\mathbf{X}}_{[i]})} = \lim_{\Delta\theta(\hat{\mathbf{X}}_{[i]})\to 0} \frac{\Delta F(\mathbf{n})}{\Delta\theta(\hat{\mathbf{X}}_{[i]})} = \frac{2}{M}(\theta(\hat{\mathbf{X}}_{[i]}) - \theta(\mathbf{X}))$

Lemma 1: $\frac{\partial \hat{x}_{k[i]}}{\partial n_k} = \lim_{\Delta n_k \to 0} \frac{\Delta \hat{x}_{k[i]}}{\Delta n_k} \approx -\frac{Z\sigma_k}{2(n_{ok}+n_k)^{3/2}}$

Lemma 2: $\frac{\partial \theta(\hat{\mathbf{X}}_{[i]})}{\partial \hat{x}_{k[i]}} = \lim_{\Delta \hat{x}_{k[i]}\to 0} \frac{\Delta\theta(\hat{\mathbf{X}}_{[i]})}{\Delta \hat{x}_{k[i]}} \approx \left(\theta(\hat{\mathbf{X}}_{[i]}) - \lambda'_{[i]}\right)\pi_{k[i]}$ if $k \in S$, and

$$\frac{\partial \theta(\hat{\mathbf{X}}_{[i]})}{\partial \hat{x}_{k[i]}} = \lim_{\Delta \hat{x}_{k[i]}\to 0} \frac{\Delta\theta(\hat{\mathbf{X}}_{[i]})}{\Delta \hat{x}_{k[i]}} \approx \left(1 - \lambda'_{[i]}\right)\pi_{k[i]} \text{ if } k \in R$$

---

[3] Chain rule concept (Apostol, 1974) : if f(u)=h(g(u)), then f'(u)=h'(g(u)) g'(u). In intuitive terms, if a variable f(u), depends on a second variable, h(g(u)), which in turn depends on a third variable, g(u), then the rate of change of f(u) with respect to u can be computed as the rate of change of f(u) with respect to h(g(u)) multiplied by the rate of change of g(u) with respect to u.

$$\nabla \mathbf{f_n} = \left[\nabla f_{n_k}\right]_{k \in K}$$

Using chain rule, $\nabla f_{n_k} = \dfrac{\partial F(\mathbf{n})}{\partial n_k} = \sum_{i=1}^{M}\left[\left(\dfrac{\partial F(\mathbf{n})}{\partial \theta(\hat{\mathbf{X}}_{[i]})}\right)\left(\dfrac{\partial \theta(\hat{\mathbf{X}}_{[i]})}{\partial \hat{x}_{k[i]}}\right)\left(\dfrac{\partial \hat{x}_{k[i]}}{\partial n_k}\right)\right]$

Hence,

$$\nabla f_{n_k} \approx \frac{2}{M}\sum_{i=1}^{M}\left[\left(\theta(\hat{\mathbf{X}}_{[i]}) - \theta(\mathbf{X})\right)\cdot\left(\theta'(\hat{\mathbf{X}}_{[i]}) - \lambda'_{[i]}\right)\pi_{k[i]}\cdot\left(-\frac{Z\sigma_k}{2(n_{ok}+n_k)^{3/2}}\right)\right] \text{ if } k \in S$$

Otherwise $\nabla f_{n_k} \approx \dfrac{2}{M}\sum_{i=1}^{M}\left[\left(\theta(\hat{\mathbf{X}}_{[i]}) - \theta(\mathbf{X})\right)\cdot\left(1 - \lambda'_{[i]}\right)\pi_{k[i]}\cdot\left(-\dfrac{Z\sigma_k}{2(n_{ok}+n_k)^{3/2}}\right)\right]$ if $k \in R$

This completes the proof. □

5.3 First phase (Hill-climbing algorithm)

This section explains the first phase of the two-phase gradient technique. In this first phase, which is based on the hill-climbing algorithm, there are two important concepts. The first is to find a good direction and the second is to determine how far to move along that direction. Based on these two concepts, the overall idea for our hill climbing algorithm is as follows. Given a starting/current design, we find the direction to move based on the gradient information. Next, we decide how far to move so that a lower value of MSE can be obtained when we move. With these information, we will be able to obtain a new design. After that, we will move to the new design, set it as the current design and repeat the process until the entire budget is used up, which is $\sum_{k \in K} n_k = N$.

We will now discuss the details of the first phase. Here, let us denote $\mathbf{n}_{(t)}$ as the starting allocation design for iteration t (initially, we set t = 1). The gradient $\nabla \mathbf{f}_{\mathbf{n}_{(t)}}$ is used to identify the move direction. Let $\mathbf{d}_{(t)} = [d_{k(t)}]_{k \in K}$ denote the move direction used during iteration t. Since the objective function is to minimize the MSE, the direction chosen should be at the most descent gradient, i.e. $\mathbf{d}_{(t)} = -\nabla \mathbf{f}_{\mathbf{n}_{(t)}}$ (Winston, 2003). Thus, the new allocation design (point) we intend to evaluate in iteration t + 1 is given by

$$\mathbf{n}_{(t+1)} = \mathbf{n}_{(t)} + \delta \mathbf{d}_{(t)} = \mathbf{n}_{(t)} - \delta \nabla \mathbf{f}_{\mathbf{n}_{(t)}} \tag{5.2}$$

where $\delta > 0$ is the selected step size.

For our research problem since we are minimizing the MSE, the move direction should be the most descent gradient i.e. negative gradient. However, due to uniqueness of our problem, the gradient can be non-negative. Therefore, in implementing our hill climbing algorithm, apart from considering the direction which can improve the performance, we also have to control the move direction. This is to ensure that that the hill-climbing will move in the direction that gradually increases the total number of allocation and eventually reaches the budget. Therefore, we need to address the followings: (i) the move direction must be determined from negative gradient, (ii) the allocations must be rounded off to maintain integrality and budget requirements, and (iii) how to determine an appropriate step size. We will elaborate each of the issues and describe how to tackle each of these in details.

5.3.1 Negative Gradient

From Eq. (5.2), for each number of allocation evaluated at iteration t, $n_{k(t)}$, we can find the updated number of allocation as follows:

$$n_{k(t+1)} = n_{k(t)} - \delta \nabla f_{n_{k(t)}} \quad \forall k \in K \tag{5.3}$$

Similarly, for the summation of the total number of allocation

$$\sum_{k \in K} n_{k(t+1)} = \sum_{k \in K} n_{k(t)} - \delta \sum_{k \in K} \nabla f_{n_{k(t)}} \tag{5.4}$$

In order for the total number of allocation to increase at each move, i.e. $\sum_{k \in K} n_{k(t+1)} > \sum_{k \in K} n_{k(t)}$, we need to have $\sum_{k \in K} \nabla f_{n_{k(t)}} < 0$. This means the gradient has to be negative. However, due to the nature of our problem, the gradient can be non-negative. Let us define $\sum_{+} \nabla f_{n_{k(t)}}$ as the total sum of the positive gradients and $\sum_{-} \nabla f_{n_{k(t)}}$ as the total sum of the negative gradients. That is,

$$\sum_{k \in K} \nabla f_{n_{k(t)}} = \sum_{+} \nabla f_{n_{k(t)}} + \sum_{-} \nabla f_{n_{k(t)}} \tag{5.5}$$

In order to guarantee that $\sum_{k \in K} \nabla f_{n_{k(t)}} < 0$, we must have $\sum_{+} \nabla f_{n_{k(t)}} < \sum_{-} \nabla f_{n_{k(t)}}$.

If $\sum_{+} \nabla f_{n_{k(t)}} \geq \sum_{-} \nabla f_{n_{k(t)}}$ we can reduce the value of $\sum_{+} \nabla f_{n_{k(t)}}$ while maintain the value of $\sum_{-} \nabla f_{n_{k(t)}}$ by multiplying all the positive gradient $\nabla f_{n_{k(t)}}$ with a ‘factor’, denoted by β, which will make $\beta \sum_{+} \nabla f_{n_{k(t)}} < \sum_{-} \nabla f_{n_{k(t)}}$. In this study, we select the value of β such that $\beta \sum_{+} \nabla f_{n_{k(t)}} - \sum_{-} \nabla f_{n_{k(t)}} = \varepsilon$. Note that, based on some experiments at the preliminary stage, we select ε = 0.001 for our study. That is assuming $\sum_{+} \nabla f_{n_{k(t)}}$ > 0, we set

$$\beta = \frac{\left(-\sum_{-} \nabla f_{n_{k(t)}} - \varepsilon\right)}{\sum_{+} \nabla f_{n_{k(t)}}} \tag{5.6}$$

Thus, in the case that $\sum_{k \in K} \nabla f_{n_{k(t)}} \geq 0$, the adjusted direction is

$$d_{k(t)} = \begin{cases} -\nabla f_{n_{k(t)}} & \text{if } \nabla f_{n_{k(t)}} < 0 \\ -\beta \nabla f_{n_{k(t)}} & \text{otherwise} \end{cases} \tag{5.7}$$

where β is given by (5.6).

5.3.2 Round off

Recall that a feasible design must only be consisted of non-negative integer numbers of allocation. It is likely that equation (5.4) will result in a feasible design. Thus, it is necessary to round off each number of allocation to the nearest integer. In additions, if the number of allocation is negative, we will set it equal to zero.

Observe that it is also possible that $\sum_{k \in K} n_{k(t+1)} > N$. In other words, as we are advancing, we may move to a design where the total number of allocation exceeds our budget (i.e., $\sum_{k \in K} n_{k(t+1)} > N$). Thus, if $\sum_{k \in K} n_{k(t+1)} > N$ after rounding off, we need to adjust some numbers of allocations down. Note that we will only consider adjusting those numbers of allocation that are positive. For the ease of explanation, we will use $n_{k(t+1)}$ as the original number of allocation obtained by equation (5.4) and $n^r_{k(t+1)}$ as the number of allocation after being rounded off. It is easy to see that the largest value of $n^r_{k(t+1)} - n_{k(t+1)}$ indicates that this number of allocation is the least deserved to be rounded up. Our adjusting approach is to decrease the numbers of allocation one by one starting from that with the largest value of $n^r_{k(t+1)} - n_{k(t+1)}$ until the total number of allocation is equal to the total budget N. Note that if the total number of allocation still exceeds the budget after all numbers of allocation have been decreased by one, we will continue decreasing them in the same order.

To briefly illustrate, let say, $N = 5$ and $\mathbf{n}_{(t+1)} = [3.7, 1.1, 2.2, 0.6, 0.3]$. After rounded off, $\mathbf{n}^r_{(t+1)}$ = [4, 1, 2, 1, 0]. That is, $\sum_{k \in K} n_{k(t+1)} = 8 > 5 = N$. The differences $\mathbf{n}^r_{(t+1)} - \mathbf{n}_{(t+1)}$ = [0.3, -0.1, -0.2, 0.4, -0.3]. Note that we will not decrease the number of allocation for $k = 5$ as it is already zero. Thus, we will decrease the numbers of allocation starting from that with the largest value of $n^r_{k(t+1)} - n_{k(t+1)}$ to the smallest (i.e., in the order of $k = 4, 1, 2, 3$). As the total number of allocation exceeds the budget by three, we will decrease the numbers of allocation by one for $k = 4$, 1, and 2. Therefore, it results in the design [3, 0, 2, 0, 0] which fulfils the budget constraint.

5.3.3 Step size

Now, we assume that $\sum_{k \in K} \nabla f_{n_{k(t)}} < 0$; or otherwise, they have been adjusted so that the move will increase the total number of allocation. Next, we discuss how to determine the appropriate step size value. To move in a direction that gradually increases the total number of allocation, we need to determine the appropriate step size δ to be used at each move. The step size cannot be too small (it may not move at all) nor too large (it may move too far away and may miss out some better designs that lie in between the path). To overcome this problem, we adopt the following approach.

First, we ensure that the step size is not too large. As mentioned earlier, we will select the step size δ that gradually increases the number of allocation. The approach we use is to set the increment number of allocation to a fraction ϕ of the total budget N (e.g. ϕ =10%). The increment number of allocation during iteration t is given by $\delta\left(-\sum_{k \in K} \nabla f_{n_{k(t)}}\right)$. Hence, an appropriate value of δ can be obtained by letting

$$\delta\left(-\sum_{k \in K} \nabla f_{n_{k(t)}}\right) = \phi N$$

That is,

$$\delta = \frac{\phi N}{\left(-\sum_{k \in K} \nabla f_{n_{k(t)}}\right)} \tag{5.8}$$

Note that we set ϕ = 10% in the experiment. This is only a target fraction and the final movement might be more or less than this fraction due to round off. This value of ϕ

seem to suit our problem based on our results obtained from running some preliminary experiments.

Next, we ensure that the step size is not too small. We make sure that, $n_{k(t+1)} > n_{k(t)}$ for at least one $k \in K$. Before rounding off, $n_{k(t+1)} - n_{k(t)} = -\delta \nabla f_{n_{k(t)}}$. As $n_{k(t)}$ is an integer, it is required that $n_{k(t+1)} - n_{k(t)} > 0.5$ before rounding off so that $n_{k(t+1)} > n_{k(t)}$ after rounding off. Hence, it is required that $\max\left(-\delta \nabla f_{n_{k(t)}}\right) \geq 0.5$; that is,

$$\delta \geq \frac{0.5}{\max\{-\nabla f_{n_{k(t)}}\}} \tag{5.9}$$

Therefore, the appropriate step size can be determined by choosing the greater value between the two choices, as shown in Eq. (5.10).

$$\delta = \max\left\{\frac{0.5}{\max\{-\nabla f_{n_{k(t)}}\}}, \frac{\phi N}{\left\{-\sum_{k \in K} \nabla f_{n_{k(t)}}\right\}}\right\} \tag{5.10}$$

The pseudo code of the algorithm is shown below.

Algorithm 1: First phase (hill-climbing)

Step 0: Initialization: Set $\mathbf{n}_{(1)} = \mathbf{0}$ and $t = 1$.

Step 1: Gradient: Compute $\nabla f_{n_{k(t)}}$ using Lemmas 1-3.

Step 2: Direction: If $\sum_{k \in K} \nabla f_{n_{k(t)}} < 0$ then set the direction $\mathbf{d}_{(t)} = -\nabla \mathbf{f}_{\mathbf{n}_{(t)}}$; otherwise determine the direction using (5.6) and (5.7).

Step 3: Step Size: Determine the step size δ using (5.10).

Step 4: New Design: Set $\mathbf{n}_{(t+1)} = \mathbf{n}_{(t)} - \delta\mathbf{d}_{(t)}$. Round off $\mathbf{n}_{(t+1)}$ to obtain the new design using the approach described in section 5.3.2.

Step 5: Termination: If $\sum_{k\in K} n_{k(t+1)} < N$, set $t = t + 1$ and return to Step 1; otherwise, stop.

The above proposed algorithm is able to find feasible solutions for Model 4.2. However, up to this stage of the technique, there are still some issues. The algorithm may have difficulties in finding good solutions. This is because, due to the nature of the gradient search technique, this algorithm will stop once the total number of allocation reaches the total budget (boundary). In others words, once the constraint $\sum_{k\in K} n_k = N$ is met, the algorithm terminates. The solution at this point, though is feasible, may not be a good solution. Therefore, in order to explore the other points on the boundary which can give better solutions, an improvement stage is needed.

5.4 Second phase (Gradient Improvement Stage)

Even though we managed to find a feasible solution in the first phase, we have not actually explored the neighbourhood yet. In this second stage, which is called the Gradient Improvement Stage (GIS), it aims to explore the neighbourhood of the feasible solution from the first phase. Neighbourhood here is defined as the set of feasible solutions which are near to the current design/point.

5.4.1 Overall concept

The first phase of the two-phase gradient technique is just to find a design at the boundary and this design might not be good. Hence, it is important to perform some local or neighbourhood search around this design so as to further improve the solution quality.

The overall concept of GIS is that, given a current design, we will first identify the feasible neighbourhood. Then, we will select which design/point from the neighbourhood that we should move to. This design will then be updated as the current design. After that, the entire process (identifying neighbourhood and selection) will be repeated until the best design, which is the design with the lowest MSE, is found.

Before we explain how to define this neighbourhood, we first have to find the direction, as analogous to the hill-climbing concept. Our desire is to find a direction such that it has a good potential to improve the objective function. Since we are minimizing MSE, the improving direction should also be at the most descent gradient. With the improvement in performance given by $\sum_{k \in K} \nabla f_{n_k} d_k$; hence we want to find the direction such that it minimizes $\sum_{k \in K} \nabla f_{n_k} d_k$. We also set the bound of d_k to be within -1 and 1. In order to maintain integrality of the solution, d_k must also be integer; hence, d_k = -1, 0 or 1. We must maintain the feasibility of the solution after the move. The number of allocation must not be negative; thus, it is required that $n_k + d_k \geq 0$ for all $k \in K$. We also must maintain the total number of allocation; thus, it is required that

$\sum_{k \in K} d_k = 0$. Note that with these requirements, the number of d_k's having the value of +1 must equal to those having the value of -1.

Observe that, ignoring the constraints $n_k + d_k \geq 0$ for all k, the direction that minimizes $\sum_{k \in K} \nabla f_{n_k} d_k$ is determined by setting half of the d_k's to +1 and another half to -1. For our desired direction, we also want to control the number of d_k's that have a nonzero value. With this reason, we impose a constraint $\sum_{k \in K} |d_k| \leq 2L$, where L is the maximum number of pairs of +1 and -1 direction. The mathematical model for finding a direction for the second phase is as follows:

$$\begin{aligned}
\min \quad & \sum_{k \in K} \nabla f_{n_k} d_k \\
\text{s.t.} \quad & \sum_{k \in K} d_k = 0 \\
& \sum_{k \in K} |d_k| \leq 2L \\
& n_k + d_k \geq 0, \qquad k \in K \\
& d_k \in \{-1, 0, 1\} \qquad k \in K
\end{aligned} \tag{5.11}$$

Let $\mathbf{d}^*$ denote the optimal direction obtained from above model. Note that it is not difficult to develop an efficient algorithm to solve model (5.11). However, in this thesis, we solve the model using a commercial solver.

As GIS is also an iterative approach, we will reintroduce the iteration index t. We use the solution obtained from the first phase as the starting point/design $n_{(1)}$. In iteration t, after an improving direction $\mathbf{d}_{(t)}$ is found using the model (5.11), we use it to construct the neighbourhood. In other words, we find all the possible points $\mathbf{n}_{(t)} + \gamma$

$\mathbf{d}_{(t)}$, $\gamma = 1, 2, ...$, for which improvement in performance is expected. (Note that γ is similar to the step size in the hill climbing concept). In order to maintain feasibility in the move, it is required that $\gamma \leq \min_{k(t)} \{n_{k(t)} : d_{k(t)} = -1\}$ so that $\mathbf{n}_{(t)} + \gamma\,\mathbf{d}_{(t)} \geq 0$. Let $A_{(t)}$ be the set of feasible neighbourhood for design $\mathbf{n}_{(t)}$ which can potentially improve the current solution $\mathbf{n}_{(t)}$. Hence, the neighbourhood is given by Eq. (5.12) below.

$$A_{(t)} = \left\{ \mathbf{n}_{(t)} + \gamma \mathbf{d}_{(t)} : \gamma = 1, 2, ..., \min_{k(t)} \{n_{k(t)} : d_{k(t)} = -1\} \right\} \tag{5.12}$$

After the neighbourhood $A_{(t)}$ is identified, we evaluate all the designs in it. After evaluation, we select which design in $A_{(t)}$ that we should move to. To select the designs, we not only consider whether the designs have the best performance (lowest MSE), but we also consider its potential. We will explain the meaning of 'potential' in the following discussion. Let $M(\mathbf{n})$ denotes the MSE of design $\mathbf{n}$. Let α be a given constant, potential of the design $\mathbf{n}$ is defined as

$$V(\mathbf{n}) = M(\mathbf{n}) + \alpha \nabla \mathbf{f}_{\mathbf{n}} \mathbf{d} \tag{5.13}$$

Recall that for each design $\mathbf{n}$, $M(\mathbf{n})$ is calculated using the Monte Carlo method. The improvement in performance is represented by $\nabla \mathbf{f}_{\mathbf{n}} \mathbf{d}$ as previously described in Model (5.11). We use a constant α to form a linear relationship between the potential and the improvement in performance. Note that this is analogous to the linear function $y = mx + c$, i.e. $y = V(\mathbf{n})$, $m = \alpha$, $x = \nabla \mathbf{f}_{\mathbf{n}} \mathbf{d}$ and $c = M(\mathbf{n})$. The reason we consider $V(\mathbf{n})$ is that we want to advance to a design which not only have the lowest MSE, but also with great potential (good future). The greater the potential of the design, the more improvement in performance (reduction in MSE) may be expected from the design for the future move. Next, we illustrate the full details of the GIS algorithm. Here, we use the term design and point interchangeably.

5.4.2 GIS algorithm

Let us define additional notation necessary to describe the algorithm of GIS.

$\mathbf{n}_{\text{best}}$ = best point/design that we have found (the lowest MSE)

$\mathbf{n}_{\text{bv}} = \arg\min_{\mathbf{n}\in A_{(t)}} V(\mathbf{n})$ = point with best potential in $A_{(t)}$

$\mathbf{n}_{\text{bm}} = \arg\min_{\mathbf{n}\in A_{(t)}} M(\mathbf{n})$ = point with best MSE in $A_{(t)}$

Our approach is to explore the most potential point first and keep the best point to be explored later. We use a flag called 'unexplore' to indicate whether or not the best point kept has already been explored. If unexplore flag = 1, this means that the best point kept has not been explored yet. In our algorithm, every time we have found a new best point, the unexplore flag will be set to 1. The unexplore flag will be set to 0 if the best point will be explored in the next iteration.

Suppose that we are currently in iteration t. We discuss selection process to identify the point to be explored in the next iteration (i.e., $\mathbf{n}_{(t+1)}$). There are four cases to be considered.

<u>Case 1</u>: $V(\mathbf{n}_{\text{bv}}) < V(\mathbf{n}_{\text{best}})$ and $M(\mathbf{n}_{\text{bm}}) < M(\mathbf{n}_{\text{best}})$

In this case, we have found the new best point $\mathbf{n}_{\text{bm}}$; that is, we set $\mathbf{n}_{\text{best}} = \mathbf{n}_{\text{bm}}$ and unexplore flag = 1. The best potential point is $\mathbf{n}_{\text{bv}}$ which will be explored next; that is, we set $\mathbf{n}_{(t+1)} = \mathbf{n}_{\text{bv}}$.

Case 2: $V(\mathbf{n}_{\text{bv}}) \geq V(\mathbf{n}_{\text{best}})$ and $M(\mathbf{n}_{\text{bm}}) < M(\mathbf{n}_{\text{best}})$

In this case, we also have found the new best point $\mathbf{n}_{\text{bm}}$ while the current best point is the most potential. If the current best point has not been explored (i.e., unexplore = 1), it will be the next point to consider ($\mathbf{n}_{(t+1)} = \mathbf{n}_{\text{best}}$). Otherwise, if the current best point has been explored (i.e., unexplore = 0), then the next most potential point is $\mathbf{n}_{\text{bv}}$; thus, we set $\mathbf{n}_{(t+1)} = \mathbf{n}_{\text{bv}}$. After that, we update $\mathbf{n}_{\text{best}} = \mathbf{n}_{\text{bm}}$ and set unexplore flag = 1.

Case 3: $V(\mathbf{n}_{\text{bv}}) < V(\mathbf{n}_{\text{best}})$ and $M(\mathbf{n}_{\text{bm}}) \geq M(\mathbf{n}_{\text{best}})$

In this case, the best potential point is $\mathbf{n}_{\text{bv}}$, therefore we set $\mathbf{n}_{(t+1)} = \mathbf{n}_{\text{bv}}$. The best point, $\mathbf{n}_{\text{best}}$, remains unchanged.

Case 4: $V(\mathbf{n}_{\text{bv}}) \geq V(\mathbf{n}_{\text{best}})$ and $M(\mathbf{n}_{\text{bm}}) \geq M(\mathbf{n}_{\text{best}})$

In this case, the best point remains unchanged and it also indicates that the current neighbourhood is not good at all. If the best point has not been explored (i.e., unexplore = 1), we set $\mathbf{n}_{(t+1)} = \mathbf{n}_{\text{best}}$; otherwise if unexplore = 0, we stop.

In general, when it happens that $\mathbf{n}_{(t+1)} = \mathbf{n}_{\text{best}}$, this means that the local optimal solution is nearby. To ensure that we can keep exploring and to further exploit the neighbourhood, we can reduce α (of Eq. (5.13)) proportionately. This is analogous to reducing the step size in the gradient search when we are near to the optimal solution in order to find a better solution.

The algorithm will terminate by itself once the optimal solution has been found. In addition, to avoid the same points being explored again, we set $A_{(t)}$ such that it always contains those neighbourhood points which have not been explored before. Let S be the updated set of all the neighbourhood points of $A_{(t)}$; initially, $S = \{\mathbf{n}_{(1)}\}$; as iteration proceeds, $S \leftarrow S \cup A_{(t)}$. Hence, $A_{(t)}$ can be written equivalently as:

$$A_{(t)} = \left\{ \mathbf{n}_{(t)} + \gamma \mathbf{d}_{(t)} : \gamma = 1, 2, \ldots, \min_{k(t)} \left\{ n_{k(t)} : d_{k(t)} = -1 \right\} \text{ and } \mathbf{n}_{(t)} + \gamma \mathbf{d}_{(t)} \notin S \right\} \tag{5.14}$$

The detail pseudo-code for the GIS algorithm is shown below.

Algorithm 2: GIS

Step 1: Initialization

Set $t = 1$, $\mathbf{n}_{(t)}$ = the solution obtained from first phase (see Section 5.3), $\mathbf{n}_{\text{best}} = \mathbf{n}_{(t)}$, unexplore = 0, and $S = \{\mathbf{n}_{(t)}\}$.

Step 2: Identify Neighbourhood

Determine $\mathbf{d}^*_{(t)}$ using Model (5.11) for given $\mathbf{n}_{(t)}$, and determine the neighbourhood $A_{(t)}$ using equation (5.14).

```
if A(t) ≠ ∅
        go to Step 3
else
        if unexplore = 1
                n(t+1) = nbest  and go to Step 4
        else
                Stop
        end
end
```

Step 3: Selection

3.1 Estimate MSE and V for $\forall$ $\mathbf{n} \in A_{(t)}$ using Monte Carlo
Determine $\mathbf{n}_{\mathrm{bv}}$ and $\mathbf{n}_{\mathrm{bm}}$

3.2 Select next point, $\mathbf{n}_{(t+1)}$

Case 1: $V(\mathbf{n}_{\mathrm{bv}}) < V(\mathbf{n}_{\mathrm{best}})$ and $M(\mathbf{n}_{\mathrm{bm}}) < M(\mathbf{n}_{\mathrm{best}})$
$\mathbf{n}_{(t+1)} = \mathbf{n}_{\mathrm{bv}}$
$\mathbf{n}_{\mathrm{best}} = \mathbf{n}_{\mathrm{bm}}$
unexplore = 1

Case 2: $V(\mathbf{n}_{\mathrm{bv}}) \geq V(\mathbf{n}_{\mathrm{best}})$ and $M(\mathbf{n}_{\mathrm{bm}}) < M(\mathbf{n}_{\mathrm{best}})$
if unexplore = 1
$\mathbf{n}_{(t+1)} = \mathbf{n}_{\mathrm{best}}$
else
$\mathbf{n}_{(t+1)} = \mathbf{n}_{\mathrm{bv}}$
end
$\mathbf{n}_{\mathrm{best}} = \mathbf{n}_{\mathrm{bm}}$
unexplore = 1

Case 3: $V(\mathbf{n}_{\mathrm{bv}}) < V(\mathbf{n}_{\mathrm{best}})$ and $M(\mathbf{n}_{\mathrm{bm}}) \geq M(\mathbf{n}_{\mathrm{best}})$
$\mathbf{n}_{(t+1)} = \mathbf{n}_{\mathrm{bv}}$

Case 4: $V(\mathbf{n}_{\mathrm{bv}}) \geq V(\mathbf{n}_{\mathrm{best}})$ and $M(\mathbf{n}_{\mathrm{bm}}) \geq M(\mathbf{n}_{\mathrm{best}})$
if unexplore = 1
$\mathbf{n}_{(t+1)} = \mathbf{n}_{\mathrm{best}}$
else
Stop
end

Step 4: Advance

if $\mathbf{n}_{(t+1)} = \mathbf{n}_{\mathrm{best}}$
unexplore = 0
$\alpha \leftarrow \alpha/2$
end
set $S \leftarrow S \cup A_t$, $t \leftarrow t+1$ and go to Step 2.

5.5 Summary

In this chapter, we have presented the two-phase gradient technique. This technique consists of two phases/stages. In the first phase, it uses the IPA to find the gradient and then applies the hill-climbing technique to find the solutions. In the second phase, which is called the gradient improvement stage (GIS), it explores the neighbourhood of the solutions. This results in an improvement in the final solutions, where better designs with lower MSE values can be obtained. Next, we will present the hybrid GA technique and the combination of other techniques.

CHAPTER 6

GA TECHNIQUE AND COMBINATIONS OF OTHER TECHNIQUES

In this chapter, first we present the hybrid GA technique, followed by the combination of other techniques.

6.1 Background Information

In this chapter, we will explain how to use GA (Genetic Algorithm) to search for the optimal solutions for Model (4.2). Due to the problems that the objective function is non convex and does not have an explicit expression, it might not be easy to solve by traditional optimization methods. Hence, GA is a good approach as it does not need to have an explicit objective function. We chose GA over other metaheuristics (i.e. tabu search, simulated annealing and etc.) because it offers several advantages over these techniques. GA keeps track of multiple independent solutions to the problem, so it easily lends itself to parallel computing possibilities. While heuristic search algorithms requires the users to write very problem-specific code to come up with a good solution, GA relies on the forces of random mutation and the process of natural selection to guide the solution of the problem. Another advantage of GA is its broad searching capabilities; it is able to conduct a broader search of the area, exploring many local optima. However, due to the problem that we are solving, the objective function needs to be estimated using Monte Carlo method to sample as many data as possible for the evaluation of the designs. Thus, this requires a large number of simulation replications and there exists a simulation allocation problem. If the simulation replications are to be allocated uniformly among the designs, the total simulation cost and computational time can be exhaustively high. In this chapter we will explain how to use OCBA (Optimal Computing Budget Allocation) to improve the simulation efficiency. By using OCBA, it can efficiently allocate the simulation budget of the Monte Carlo runs by optimally determine the number of simulation replications needed for each design alternative while identifying the single best design with high confidence. Note that the budget here refers to the computational budget for running simulation not the budget for collecting additional data used throughout this thesis. Next section discusses the GA technique and the incorporation of OCBA with GA.

6.2 Genetic Algorithm

The theoretical foundations of GA were originally developed by Holland (1975) based on the evolutionary process of biological organisms in nature. GA has been widely applied in many fields. GA works with a finite population, which evolves from one generation to the next, governed by the principles of natural selection and survival of the fittest among the individuals. Each generation consists of a population of chromosomes representing the possible solutions. Based on a random generated initial population, at every generation, GA evaluates the chromosomes and ranks them according to their fitness. The fitter chromosomes are selected to generate new offsprings by recombination and mutation operators. This evaluation-selection-reproduction cycle is repeated until a satisfactory solution is found.

6.3 Mechanisms

In this section, we discuss the mechanism of our GA algorithm which comprises of the all parameter settings and conditions used.

6.3.1 Integer encoding scheme

A standard chromosome is an array of bits. In our budget allocation problem, without loss of generality, we let the set of inputs/outputs $K = \{1, \ldots, D\}$. We use an array of integers to represent a solution as illustrated in Figure 6.1. To relate the chromosome to a feasible design, we limit the sum of the integers to equal to the budget, which is $\sum_{k \in K} n_k = N$.

n_1	n_2	$\ldots$	n_D

Figure 6.1: A chromosome representation

6.3.2 Feasibility

The chromosome represents a feasible solution if it satisfies condition $\sum_{k \in K} n_k = N$. If $\sum_{k \in K} n_k > N$, the chromosome cannot represent a feasible solution. In this case, we attempt to repair the chromosome by 'randomly' select one position and reduce it by one, and repeat it until a feasible solution is found. If the selected position/gene has a value of zero, it will be omitted and we will select another position. Alternatively, if $\sum_{k \in K} n_k < N$ we randomly select one position and increase it by one, and repeat it until a feasible solution is found. Note that only feasible chromosomes are kept in the population.

6.3.3 Fitness value

We relate the fitness of a chromosome to the objective value of Model 4.3. Hence, the calculation of fitness is as follows.

$$\text{fitness} = MSE = \frac{1}{M}\sum_{i=1}^{M}\left(\theta(\hat{\mathbf{X}}_{[i]}) - \theta(\mathbf{X})\right)^2$$

(6.1)

The chromosome which has the lowest MSE value is deemed to be the fittest. We use Monte Carlo simulation to estimate the fitness.

6.3.4 Population initialization

We use different ways to construct the initial population in order to ensure that the starting points are diversified. First, based on extreme allocation, we allocate the entire budget to one particular gene. For instance, when N = 5, D = 5, $\mathbf{n} \in \{(5, 0, 0, 0, 0), (0, 5, 0, 0, 0), ..., (0, 0, 0, 0, 5)\}$. This method will produce D solutions. Second, we allocate the budget equally among the genes. For example, when N = 10, D = 5, $\mathbf{n}$ = (2, 2, 2, 2, 2). Note that this is only applicable to the case where N/D is an integer.

The remaining of the initial population will be generated equally using the following two methods. We select two genes, then assign randomly to them; one with the value e.g. 'C' and one with 'N - C'. Another method is we randomly select a gene, and then assign '1' to it and repeat it until the total number of allocation reaches N.

6.3.5 Selection and reproduction

As we aim to balance towards more 'exploitation' and less 'exploration', we use 'tournament selection'. It gives faster initial convergence and less computation time compared to other types of selection. We randomly select some number of parents from the generation to form a tournament (sub-population) and then select the individual with the best fitness in this sub-population, effectively winning the tournament. We repeat the process several times until the required number of winners is chosen. In our experiment, we use a tournament size of two and selection with replacement. The reproduction of two offsprings is obtained by a two-position crossover on two parents as illustrated in Figure 6.2, where the two positions are generated randomly. Note that if the offsprings are not feasible (i.e. $\sum_{k \in K} n_k > N$ or $\sum_{k \in K} n_k < N$), we used the technique mentioned in Section 6.3.2 to repair them.

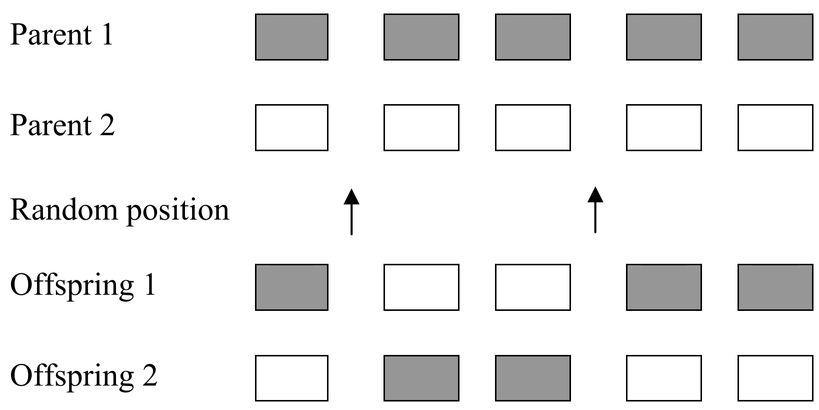

Figure 6.2: Two-position crossover

Each offspring is assigned a small probability of mutation to create more diversification to the solutions. Unlike crossover, the operation alters or mutates one or more genes within an individual chromosome rather than across a pair of chromosomes. Here, we randomly select two genes from the offspring and exchange their values. Figure 6.3 shows the illustrations of mutation for one particular offspring.

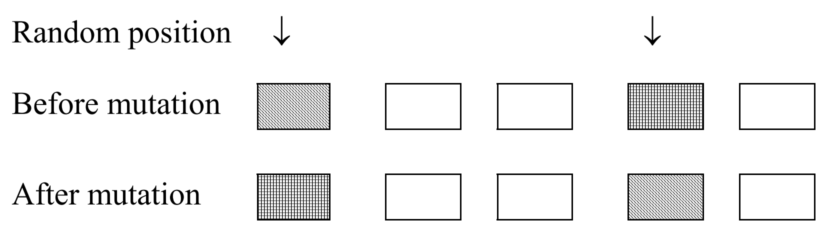

Figure 6.3: Mutations

6.3.6 Values of parameters and the termination condition

In summary, the parameters and conditions used in our GA evaluation are as follows.

Population size: 100

Maximum number of generations: 200

Crossover strategy: two points

Crossover rate: 1

Mutation rate: 0. 01

Stopping criteria: Either the best solution does not improve for 20 generations or maximum number of generations has been generated.

Selection strategy: Tournament, size = 2

Percentage of best solutions to be retained in the new generation: 20%.

Note: The values of these parameters are determined according to some preliminary experiments conducted using a base design. They are the best choice for the GA method to solve this research problem.

6.4 Issues

As GA evolves from one generation to the next, it extensively searches the solution space which involves the evaluation of the fitness of a large number of solutions. In addition, the fitness of each solution has to be estimated with high accuracy in order to ensure the survival of the fittest. At such, this requires a large number of simulation replications. Therefore the key difficulty with the GA technique in association to our problem is how to improve the simulation efficiency.

Our preliminary results showed that the computational time needed to find the solutions is very long. We will overcome this problem by using OCBA (Optimal Computing Budget Allocation). In our GA technique, as we need to retain certain percentage of the best design, the original OCBA procedure which only selects the single best design could not be applied directly. Slight modifications have to be made to the OCBA procedure to cater for this criterion, which is to select the top *m*-design. Chen et al. (2008) developed the OCBA-*m* technique to provide an efficient allocation of simulations runs among design alternatives while selecting the *m* best design. A brief description of OCBA and OCBA-*m* procedure is explained next.

6.5 OCBA

The basic idea of OCBA is to optimally determine the number of simulation runs for all designs to maximize simulation efficiency with a given computing budget or to attain a desired simulation decision quality using a minimum computing budget. For our problem, we apply OCBA to efficiently allocate the simulation runs among the designs in order to maximize the simulation efficiency with a given computing budget. In contrast to the common way used in simulation which is equal allocation of runs to all designs, the concept of OCBA is the unequal allocation of runs to different designs, favouring the better designs with more runs. In procedure, OCBA sequentially determines which design alternatives need more simulation runs and how many additional runs are needed. By doing this, the overall simulation efficiency can be improved as less computational effort is spent on simulating non-critical alternatives and more is spent on critical alternatives. While the run allocation given by OCBA may not be an optimal allocation when the simulation budget is finite, the numerical testing demonstrates that OCBA is a very efficient approach and can dramatically reduce simulation time. Next, we will explain the OCBA-*m* procedure in detail.

6.5.1 OCBA-*m* Allocation Procedure

The following notations are used.

H = number of simulation replications (simulation budget).

l = total number of designs need to be evaluated.

m = number of top designs to be selected in the optimal subset.

S_m = set of m (distinct) indices indicating designs in selected subset.

T_i = number of simulation replications allocated to design *i*.

B_i = the unknown true value/mean of MSE for design *i*.

$\hat{B}_{ij}$ = the observed MSE for design *i* in the *j*-th simulation replication.

$\overline{B}_i$ $= \frac{1}{T_i}\sum_{j=1}^{T_i}\hat{B}_{ij}$, sample mean of the MSE for design *i*.

σ_i^2 = variance of the MSE for design *i*.

t_o = the number of initial simulation replications for each design.

t = iteration number

Δ = increment of the computing budget (i.e., a pre-specified number of replications to be added to the existing computing budget at each iteration).

Chen et al. (2008) explained that the objective of OCBA-*m* is to find a simulation budget allocation that maximizes the probability of selecting the optimal subset, defined as the set of *m* ($m < l$) best designs, subject to a constraint on the computing budget *H*. Note that rank order within the subset is not part of the objective. In this thesis, we will take S_m to be the *m* designs with the smallest sample means of the MSE. Let $\overline{B}_{i_r}$ be the *r*-th smallest (order statistic) of $\{\overline{B}_1, \overline{B}_2, ..., \overline{B}_l\}$ i.e. $\overline{B}_{i_1} \leq \overline{B}_{i_2} \leq \cdots \leq \overline{B}_{i_l}$. Then, the selected subset is given by

$$\mathrm{S}_m \equiv \{i_1, i_2, ..., i_m\}.$$

Without loss of generality, we will take the *m* best designs as those designs with the *m* smallest MSE (i.e. $\overline{B}_i$). Let CS_m be the correct selection of the event where S_m actually contains the *m* best designs. Let the correct selection probability $P\{\mathrm{CS}_m\} \equiv P\{$The selected optimal subset, S_m actually contains all of the *m* smallest MSE designs$\}$. Hence, the OCBA-*m* problem formulated by Chen et al. (2008) is as follows:

$$\begin{aligned} &\max_{T_1,\dots,T_l} P\{\mathrm{CS}_m\} \\ &\text{s.t. } \sum_{i=1}^{l} T_i = H \end{aligned} \tag{6.2}$$

Model (6.2) aims to choose T_1, T_2, ..., T_l such that $P\{\mathrm{CS}_m\}$is maximized, subject to a limited computing budget H. Here, T_i is the set of non-negative integers and $\sum_{i=1}^{l} T_i$ denotes the total computational cost assuming the simulation execution times for different designs are roughly the same. This formulation implicitly assumes that the computational cost of each replication is constant across designs.

Chen et al. (2008) explained that the OCBA-m problem (6.2) can be solved by approximating $P\{\mathrm{CS}_m\}$ using a lower bound based on the Bayesian setting. The true value/mean of the MSE for each design, B_i, is assumed unknown and treated as a random variable, whose posterior distribution is updated after observed the simulation output $\{\hat{B}_{ij}$, j = 1, ..., $T_i^t\}$. Assuming that B_i has a conjugate normal prior distribution and non-informative prior distribution and following DeGroot (1970), the posterior distribution of B_i will be $B_i' \sim N\left(\overline{B}_i, \frac{\sigma_i^2}{T_i^t}\right)$ where $\overline{B}_i = \frac{1}{T_i^t}\sum_{j=1}^{T_i^t} \hat{B}_{ij}$ is the sample mean of the observed MSE for design i and σ_i^2 is the true variance of the MSE which can be approximated by the sample variance of the MSE $\sigma_i^2 \approx \frac{1}{T_i^t - 1}\sum_{j=1}^{T_i^t}\left(\hat{B}_{ij} - \overline{B}_i\right)^2$. Based on the updated values of these quantities (i.e. the sample means and variances), the lower bound can be asymptotically maximized following the relationship (6.3) below.

<u>From Chen et al. (2008):</u>

$$\frac{T_i^{t+1}}{T_j^{t+1}} = \left(\frac{\sigma_i / \delta_i}{\sigma_j / \delta_j}\right)^2, \quad i, j \in \{1,2,....,l\} \quad \text{and} \quad i \neq j \tag{6.3}$$

where $\delta_i = \overline{B}_i - (\overline{B}_{i_m} + \overline{B}_{i_{m+1}})/2$.

The idea behind this approach is to allocate the replications runs in such a way that those designs that are in the boundary of the optimal set (i.e., the designs nearer to *m*) will be assigned more replications runs. As the target is to get the optimal set, allocating the replication runs this way will provide a high confidence of choosing the correct set. Note also that the allocation given by (6.3) assumes known variances and independence of estimated sample means of the MSE across designs.

In practice, a boundary condition needs to be imposed in order to solve the allocation (i.e. to find the values of T_i^{t+1} in (6.3)). The condition is given as

$$\sum_{i=1}^{l} T_i^{t+1} = \Delta + \sum_{i=1}^{l} T_i^{t} \tag{6.4}$$

The resulting T_i^{t+1} is a continuous number that must be rounded to an integer; in the numerical experiments, T_i^{t+1} is rounded to the nearest integer. As T_i^{t+1} is the number of replication, it may happens that $T_i^{t+1} < T_i^t$; hence, in our thesis, we update T_i^{t+1} with $\max(T_i^{t+1}, T_i^t)$ and then perform the additional $T_i^{t+1} - T_i^t$ simulations. To briefly illustrate, suppose there are two designs, $\Delta = 10$, $T_1^t = 10$ and $T_2^t = 10$. Let say, from (6.3), $T_1^{t+1} = 2T_2^{t+1}$; and from (6.4) and after rounded to the nearest integer, $T_1^{t+1} = 17$ and $T_2^{t+1} = 8$. Thus, we set $T_1^{t+1} = \max\{17,10\} = 17$ and $T_2^{t+1} = \max\{8,10\} = 10$. Therefore, we run 7 additional simulation replications for design 1 and none for design 2.

The sequential allocation procedure of OCBA-*m* is summarized as follows. We assume total absence of knowledge about any design considered and any other basis for allocating computing budget at the beginning of the experiment. We first simulate all l designs with t_o replications to get some information about the performance of each design during the first stage. As simulation proceeds, the sample means and sample variances of each design are computed from the data already collected up to that stage. The computing budget is then increased by Δ and Equation (6.3) and (6.4) are applied to determine the new budget allocation. Further simulation replications are then performed based on the allocation and the procedure is repeated until the total budget H is exhausted. Note that we do not really need to make sure that the total summation of the simulation runs exactly equals to H, as long as it exceeds H, we stop. This is because, ideally, each new replication should bring us closer to the optimal solution. The algorithm is summarized as follows.

Algorithm 1: OCBA-m Allocation Procedure

Step 1: **Initialize:** Set $t = 1$ and perform t_o simulation replications for all designs;

$$T_1^t = T_2^t = \cdots = T_l^t = t_o$$

Step 2: a. **Update:** Calculate $\overline{B}_i$, σ_i and δ_i for $i = 1, ..., l$.

b. **Allocate**: Increase the computing budget by Δ and calculate the new budget allocation T_1^{t+1}, T_2^{t+1}, ..., T_l^{t+1} according to (6.3) and (6.4). Round off T_i^{t+1} and set $T_i^{t+1} \leftarrow \max\{T_i^t, T_i^{t+1}\}$.

c. **Simulate:** Perform additional $T_i^{t+1} - T_i^t$ simulations for design i, $i = 1, ..., l$.

Step 3: Termination: If $\sum_{i=1}^{l} T_i^t < H$, set $t \leftarrow t + 1$ and return to Step 2; otherwise, stop.

In the above algorithm, we need to select the initial number of simulation, t_o and one-time increment Δ. Chen et al. (1999) offers detailed discussions on the selection. It is well understood that t_o cannot be too small to avoid poor estimation at the beginning. A suitable choice for t_o is between 5 and 20 (Law and Kelton, 1991). Also, a large Δ can result in waste of computation time to obtain an unnecessarily high confidence level. On the other hand, if Δ is small, we need to do the computation procedure in step 2 many times. A suggested choice for Δ is a number bigger than 5 but smaller than 10% of the simulated designs. In particular, we set $t_o = 10$ and $\Delta = 10$ in our experiment. The settings for other parameters are as follows. We set $H = 5000$, the results obtained in this study indicate that the procedure works well for given value of H. We set $m = 20$, which means we retain the top 20% of the population. Total number of designs, l which corresponds to the population size, is set to 100.

In GA, we use OCBA-*m* to allocate the simulation runs for the computation of the fitness values (i.e. the MSE). The top-*m* solutions selected by OCBA-*m* are then used in the replacement stage to update the subsequent population in the next generation (i.e., we retain the top ($m/l \times 100$) % of the population and replace the rest of the population with the offsprings). Note that during the reproduction stage i.e. crossover, all the l designs are used for selecting parents. The pseudo-code for the GA+OCBA-*m* algorithm is given in Appendix C. Next, we explain other existing heuristic techniques as well as the combinations of the techniques.

6.6 Other Algorithms and Combination of the Techniques

This section discussed other existing heuristic techniques and the combinations of the GIS with GA and the other algorithms. The purpose of the incorporation is to examine whether GIS can further improve the prior solutions. We will provide some brief explanations for each of the techniques.

Other algorithms

a. Greedy

The concept of the greedy search is equivalent to iteratively construct a rooted tree. The idea is to allocate additional data one by one until the total budget N is reached. There are two basic steps in the greedy search algorithm. First, we find feasible designs that can be formed by increase the number of allocation of an input/output by one; this results in W new designs. Secondly, we evaluate these designs and choose the best one, in a greedy sense. The algorithm stops when the total budget N is reached. Refer Appendix C for the pseudo-code of the algorithm.

b. Batch

The concept of this algorithm is to divide the total allocation budget N into various batches. Then, we find all the possible designs from these batches and select the best one. Let ω be the batch size. In this study, we assume that the total allocation budget N is a multiple of the batch size ω. For example, if $N = 20$, $D = 2$, $\omega = 10$, the possible allocation designs are (20, 0), (10, 10) and (0, 20). The total number of possible designs is given by $[(N/\omega) + (D - 1)]!/[(N/\omega)!(D - 1)!]$. Note that the number of designs can be large. In our study, we select the batch size such that a total number of designs is less than 5000. This is to ensure that we can complete the simulation in a reasonable computational time. Otherwise, the simulation runs will take a very long time. Refer Appendix C for the pseudo-code of the algorithm.

Next, we discuss the combination of GIS with all the algorithms.

Combination of GIS with all the algorithms

The combination techniques are as illustrated below.

a. GA+GIS

We use only the ‘best solution’ from GA and apply the GIS at every generation. We use the *elitism* option to retain the best chromosome (point) generated at every generation. The elite (best) point will always be put back into the population in every generation. When applying the GIS to the elite point, we set some conditions. We will apply GIS to the chromosome with the best fitness which not been applied GIS in the earlier generation. By doing this, we add more diversification to the solution space and we can prevent the solution from being easily stuck at a local optima.

b. Gradient+GIS

This is the two-phase gradient technique as discussed in Chapter 5. In the first phase, we will use 10 starting points instead of a single starting point. The reason of doing this is because of the non-convex and non linear characteristics of the objective function. Multiple start points are able to yield better solutions compared to single start point. All the ten starting points are generated randomly i.e. we select one (or few) input/output at random and then assign a value less than the budget to its number of allocation. Note that these starting points must have the total number of allocations to be less than the budget N so that we can apply the gradient method (first phase) on them. To briefly illustrate, says, $N = 10$, a starting point can be $n_3 = 2$, $n_5 = 1$, and $n_k = 0$ for $k \neq 3, 5$; that is $\sum_{k \in K} n_k = 3 < 10 = N$. The first-phase gradient technique will apply to this starting point to obtain a feasible solution; and GIS will then be applied to find an improved solution.

c. Greedy+GIS

This is the continuation of the greedy technique. We use the final solution from greedy and apply the GIS.

d. Batch + GIS

This is the continuation of the Batch technique. We select the top ten solutions from batch and apply the GIS.

e. Constructive random + GIS

We construct ten starting points, which are boundary points, in a similar fashion as the generation of the initial population in GA as mentioned in Section 6.3.4. Then, we apply GIS on all the ten points. After going through GIS on each of them, we will obtain ten solutions. From there, we pick the best one.

6.7 Summary

In this chapter, we developed a hybrid GA technique, where GA is used to find the solutions, and is enhanced with simulation optimization technique which is OCBA to improve the simulation efficiency. Other existing heuristics techniques such as the greedy and batch techniques are also presented to solve the budget allocation problem. Lastly, the combination of the techniques, which is the incorporation of the GIS technique with all the techniques are being explored. The performance of these techniques will be examined in the next chapter.

CHAPTER 7

EXPERIMENTS SETUP, RESULTS & DISCUSSION

In this chapter, we compare the performances of all the techniques which were presented in Chapter 5 and 6. First, we explain the setup of the experimental runs which includes the parameter settings and the variables used in the study. Then, it will be followed by results and discussions.

7.1 Introduction

There will be two parts of the experiment. The first part introduces methods which can find good solutions. The second part incorporates the GIS with the methods proposed from the first part to further improve the prior solutions. All these methods will be compared against the 'uniform' allocation method, so as to find out whether is it better to use sophisticated way for data collection or vice versa. The uniform method is the simplest way to allocate and has been widely applied in real practice, whereby the budget is equally allocated to all the random variables. The performance of equal allocation will serve as a benchmark for comparison. The first part will cover the algorithms such as GA, Gradient (hill-climbing), Greedy and Batch. The second section covers the incorporation of the GIS with the methods explored in the first part, which includes GA+GIS, Gradient+GIS (Two-Phase Gradient), Greedy+GIS, Batch+GIS and Constructive random+GIS. We will compare a total of 10 algorithms. Next, we will explain about the parameter settings of the experiments.

7.2 Parameter settings

The purpose of the experiment is to investigate how the different algorithms perform under different scenarios. The parameters that we choose to cast the scenarios are the N (budget), D (dimension or total number of input/output with unknown true means), CV (coefficient of variation) and the initial number of data collected (n_{ok}). We choose the base setting to be (N = 90, D = 10, n_{ok} = 4 and CV = 1). To set the different sizes of problems, we vary N and D; we use 5 levels of N and 3 levels of D. There are altogether 15 different settings of N and D and we apply the entire settings to all the algorithms. Note that for these settings, we fix n_{ok} = 4 and CV = 1. To check the effects of data variations in the systems, we vary CV and n_{ok}. We fix (N = 90, D = 10, n_{ok} = 4) to analyze the effects of CV; we use 4 different levels of CV. Similarly, when we analyze the effects of n_{ok}, we fix (N = 90, D = 10, CV = 1); there are also 4 different levels of n_{ok}. Hence, there are a total of (15 + 4 + 4 = 23) sets of experiment. Each experiment is repeated 10 times independently using different random seeds.

Table 7.1 shows the parameters and the values used for the simulation model of our research problem.

Table 7.1: Simulation Setup

Parameters	**Values**
N (Budget)	30, 60, 90, 120, 150
D (Dimension)	5, 10, 15
CV (Coefficient of variation)	0.5, 0.75, 1, 1.5, 2
n_{ok} (Initial number of data collected)	2, 4, 6, 8, 10

7.3 Data used in the study

Experiments are performed using the data sets and the supply chain model from Chapter 3. The inputs/outputs which are considered to be stochastic are determined based on sensitivity analysis, i.e. starting from the most influential till the least influential on the efficiency. Note that D represents the total number of stochastic inputs/outputs. The details of the sensitivity analysis are shown in Appendix D.

Table 7.2 shows the list of variables according to their degree of impact on the efficiency. A brief explanation is given next to each of the variable. For further details and explanations of the variables used, please refer to Section 3.5.1.

Table 7.2**:** Input/output variables used in the study

Rank	Input/output variable	Definition
1.	Retailer revenue	Profit obtained by the retailer.
2.	Fill rate (manufacturer)	Inventory holding cost incurred by the manufacturer.
3.	Supplier revenue	Profit obtained by the supplier.
4.	Supplier cost	Cost incurred by the supplier.
5.	Fill rate (distributor)	Inventory holding cost incurred by the distributor.
6.	Customer response time	The velocity at which the supply chain provides products to the customers.
7.	Supplier labour	Labour cost incurred by the supplier.
8.	Cycle time	This is an attribute for production flexibility i.e. the agility of a supply chain in responding to customer demands.
9.	Retailer cost	Cost incurred by the retailer.
10.	Manufacturer cost	Cost incurred by the manufacturer.
11.	On-time delivery (manufacturer→distributor)	This refers to the performance of the manufacturer in delivering the correct product, at the correct time, condition, and quantity to the distributor.
12.	Distributor cost	Cost incurred by the distributor.
13.	On-time delivery (distributor→retailer)	This refers to the performance of the distributor in delivering the correct product, at the correct time, condition, and quantity to the retailer.
14.	Fill rate (retailer)	Inventory holding cost incurred by the retailer.
15.	On-time delivery (manufacturer→retailer)	This refers to the performance of the manufacturer in delivering the correct product, at the correct time, condition, and quantity to the retailer.

Note: ‘Rank’ refers to the degree of impact on the efficiency, i.e. if Rank=1, it has the highest impact on the efficiency.

Recall that, we are only investigating for one particular DMU j_o. For instance, if $D = 5$, this denotes that for DMU j_o, there are 5 inputs/outputs which true mean values are unknown, while the remaining inputs/outputs for DMU j_o as well as all inputs/outputs for other DMUs are deterministic.

For these inputs/outputs x_k's with unknown true means, without loss of generality, we assume that the belief for the true means are normally distributed with given mean values μ_k (refer to Table 3.3 in Chapter 3) and the standard deviation values are given by $\sigma_k = CV \cdot \mu_k$. Note that we also assume that the inputs/outputs are not correlated.

In assessing the performance of the algorithms, our aim is to compare the final MSE of the efficiency score after collecting the data based on allocation design **n**. The approach we use is we apply the algorithms (e.g. GA, Gradient and etc.) to find the allocation design **n**. Note that in order to find the **n**, we have to use Monte Carlo DEA method to estimate the MSE for the allocation design. After that, based on the solution obtained, we collect the data and update the belief for the true mean of the variables. Finally, we recompute the final MSE. This will be the 'real MSE' which we will use for comparison among the performances of each algorithm. The experimental flow is shown in Figure 7.1 below. A brief explanation is provided as follows.

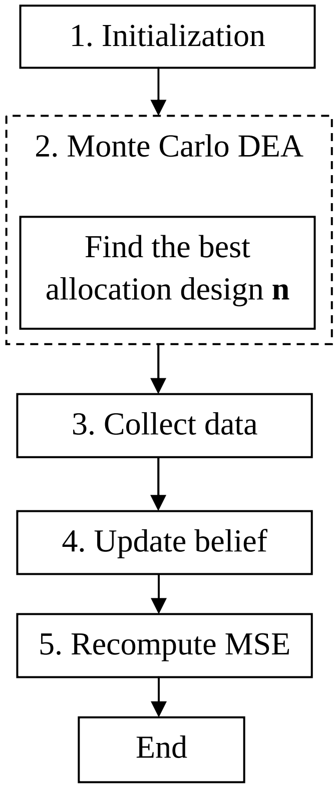

Figure 7.1: Experimental flow

1. Initialization - we generate n_{ok} data for input/output x_k by assuming that $x_k \sim N(\mu_k, \sigma_k)$.

2. We use the required technique to search for the **n** i.e. first we generate some designs (i.e. different allocations **n)**, then we run M monte carlo replications to generate the distributions for the inputs/outputs. We set M = 200 in the experiment. From the distributions of inputs/outputs, we find the distribution of the efficiency score and estimate the MSE. After that we repeat this entire process until we find the best MSE and the associated **n**.

3. After that, we collect the required number of data according to **n.** Note that we use simulation to generate the data.

4. Next, we update the belief for the true means, in other words, recalculate the sample means and sample variances of those inputs/outputs.

5. Lastly, we use the Monte Carlo DEA way again to recompute the final MSE.

Next, we will discuss the computational results and insights obtained from the experiments.

7.4 Results and Discussion

In this section, we test the effectiveness of the proposed methods. All models and algorithms are coded in Matlab (version 6.5) and tested on an Intel Pentium IV 2.6 GHz CPU with 512 MB RAM under the Microsoft Windows XP Operating System. The parameters of the proposed algorithms are chosen to ensure a compromise between the computational time and the solution quality. The values of the parameters used in the computational study are summarized as follows: a) for the two-phase gradient technique: L (of Model (5.7)) = 1 and α (of Eq. (5.13)) = 0.5; b) for the hybrid GA algorithm: GA parameters (refer section 6.3.6) and OCBA-*m* parameters (refer section 6.5.1). Note that for the two-phase gradient technique, we use the term 'Gradient' to represent the first phase only. The term 'Gradient+GIS' refers to the full technique, which is the two-phase gradient.

This section is divided into two main parts. The first part provides the main insight of the experiment. The second part discusses the performances of all the methods. The results concluded from these two parts will address the research question which is '*Is it better to collect the data using intelligent methods or collect the data naively?*'.

7.4.1 Main insights

In this section, the ultimate aim is to show the importance of allocating the budget intelligently. We will provide a comparison of the MSE obtained using intelligent method and non-intelligent method and then show how much savings in terms of budget that we can achieved. We will analyze the savings of the budget from the perspective of the size of the problem.

Figure 7.2 shows the comparison of the MSE obtained using GA+GIS and uniform for the case of D = 5. The results show that convergence is rapid at the beginning and when the number of additional allocation increases, the improvement in the solution decreases. We used the GA+GIS as the base method for comparison. Savings in the budget are calculated using the formula: Savings = (The required allocation budget by uniform to achieve the same performance as the base method) / (The required allocation budget by base method). For instances, when the required allocation budget is 30, GA+GIS results in a solution with MSE = 0.1269, where as it requires the allocation budget of 120 for the uniform method. (See the black dotted line in Figure 7.2). Hence, savings = 120/30 = 4. This means that, we can save four times the budget if we use GA+GIS compared to uniform allocation.

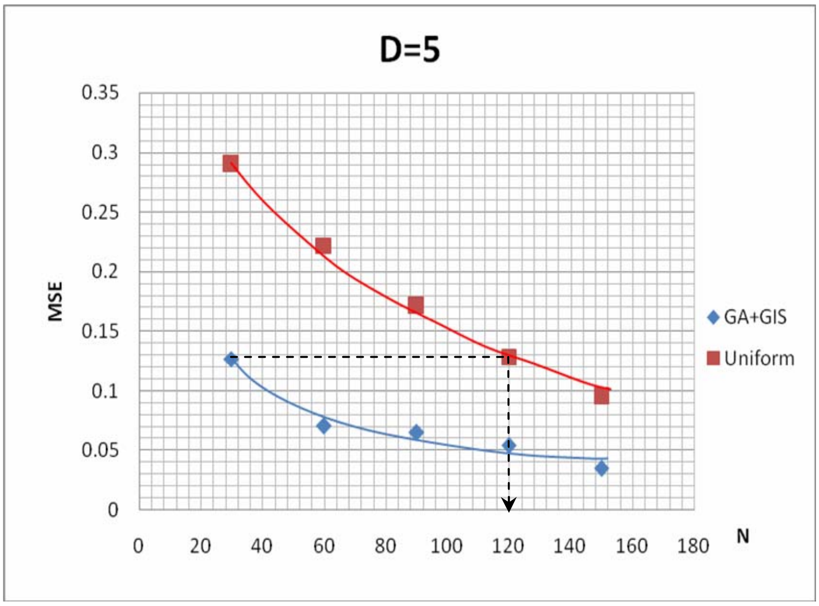

Figure 7.2: Comparison between GA+GIS and Uniform

Table 7.3-7.5 show the corresponding total number of additional allocation required by uniform method to achieve the same performance as the base method (GA+GIS) where the savings indicate the ratio required budgets for uniform method to those of GA+GIS.

Table 7.3: Comparison of N and savings when D = 5

MSE	GA+GIS	Uniform	Savings
0.12694	30	120	4.00
0.07080	60	425	7.08
0.06503	90	695	7.72
0.05563	120	1860	15.50

Table 7.4: Comparison of N and savings when D = 10

MSE	GA+GIS	Uniform	Savings
5.98E-03	30	1570	52.33
3.30E-03	60	3560	59.33
2.31E-03	90	4930	54.78
1.46E-03	120	8160	68.00

Table 7.5: Comparison of N and savings when D = 15

MSE	GA+GIS	Uniform	Savings
1.19E-05	30	330	11.00
5.56E-06	60	1635	27.25
4.23E-06	90	7185	79.83
3.75E-06	120	13905	115.88

When the size of the problem increases, the more savings we can achieve if we use sophisticated methods for budget allocation. The results showed that the savings are very significant (we can save more than 100 times the budget if we use sophisticated methods – see Table 7.5). Thus this means that it is very important for the users to use sophisticated methods for allocation and not resort to simplistic way such as equal/uniform allocation in allocating the budget for data collection.

7.4.2 Performances comparison

Table 7.6 shows the comparison of the performance for each algorithm in terms of the quality of the solution. The uniform method is set as the benchmark to compare whether intelligent or non-intelligent methods are better in allocating the budget. We use root mean square (RMSE) as the performance metric and compare how much improvement can be obtained using intelligent methods. Note that the minimization of MSE is actually the minimization of RMSE. As the values of MSE are very small, we present the percentage improvement in terms of RMSE for better illustration purposes.

Table 7.6: Comparison of RMSE and percentage improvement

Setting	Uniform (benchmark)	GA	% Imp	Gradient	% Imp	Greedy	% Imp	Batch	% Imp
***D*=5**									
***N*=30**	0.538926	0.356295	33.89	0.502869	6.69	0.457972	15.02	0.465562	13.61
***N*=60**	0.481585	0.290743	39.63	0.441677	8.29	0.407643	15.35	0.416415	13.53
***N*=90**	0.418319	0.257822	38.37	0.389521	6.88	0.360621	13.79	0.370507	11.43
***N*=120**	0.358819	0.237388	33.84	0.352121	1.87	0.329205	8.25	0.329205	8.25
***N* =150**	0.308822	0.177081	42.66	0.306476	0.76	0.306476	0.76	0.294449	4.65
***D*=10**									
***N* =30**	0.112322	0.081440	27.49	0.102706	8.56	0.091942	18.14	0.083283	25.85
***N* =60**	0.110741	0.058274	47.38	0.083283	24.80	0.083283	24.80	0.069054	37.64
***N* =90**	0.109184	0.048829	55.28	0.060702	44.40	0.057020	47.78	0.056382	48.36
***N* =120**	0.106202	0.038952	63.32	0.057020	46.31	0.049564	53.33	0.054426	48.75
***N* =150**	0.103818	0.035456	65.85	0.055738	46.31	0.045774	55.91	0.053083	48.87
***D*=15**									
***N* =30**	0.010085	0.003447	65.82	0.005376	46.69	0.005239	48.05	0.003872	61.61
***N* =60**	0.008204	0.002359	71.25	0.004014	51.08	0.002637	67.86	0.003136	61.78
***N* =90**	0.006617	0.002286	65.46	0.003607	45.50	0.002574	61.10	0.002848	56.96
***N* =120**	0.005579	0.002155	61.37	0.003292	40.99	0.002508	55.04	0.002774	50.28
***N* =150**	0.004759	0.002055	56.81	0.003237	31.99	0.002325	51.15	0.002537	46.69

Note: %Imp = % Improvement

	GA	Gradient	Greedy	Batch
Average (% Improvement)	51.23	27.41	35.76	35.89
Max (% Improvement)	71.25	51.08	67.86	61.78

The percent improvement is calculated using the formula: percent improvement = (RMSE benchmark - RMSE value obtained by algorithm) / (RMSE benchmark) × 100. All the intelligent methods perform better than the uniform method. The performance of the two proposed methods has been encouraging. The results show that, if we put in more effort in allocating, i.e. using more sophisticated technique such as the GA technique, we can obtain a better estimate of the efficiency. This signifies that it is important to allocate intelligently rather than allocate naively. GA has the highest improvement (average % improvement is 51.23%), followed by Batch, Greedy and Gradient whose average % improvement is 35.89%, 35.76% and 27.41% respectively.

Our results show that with the incorporation of the GIS technique, in most of the cases, the solutions of GA can be further improved. In average, the solution can be improved by approximately 3.71%. (See Table 7.7).

Table 7.7: Comparison of RMSE of GA and GA+GIS and percentage improvement

Setting	GA	GA+GIS	% Improvement
***D*=5**			
***N* =30**	0.356295	0.356295	0.00
***N* =60**	0.290743	0.266093	8.48
***N* =90**	0.257822	0.255000	1.09
***N* =120**	0.237388	0.235867	0.64
***N* =150**	0.177081	0.172108	2.81
***D*=10**			
***N* =30**	0.081440	0.077321	5.06
***N* =60**	0.058274	0.057455	1.40
***N* =90**	0.048829	0.048084	1.53
***N* =120**	0.038952	0.038165	2.02
***N* =150**	0.035456	0.034448	2.84
***D*=15**			
***N* =30**	0.003447	0.003447	0.00
***N* =60**	0.002359	0.002359	0.00
***N* =90**	0.002286	0.002056	10.04
***N* =120**	0.002155	0.001936	10.14
***N* =150**	0.002055	0.001857	9.66
Average (% improvement)			3.71

Note: The percent improvement is calculated using GA as the benchmark.

The results show that, by incorporating a local search (GIS) in the GA method, it explores the neighbourhood of the current solution to find a better solution. Hence, in most of the cases, the solution improves.

Table 7.8: Comparison of RMSE and % improvement with incorporation of GIS

Setting	Uniform (B)	GA+GIS	% Imp.	Greedy+ GIS	% Imp	Batch+GIS	% Imp
***D*=5**							
***N* =30**	0.538926	0.356290	33.89	0.456400	15.31	0.465550	13.62
***N* =60**	0.481585	0.266090	44.75	0.385420	19.97	0.404390	16.03
***N* =90**	0.418319	0.255000	39.04	0.319630	23.59	0.319800	23.55
***N* =120**	0.358819	0.235870	34.26	0.265680	25.96	0.245820	31.49
***N* =150**	0.308822	0.172110	44.27	0.214530	30.53	0.195180	36.80
***D*=10**							
***N* =30**	0.112322	0.077321	31.16	0.087513	22.09	0.081730	27.24
***N* =60**	0.110741	0.057455	48.12	0.081529	26.38	0.064784	41.50
***N* =90**	0.109184	0.048084	55.96	0.053083	51.38	0.051000	53.29
***N* =120**	0.106202	0.038165	64.06	0.048829	54.02	0.046768	55.96
***N* =150**	0.103818	0.034448	66.82	0.044978	56.68	0.042808	58.77
***D*=15**							
***N* =30**	0.010085	0.003447	65.82	0.004986	50.56	0.003705	63.26
***N* =60**	0.008204	0.002359	71.25	0.002637	67.86	0.002413	70.59
***N* =90**	0.006617	0.002056	68.93	0.002534	61.71	0.002364	64.28
***N* =120**	0.005579	0.001936	65.29	0.002357	57.75	0.002246	59.74
***N* =150**	0.004759	0.001857	60.99	0.002218	53.40	0.002159	54.63

Setting	Uniform (B)	Random+GIS	% Imp	Gradient+GIS	% Imp
***D*=5**					
***N* =30**	0.538926	0.488290	9.40	0.468220	13.12
***N* =60**	0.481585	0.416420	13.53	0.380500	20.99
***N* =90**	0.418319	0.360620	13.79	0.334490	20.04
***N* =120**	0.358819	0.306480	14.59	0.276000	23.08
***N* =150**	0.308822	0.224890	27.18	0.240270	22.20
***D*=10**					
***N* =30**	0.112322	0.091152	18.85	0.085411	23.96
***N* =60**	0.110741	0.061294	44.65	0.061546	44.42
***N* =90**	0.109184	0.054426	50.15	0.053759	50.76
***N* =120**	0.106202	0.048458	54.37	0.053406	49.71
***N* =150**	0.103818	0.045684	56.00	0.051629	50.27
***D*=15**					
***N* =30**	0.010085	0.004132	59.03	0.004498	55.40
***N* =60**	0.008204	0.002779	66.13	0.003292	59.87
***N* =90**	0.006617	0.002468	62.71	0.002485	62.44
***N* =120**	0.005579	0.002239	59.87	0.002479	55.57
***N* =150**	0.004759	0.002197	53.83	0.002374	50.12

Note: B = Benchmark, %Imp = % improvement

	GA+GIS	Greedy+GIS	Batch+GIS	Random+GIS	Gradient+GIS
Average (% Improvement)	52.97	41.15	44.72	40.27	40.13
Max (% Improvement)	71.25	67.86	70.59	66.13	62.44

Table 7.8 shows that with the incorporation of GIS, the solution for all the methods improves further. The gradient, greedy and batch methods are able to find some feasible design/solutions but the designs may not be good. By incorporating GIS, this helps to perform some local or neighbourhood search around the feasible design/solution so as to further improve the solution quality. Hence, as can be seen from the results of the experiments, there is a significant improvement in the solution of the greedy, batch and gradient algorithms after the incorporation of GIS. The performance of Greedy+GIS, Gradient+GIS, Constructive random+GIS and Batch+GIS are almost similar (the average percentage improvement ranges between 40.13% to 44.72%). Overall the results showed that there is no significant rule connected to the dimension (i.e. D) and the total number of additional allocation budget (i.e. N) that can favor one method to the other. The only thing noticeable is that all the methods perform better than uniform, and when the total number of additional allocation budget increases, regardless of dimension, the MSE decreases. Overall, the best performing algorithm is GA+GIS.

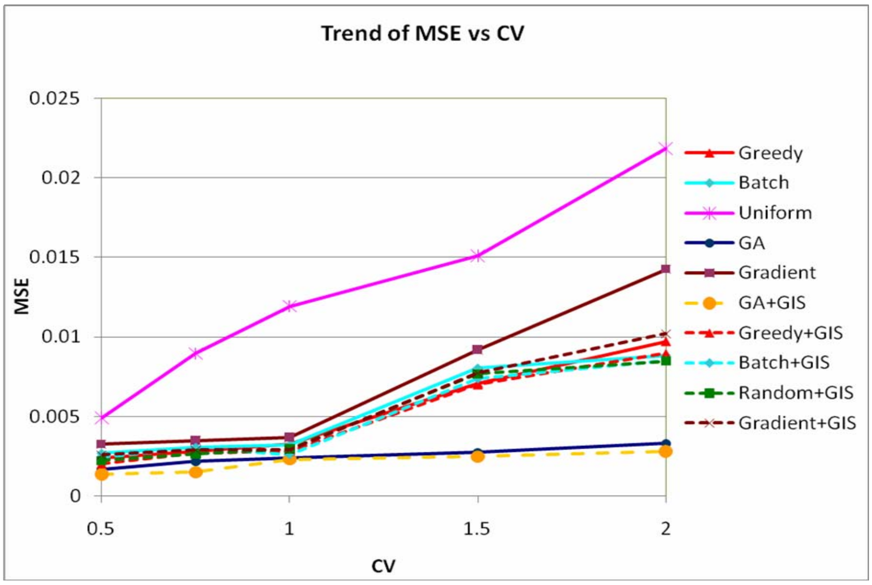

Figure 7.3**:** Comparison of MSE at different *CV* values**.**

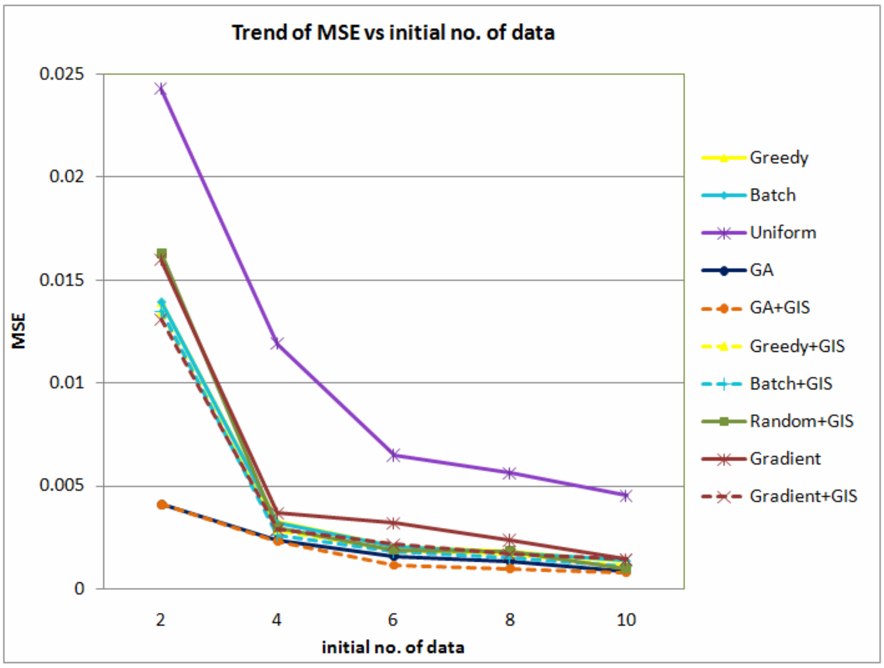

Figure 7.4**:** Comparison of MSE at different initial number of data.

Figure 7.3 and Figure 7.4 show that when the noise level increases, the GA and GA+GIS methods are still capable to locate the optimal design. The reason maybe due to the parameter settings in GA which enables it to find a good neighbourhood structure. Hence, GA can find the answer accurately even though randomness in the data increases. Nevertheless, the results are also very much dependent on the data used in the experiments.

Table 7.9 shows the average CPU time (run time) spent by each method in finding the best solution. The gradient technique is the fastest among all the sophisticated methods. The incorporation of GIS into GA helps to speed up convergence of the solutions in some cases. However, on average, it takes slightly longer time than GA. This may be due to the additional time requires by GIS to perform the neighbourhood search on the feasible solution. Nevertheless, the additional time taken is not large compared to the average time taken by GA. Hence, GA+GIS is still an efficient technique. The greedy and batch techniques take long computational time when the size of the problem is large. The incorporation of GIS into the batch, gradient and greedy techniques improve the solutions in a reasonable computational time.

Table 7.9: Average CPU time

Setting	GA	Gradient	Greedy	Batch	GA+GIS
***D*=5**					
***N*=30**	6566	290	792	1294	6596
***N* =60**	8687	598	1620	7610	8630
***N* =90**	8980	1069	3191	8837	9959
***N* =120**	9797	1267	6207	8956	10973
***N* =150**	10051	1289	11140	9080	13663
***D*=10**					
***N* =30**	4369	491	1959	704	6799
***N* =60**	9472	1030	4874	12276	9402
***N* =90**	14068	1923	8712	12538	20627
***N* =120**	14975	2079	24435	13133	23307
***N* =150**	16474	2308	42586	13548	25740
***D*=15**					
***N* =30**	7362	155	2374	27276	6683
***N* =60**	7989	903	4623	28068	7690
***N* =90**	8630	1563	9609	41247	9481
***N* =120**	9356	1868	20479	42182	10538
***N* =150**	10721	1902	55103	45668	11236
Average	9833	1249	13180	18161	12088
Average(hrs)	2.7	0.3	3.7	5.0	3.4
max(hrs)	4.6	0.6	15.3	12.7	7.2

Setting	Greedy+GIS	Batch+GIS	Random+GIS	Gradient+GIS
***D*=5**				
***N* =30**	1672	1672	374	614
***N* =60**	4216	3036	703	1290
***N* =90**	9990	7348	1590	2599
***N* =120**	22141	15510	2943	2747
***N* =150**	41130	18810	3979	3272
***D*=10**				
***N* =30**	3131	2640	897	910
***N* =60**	6749	6435	1106	1818
***N* =90**	8910	13860	2035	3573
***N* =120**	26041	27170	4283	6380
***N* =150**	48869	47520	8103	8270
***D*=15**				
***N* =30**	4694	8800	1187	636
***N* =60**	4719	8866	1400	1826
***N* =90**	7513	26950	5224	2801
***N* =120**	14940	47410	5543	3596
***N* =150**	28011	57519	14553	3805
Average	15515	19570	3595	2942
Average(hrs	4.3	5.4	1.0	0.8
max(hrs)	13.6	16.0	4.0	2.3

Lastly, a summary of the strengths and weaknesses of each technique is presented in Table 7.10 below.

Table 7.10: Strengths and weaknesses of the techniques.

Techniques	Strengths	Weaknesses
1. GA	Able to find a good solution in most cases.	Required relatively long time to converge to a good solution.
2. Gradient	Quick at finding a feasible solution.	The solution may not be a good one. It did not explore the neighbourhood of the feasible solution.
3. Greedy	Simple, easy to implement, Short computational time for small size problem.	The solution may not be good. No neighbourhood search for the feasible solution. Long computational time for large size problem.
4. Batch	Simple, easy to implement.	The solution may not be good. No neighbourhood search for the feasible solution. Long computational time for large size problem.

For all the above techniques, the incorporation of GIS will find a better solution at the expense of more computational time. On average, GIS improves the solution for each of the techniques almost equally.

7.5 Conclusion

In this chapter, we investigate the performance of the proposed techniques. The techniques had been applied on the supply chain data sets for solving the budget allocation problem in data collection to predict a good estimate of the efficiency score. We first apply the basic techniques to find good design allocations. We then incorporate the GIS technique to investigate whether the solutions can be further improved. The results showed that the two proposed techniques performed well in allocating the budget for data collection. The GA based technique is effective in finding the best solution, while the two-phase gradient method is fast and efficient in finding reasonably good solutions. In addition, the GIS technique can be incorporated with other existing methods and it can further improve the solutions efficiently and effectively.

Overall, the results showed that, it is important for the users to allocate the budget wisely when he/she is conducting the data collection. The users can achieve tremendous savings (as big as 100 times!) in the budget as well as improvements in the efficiency results when they use sophisticated way to allocate compared to allocating naively. By trading off between efficiency and effectiveness, the users may choose which method they want to use for budget allocation in data collection.

In the next chapter, we will recapitulate the findings and discuss the limitations and suggestions for future research.

CHAPTER 8

CONCLUSIONS AND FUTURE RESEARCH

In this chapter we shall summarize and discuss the main results of our research work as described in previous chapters. Possible future research will also be presented.

8.1 Summary of results

In chapter 3, we addressed the first part of our thesis, in which, we developed the Monte Carlo DEA approach to handle data uncertainties in DEA and we apply this to measure supply chain efficiency. A tentative model to measure supply chain efficiency is developed based on the basic CCR DEA model. This model removes the indirect effect of one's channel performance which affects the efficiency status of another channel. The Monte Carlo DEA method provides an alternative to measure the efficiency in stochastic environment. It is simple and easy to implement. It is able to provide statistical inferences on the efficiency and give additional information and insights to managers (e.g., the confidence interval) compare with other methods of measuring supply chain efficiency (e.g., the conventional way of using average data values to calculate a single value of efficiencies). Using average data to calculate the efficiency may leads to erroneous efficiency measures if the sample means of the data are very different from the true mean values; hence, this may not provide a strong base for making decision. Alternatively, using Monte Carlo DEA to obtain the efficiency distributions is more informative than efficiency scores for drawing appropriate conclusions.

In Chapter 4 to 7, we addressed the second part of the thesis. We provided an approach on how to get a better estimate for the efficiency score in the circumstances where there are uncertainties or variations in the data. We addressed the problem through the context of data collection, which is a norm in efficiency measurement whereby the users would have to collect the data in order to calculate the efficiency score using the DEA model. We provided a mathematical model in Chapter 4 to solve our research problem which is to find out how to allocate the budget for effective data collection in order to get a good estimate of the efficiency score. As the problem is very tough, we developed two sophisticated techniques which are the two-phase gradient technique and the GA technique to solve the model. Chapter 5 discussed the two-phase gradient technique and Chapter 6 discussed the GA technique, other existing techniques and combinations of the techniques.

Numerical experiments were conducted using the supply chain data sets. Experiments constituted of varying problem sizes and different noise levels in the data were examined to investigate how these algorithms perform under different scenarios. The performances of all the techniques are compared with the non-intelligent method, which is uniform allocation. The performances of the two-phase gradient technique and the GA technique have been encouraging. The numerical results show that the two proposed methods are effective and efficient in handling the budget allocation problem. The two-phase gradient technique is very efficient and capable of finding good solutions. The second phase of the gradient technique which is the GIS (Gradient Improvement Stage) is very flexible. It can be incorporated with any other existing techniques and it can efficiently improve the solutions. The hybrid GA algorithm yields very good solutions within reasonable amount of computational time. The combination of both, which is the GA+GIS is the most effective way in finding the best solution for the research problem.

This research also provides the insights that it is important to conduct the data collection wisely. By using sophisticated techniques to allocate the budget for data collection, this can provide a better estimate of the efficiency score and achieve greater savings in the budget. Managers can decide which methods that they want to adopt in allocating the budget for data collection, by making use of their experience, expertise and actual operational condition to handle the trade-off between practicability and optimality.

To sum up, the contributions of this research are three-folds. First, we develop a tentative DEA model to measure supply chain efficiency and provide an alternative approach to treat stochastic variations in data, which is the Monte Carlo DEA. Second, when data collection is needed and expensive, we provide a way on how to intelligently allocate the resources in data collection. Third, by developing method to solve this difficult problem (i.e. using IPA to estimate the gradient and using OCBA to improve the simulation efficiency of GA), it is innovative and provides a potential methodological contribution in the operational research field.

Finally, the research problem that we solve in this thesis is a pretty generic stochastic linear programming problem. Besides, contributing to DEA, it also offered an effective approach for sampling/computing budget allocation in stochastic LP problems.

8.2 Limitations of the Research

Despite the contributions described above, the work reported in this thesis has inevitably some limitations. In the first part of the research, the supply chain efficiency is taken as the weighted efficiencies of all the supply chain members i.e., by summing the weighted combination; this may have some double counting effect on the performance of the entire supply chain. The research problem in the second part of the thesis is addressed by assuming that only one DMU has uncertain inputs/outputs data. This is analogous to one's own organization, where the users are uncertain about the data and have to spend some effort in collecting those data. On the other hand, the data for the other organizations are assumed to be certain and deterministic. In addition, the constraint in the mathematical model which is rather simplified may not reflect the real application in actual industry practices. There had been no allowances made on the cost of collecting the data.

The data collection is examined using the non-sequential approach, which means that the users collect the data after making the final decision and the decision cannot be changed halfway. In reality, the users may use a different approach in collecting data, such as doing it sequentially. They may collect a few data first, then update their decision and collect the subsequent data. Hence, the results obtained may not be generalized as it depends on the approach used in collecting data.

Lastly, the application study was conducted using industries located in Penang, Malaysia. There may be systemic factors (e.g., economic environment, government regulations, financial system, market risks etc) that are not captured in the study. Therefore, the results may not be generalized.

8.3 Suggestions for future research

Based on the limitations mentioned above, some of the suggestions for future research work are:

1. To enhance the DEA supply chain model by analyzing how different setting of weights affect the supply chain performance.

2. To compare the effects of different data collection procedures.

3. To test the proposed algorithms in a more complex environment such as considering multiple DMUs and enhance the technicalities of the models (e.g., considering the cost of data collection).

4. To consider how to allocate the budget if we want to determine which DMU is the most (or least) efficient.

Lastly, apart from the suggestions mentioned above, as our problem can be viewed as generic stochastic LP problem, we can also extend the research to the development of more effective allocation procedures for general stochastic LP problems.

BIBLIOGRAPHY

Aigner, D.J., Chu, S.F., 1968. On estimating the industry production function, American Economic Review 58, 826-839.

Al-Faraj, T.N, 1993. Evaluation of bank branches by means of data envelopment analysis. International Journal of Operations and Production Management 13(9), 45-52.

Apostol. T. M., 1974. Mathematical Analysis, 2nd edition, Addison-Wesley, U.S.

Athanassopoulos, A.D., 1995. Performance improvement decision aid systems (PIDAS) in retailing organizations using data envelopment analysis. Journal of Productivity Analysis 6(2), 153-170.

Banker, R., Conrad, R.F., Strauss, R., 1986. A comparative application of DEA and translog methods: an illustrative study of hospital production. Management Science 32(1), 30-44.

Banker, R.D., Morey, R.C., 1986. The use of categorical variables in data envelopment analysis. Management Science 32(12), 1613-1627.

Banker, R.D., Charnes, A. Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30(9), 1078-1102.

Banker, R.D., 1993. Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation. Management Science 39, 1265-1273.

Banker, R.D., Gadh, W.M., Gorr, W.L., 1993. A Monte Carlo Comparison of Two Production Frontier Estimation Methods: Correlated Ordinary Least Squares and Data Envelopment Analysis. European Journal of Operational Research 69, 332-343.

Banker, R.D., Maindiratta, A., 1992. Maximum Likelihood Estimation of Monotone Convex Production Frontiers. Journal of Productivity Analysis 3, 401-415.

Barr, R.S., Seiford, L.M., Siems, T.F., 1993. An envelopment analysis approach to measuring the managerial quality of banks. Annals of Operations Research 45, 1-19.

Bauer, P.W., 1990. Recent developments in the econometric estimation of frontiers. Journal of Econometrics 46, 39-56.

Bazaraa, M.S., Sherali, H.D., Shetty, C.M., 1993. Nonlinear Programming: Theory and Algorithms, 2nd edition, John Wiley & Sons Inc., New York.

Beasley, J.E., 1995. Determining teaching and research efficiencies. Journal of the Operational Research Society 46(4), 441-452.

Berger, J.O., 1999. Statistical Decision Theory and Bayesian Analysis, 2nd edition, Springer Verlag, New York.

Bernardo, J.M., Smith, A.F.M., 1994. *Bayesian Theory*. J. Wiley, Chichester, UK.

Borden, J.P., 1988. An assessment of the impact of diagnosis related group (DRG)-based reimbursement on the technical efficiency of New Jersey hospitals using data envelopment analysis. Journal of Accounting and Public Policy 7(2), 77-96.

Bowlin, W.F.,1995. A characterization of the financial condition of the United States' aerospace-defense industrial base. Omega 23(5), 539-555.

Carlin, B.P., Louis, T.A., 2008. Bayesian Methods for Data Analysis, 3rd edition, Chapman & Hall/CRC, Boca Raton, Florida.

Cao, X.R., 1985. Convergence of Parameter Sensitivity Estimates in a Stochastic Experiment. IEEE Trans. Automatic Control 30, 845-853.

Chan, P.S., Sueyoshi, T., 1991. Environmental change, competition, strategy, structure and firm performance: an application of data envelopment analysis in the airline industry. International Journal of Systems Science 22(9), 1625-1636.

Charnes A., Cooper, W.W., 1963. Deterministic equivalents for optimizing and satisficing under chance constraints. Management Science 11, 18-39.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the inefficiency of Decision Making Units. European Journal of Operational Research 2(6), 429-444.

Charnes A., Cooper, W.W., Rhodes, E., 1981. Data envelopment analysis as an approach for evaluating program and managerial efficiency- with an illustrative application to the program follow through experiment in US public school education. Management Science 27(6), 668-697.

Charnes, A., Clarke, R., Cooper, W.W., 1989. Comparisons of DEA and existing ratio and regression systems for effecting efficiency evaluations of regulated electric cooperatives in Texas. Research in Governmental and Nonprofit Accounting, JAI Press, 5, 187-210.

Charnes A. Cooper, W.W., Lewin, A.Y., Seiford, L.M., 1994. Data Envelopment Analsis: Theory, Methodology and Applications. Kluwer Academic Publishers, Boston, UK.

Chen, C. H., Lin, J., Yücesan, E., Chick, S. E., 2000. Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization. Journal of Discrete Event Dynamic Systems: Theory and Applications 10, 251-270.

Chen, C. H., Yücesan, E., 2005. An Alternative Simulation Budget Allocation Scheme for Efficient Simulation. International Journal of Simulation and Process Modeling 1(1/2), 49-57.

Chen, C. H., 1996. A Lower Bound for the Correct Subset-Selection Probability and Its Application to Discrete Event System Simulations. IEEE Transactions on Automatic Control 41(8), 1227-1231.

Chen, H. C., Chen, C. H. , Dai, L., Yücesan, E., 1997. New Development of Optimal Computing Budget Allocation For Discrete Event Simulation, Proceedings of the 1997 Winter Simulation Conference, pp. 334-341.

Chen, H. C., Chen, C. H., Yücesan, E., 2000. Computing Efforts Allocation for Ordinal Optimization and Discrete Event Simulation. IEEE Transactions on Automatic Control 45(5), 960-964.

Chen, C.H., He, D., Fu, M., Lee, L.H., 2008. Efficient Simulation Budget Allocation for Selecting an Optimal Subset. INFORMS Journal on Computing 20(4), 579-595.

Chen, C. H., He, D., 2005. Intelligent Simulation for Alternatives Comparison and Application to Air Traffic Management. Journal of Systems Science and Systems Engineering 14(1), 37-51.

Chen, C. H., Donohue, K., Yücesan, E., Lin, J., 2003. Optimal Computing Budget Allocation for Monte Carlo Simulation with Application to Product Design. Journal of Simulation Practice and Theory 11(1), 57-74.

Chen, C. H., Wu, S. D., Dai, L., 1999. Ordinal Comparison of Heuristic Algorithms Using Stochastic Optimization. IEEE Transactions on Robotics and Automation 15(1), 44-56.

Chen, C.H., 1995. An effective approach to smartly allocate computing budget for discrete event simulation. Proceedings of the 34th IEEE Conference on Decision and Control, pp.2598-2605.

Chen, H., Schmeiser, B., 1995. Monte Carlo estimation for guaranteed-coverage non-normal tolerance intervals. Journal of Statistical Computation and Simulation 51(2), 223-238.

Chen, Y., Zhu, J., 2004. Measuring information technology's indirect impact on firm performance. Information Technology & Management Journal 5(1), 9-22.

Cheng, R.C.H., 1999. Regression metamodelling in simulation using Bayesian methods. In: Farrington, P.A., Nembhard, H.B., Sturrock, D., Evans, G. (eds.), Proceedings of the Winter Simulation Conference, Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc., pp. 330-335.

Cheung, K.L., Hansman, W.H., 2000. An exact performance evaluation for the supplier in a two-echelon inventory system. Operations Research 48, 646-653.

Chick, S.E., 2006. Bayesian Ideas and Discrete Event Simulation: Why, What and How. In: Perrone, L.F., Wieland, L.F., Liu, J., Lawson, B.G., Nicol, D.M., Fukimoto, R.M. (eds.), Proceedings of the 2006 Winter Simulation Conference, Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc., pp.96-106.

Chick, S.E., 1997. Bayesian analysis for simulation input and output. In: Andradottir, S., Healy, K., Withers, D., Nelson, B. (eds.), Proceedings of the Winter Simulation Conference, NJ: Institute of Electrical and Electronics Engineers, Inc., pp. 253-260.

Chick, S.E., Inoue, K., 2001a. New two-stage and sequential procedures for selecting the best simulated system. Operations Research 49(5), 732-743.

Chick, S.E., Inoue, K., 2001b. New procedures for identifying the best simulated sysem using common random numbers. Management Science 47(8), 1133-1149.

Chick, S.E., 2001. Input distribution selection for simulation experiments: Accounting for input uncertainty. Operations Research 49(5), 744-758.

Chu, X., Fielding, G.J., 1992. Measuring transit performance using data envelopment analysis. Transportation Research, Part A (Policy and Practice), 26A(3), 223-230.

Clarke, R., 1992. Evaluating USAF vehicle maintenance productivity over time: an application of data envelopment analysis. Decision Sciences 23(3), 376-384.

Clarke, R.L., Gourdin, K.N., 1991. Measuring the efficiency of the logistics process. Journal of Business Logistics 12(2), 17-33.

Cook, W.D., Kress, M., Seiford, L.M., 1996. Data envelopment analysis in the presence of both quantitative and qualitative factors. Journal of Operational Research Society 47, 945-953.

Cooke, R.M., 1994. Uncertainty in dispersion and deposition accident consequence modelling assessed with performance-based expert judgement. Reliability Engineering and System Safety 45, 35-46.

Cooper, W.W., Park, K.S., Yu, G., 1999. IDEA and AR-IDEA: Models for dealing with imprecise data in DEA. Management Science 45, 597-607.

Cooper, W.W., Seiford, L.M., Tone, K., 2006. Introduction to Data Envelopment Analysis and its Uses: with DEA-Solver Software and Reference, Springer, New York.

Cooper, W.W., Huang, Z.M., Li, S.X., 1996. Satisficing DEA models under chance constraints. Annals of Operations Research 66, 279-295.

Cooper, W.W., Huang, Z., Lelas, V., Li, S.X., Olesen, O.B., 1998. Chance constraint programming formulations for stochastic characterization of efficiency and dominance in DEA. Journal of Productivity Analysis 9, 53-79.

Day, D.L., Lewin, A.Y., Li, H., 1995. Strategic leaders or strategic groups: a longitudinal data envelopment analysis of the US brewing industry. European Journal of Operational Research 80(3), 619-38.

de Finetti, B., 1990. Theory of probability. John Wiley & Sons, Inc., New York.

DeGroot, M.H., 1970. Optimal Statistical Decisions, McGraw-Hill, Inc., New York.

Despotis, D.K., Smirlis, Y.G., 2002. Data envelopment analysis with imprecise data. European Journal of Operational Research 140, 24-36.

Deprins, D., Simar, L. Tulkens, H., 1984. Measuring labour efficiency in post offices. In: Marchand, M., Pestieu, P., Tulken, H. (eds.), The Performance of Public Enterprises. North-Holland, Amsterdam, pp.243-267.

Desai, A., Schinar, A.P., 1987. Stochastic Data Envelopment Analysis. Working Paper, College of Business, Ohio State University, Columbus, Ohio.

Desai, A., Ratick, S.J., Schinnar, A.P., 2005. Data envelopment analysis with stochastic variations in data. Socio-Economic Planning Sciences 39, 147-164.

Dudewicz, E.J., Dalal, S.R., 1975. Allocation of observations in ranking and selection with unequal variances. The Indian Journal of Statistics 37B, 28-78.

Easton, L., Murphy, D.J., Pearson, J.N., 2002. Purchasing performance evaluation: with data envelopment analysis. European Journal of Purchasing & Supply Management 8, 123-134.

Emrouznejad, A. Parker, B., Tavares, G., 2008. Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. Journal of Socio-Economics Planning Science 42(3), 151-157.

Färe, R., Grosskopf, S., 2000. Network DEA. Socio-Economic Planning Sciences 34, 35-49.

Färe, R., Grosskopf, S., Lovell, C.A.K., 1994a. Production frontiers. Cambridge University Press, Cambridge.

Farrell, M.J., 1957. The Measurement of Productive Efficiency. Journal of the Royal Statistical Society Series A, General 120(3), 253-282.

Ferrier G.D., Hirschberg, J.G., 1997. Bootstrapping confidence intervals for linear programming efficiency scores: with an illustration using Italian bank data. Journal of Productivity Analysis 8, 19-33.

Fishman, G.S., 1995. Monte Carlo: Concepts, Algorithms, and Applications, Springer Verlag, New York.

Forker, L.B., Mendez, D., J. Hershauer, 1997. Total quality management in the supply chain: What is its impact on performance ?. International Journal of Production Research 35, 1681-1702.

F∅rsund, F.R., Sarafoglou, N., 2005. The tale of two research communities: the diffusion of research on productive efficiency. International Journal of Production Economics 98, 17-40.

Fried, H., Lovell, C.A.K., Schmidt, S., Yaiswarng, S., 2002. Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis. Journal of Productivity Analysis 17(1/2), 157-174.

Fu, M.C., Hu, J.Q., Chen, C.H., Xiong, X., 2007. Simulation Allocation for Determining the Best Design in the Presence of Correlated Sampling. Informs Journal on Computing 19(1), 101–111.

Gattoufi, S., Oral, M., Kumar, A., Reisman, A., 2004. Epistemology of data envelopment analysis and comparison with other fields of OR/MS for relevance to applications. Socio-Economic Planning Sciences 38(2-3), 123-140.

Gattoufi, S., Oral, M., Reisman, A., 2004. Data Envelopment Analysis literature: a bibliography update (1996–2001). Socio-Economic Planning Sciences 38(2–3), 122–159.

Gentle, J.E., 2003. Random Number Generation and Monte Carlo Methods, 2nd edition, Springer-Verlag, New York.

Giokas, D.I., 1991. Bank branch operating efficiency: a comparative application of DEA and the loglinear model. Omega 19(6), 549-557.

Glynn, P., 1986. Problems in Bayesian analysis of stochastic simulation. In: Wilson, J.R., Henriksen, J.O., Roberts, S.D. (eds.), Proceedings of the Winter Simulation Conference, Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc., pp.376-383.

Golany, B., Hackman, S.T., Passy, U., 2006. An Efficiency Measurement Framework for Multi-Stage Production Systems. Annals of Operations Research 145(1), 51-68.

Gong, W.B., Ho, Y.C., 1987. Smoothed Perturbation Analysis of Discrete Event Dynamic Systems. IEEE Trans. Automatic Control 32, 858-866.

Gong, L., Sun, B., 1995. Efficiency measurement of production operations under uncertainty. International Journal of Production Economics 39, 55-66.

Grosskopf, S., 1996. Statistical Inference and Nonparametric Efficiency: A Selective Survey. Journal of Productivity Analysis 7(2/3), 139-160.

Gstach, D., 1998. Another Approach to Data Envelopment Analysis in Noisy Environments: DEA+. Journal of Productivity Analysis 9, 161-176.

Ho, Y.C., Sreenivas, R.S., Vakili, P., 1992. Ordinal Optimization of DEDS. Journal of Discrete Event Dynamic Systems 2(2), 61-88.

Ho, Y.C., Cao, X.R., 1991. Discrete Event Dynamic Systems and Perturbation Analysis, Kluwer Academic Publishers, Boston, UK.

Ho, Y.C., Cao, X.R., 1983. Perturbation Analysis and Optimization of Queueing Networks. Journal of Optimization Theory and Application 40(4), 559-582.

Ho, Y.C., Li, S., 1988. Extensions of Infinitesimal Perturbation Analysis. IEEE Trans. Automatic Control 33, 427-438.

Holland, J.H., 1975. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence. Michigan: University of Michigan Press.

Howard, L.H., Miller, J.L., 1993. Fair pay for fair play: estimating pay equity in professional baseball with data envelopment analysis. Academy of Management Journal 36(4), 882-94.

Hsieh, B.W., Chen, C.H., Chang, S.C., 2007. Efficient Simulation-based Composition of Dispatching Policies by Integrating Ordinal Optimization with Design of Experiment. IEEE Transactions on Automation Science and Engineering 4(4), 553-568.

Hsieh, B.W., Chen, C.H., Chang, S.C., 2001. Scheduling Semiconductor Wafer Fabrication by Using Ordinal Optimization-Based Simulation. IEEE Transactions on Robotics and Automation 17(5), 599-608.

Humphreys, P., Huang, G., Cadden, T., 2005. A web-based supplier evaluation tool for the product development process. Industrial Management & Data Systems 105(2), 147-163.

Kao, C., Lie, S.T., 2000. Fuzzy efficiency measures in data envelopment analysis. Fuzzy Sets and Systems 113, 427-437.

Kleinsorge, I., Schary, P., Tanner, R., 1989. Evaluating logistics decisions. International Journal of Physical Distribution and Materials Management 19(12), 3-14.

Kleinsorge, I., Schary, P., Tanner, R., 1991. The shipper-carrier partnership: a new tool for performance evaluation. Journal of Business Logistics 12, 35-37.

Kniep, A., Simar, L., 1996. A General Framework for Frontier Estimation with Panel Data. Journal of Productivity Analysis 7, 187-212.

Koenig, L.W., Law, A., 1985. A procedure for selecting a subset of size *m* containing the *l* best of *k* independent normal populations. Communication in Statistics – Simulation and Communication 14, 719-734.

Koopmans, T.C., 1951. An Analysis of Production as an Efficient Combination of Activities. In: Cowles Commission for Research in Economics (eds.), Activity Analysis of Production and Allocation, Monograph No.13, Wiley, New York.

Kuosmanen, T., Post, G.T., Scholtes, S., 2007. Testing for Productive Efficiency in Case of Errors-in-Variables. Journal of Econometrics 136, 131-162.

Land, K.C., Lovell, C.A., Thore, S., 1988.Chance-Constrained Efficiency Analysis, Working Paper, Department of Economics, University of North Carolina, Chapel Hill, N.C.

Law, A.M., Kelton, W.D., 1991. Simulation modelling and analysis, McGraw-Hill, New York.

Lee. L.H., Chew, E.P., Teng, S.Y., Goldsman, D., 2004. Optimal computing budget allocation for multi-objective simulation models. Proceedings of the 2004 Winter Simulation Conference.

Lee, L.H., Chew, E.P., Teng, S.Y., 2008. Finding the non-dominated pareto set for multi-objective simulation models. IIE Transactions.

Lindley, D.V., 1972. *Bayesian statistics, a review*. Society of Industrial and Applied Mathematics, Philadelphia.

Metropolis, N., Ulam, S., 1949. The Monte Carlo Method. Journal of the American Statistical Association 44 (247), 335–341.

Metzger, L.M, 1993. Measuring quality cost effects on productivity using data envelopment analysis. Journal of Applied Business Research 9(3), 69-79.

Miliotis, P.A., 1992. Data envelopment analysis applied to electricity distribution districts. Journal of the Operational Research Society 43(5), 549-555.

Nelson, B.L., Schmeiser, B.W., Taaffe, M.R., Wang, J., 1997. Approximation-assisted point estimation. Operations Research Letters 20(3), 109-118.

Mitchell M., 1996. An introduction to genetic algorithms, MIT Press, Cambridge, MA.

O'Donnell, C.J., Coelli, T.J., 2005. A Bayesian approach to imposing curvature on distanc functions. Journal of Econometrics 126(2), 493-523.

Olesen, O.B., Petersen, N.C., 1995. Chance constrained efficiency evaluation, Management Science 41, 442-457.

Olesen, O.B., Thore, S., 1990. Two-Stage DA Under Uncertainty, Working Paper, IC Institute, University of Texas, Austin, Texas.

Oral, M.O., 1992. An empirical study analyzing the productivity of bank branches. IIE Transactions 24(5), 166-76.

Park, B.U., Simar, L., 1994. Efficient Semiparametric Estimation in a Stochastic Frontier Model. Journal of the American Statistical Association 89, 929-36.

Petersen, N.C., 1990. Data envelopment analysis on a relaxed set of assumptions. Management Science 36(3), 305-314.

Petersen, N.C., Olesen, O.B., 1989. Chance Constrained Efficiency Evaluation, Working Paper, Department of Management, Odense University, Odense, Denmark.

Pina, V., Torres, L., 1992. Evaluating the efficiency of nonprofit organizations an application of data envelopment analysis to the public health service. Financial Accountability & Management 8(3), 213-24.

Ramanathan, R., 2003. An Introduction to Data Envelopment Analysis: a Tool for Performance Measurement. Sage Publications, New Delhi.

Rardin, R.L., 1998. Optimization in operations research, Prentice Hall, New Jersey.

Ray, S.C., Kim, H.J., 1995. Cost efficiency in the US steel industry: a nonparametric analysis using data envelopment analysis. European Journal of Operational Research 80(3), 654-71.

Reeves, C.R., 1997. Genetic algorithms for the operations researcher. INFORMS Journal on Computing 9, 231-50.

Rickards, R., 2003. Setting benchmarks and evaluating balanced scorecards with Data envelopment analysis. Benchmarking: An International Journal 10(3), 226-245.

Romero, V. J., Ayon, D.V., Chen, C. H., 2006. Demonstration of Probabilistic Ordinal Optimization Concepts to Continuous-Variable Optimization Under Uncertainty. Optimization and Engineering 7(3), 343-365.

Rubinstein, R.Y., Kroese, D.P., 2007. Simulation and the Monte Carlo Method, 2nd edition, John Wiley & Sons, New York.

Sarkis, J., 2000. An analysis of the operational efficiency of major airports in the United States. Journal of Operations Management 18, 335-351.

Savage, L.J., 1972. The foundations of statistics, Dover Publications Inc., New York.

Schefczyk, M., 1993. Operational performance of airlines: an extension of traditional measurement paradigms. Strategic Management Journal 14, 301-317.

Sarkis, J., Talluri, S., 2002. A model for performance monitoring of suppliers. International Journal of Production Research 40(16), 4257-4269.

Seiford, L.M., 1996. Data envelopment analysis: the evolution of the state of the art (1978-1995). Journal of Productivity Analysis 7, 49-78.

Seiford, L.M., Zhu, J., 1999. Profitability and marketability of the top 55 US commercial banks. Management Science 45(9), 1270-1288.

Sengupta, J.K., 1982. Efficiency measurement in stochastic input-output systems. International Journal of System Science 13, 273-87.

Sengupta, J.K., 1987. Data envelopment analysis for efficiency measurement in the stochastic case. Computer Operations Research 14, 117-129.

Sengupta, J.K., 1988. Robust efficiency measures in a stochastic efficiency. International Journal of Systems Science 19, 779-791.

Sengupta, J.K., 1989. Data envelopment analysis with maximum correlation. International Journal of System Science 20, 2085-2093.

Sengupta, J.K., 1995. Dynamics of Data Envelopment Analysis. Kluwer Academic Publishers, Netherlands.

Schmidt, P., 1985. Frontier production functions. Econometric Reviews 4, 289-328.

Shafer, S.M., Bradford, J.W., 1995. Efficiency measurement of alternative machine component grouping solutions via data envelopment analysis. IEEE Transactions on Engineering Management 42(2), 159-165.

Sherman, H.D., Ladino, G., 1995. Managing bank productivity using data envelopment analysis (DEA). Interfaces 25(2), 60-73.

Shi, L., Chen, C.H., 2000. A New Algorithm for Stochastic Discrete Resource Allocation Optimization. Journal of Discrete Event Dynamic Systems: Theory and Applications 10, 271-294.

Simar, L., Wilson, P., 2000. Statistical Inference in Nonparametric Frontier Models: The State of the Art. Journal of Productivity Analysis 13, 49-78.

Stewart, G., 1997. Supply-chain operations reference model (SCOR): the first cross-industry framework for integrated supply chain management. Logistics Information Management 10(2), 62-67.

Sueyoshi, T., 2000. Stochastic DEA for restructure strategy: an application to a Japanese petroleum company. Omega 28, 385-398.

Suri, R., 1987. Infinitesimal Perturbation Analysis for General Discrete Event Systems. Journal Of ACM, 686-717.

Suri, R., Zazanis, M.A., 1988. Perturbation Analysis Gives Strongly Consistent Estimates for the M/G/1 Queue. Management Science 34, 39-64.

Triantis, L., Girod, O., 1998. A Mathematical Programming Approach for Measuring Technical Efficiency in a Fuzzy Environment. Journal of Productivity Analysis 10, 85-102.

Varian, H.R., 1985. Nonparametric Analysis of Optimizing Behavior with Measurement Error. Journal of Econometrics 30, 445-458.

Weber, C.A., 1996. Data envelopment analysis approach to measuring vendor performance. Supply Chain Management 1(1), 28-39.

Weber, C.A., Desai, A., 1996. Determinants of paths to vendor market efficiency using parallel coordinates representation: a negotiation tool for buyers. European Journal of Operational Research 90, 142-155.

Winkler, R.L., 1972. Introduction to Bayesian inference and decision, Wiley, New York.

Winston, W.L., 2003. Introduction to mathematical programming: operations research, Thompson Pacific Grove, CA.

Wong,W.P., Jaruphongsa, W., Lee, L.H., 2008. Supply Chain Measurement System – A Monte Carlo DEA based approach. International Journal of Industrial and Systems Engineering 3(2), 162-188.

Wong, W.P., Wong, K.Y., 2008. A review on benchmarking of supply chain performance measures. Benchmarking: an International Journal 15(1), 25-51.

Yang, T., Kuo, C., 2003. A hierarchical AHP/DEA methodology for the facilities layout design problem. European Journal of Operational Research 147, 128-136.

Yu, C., 1998. The effects of exogenous variables in efficiency measurement – A Monte Carlo study. European Journal of Operational Research 105, 569-580.

Zazanis, M.A., 1986. Statistical Properties of Perturbation Analysis Estimates for Discrete Event Systems, Ph.D. thesis, Harvard University.

Zhang, Y., Bartels, R., 1998. The Effect of Sample Size on the Mean Efficiency in DEA with an application to Electricity Distribution in Australia, Sweden and New Zealand. Journal of Productivity Analysis 9, 187-204.

APPENDIX A: SUMMARY OF PAST LITERATURE SURVEYS

Table A.1: Summary of previous literature surveys on supply chain performance measures

Category: Theoretical (Performance Measures)					
Title	Author	Type of publications	Year of publication	Published in	Focus objectives
Supply-chain performance benchmarking study reveals keys to supply chain excellence	Stewart, G.	Article	1995	Logistics Information Management, Vol.8 No.2, pp.38-44	This paper suggested that best-in-class supply chain was characterized by the best achievement of both internal-facing measures and customer-facing measures.
Logistics and the Extended Enterprise: Benchmarks and Best Practices for the Manufacturing Professional	Boyson, S., Corsi, T.M., Dresner, M.E. and Harrington, L.H	Book	1999	Wiley, NY	This paper discussed the set of performance targets in benchmarking and possible methods to implement improvement solutions

Table A.2: Summary of previous literature surveys on supply chain integration

Category: Theoretical (Concepts on supply chain integration / inter organizational level)

Title	Author	Type of publications	Year of publication	Published in	Focus objectives
Logistics partnerships and supply chain restructuring: survey results from the US computer industry	Kopczak, L.R.	Article	1997	Production and Operations Management, Vol.6, No.3, pp.226-247	These papers revealed that the core of supply chain management is the improvement process at the interorganizational level.
Benefits of interfirm coordination in food industry supply chain	Stank, T.P., Crum, M.R. and Arango, M.	Article	1999	Journal of Business Logistics, Vol.20 No.2, pp.21-41	
Logistics and Supply Chain Management	Christopher, M.	Book	1998	Financial Times Management, Pitman Publishing, London.	The author explained that supply chain benchmarking includes joint practices and achievements of the chain members in the supply chain
Benchmarking supply chain operations	Gilmour, P.	Article	1999	Benchmarking for Quality Management and Technology, Vol.5, No.4, pp.283-290.	The author proposed a set of benchmark measures based on a set of capabilities which consists of process, information technology and organization.
How supply chain competency leads to business success	Bowersox, D.J., Closs, D.J., and Keller, S.B.	Article	2000	Supply Chain Management Review, Vol.4 No.4, pp.70-78.	The authors found that best practice in supply chain management resulted in better performance compared to companies with less integrated supply chain practices

Table A.2: Summary of previous literature surveys on supply chain integration [continued]

Category: Theoretical (Concepts on supply chain integration / inter organizational level)					
Title	Author	Type of publications	Year of publication	Published in	Focus objectives
Chain or shackles: understanding what drives supply-chain performance	Ramdas, K. and Spekman, R.E.	Article	2000	Interfaces, Vol.30, No.4, pp.3-21	The author used system-wide revenues and costs to examine collaborative practices between high performers among innovative-product supply chains and high performers among functional-product supply chains.
The nature of interfirm partnering in supply chain management	Metnzer, J.M., Min, S. and Zacharia, Z.G.	Article	2000	Journal of Retailing, Vol.76 No.4, pp.549-568	The authors discussed that companies became involved in the progressive process of collaboration as they moved toward closer arrangements with their partners.
A benchmarking scheme for supply chain collaboration	Simatupang, T.M., Sridharan, R.	Article	2004a	Benchmarking: An International Journal, Vol.11, No.1, pp.9-30.	This paper highlighted that supply chain collaboration shifted the focus of benchmarking from a single company level to an inter organizational level. The authors also recommended an integrated benchmarking scheme for supply chain collaboration that consists of enabling practices and collaborative performance system.

Table A.3: Summary of previous literature surveys on supply chain model/framework

Category: Practical (Model./ Framework)					
Title	Author	Type of publications	Year of publication	Published in	Focus objectives
Supply-chain operations reference model (SCOR): the first cross-industry framework for integrated supply chain management	Stewart, G., (1997)	Article	1997	Logistics Information Management, Vol.10 No.2, pp.62-67	The author provided the development of the supply chain operations reference (SCOR) model as the first cross-industry framework for evaluating and improving extended supply chain performance.
Advanced Supply Chain Management	Poirier, C.C.	Article	1999	Berret-Koehler Publishers, San Francisco, CA.	The author proposed a progressive framework consisting of four levels of supply chain optimization. The first two levels of progress are internally focused - "sourcing and logistics" and "internal excellence". The last two levels are “network construction and industry leadership".
What it means to be best in class	Geary, S. and Zonnenberg, J.P.	Article	2000	Supply Chain Management Review, Vol.4 No.3, pp.42-48.	The author employed the SCOR model to show that the best-in-class performers gained considerable financial and operating advantages over the rest of the respective groups.
Benchmarking a logistical operations based on causal model	van Landeghem, R. and Persoons, L.	Article	2001	International Journal of Operations and Production Management, Vol.21 No 1/2 pp.254-266.	The authors developed a causal model as a mean for identifying possible initiatives to bridge the performance gap between a company and best-in-class performers.

Table A.3: Summary of previous literature surveys on supply chain model/framework [continued]

Category: Practical (Model./ Framework)					
Title	Author	Type of publications	Year of publication	Published in	Focus objectives
Measuring the success of collaboration across the virtual supply chain through performance measurement systems and benchmarking	Polese, W.T.	Research paper	2002	Paper presented at the Supply Chain World Conference and Exposition, New Orleans, LA, 23 April.	The author developed a supply chain maturity model that reflects how companies progress in terms of operational capability. There are four stages: the first two are functional focus and internal integration. Collaboration is the key ingredient to reach stage 3 (external integration) and stage four (cross-enterprise collaboration). In conjunction with the SCOR model, the maturity model can be used to measure fact-based benchmarking for determining best-in-class performance opportunities.

.

Table A.4: Past literature survey on supply chain case study

Category: Practical (Case Study)					
Title	Author	Type of publications	Year of publication	Published in	Focus objectives
Benchmarking supply chain management practice in New Zealand	Basnet, C., Corner, J., Wiense, J. and Tan,K.	Article	2003	Supply Chain Management: An International Journal, Vol.8, No.1, pp.57-64	This paper illustrated an empirical study of benchmarking on supply chain practices in New Zealand companies.

Table A.5 Studies of DEA with their specific features

Year	Study	Source of Publication	Type	Journal Type	Application scheme
1980	Bessent (1980)	Educational Administration Quarterly	A	O	E
1980	Charnes and Cooper (1980)	Journal of Enterprise Management	T/A	O	O
1983	Lewin (1983)	Health Services Research	A	O	H
1983	Shaku et al. (1983)	Journal of General Systems	A	O	P
1984	Fare (1984)	Journal of Economics	T	E	-
1984	Lewin (1984)	Book	A	O	P
1984	Weining and Wong (1984)	Agricultural Production	A	E	I
1985	Charnes et al. (1985)	Journal of Econometrics	T	E	-
1985	Fare (1985)	European Journal Operational Research	T	M	-
1985	Fare e al. (1985)	Resources and Energy	T/A	O	U
1985	Miller (1985)	American Political Science Review	A	O	P
1986	Sexton (1986)	Books	T	O	-
1987	Macmillan (1987)	Environment and Planning	A	O	P
1987	Sengupta (1987)	International Journal of Systems Science	T	M	-
1988	Fare (1988)	Books	T	M	-
1988	Kamakura (1988)	Management Science	T	M	-
1988	Learner et al. (1988)	Conference Paper	T/A	O	O
1989	Jesson and Mayston (1989)	Policy Journals	A	O	E
1989	Nyman and Bricker (1989)	Review of Economics and Statistics	A	E	H
1989	Sengupta (1989)	Books	T	O	-
1989	Spanjers(1989)	Journal of Operational Research	T	M	-
1990	Desai and Schinnar (1990)	Socio-Economic Planning Sciences	T/A	E	-
1990	Kamis (1990)	Health Services Research	A	O	H
1990	Oral and Yolalan (1990)	European Journal of Operational Research	A	M	B
1990	Seiford (1990)	Computers, environment and Urban Systems	T	O	-
1990	Seiford and Thrall (1990)	Journal of Econometrics	T	E	-
1990	Sueyoshi (1990)	Journal of the Operational Research Society	T	M	-
1991	Boussofiance et al. (1991)	European Journal of Operational Research	T	M	-
1991	Giokas (1991)	Omega	A	M	B
1991	Mahajan (1991)	European Journal of Operational Research	T/A	M	-

Table A.5 Studies of DEA with their specific features [continued]

Year	Study	Source of Publication	Type	Journal Type	Application scheme
1991	Parkan (1991)	International Journal of Production Economics	A	E	O
1992	Bjurek et al. (1992)	Scandinavian Journal of Economics	A	E	P
1992	Chang (1992)	Journal of Productivity Analysis	A	E	P
1992	Dismuke and Sena (1999)	Health Care Management Science	A	O	H
1992	Haag et al. (1992)	Applied economics	A	E	I
1992	Kao and Yang (1992)	European Journal of Operational Research	A	M	I
1992	Morey et al. (1992)	Medical Care	A	O	H
1992	Sueyoshi (1992)	Journal of the Operational Research Society	T	M	-
1992	Thompson et al. (1992)	Computer and Operations Research	T/A	O	U
1993	Burgess and Wilson (1993)	Book	A	O	H
1993	Caulkins et al. (1993)	Operations Research	A	M	I
1993	Grizzle (1993)	International Journal of Public Administration	A	O	-
1993	Lee and Schmidt (1993)	Book	T	E	-
1993	Roll and Hayuth (1993)	Maritime Policy and Management	A	O	I
1993	Thanassoulis (1993)	Journal of the Operational Research Society	T	M	-
1993	Thompson et al. (1993)	Journal of Productivity Analysis	T	E	-
1994	Fuss (1994)	Canadian Journal of Economics Review	A	E	I
1994	Sueyoshi	European Journal of Operational Research	T/A	M	-
1994	Sueyoshi (1994)	European Journal of Operational Research	A	M	U
1994	Yaisawarng and Klein (1994)	Review of Economics and Statistics	A	E	U
1995	Athanassopoulos and Thanassoulis (1995)	International Journal of Production Economics	T/A	E	O
1995	Cooper et al. (1995)	European Journal of Operational Research	T/A	M	I
1995	Dula (1995)	International Journal of Systems Science	T	M	-
1995	Johnes (1995)	Economics of Education Review	A	E	U
1995	Lewin and Lovell (1995)	European Journal of Operational Research	T	M	-
1995	Li and Liu (2005)	Journal of South China University of Technology	A	E	I
1995	Majumdar (1995)	Journal of Economic Behaviour and Organization	T/A	E	I
1995	Olesen (1995)	Books	T	O	-

Table A.5 Studies of DEA with their specific features [continued]

Year	Study	Source of Publication	Type	Journal Type	Application scheme
1996	Boyd et al. (1996)	Working Paper	T	O	-
1996	Charnes et al. (1996)	European Journal of Operational Research	T/A	M	O
1996	Deborger and Kerstens (1996)	Journal of Productivity Analysis	T/A	E	O
1996	Fried et al. (1996)	Computer and Operations Research	A	M	E
1996	Guangfu (1996)	Annals of Operations Research	A	M	I
1996	Kersten (1996)	Transportation Research	A	O	I
1996	Mahmood et al. (1996)	Decision Sciences	A	M	O
1996	Nolan (1996)	Logistics and Transportation Review	A	O	I
1996	Piesse et al. (1996)	Journal of International Development	A	O	I
1996	Retzlaff (1996)	Computer and Operations Research	B	O	-
1996	Sengupta (1996)	International Journal of Systems Science	T	M	-
1996	Soterious and Stavrinids (1996)	Conference Paper	A	O	B
1996	Sueyoshi (1996)	Management Science	A	M	I
1996	Tyteca (1996)	Journal of Environmental Management	A	O	U
1997	Athanassopoulos (1997)	European Journal of Operational Research	T/A	M	B
1997	Athanassopoulos (1997)	Journal of the Operational Research Society	B	M	-
1997	Boussofiance et al. (1997)	Applied Economies	A	E	P
1997	Briec (1997)	Journal of Productivity Analysis	T	E	-
1997	Chang (1997)	European Journal of Operational Research	T	M	-
1997	Cooper and Tone (1997)	Journal of Operational Research	T	M	-
1997	Giokas (1997)	Journal of the Operational Research Society	B	M	-
1997	Mu and Du (1997)	Conference Paper	A	O	I
1997	Tyteca (1997)	Journal of Productivity Analysis	T/A	E	U
1998	Brockett et al. (1998)	International Journal of Systems Science	A	M	I
1998	Cummins and Zi (1998)	Journal of Productivity Analysis	T/A	E	I
1998	Grifell et al. (1998)	Journal of Productivity Analysis	T	E	-
1998	Hashimoto (1998)	Journal of Operations Research Society of Japan	T	M	-
1998	Ozcan et al. (1998)	Journal of Medical Systems	A	O	H
1998	Pitaktong et al. (1998)	European Journal of Operational Research	T	M	-

Table A.5 Studies of DEA with their specific features [continued]

Year	Study	Source of Publication	Type	Journal Type	Application scheme
1998	Ray and Mukherjee (1998)	International Journal of Systems Science	A	M	B
1998	Rosen et al. (1998)	Journal of Productivity Analysis	T	E	-
1999	Avkiran (1999)	Journal of Bank Marketing	A	O	B
1999	Camanho et al. (1999)	Journal of the Operational Research Society	A	M	B
1999	Dinc and Haynes (1999)	Annals of Regional Science	T/A	O	O
1999	Gropper et al. (1999)	Journal of Productivity Analysis	T	E	-
1999	Kao et al. (1999)	International Journal of Libraries and Information Services	A	O	E
1999	Lee and Barua (1999)	Journal of Productivity Analysis	T/A	E	O
1999	Lothgren and Tambour (1999)	Applied Economics	T	E	-
1999	Metters and Vargas (1999)	Production and Operations Management	A	E	I
1999	Mota et al. (1999)	International Journal of Technology Management	T/A	O	O
1999	Ramanathan (1999)	Indian Journal of Transport Management	A	O	I
2000	Deng et al. (2000)	Computer and Operations Research	T/A	M	B
2000	Devinney et al.	Organization Science	A	O	O
2000	Fare and Grosskopf (2000)	Socio-Economic Planning Sciences	T	E	-
2000	Halme and Korhonen (2000)	European Journal of Operational Research	T	M	-
2000	Hong et al. (2000)	International Journal of Systems Science	T	M	-
2000	Lai et al. (2006)	Journal of Risk and Insurance	T/A	E	I
2000	McCallion et al. (2000)	Applied Economics	A	E	H
2000	Nold et al. (2000)	Journal of Regional Science	A	O	E
2000	Odeck (2000)	European Journal of Operational Research	A	M	I
2000	Sarkis (2000)	Journal of Operations Management	A	O	I
2000	Simar and Wilson (2000)	Journal of Applied Statistics	T	E	-
2000	Tybout (2000)	Journal of Economic Literature	B	E	-
2000	Uri (2000)	Telecommunications Policy	A	O	I
2000	Worthington and Dollery (2000)	Local Government Studies	A	O	E
2000	Yeboon et al. (2000)	Transactions of the Society of Instrument and Control Engineers	T	O	-
2001	Brockett et al. (2001)	Engineering Economist	A	O	U
2001	Cooper et al. (2001)	Journal of Productivity Analysis	T	E	-

Table A.5 Studies of DEA with their specific features [continued]

Year	Study	Source of Publication	Type	Journal Type	Application scheme
2001	Grundy and Merton (2001)	Journal of finance	A	E	B
2001	Staat (2001)	Journal of Productivity Analysis	B	E	-
2001	Steinmann and Zweifel (2001)	Journal of Productivity Analysis	T	E	-
2001	Tone (2001)	European Journal of Operational Research	T	M	-
2001	Valdmanis (2001)	Socio-Economic Planning Sciences	A	E	H
2001	Weber (2001)	Review of Economics and Statistics	A	E	I
2001	Worthington and Dollery (2001)	Policy Studies Journal	A	O	E
2002	Camanho and Dyson (2002)	Book	A	O	B
2002	Fare et al. (2002)	Mathematical and Computer Simulation	T/A	O	O
2002	Hofmarcher et al. (2002)	Health Care Management Science	A	O	H
2002	Lau and Lam (2002)	Journal of the Operational Research Society	T/A	M	O
2002	Li and Yan (2002)	System Engineering Theory and Practice	T	M	-
2002	Lozano et al. (2002)	Journal of the Operational Research Society	T/A	M	O
2002	Manandhar and Tang (2002)	Journal of High Technology Management Research	T/A	O	B
2002	Weber (2002)	Decision Sciences	A	M	I
2002	Yan et al. (2002)	European Journal of Operational Research	T	M	-
2003	Birman et al. (2003)	Mathematical and Computer Modelling	A	O	H
2003	Calara and Cabanda (2004)	Conference Paper	A	O	O
2003	Guo et al. (2003)	Journal of Tianjin University Science and Technology	T	O	-
2003	Kruger (2003)	Oxford Economic Papers	A	E	O
2003	Lovell (2003)	Journal of Productivity Analysis	T	E	-
2004	Amin and Toloo(2004)	Computer and Operations Research	T	O	-
2004	Bernardes and Pinillos (2004)	Conference Paper	B	O	-
2004	Bowlin (2004)	European Journal of Operational Research	A	M	B
2004	Cielen et al. (2004)	European Journal of Operational Research	T/A	M	B
2004	Cummin et al. (2004)	Conference Paper	T/A	O	I
2004	Dmitry and Balash (2004)	Conference Paper	A	O	B
2004	Hof et al. (2004)	Forest Science	A	O	I
2004	Holvad et al. (2004)	Transportation	T/A	O	I

Table A.5 Studies of DEA with their specific features [continued]

Year	Study	Source of Publication	Type	Journal Type	Application scheme
2004	Jahanshahloo et al. (2004)	Applied Mathematics and Computation	T	O	-
2004	Jahanshallo et al. (2004)	Applied Mathematics and Computation	T	O	-
2004	Joro (2004)	Journal of the Operational Research Society	T	M	-
2004	Korhonen (2004)	Management Science	T/A	M	O
2004	Korhonen nd Luptacik (2004)	European Journal of Operational Research	T/A	M	U
2004	Lozano and Villa (2004)	Journal of Productivity Analysis	T/A	E	O
2004	Neves et al. (2004)	International Journal of Management and Decision Making	A	O	I
2004	Ozgen and Ozcan (2004)	Health Care Management Science	A	O	H
2004	Ruggiero (2004)	Journal of the Operational Research Society	T	M	-
2004	Shao and Shu (2004)	Journal of the Operational Research Society	A	M	O
2004	Sowlati and Paradi (2004)	Omega	T	M	-
2004	Tuzkaya and Ertay (2004)	Conference Paper	B	O	-
2004	Vaninsky (2004)	Journal of Information and Optimization Sciences	A	O	O
2004	Wu and Xuan (2004)	System Engineering Theory and Practice	B	O	O
2005	Barth and Staat (2005)	Journal of Business Performance Management	T/A	O	B
2005	Bhat (2005)	European Journal of Health Economics	A	E	H
2005	Camanho and Dyson (2005)	Journal of the Operational Research Society	T/A	M	B
2005	Coelli and Rao (2005)	Agricultural Economics	T/A	E	I
2005	Costantino et al. (2005)	Conference Paper	T/A	O	I
2005	Donthu et al. (2005)	Journal of Business Research	A	O	I
2005	Ertuayrul and Mehmet (2005)	Emerging Markets Finance and Trade	A	O	B
2005	Garcia et al. (2005)	Progress in Nuclear Energy	B	O	-
2005	Hong et al. (2005)	Construction innovation	A	O	I
2005	Jahanshahloo et al. (2005)	Applied Mathematics and Computation	T	O	-
2005	Kitayama et al. (2005)	Transactions of the Japan Society of Mechanical Engineers	A	O	I
2005	Li et al. (2005)	Journal of the North china Electric Power University	A	O	U
2005	Munksgaard et al. (2005)	Energy Policy	A	O	U
2005	Ramanathan (2005)	International Journal of Operations and Production Management	A	O	H

Table A.5 Studies of DEA with their specific features [continued]

Year	Study	Source of Publication	Type	Journal Type	Application scheme
2005	Saen et al. (2005)	Applied Mathematics and Computation	T/A	O	B
2005	Saen et al. (2005)	Applied Mathematics and Computation	T	O	-
2006	Arestis et al. (2006)	International Review of Applied Economics	T/A	E	U
2006	Bian and Tang (2006)	Working Paper	T/A	O	O
2006	Camanho and Dyson (2006)	Journal of Productivity Analysis	T	E	-
2006	Damar (2006)	Applied Economics	T/A	E	B
2006	Kirkparick et al. (2006)	World Bank Economic Review	A	E	I
2006	Lee et al. (2006)	Lecture notes in Artificial Intelligent	B	O	-
2006	Ma and Zhang (2006)	Systems Engineering and Electronics	T	O	-
2006	Mabert et al. (2006)	Mathematical and Computer Modelling	A	O	O
2006	Newman and Matthews (2006)	Journal of Productivity Analysis	T/A	E	I
2006	Prior (2006)	Annals of Operations Research	T/A	M	H
2006	Ramanathan (2006)	Socio-Economic Planning Sciences	A	E	O
2006	Soleimani et al. (2006)	Applied Mathematics and Computation	T	O	-
2006	Soteriou and Hadjicostas (2006)	European Journal of Operational Research	T	M	-
2006	Wang et al. (2006)	Journal of American Society for Horticultural Science	A	O	I
2006	Wang et al. (2006)	System Engineering Theory and Practice	A	O	O
2006	Xu et al. (2006)	European Journal of Operational Research	B	M	-
2006	Yang and Lu (2006)	IEEE Transactions on Power Systems	A	O	U
2007	Amin (2007)	International Journal of Operations Research	T	M	-
2007	Cook and Bala (2007)	Omega	T	M	-
2007	Garcia et al. (2007)	Applied Economics	T/A	E	O
2007	Podinovski and Thanassoulis (2007)	Journal of Productivity Analysis	T	E	-

Note: T/A: Theoretical and application; T:Theoretical; A:Application; B:Bridging with other theoretical discipline; M:OR Journal; E:Economics Journal; O:Other journals; B:Banking/finance; U:Utilies; I:Industry; O:Others; H:Healthcare; E:Education; P:Public sector;

APPENDIX B : SUPPLEMENTARY RESULTS FOR THE MONTE CARLO DEA APPLICATION STUDY

Table B.1: Target optimal values for inputs, outputs and intermediate of each DMU

DMU		Supplier-cost	Supplier-revenue	Manufacturing cost	Manufacturing time	Distributor cost	Customer response time	Fill rate	On time delivery	Retailer cost
1	Original value	130	20	125	3	90	3	0.7	0.96	100
	Optimal	119.37	21.22	121.26	2.91	75.91	2.53	0.91	0.96	100.00
	% change	-8.18	6.11	-2.99	-2.99	-15.65	-15.65	29.92	0.00	0.00
2	Original value	150	21	120	2	100	3	0.9	0.95	110
	Optimal	89.97	14.30	74.94	1.25	46.50	1.40	0.54	0.58	78.57
	% change	-40.02	-31.90	-37.55	-37.55	-53.50	-53.50	-39.75	-38.47	-28.57
3	Original value	165	23	110	3	80	2	0.78	0.97	130
	Optimal	118.84	17.18	82.18	2.24	59.76	1.49	0.58	0.72	123.21
	% change	-27.97	-25.29	-25.29	-25.29	-25.29	-25.29	-25.29	-25.29	-5.22
4	Original value	170	24	150	4	70	4	0.88	0.89	125
	Optimal	85.53	13.88	66.99	1.79	48.33	2.76	0.48	0.59	82.82
	% change	-49.69	-42.19	-55.34	-55.34	-30.96	-30.96	-45.85	-34.02	-33.74
5	Original value	200	27	146	2	85	2	0.73	0.99	140
	Optimal	137.58	20.75	112.18	1.54	65.31	1.54	0.56	0.76	133.51
	% change	-31.21	-23.17	-23.17	-23.17	-23.17	-23.17	-23.17	-23.17	-4.64
6	Original value	185	25	115	3	77	2	0.95	0.89	135
	Optimal	97.85	14.11	66.58	1.74	48.14	1.25	0.49	0.59	96.85
	% change	-47.11	-43.57	-42.11	-42.11	-37.48	-37.48	-48.91	-33.99	-28.26
7	Original value	135	24	105	2	78	1	0.89	0.93	125
	Optimal	135	24	105	2	78	1	0.89	0.93	125
	% change	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	Original value	190	30	100	2	90	3	0.87	0.88	155
	Optimal	190	30	100	2	69.30	2.31	0.87	0.88	155
	% change	0.00	0.00	0.00	0.00	-23.00	-23.00	0.00	0.00	0.00

Table B.1: Target optimal values for inputs, outputs and intermediate of each DMU [continued]

DMU		**Supplier-cost**	**Supplier-revenue**	**Manufacturing cost**	**Manufacturing time**	**Distributor cost**	**Customer response time**	**Fill rate**	**On time delivery**	**Retailer cost**
9	Original value	185	28	135	4	78	2	0.95	0.99	135
	Optimal	128.72	19.60	93.74	2.78	68.17	1.75	0.66	0.83	117.12
	% change	-30.42	-30.00	-30.56	-30.56	-12.60	-12.60	-30.03	-16.50	-13.24
10	Original value	190	25	120	3	68	1	0.9	0.83	130
	Optimal	190	25	116.18	2.90	68	1	0.9	0.83	130
	% change	0.00	0.00	-3.18	-3.18	0.00	0.00	0.00	0.00	0.00

Table B.2: The distribution of the Monte Carlo efficiency scores.

DMU No	Mean	Median (50%)	5%	10%	25%	75%	90%	95%
DMU1	0.8923	0.9330	0.6941	0.7171	0.9330	0.9330	0.9330	0.9330
DMU2	0.6021	0.5973	0.4867	0.5140	0.5474	0.6668	0.6974	0.7092
DMU3	0.7933	0.7954	0.6459	0.6676	0.7230	0.8655	0.9214	0.9300
DMU4	0.5661	0.5618	0.4653	0.4717	0.5057	0.6192	0.6677	0.6821
DMU5	0.7911	0.7976	0.6427	0.6557	0.7222	0.8626	0.9179	0.9369
DMU6	0.5934	0.5892	0.4975	0.5007	0.5279	0.6470	0.7102	0.7175
DMU7	0.9543	1	0.8225	0.8375	0.9037	1	1	1
DMU8	0.8839	0.9425	0.7402	0.7588	0.8023	0.9425	0.9425	0.9425
DMU9	0.7846	0.7741	0.6375	0.6436	0.6871	0.8874	0.9474	0.9484
DMU10	0.9527	0.9920	0.8039	0.8469	0.9018	0.9920	0.9920	0.9920

APPENDIX C : ALGORITHMS FOR THE GA AND OTHER TECHNIQUES.

Algorithm for GA+OCBA-m technique

```
Step 1: Initialization
        i.e. set N=budget,  number of initial_data, CV, D,
        pop_size=100, max_generation=200, crossover_prob=1, mutat_prob=0.01,
        tsize=2, popcount=1, noimprovement=0

Step 2: Generate initial population

Step 3: Evaluation-Selection-Reproduction cycle

        3.1 Evaluate fitness of individuals in the population
                Apply OCBA-m procedure here.
                /Calculate the MSE and select the top-m solutions.
                        Set bestMSE=individual with best fitness
                        Arrange the solutions (from the fittest to the least fit)

        3.2 Create next generation
                /*Loop 40 times (percentage of best solutions to be retained 20%) in
                order to generate 80 children*/
                For i=1:40
                        /*Selection of parents*/
                        k=1;
                        While k<=2
                                Select two individuals randomly
                                Compare MSE of the two
                                Set winner as parent_k;
                        End
                        /*Two-point crossover*-  Exchange of parent's vector/
                        Do  crossover using method described in 6.3.5
                End
                /*Check feasibility of all children*/
                For i=1 to 80
                        Calculate sum_gene  of child [i]
                        If sum_gene>Budget
                                surplus=sum_gene-Budget
                                For i=1 to surplus
                                        Randomly select a position or gene
                                        Substract 1 from the position
                                End
                        Else if sum_gene<Budget
                                slack=Budget-sum_gene
                                For i=1 to slack
                                        Randomly select a position or gene
                                        Add 1 to the position
                                End
                        End
                End
```

```
/*Mutation*/
For i=1 to 80
        If rand()>mutat _prob
                Randomly select two genes from child [i]
                Randomly exchange the values of the genes
        end
End
```

3.2 Check whether termination condition is satisfied

```
If popcount>1
        If (bestMSE(popcount)>=bestMSE(popcount-1))
                noimprovement=noimprovement+1;
        End
        If (noimprovement>20) or  (popcount > max_generation)
                STOP
        End
End
```

3.3 Replacement of the populations by the children.

Use the selected top-*m* solutions from OCBA-*m* to update the subsequent population for next iteration i.e. retain the top *m* solutions & replace the remainder with the children; popcount=popcount+1; Go to Step 3.1.

Figure C.1: Pseudo-code for the GA+OCBA-*m* algorithm

Algorithm for Greedy technique

Step 1: Initialization

i.e. set budget, number of initial_data, D, CV, N=0, ΔN=1

Step 2: Increment N by $+\Delta N$

set $N=N+\Delta N$

Step 3: Evaluate the designs and choose the best one

While $N \leq$ Budget

Find feasible design allocations n

Calculate MSE for all designs

Determine best MSE and the associated **n**

Set best_design = **n** which has the best MSE

$N=N+\Delta N$

End

Recall: $\sum_{k\in K} n_k = N$ and $\mathbf{n} = [n_k]_{k\in K}$

Figure C.2: Pseudo-code for the Greedy algorithm

Algorithm for Batch technique

Step 1: Initialization

i.e. set N=budget, D, number of initial_data, CV

Step 2: Find feasible factors

List all possible factors (ω)

Total number of design = $[(N/\omega)+D-1]!/[(D-1)!*(N/\omega)!]$.

If total number of design <5000

feasible_factors=factors.

End

Step 3: Evaluate the designs

For all feasible_factors

Find all possible design allocations **n**

Calculate MSE

Determine the top 10 best MSE and the corresponding designs **n**

End

Figure C.3: Pseudo-code for the Batch algorithm

APPENDIX D: SUPPLEMENTARY TABLES AND FIGURES FOR CHAPTER 7

Table D.1: Variance of the efficiency as affected by the range of values of the inputs/outputs/variables

Input /output variable	**Input Values**		**Efficiency**		**Variance of the efficiency**
	Min	Max	Min	Max	
Retailer revenue	155.9	467.7	0.35021988	1	0.422214
Fill rate (mfg)	42.58%	100.00%	0.908146899	0.648728342	0.067298
Supplier-revenue	12.78	38.34	0.845703801	0.597763662	0.061474
Supplier-cost	81	243	0.839970084	0.600661764	0.057268
customer res time	1.22	3.66	0.803319102	0.620919889	0.033269
fill rate (DC)	40.19%	100%	0.88081469	0.668728342	0.044981
Supplier-labor	78.9	236.7	0.763681497	0.64687127	0.013645
cycle time	1.42	4.26	0.708946597	0.692134652	0.000283
Retailer cost	65.1	195.3	0.706281897	0.694693477	0.000134
mfg cost	60.56	181.68	0.704353693	0.696569082	6.06E-05
on-time delivery (mfg-dc)	46.02%	100.00%	0.701826472	0.700200457	2.64E-06
DC cost	40.28	120.84	0.70147188	0.700407641	1.13E-06
on-time delivery (dc-ret)	51.50%	99.90%	0.701239759	0.700439759	6.4E-07
fill rate (ret)	57.50%	100%	0.700839759	0.700339759	1.6E-07
on-time delivery (mfg-ret)	55.80%	99.99%	0.700439759	0.700139759	0

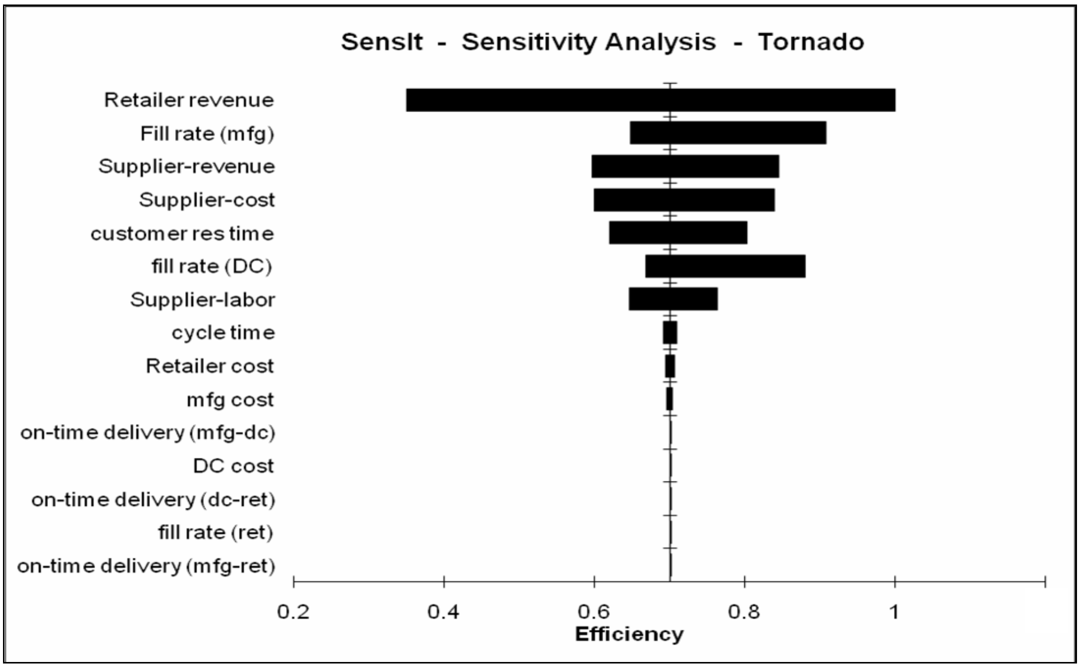

Figure D.1: Sensitivity Analysis

From the sensitivity analysis study conducted by Wong et al. (2008), the overall average supply chain efficiency is most sensitive to 'retailer revenue' followed by 'fill rate (mfg)', 'supplier revenue' and 'supplier cost'. These are the top 4 most influential variables in the efficiency study of supply chain. The results are as shown in Figure D.1 and Table D.1. In conducting the sensitivity analysis, we vary one variable at a time. For example, when we want to analyze the impact of 'retailer revenue' on efficiency, 'retailer revenue' will be varied while the other variables are kept constant. The purpose of conducting the sensitivity analysis is to enable the user to find out the influential variables which have significant impact on the average supply chain efficiency.