Journal of Applied Mathematics

Operational Research

Guest Editors: M. Khodabakhshi, F. Hosseinzadeh Lotfi, X. Zhang, S. Li,
S. H. Nasseri, K. Aryavash, and R. Tavakkoli-Moghaddam

Operational Research

Journal of Applied Mathematics

Operational Research

Guest Editors: M. Khodabakhshi, F. Hosseinzadeh Lotfi, X. Zhang, S. Li, S. H. Nasseri, K. Aryavash, and R. Tavakkoli-Moghaddam

Editorial Board

Contents

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 318439, 2 pages
http://dx.doi.org/10.1155/2013/318439

Editorial

Operational Research

M. Khodabakhshi,[1] F. Hosseinzadeh Lotfi,[2] X. Zhang,[3] S. Li,[4] S. H. Nasseri,[5] K. Aryavash,[6] and R. Tavakkoli-Moghaddam[7]

[1] *Department of Mathematics, Shahid Beheshti University, G.C., Tehran, Iran*
[2] *Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran*
[3] *School of Economics & Management, Tongji University, Shanghai, China*
[4] *Faculty of Management, McGill University, Montreal, Canada*
[5] *Department of Mathematics, University of Mazandaran, Babolsar, Iran*
[6] *Department of Mathematics, Faculty of Science, Lorestan University, Khorram Abad, Iran*
[7] *School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran*

Correspondence should be addressed to M. Khodabakhshi; mkhbakhshi@yahoo.com

Received 7 November 2013; Accepted 7 November 2013

1. Introduction

Operations research or operational research (OR) is a discipline that deals with the application of advanced analytical methods to help make better decisions. Employing techniques from other mathematical sciences, such as mathematical modelling, statistical analysis, mathematical optimization, and operations research arrives at optimal or near-optimal solutions to complex decision-making problems. The main focus of this special issue will be on the new research ideas and results for the OR. Considering the wide applications of OR, it seems natural that this journal selected it as the theme of its special issue.

2. Overview of Works Presented in This Special Issue

This special issue includes 19 high-quality papers that deal with different fields in OR. These papers have been accepted among 54 manuscripts. These papers contain some new, novel, and innovative techniques and ideas which can be developed and extended in the further scientific works.

The subjects in data envelopment analysis (DEA) have occupied three contributions. The DEA is a nonparametric efficiency estimating technique. One of these papers proposed is "Modified Malmquist productivity index based on present time value of money." The other papers were devoted to interval scale of efficiency and ranking subjects. Also, two papers have studied road congestion pricing problem. There are also three contributions on the applications of fuzzy set theory. Furthermore, five works have been presented about the different methods of optimization. Also, two papers presented in this special issue are closely related to the risk models. A new bandwidth allocation model is studied in one of the published papers. One of the published papers proposed a model for estimating car delays at bus stops. One of the contributions reviewed the theories of process control. Finally, one of the accepted papers considered stochastic P-function, stochastic P_0-function, and stochastic uniformly P-function.

3. Conclusions

The operational research and its applications are one of the most important fields of applied mathematics. So, this special issue focuses on this subject. The aim of this special issue is to present and extend the applications of the relatively new approaches and theories for the OR and its applications. The

editors hope that the special issue will provide new ideas in the development of OR.

Acknowledgments

In closing this editorial, we would like to express our deepest gratitude to many reviewers, whose useful comments guaranteed the high quality of the selected papers.

M. Khodabakhshi
F. Hosseinzadeh Lotfi
X. Zhang
S. Li
S. H. Nasseri
K. Aryavash
R. Tavakkoli-Moghaddam

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 636259, 12 pages
http://dx.doi.org/10.1155/2013/636259

Research Article

Risk-Averse Newsvendor Model with Strategic Consumer Behavior

Tie Wang[1,2] and Qiying Hu[3]

[1] *School of Mathematics, Liaoning University, Shenyang, Liaoning 110036, China*
[2] *School of Management, Shanghai University, Shanghai 200444, China*
[3] *School of Management, Fudan University, Shanghai 200443, China*

Correspondence should be addressed to Tie Wang; wangt_1997@163.com

Received 1 February 2013; Revised 8 July 2013; Accepted 20 July 2013

Academic Editor: Mohammad Khodabakhshi

The classic newsvendor problem focuses on maximizing the expected profit or minimizing the expected cost when the newsvendor faces myopic customers. However, it ignores the customer's bargain-hunting behavior and risk preference measure of the newsvendor. As a result, we carry out the rational expectation (RE) equilibrium analysis for risk-averse newsvendor facing forward-looking customers who anticipate future sales and choose purchasing timing to maximize their expected surplus. We propose the equations satisfied by the RE equilibrium price and quantity for the risk-averse retailer in general setting and the explicit equilibrium decisions for the case where demand follows the uniform distribution and utility is a general power function. We identify the impacts of the system parameters on the RE equilibrium for this specific situation. In particular, we show that the RE equilibrium price for some risk-averse newsvendors is lower than for a risk-neutral retailer and the RE equilibrium stocking quantity for some risk-averse newsvendors is higher than for a risk-neutral retailer. We also find that the RE equilibrium sale price for a risk-averse newsvendor is decreasing in salvage price in some situations.

1. Introduction

1.1. Motivation. Strategic consumer behavior is widely acknowledged by the retailer and deeply influenced ordering, pricing, and other marketing decisions for the retailer. A strategic consumer chooses between a purchase at the initial full price with the possibility, if inventory remains, of a later purchase at a salvage price. Recently, it has attracted much attention by the researchers from supply chain management and revenue management when the decision maker is risk neutral. Evidently, not all decision makers are risk neutral. Indeed, some experimental evidence suggests that for some products, the so-called high-profit products, the decision makers are risk averse; see Schweitzer and Cachon [1] for more details. However, according to our knowledge, few considered the combined impacts of strategic consumer behavior and risk aversion on the pricing and ordering decisions for the newsvendor.

In this paper we study a risk averse retailer's stocking and pricing in the presence of strategic consumers. This paper has three main objectives. First, we obtain the rational expectation (RE) equilibria under the rational expectations hypothesis firstly proposed by Muth [2] for the risk-averse retailer. It states that economic outcomes do not dier systematically from what people expect them to be. We begin with the classic newsvendor setting, which is a fundamental building block in the literature, and proceed to incorporate strategic demand and risk aversion into the model. Second, we would like to introduce a specific risk averse utility-power utility to investigate explicitly the impacts of strategic consumer behavior and risk aversion on newsvendor decisions. The third objective is to study the impact of systems parameters on the RE equilibria.

1.2. The Literature Review. The classic newsvendor problem is a crucial building block of the stochastic inventory theory because of its simple and elegant structure as well as its rich managerial insights. It assumes that if any inventory remains at the end of the period, a discount is used to sell it or it is disposed of. If the order quantity is smaller than

the realized demand, the newsvendor loses some profit. The classic newsvendor model reflects many real situations and is often used to help make decisions in many industries, such as fashion and sporting, both at the manufacturing and sale level. It has been extensively studied over decades with extensions including different objectives and utility functions, multiple products with substitution, multiple locations, different pricing, and marketing strategies. Khouja [3] builds a taxonomy of the single-period problem literature and delineates the contribution of the different extensions and suggests some future directions for research.

One important extension to the classic newsvendor model is the interface of marketing and operation that provides an important tool for examining how operational problem interacts with marketing issues to influence decision-making at the firm level. There are three research streams about this extension, one is the newsvendor problem with pricing whose demand is stimulated by sale price, one is the newsvendor problem with inventory whose demand is driven by inventory, and the other is the newsvendor problem with marketing whose demand is stimulated by other marketing instruments. Petruzzi and Dada [4] provided a good review about the newsvendor problem with pricing that synthesizes existing results and develops additional results to enrich existing knowledge base. Yao et al. [5] extended this model to increasing price elasticity (IPE) mean demand situation and used this condition to investigate the newsvendor problem with pricing under stochastic demand distribution class with increasing generalized failure rate (IGFR) that was presented in Lariviere and Porteus [6] to study supply coordination with wholesale price contract based on the classic newsvendor model.

The first paper that takes the inventory as a marketing measure is by Gerchak and Wang [7]. They extended the classical newsvendor model to include endogenous, inventory-dependent demand for aexogenous price. The demand form in their paper was multiplicative demand that models actual demand as a deterministic multiple of a base random variable with a fixed probability distribution. Balakrishnan et al. [8] investigated the deterministic counterpart for the newsvendor problem with inventory. Dana and Petruzzi [9] extended the classic newsvendor model by assuming that expected utility maximizing consumers choose between visiting the firm and consuming an exogenous outside option. They modeled the stochastic demand as multiplicative form depending inventory and price and found that the firm holds more inventories, provides a higher fill rate, attracts more customers, and earns higher profits when it internalizes the effect of its inventory on demand. Recently, Balakrishnan et al. [10] extended the newsvendor problem to general inventory-dependent demand distribution with given price and showed that demand stimulation has the effect of increasing the target service level beyond the classical newsvendor model's critical ratio. Similar to Dana and Petruzzi [9], they also addressed the problem of jointly optimizing both stocking quantity and price for demand-stimulating products using a multiplicative model to represent the influence of price and stocking quantity on the demand distribution. For this model, they showed that the pricing and stocking decision can be determined sequentially, with the optimal policy setting higher prices and stock levels than both the functional policies (demand-driven and critical fractile). Liu et al. [11] study the impact of supply reliability on a retail firm's performance under joint marketing and inventory decisions. They established a necessary and sufficient condition under which the maximum unit cost a firm is willing to pay to improve supply reliability increases in product price and showed that for two products with the same price, a firm is willing to pay more to improve supply reliability for the product with higher product cost. They also found that a product with a lower marketing cost function always benefits more from improved supply reliability than a product with a higher marketing cost function. Taylor [12] showed that when demand is influenced by retailer sales effort, a properly designed target rebate and return contract achieves coordination and a win-win outcome. Krishnan et al. [13] investigated similar problem as Taylor's [12]. But the retailer chooses inventories ex ante and promotional effort ex post in their paper.

Taking advertising as a special marketing instrument, Gerchak and Parlar [14] studied the newsvendor model when multiplicative demand has a distribution with a mean that is specific concave and increasing in advertising expenditure. They developed a mixed optimization technique which combines simulation with the first order condition to solve the previous problem. Khouja and Robbins [15] extended the model presented by Gerchak and Parlar [14] to three cases of demand variation as a function of advertising expenditure: (1) demand has constant variance, (2) demand has constant coefficient variation, and (3) demand has an increasing coefficient variation. They investigated the newsvendor problem with advertising under multiplicative demand and obtained the optimal advertising premium and ordering quantity by maximizing the expected profit or maximizing the probability of achieving a target profit under the previous three situations using particular mean demand and discussed that the optimal advertising decisions for maximizing profit is increased with the profit margin. Recently, Wang and Zhou [16] discussed the supply chain coordination with newsvendor under advertisement sensitive demand and proposed an improved revenue-sharing contract to achieve the supply chain coordination. Wang et al. [17, 18] investigated the supply chain coordination with a newsvendor under specific advertisement and price sensitive demand using improved revenue sharing contract and combined return and sales rebate/penalty contract.

It is natural to incorporate the decision maker's risk attitude in newsvendor model because anybody has his own preference when he makes a decision. Many planners are willing to trade off lower expected profit for downside protection against possible losses. The literature about risk-averse newsvendors is fewer than those about risk-neutral newsvendors. Lau [19] analyzed the classical newsvendor model under two different objectives. In the first objective, the focus is on maximizing the decision maker's expected utility of total profit. The second objective is the maximization of the probability of achieving a certain level of profit. Eeckhoudt et al. [20] explored the newsvendor model with

increasing, concaves and thrice differentiable utility function, and showed that the optimal ordering quantity in their environment is smaller than the risk-neutral optimal ordering quantity and decrease in risk-aversion level. Keren and Pliskin [21] derived the first order conditions for optimality for the problem of a risk-averse expected-utility maximizer newsvendor. They solved a special case where the utility function is any increasing differentiable function, and the random demand is uniformly distributed by using these conditions. Agrawal and Seshadri [22] explored the effect of risk aversion on the pricing and ordering decision with emergency ordering. They modeled the risk aversion as a twice continuous differentiable and increasing concave utility function. Chen and Federgruen [23] analyzed the mean-variance trade-offs in newsvendor model as well as some standard infinite horizon inventory problems. Choi et al. [24] studied the same problem, but they explore all kinds of risk attitudes and focus on the profit and instantiate the case with shortage penalty and safety-first objective.

Some research papers use the conditional value at risk (CVaR) as a special risk-averse metrics to investigate the newsvendor problem. Gotoh and Takano [25] analyzed the minimization of the CVaR for single and multiple products cases. Following the condition for mean demand presented in Yao et al. [5], Chen et al. [26] investigated the joint pricing and ordering strategies for newsvendor under CVaR decision criteria which is based on the maximization of the CVaR. They provided sufficient conditions for the existence and uniqueness of optimal policies for both additive and multiplicative demands and performed the comparative statics which show the monotone properties and other characteristics of the optimal pricing and ordering decisions. They also compared their results with those for a risk-neutral newsvendor. In addition, Ahmed et al. [27] used the general coherent risk measure as a risk-averse preference to investigate both the single period and multiperiod inventory problems.

Behavioral operations management is an emerging area to the study of operations that explicitly incorporates social and cognitive psychological theory. It is the study of human behavior and cognition and their impacts on operating systems and processes. So far, there are three review papers about behavior operation. Gino and Pisano [28] explored the theoretical and practical implications of incorporating behavioral and cognitive factors into models of operations management and suggest fruitful avenues for research in behavioral operations. Bendoly et al. [29] highlighted theoretical constructs and empirical phenomena from behavioral economics/judgment and decision making, industrial and organizational psychology, group dynamics, and system dynamics and provided a guide for where to go to learn more about each body of knowledge. Shen and Su [30] reviewed current models of customer behavior in the revenue and suction literature and suggested several future research directions. Su [31] proposed a decision framework of bounded rationality, in which decision makers are prone to errors and biases. He applied this framework to the classic newsvendor model and characterized the ordering decisions made by a roundedly rational decision maker. He identified systematic biases and offered insight into when overordering and underordering may occur. Su [32] studied the dynamic pricing of finite inventories with a heterogeneous population of strategic as well as myopic customers and showed that depending on the customer composition, optimal price paths could involve either markups or markdowns. Su and Zhang [33] investigated the impact of strategic customer behavior on supply chain performance. Applying rational expectations hypothesis to the newsvendor model, they analyzed the previous model by looking for rational expectations (RE) equilibria, which satisfies (i) given their expectations of future availability, consumers make their purchase (or waiting) decisions, (ii) given his expectations of consumers' willingness to pay, the newsvendor makes his pricing and stocking decisions, and (iii) everyone's expectations are consistent with actual outcomes. They show that in RE equilibrium, the newsvendor will invest in less inventory and charge a lower regular sales price. The newsvendor's performance is substantially affected by the consumers' waiting behavior. To alleviate this impact, they study two mechanisms—quantity commitment and price commitment—embedded in supply chain management. Su and Zhang [34] studied the role of product availability in attracting consumer demand. They base on a newsvendor but assume that consumers must incur some search cost in order to visit the seller. The seller sets an observable price and an unobservable stocking quantity. Consumers anticipate the likelihood of stockout and determine whether to visit the seller. They characterize the RE equilibrium in this game and show that the seller can improve profits by providing inventory information or availability guarantees. Lai et al. [35] examine the impact of a posterior price matching policy in which the seller guarantees to reimburse the price difference to a consumer who buys a product before the seller marks it down on strategic consumer's purchasing behavior, a seller's pricing and inventory decisions and their expected payoffs, assuming that the seller cannot credibly commit to a price path but can implement a posterior price matching policy.

Our model differs from the previous papers in the following aspects. Firstly, we explore the case where the seller is risk averse and consumers are forward looking. Secondly, we obtain an analytical solution for a specific situation where demand follows the uniform distribution and utility is a power function to illustrate the combined impacts of strategic consumer behavior and risk aversion on newsvendor's decisions. Thirdly, we investigate the influence of system parameters on RE equilibrium and compare them to the existing results about the risk-averse newsvendor with myopic consumers and the risk-neutral newsvendor with strategic consumers. Based on the analysis of the model, we obtain the following insights: (1) the RE equilibrium ordering quantity and sale price are all lower than those for a risk-averse newsvendor with myopic customers, (2) the optimal ordering quantity for a risk-averse newsvendor facing myopic consumers is higher than the RE equilibrium stocking quantity for a risk-neutral retailer, (3) the RE equilibrium sale price for some risk-averse newsvendors is lower than for a risk-neutral retailer, and the RE equilibrium ordering quantity for some risk-averse newsvendors is higher than for a risk-neutral retailer, (4) in some situations, the

Table 1: Notions used in this paper.

c	Product cost
v	Customer valuation
p	Retail price
p_{rn}	RE equilibrium price for risk-neutral retailer
p_{ra}	RE equilibrium price for risk-averse retailer
s	Salvage price
Q	Order quantity
Q_0	Optimal ordering quantity for risk-averse newsvendor model with myopic customers
Q_{rn}	RE equilibrium ordering quantity for risk-neutral retailer
Q_{ra}	RE equilibrium ordering quantity for risk-averse retailer
$f(\cdot), F(\cdot)$	pdf and cdf of the distribution of stochastic demand ξ
$\pi(q)$	Profit function
$U(\cdot)$	Newsvendor's utility function

sale price in RE equilibrium for risk-averse newsvendor is decreasing in salvage price.

The remainder of this paper is organized as follows. Section 2 gives the model and derives RE equilibrium for a risk-averse newsvendor. Section 3 explores a specific case where demand follows the uniform distribution and utility is a power function. Section 4 presents the numerical examples to illustrate the model. Section 5 concludes the paper with future researches.

2. The General Model

In this section, we present the risk-averse newsvendor model with strategic customers. Notions used in this paper are given in Table 1 at the end of this part.

Our starting point is the classic newsvendor model. There is a single risk-averse retailer who must determine how many units of a product to order. The retailer faces a random demand $\xi \geq 0$, which has distribution F and probability density function f. We assume that f is continuous, $f(0) > 0$ and $F(0) = 0$. The risk-averse retailer has a utility function $U(\cdot)$. We assume $U(0) = 0$ and $U'(\cdot) > 0$, $U''(\cdot) < 0$, and each unit of the product costs c but is valued by customer at v. Leftover units can be sold in an exogenous salvage market at s per unit. We also assume $0 < s < c < v$. Customers choose between buying immediately at full price or waiting for the sale at salvage price because they recognize that the product may become available on the salvage market at price s. If the regular retail price is too high, customers may find it worthwhile to wait for the sale, even if the product may be sold out by then.

Our model setup is similar to that in Su and Zhang [33]. Sequence of events are as the following. First, the risk-averse retailer privately forms his beliefs over customers' reservation prices and then optimally chooses the price and quantity given these beliefs. We also assume that customers may observe the retail price but do not observe the ordering quantity. Then, customers privately form beliefs over their chances of obtaining the product on the salvage market and then form their reservation prices based on these beliefs. Next, the random demand ξ is realized. Then, sales occur at the full price p (provided that the retail price p does not exceed consumers reservation r). Finally, all remaining units are sold at the salvage price s.

We first describe the consumer's decision problem. Consider a particular consumer who forms the belief that he will obtain the product with probability ξ_{prob} if he waits for the sale. Based on these expectations, the consumer's expected surplus if he faces an actual retail price p is

$$\max\left\{v - p, (v - s)\,\xi_{\text{prob}}\right\}. \tag{1}$$

The first term is his surplus from buying at the regular price p, and the second term is his expected surplus if he waits for the sale, where there is probability ξ_{prob} that he earns surplus $v - s$ and probability $1 - \xi_{\text{prob}}$ that he earns zero surplus (if the product is out of stock). Since the consumer chooses the more attractive option between buying and waiting, he will buy at price p if and only if $v - p \geq (v - s)\xi_{\text{prob}}$. In other words, given his expectations ξ_{prob}, the consumer's reservation price for the product is

$$r\left(\xi_{\text{prob}}\right) = v - (v - s)\,\xi_{\text{prob}}. \tag{2}$$

We consider homogeneous customers who share the same beliefs ξ_{prob} and the same reservation price r and assume that consumers are risk neutral and they do not discount future payoff.

Next, we consider the retailer's decision problem to determine an ordering quantity Q and a retail price p. Suppose that the seller expects that all customers have a reservation price ξ_r. Given these beliefs, it is clear that he will choose price and quantity as follows

$$p = \xi_r, \tag{3}$$

$$Q(p) = \arg\max_{Q} E\left[U\left(\pi\left(Q, p\right)\right)\right], \tag{4}$$

where $\pi(Q, p) = p\min(Q, \xi) + s(Q - \xi)^+ - cQ = (p - c)Q - (p - s)(Q - \xi)^+$ is the profit for the risk-neutral newsvendor. Notice that given his expectations ξ_r, the retailer is essentially considering a fixed price and solving a risk-averse newsvendor problem.

The previous discussion establishes the relationship between the initial beliefs ξ_r, ξ_{prob} and the subsequent decisions p, Q, r. After demand is realized, sales are generated according to these decisions. If their reservation price r exceeds the retail price p, consumers are willing to buy at this price, so regular and salvage sales occur as in the risk-averse newsvendor model (i.e., $\min(Q, \xi)$ units are sold at price p and the remaining are salvaged at price s); otherwise, all customers prefer to wait, no regular sales occur, and all units are salvaged at price s. The final requirement of our model is that these eventual outcomes (in terms of sales) should be consistent with all initial beliefs. This will be made clear in the following.

We can obtain the following rational expectation equilibrium through the strategic interaction analysis between the risk-averse retailer and the customers.

Definition 1. A rational-expectations equilibrium for risk-averse newsvendor and risk-averse strategic consumers $(p, Q, r, \xi_{\text{prob}}, \xi_r)$ satisfies the following: (i) $v - r = (v - s)\xi_{\text{prob}}$, (ii) $p = \xi_r$, (iii) $Q = \arg\max_Q E[U(\pi(Q, p))]$, (iv) $\xi_{\text{prob}} = F(Q)$, and (v) $\xi_r = r$.

Conditions (i), (ii), and (iii) assert that under expectations ξ_{prob} and ξ_r, the seller and all consumers will rationally choose the appropriate utility-maximizing actions, as specified in (2), (3), and (4). The last two conditions require that expectations must be consistent with outcomes. In (iv), the expectations ξ_{prob} must concur with the actual probability of obtaining the product if an individual consumer waits for the sale. This actual probability can be calculated as follows. In equilibrium, the seller prices the product at consumers reservation price, so all consumers will buy the product. Therefore, if an individual consumer waits instead, this consumer will obtain the product if and only if $\xi \le Q$, which occurs with probability $F(Q)$, as shown in (iv). Here, we have implicitly assumed efficient rationing: customers who wait for the sale have the highest priority to receive the product in the salvage market. This is reasonable because customers who are interested in a particular product and eagerly waiting for a sale are also the ones who are more likely to get the product when the sale actually takes place. Finally, in (v), the seller must correctly anticipate consumer's reservation price.

For risk-neutral newsvendor, Su and Zhang [33] derived the following result:

$$\begin{aligned} p &= v - (v - s) F(Q), \\ Q &= \arg\max_Q E[\pi(Q)]. \end{aligned} \tag{5}$$

They concluded that the optimal ordering quantity and retail price for the risk-neutral newsvendor facing strategic consumers must satisfy

$$p_{\text{rn}} = s + \sqrt{(v - s)(c - s)}, \tag{6}$$

$$\overline{F}(Q_{\text{rn}}) = \sqrt{\frac{c - s}{v - s}}. \tag{7}$$

In the following proposition, we explore the relationship between the retail price and ordering quantity with the system parameters in RE equilibrium.

Proposition 2. *Under the RE equilibrium for the risk-neutral newsvendor, the retail price p_{rn} is increasing in customer valuation, ordering cost, but is decreasing in salvage price, while the ordering quantity Q_{rn} is increasing in customer valuation, salvage price but is decreasing in ordering cost.*

Proof. From (6) and (7), we immediately know that p_{rn} is increasing in v and c, q_{rn} is decreasing in c and increasing in v.

Taking the derivative of p_{rn} with respect to s leads to

$$\begin{aligned} \frac{dp_{\text{rn}}}{ds} &= 1 - \frac{v + c - 2s}{2\sqrt{(v - s)(c - s)}} \\ &\ge 1 - \frac{2\sqrt{(v - s)(c - s)}}{2\sqrt{(v - s)(c - s)}} = 0. \end{aligned} \tag{8}$$

Notice that $v > c$; therefore, $(dp_{\text{rn}}/ds) > 0$ for all $s < c$, and p_{rn} is strictly increasing in salvage value s.

Taking the derivative of $\overline{F}(q_{\text{rn}})$ with respect to s leads to

$$\frac{d\overline{F}(q_{\text{rn}})}{ds} = \frac{(c - v)}{2(v - s)^2\sqrt{(c - s)/(v - s)}} < 0. \tag{9}$$

Therefore, $\overline{F}(q_{\text{rn}})$ is decreasing in s; that is, q_{rn} is increasing in s. □

The conditions for rational expectation (RE) equilibrium in Definition 1 can be reduced to a pair of equations for only p and Q in the following proposition.

Proposition 3. *In the RE equilibrium, all consumers buy immediately, and the risk-averse newsvendor's retail price and order quantity are characterized by*

$$v - p = (v - s) F(Q), \tag{10}$$

$$\begin{aligned} &-(c - s)\int_0^Q U'((p - s)x - (c - s)Q) f(x)\,dx \\ &\quad + (p - c)\int_Q^{+\infty} U'((p - c)Q) f(x)\,dx = 0. \end{aligned} \tag{11}$$

Proof. Notice that the conditions for rational expectation (RE) equilibrium between risk-averse newsvendor and risk-neutral strategic consumers in Definition 1 can be reduced to the following pair of equations for only p and Q

$$\begin{aligned} v - p &= (v - s) F(Q), \\ Q &= \arg\max_Q E[U(\pi(Q))]. \end{aligned} \tag{12}$$

From the profit for the risk-neutral newsvendor, we can get

$$\begin{aligned} E[U(\pi(Q))] &= \int_0^Q U((p - s)x - (c - s)Q) f(x)\,dx \\ &\quad + \int_Q^{+\infty} U((p - c)Q) f(x)\,dx. \end{aligned} \tag{13}$$

Hence,

$$\begin{aligned}
&\frac{dE\left[U\left(\pi\left(Q\right)\right)\right]}{dQ} \\
&\quad = -\left(c-s\right)\int_0^Q U'\left(\left(p-s\right)x-\left(c-s\right)Q\right)f\left(x\right)dx \\
&\qquad + \left(p-c\right)\int_Q^{+\infty} U'\left(\left(p-c\right)Q\right)f\left(x\right)dx, \\
&\frac{d^2E\left[U\left(\pi\left(Q\right)\right)\right]}{dQ^2} \\
&\quad = \left(c-s\right)^2\int_0^Q U''\left(\left(p-s\right)x-\left(c-s\right)Q\right)f\left(x\right)dx \\
&\qquad + \left(p-c\right)^2\int_Q^{+\infty} U''\left(\left(p-c\right)Q\right)f\left(x\right)dx \\
&\qquad - \left(p-s\right)f\left(Q\right)U'\left(\left(p-c\right)Q\right) < 0.
\end{aligned} \tag{14}$$

Notice that

$$\left.\frac{dE\left[U\left(\pi\left(Q\right)\right)\right]}{dQ}\right|_{Q=0} > 0, \qquad \lim_{Q\to+\infty}\frac{dE\left[U\left(\pi\left(Q\right)\right)\right]}{dQ} < 0. \tag{15}$$

Therefore, there must be unique solution of $dE[U(\pi(Q))]/dQ = 0$ in $(0, +\infty)$. □

From Proposition 3, we cannot know if a risk-averse newsvendor with a uniformly more concave utility function sets his Q_{ra} to a lower value than a less risk-averse newsvendor, so a risk-averse newsvendor with a concave utility function sets Q_{ra} lower than a newsvendor who is risk neutral when the retailer faces strategic consumers. We also cannot know the relation of p_{ra} between the risk-averse newsvendors, so the relation of p_{ra} between risk-averse and risk-neutral newsvendors. In addition, it is meaningful to understand fully the interaction between strategic consumers behavior and risk aversion.

For readers' convenience, in Table 1 we list the notations used in this paper.

3. Rational Expectation Equilibrium for Power Utility and Uniform Distribution

It is instructive to compare the equilibrium price and quantity in our model with that in the risk-averse newsvendor model, where customers are not strategic and are willing to pay their valuation v for the product (so the retailer also charges v). It is also significant to compare the equilibrium price and quantity in our model with that in the classic newsvendor model or in the risk-neutral newsvendor model with strategic consumers. To derive structural results and generate managerial insights into the equilibrium decisions of the risk-averse newsvendor problem with strategic consumers, in the following, we present the specific results for the case where demand follows the uniform distribution, and retailer's utility is a power function. Without loss of generality, we assume $F(x)$ is distributed uniformly in $[0, A]$ with $A > 0$. The retailer's utility is a power function $U(x) = x^k$ for some $0 < k < 1$. The retailer is more risk averse when k is smaller.

The following proposition gives the RE equilibrium order quantity and retail price for risk-averse newsvendor in the previous situation.

Proposition 4. *Suppose that $F(x)$ is uniform distribution and $U(y) = y^k$. In the RE equilibrium, all consumers buy immediately, and the risk-averse retailer's price and quantity are*

$$p_{\text{ra}} = s + \frac{\sqrt{\left(c-s\right)^2+4k\left(c-s\right)\left(v-s\right)}-\left(c-s\right)}{2k}, \tag{16}$$

$$Q_{\text{ra}} = \frac{A\left(\sqrt{\left(c-s\right)^2+4k\left(c-s\right)\left(v-s\right)}-\left(c-s\right)\right)}{\sqrt{\left(c-s\right)^2+4k\left(c-s\right)\left(v-s\right)}+c-s}. \tag{17}$$

Proof. The RE equilibrium condition (11) and $U(x) = x^k$ $(0 < k < 1)$ reduce to

$$\begin{aligned}
&\frac{dE\left[U\left(\pi\left(Q\right)\right)\right]}{dQ} \\
&\quad = -k\left(c-s\right)\int_{((c-s)/(p-s))Q}^{Q}\left(\left(p-s\right)x-\left(c-s\right)Q\right)^{k-1}f\left(x\right)dx \\
&\qquad + k\left(p-c\right)\int_Q^{+\infty}\left(\left(p-c\right)Q\right)^{k-1}f\left(x\right)dx.
\end{aligned} \tag{18}$$

Notice that the distribution of stochastic demand ξ is uniform on $[0, A]$. Consider that $f(x) = 1/A$, and so

$$\begin{aligned}
&\frac{dE\left[U\left(\pi\left(Q\right)\right)\right]}{dQ} \\
&\quad = -k\left(c-s\right)\int_{((c-s)/(p-s))Q}^{Q}\frac{\left(\left(p-s\right)x-\left(c-s\right)Q\right)^{k-1}}{A}dx \\
&\qquad + k\left(p-c\right)\int_Q^{A}\frac{\left(\left(p-c\right)Q\right)^{k-1}}{A}dx.
\end{aligned} \tag{19}$$

According to $dE[U(\pi(Q))]/dQ = 0$, we have

$$Q = \frac{Ak\left(p-s\right)}{k\left(p-s\right)+\left(c-s\right)}. \tag{20}$$

From the RE equilibrium condition (10), we know that

$$p = v - \left(v-s\right)F\left(Q\right). \tag{21}$$

So,

$$\left(p-v\right)\left(k\left(p-s\right)+c-s\right)^2 + k^2\left(p-s\right)^2\left(v-s\right) = 0. \tag{22}$$

Let $p - v = p - s - (v - s)$. The previous equation becomes equivalently as

$$k\left(p-s\right)+\left(p-s\right)\left(c-s\right)-\left(v-s\right)\left(c-s\right) = 0. \tag{23}$$

Solving the previous equation yields the desired results. □

The impact of unit ordering cost and customer valuation on the rational equilibrium price and order quantity for risk-averse newsvendor in the case where demand follows the uniform distribution and utility is power function are summarized in the following proposition.

Proposition 5. *Suppose that $F(x)$ is uniform distribution and $U(y) = y^k$. In RE equilibrium for risk-averse newsvendor, (i) both the retail price p_{ra} and the ordering quantity Q_{ra} are increasing in customer valuation.*

(ii) The price p_{ra} in RE equilibrium for risk-averse newsvendor is increasing in ordering cost, while the ordering quantity Q_{ra} is decreasing in ordering cost.

Proof. From Proposition 4 and

$$p_{\mathrm{ra}} = s + \frac{2(v-s)}{\sqrt{1+4k((v-s)/(c-s))}+1},$$
$$Q_{\mathrm{ra}} = A\left(1 - \frac{2}{\sqrt{1+4k((v-s)/(c-s))}+1}\right), \tag{24}$$

the desired results are straight. □

This result is similar to the risk-neutral case.

The following proposition compares the RE equilibrium order quantity and price for risk-averse newsvendor with the optimal ordering quantity and retail price for risk-averse newsvendor with myopic customers in the case where demand follows the uniform distribution and utility is power function.

Proposition 6. *(i) The price p_{ra} in RE equilibrium is higher than c and lower than v.*

(ii) The ordering quantity Q_{ra} in RE equilibrium is lower than a risk-averse newsvendor model with myopic customers.

Proof. From the RE equilibrium condition (10), we know that $p_{\mathrm{ra}} < v$. The equation (16) can be reduced to

$$\frac{p_{\mathrm{ra}} - s}{c-s} = \frac{\sqrt{1+4k((v-s)/(c-s))}-1}{2k}. \tag{25}$$

To obtain the relationship between p_{ra} and c, we define a function

$$l(x) = \frac{\sqrt{1+4x((v-s)/(c-s))}-1}{2x}. \tag{26}$$

The derivative of $l(x)$ is

$$l'(x) = \frac{-1-2x((v-s)/(c-s))}{2x^2\sqrt{1+4k((v-s)/(c-s))}} < 0. \tag{27}$$

Hence,

$$\begin{aligned} l(x) &< l(0) \\ &= \lim_{k\to 0^+} \frac{\sqrt{1+4x((v-s)/(c-s))}-1}{2x} = \frac{v-s}{c-s} > 1 \end{aligned} \tag{28}$$

for $x \in (0,1)$. Therefore $p_{\mathrm{ra}} > c$.

From the proof of Proposition 3, we know that the optimal ordering quantity in the risk-averse newsvendor model with myopic customers is

$$Q_0 = \frac{Ak(v-s)}{k(v-s)+(c-s)} = A\left(1 - \frac{1}{k((v-s)/(c-s))+1}\right). \tag{29}$$

To obtain the relationship between Q_{ra} and Q_0, we define a function

$$g(x) = \frac{\sqrt{1+4x((v-s)/(c-s))}+1}{x((v-s)/(c-s))+1}. \tag{30}$$

Then we have

$$\begin{aligned} &g'(x) \\ &= \left(\frac{v-s}{c-s}\right)\left(1 - 2\left(\frac{v-s}{c-s}\right)x - \sqrt{1+4x\left(\frac{v-s}{c-s}\right)}\right) \\ &\quad \times \left(\left(x\left(\frac{v-s}{c-s}\right)+1\right)^2\sqrt{1+4x\left(\frac{v-s}{c-s}\right)}\right)^{-1} \\ &< 0. \end{aligned} \tag{31}$$

Hence, $g(x) < g(0) = 2$ for $x \in (0,1)$.

Notice that $Q_{\mathrm{ra}} = A(1-(2/\sqrt{1+4k((v-s)/(c-s))}+1))$. Therefore, $Q_{\mathrm{ra}} < Q_0$. □

The previous proposition suggests that the impact of strategic customers on risk-averse newsvendor equilibrium decisions is the same as in the risk-neutral newsvendor. In term of price, strategic consumer behavior forces the risk-averse newsvendor to price below v in order to induce the strategic consumers to purchase the product in the regular sale period. Next, in term of quantity, the equilibrium stocking quantity Q_{ra} for the risk-averse newsvendor is also lower than in the standard risk-averse newsvendor model for the purpose of increasing customers' willingness to pay.

The following proposition characterizes the impact of the level of risk aversion on the RE equilibrium price and order quantity for risk-averse newsvendor.

From (7), we know that the RE equilibrium order quantity for the risk-neutral retailer is

$$Q_{\mathrm{rn}} = A\left(1 - \sqrt{\frac{c-s}{v-s}}\right). \tag{32}$$

Proposition 7. *(i) The RE equilibrium price p_{ra} is increasing in the degree of risk aversion.*

(ii) For fixed s, v, and c, there exists a threshold $k_0 = 1 - \sqrt{(c-s)/(v-s)}$ such that the RE equilibrium price p_{ra} is not greater than p_{rn} for $k \geq k_0$.

(iii) The RE equilibrium ordering quantity Q_{ra} is decreasing in the degree of risk aversion.

(iv) For fixed s, v, and c, there exists a threshold $k_1 = k_0$ such that the RE equilibrium price Q_{ra} is not smaller than Q_{rn} for $k \geq k_1$.

Proof. The derivative of p_{ra} with respect to k can be written as

$$\begin{aligned}\frac{dp_{\mathrm{ra}}}{dk} &= \frac{c-s}{2k^2} - \frac{(c-s)^2 + 2k(c-s)(v-s)}{2k^2\sqrt{(c-s)^2 + 4k(c-s)(v-s)}} \\ &= \frac{(c-s)\sqrt{(c-s)^2 + 4k(c-s)(v-s)}}{2k^2\sqrt{(c-s)^2 + 4k(c-s)(v-s)}} \\ &\quad - \frac{(c-s)^2 + 2k(c-s)(v-s)}{2k^2\sqrt{(c-s)^2 + 4k(c-s)(v-s)}} \\ &< \frac{(c-s)^2 + 2k(c-s)(v-s)}{2k^2\sqrt{(c-s)^2 + 4k(c-s)(v-s)}} \\ &\quad - \frac{(c-s)^2 + 2k(c-s)(v-s)}{2k^2\sqrt{(c-s)^2 + 4k(c-s)(v-s)}} = 0.\end{aligned} \tag{33}$$

Therefore, p_{ra} is decreasing in k.

Note that

$$\begin{aligned}&\lim_{k\to 0^+} p_{\mathrm{ra}} \\ &\quad = \lim_{k\to 0^+} \frac{2(c-s)(v-s)}{\sqrt{(c-s)^2 + 4k(c-s)(v-s)} + c - s} + s = v, \\ &\lim_{k\to 1^-} p_{\mathrm{ra}} \\ &\quad = s + \frac{\sqrt{(c-s)^2 + 4(c-s)(v-s)} - (c-s)}{2} \\ &\quad = s + \frac{2\sqrt{(c-s)(v-s)}}{\sqrt{((c-s)/(v-s)) + 4} + \sqrt{(c-s)/(v-s)}} \\ &\quad < s + \sqrt{(c-s)(v-s)} = p_{\mathrm{rn}},\end{aligned} \tag{34}$$

we know that there exists k_0 such that the RE equilibrium price p_{ra} is not greater than p_{rn} for $k \geq k_0$ and not smaller than p_{rn} for $k \leq k_0$. Solving the equation $p_{\mathrm{ra}} = p_{\mathrm{rn}}$ yields the k_0.

To obtain the relationship between Q_{ra} and k, we define a function

$$h(k) = \sqrt{(c-s)^2 + 4k(c-s)(v-s)}. \tag{35}$$

Taking derivative of Q_{ra} with k leads to

$$\frac{dQ_{\mathrm{ra}}}{dk} = \frac{2Ah'(k)(c-s)}{(h(k) + c - s)^2}. \tag{36}$$

Since

$$h'(k) = \frac{2(c-s)(v-s)}{h(k)}, \tag{37}$$

we have

$$\frac{dQ_{\mathrm{ra}}}{dk} > 0. \tag{38}$$

Hence, Q_{ra} is increasing in k. Moreover, since $\lim_{k\to 0^+} Q_{\mathrm{ra}} = 0$ and

$$\begin{aligned}\lim_{k\to 1^-} Q_{\mathrm{ra}} &= \frac{A\left(\sqrt{1 + 4((v-s)/(c-s))} - 1\right)}{\sqrt{1 + 4((v-s)/(c-s))} + 1} \\ &= A\left(1 - \frac{2}{\sqrt{1 + 4((v-s)/(c-s))} + 1}\right) \\ &> A\left(1 - \sqrt{\frac{c-s}{v-s}}\right) = Q_{\mathrm{rn}},\end{aligned} \tag{39}$$

there exists k_1 such that the RE equilibrium ordering quantity Q_{ra} is not smaller than Q_{rn} for $k \geq k_1$ and not greater than Q_{rn} for $k \leq k_1$. Solving the equation $Q_{\mathrm{ra}} = Q_{\mathrm{rn}}$ yields the $k_1 = k_0$. □

Eeckhoudt et al. [20] investigated the classic newsvendor model with increasing, concave, and thrice differentiable utility functions and showed that the optimal ordering quantity in their environment is smaller than the risk-neutral optimal ordering quantity and decreases in risk-aversion level. Keren and Pliskin [21] derived the first order conditions for optimality for the problem of a risk-averse expected-utility maximizer newsvendor. They solved a special case where the utility function is any increasing differentiable function, and the random demand is uniformly distributed by this condition. They also claimed that a risk-averse newsvendor with a uniformly more concave utility function sets his order quantity to a lower value than a less risk-averse newsvendor, so a risk-averse newsvendor with a concave utility function sets order quantity less than a newsvendor who is risk neutral. From Proposition 7, we find that this may not be true when the strategic consumer behavior is incorporated into the classic newsvendor model. The RE equilibrium order quantity for some risk-averse newsvendor with the uniform distribution demand and power utility will order more than the risk-neutral newsvendor. In addition, together with the results in Proposition 6 we know that the optimal ordering quantity for some risk-averse newsvendor facing myopic consumers is higher than in risk-neutral retailer.

In Agrawal and Seshadri [22] it is assumed that the customers are myopic and if the realized demand is greater than the ordered quantity, the retailer can make emergency orders at a higher price ($>c$) to meet the extra demand. And if the realized demand is less than the order quantity, the leftover inventory can be salvaged at a value that is not more than the cost. So the risk is mainly stemmed from overstock and some emergency purchase cost occurs for larger demand. Agrawal and Seshadri [22] showed that for the multiplicative model, the optimal price for a risk-averse retailer is not lower than that for a risk-neutral retailer and the optimal order quantity for a risk-averse retailer is smaller than that for a risk-neutral retailer. For the additive demand model, Agrawal and Seshadri [22] claimed that the optimal price is lower

when the retailer is more risk averse, and the impact of risk aversion on the optimal order quantity is unclear. In our model, the equilibrium retail price is larger when the retailer is more risk averse, and the equilibrium order quantity is smaller when the retailer is more risk averse when he or she faces strategic consumers. The RE equilibrium retail price for some risk-averse newsvendors is lower than for the risk-neutral newsvendors.

The impacts of salvage value on the RE equilibrium price and order quantity for risk-averse newsvendor in the case where demand follows the uniform distribution and utility is power function are summarized in the following proposition.

Proposition 8. *(i) The RE equilibrium retail price p_{ra} is decreasing in salvage price when $v \geq (4k+2)c$, and there is a threshold s_0 such that p_{ra} is decreasing in s for $s_0 < s < c$ and is increasing in s for $0 < s \leq s_0$ when $v < (4k+2)c$.*

(ii) The RE equilibrium ordering quantity Q_{ra} is increasing in salvage value.

Proof. The first and second derivatives of p_{ra} with respect to s are, respectively,

$$\begin{aligned}
\frac{dp_{\mathrm{ra}}}{ds} &= 1 + \frac{1}{2k} - \frac{(2k+1)(c-s) + 2k(v-s)}{2k\sqrt{(c-s)^2 + 4k(c-s)(v-s)}}, \\
\frac{d^2 p_{\mathrm{ra}}}{ds^2} &= \frac{4k+1}{2k\sqrt{(c-s)^2 + 4k(c-s)(v-s)}} \\
&\quad - \frac{[(2k+1)(c-s) + 2k(v-s)]^2}{2k\left[(c-s)^2 + 4k(c-s)(v-s)\right]^{3/2}} \\
&= -\frac{4k^2(v-c)^2}{2k\left[(c-s)^2 + 4k(c-s)(v-s)\right]^{3/2}} < 0.
\end{aligned} \tag{40}$$

Hence, p_{ra} is concave in s. Notice that

$$\begin{aligned}
\lim_{s\to c^-} \frac{dp_{\mathrm{ra}}}{ds} &= 1 + \frac{1}{2k} \\
&\quad - \lim_{s\to c^-} \frac{(2k+1)(c-s) + 2k(v-s)}{2k\sqrt{(c-s)^2 + 4k(c-s)(v-s)}} = -\infty, \\
\lim_{s\to 0^+} \frac{dp_{\mathrm{ra}}}{ds} &= 1 + \frac{1}{2k} - \lim_{s\to 0^+} \frac{(2k+1)(c-s) + 2k(v-s)}{2k\sqrt{(c-s)^2 + 4k(c-s)(v-s)}} \\
&= 1 + \frac{1}{2k} - \frac{(2k+1)c + 2kv}{2k\sqrt{c^2 + 4kcv}} \\
&= 2k\left((4k+2)cv - v^2\right) \\
&\quad \times \Big(\sqrt{c^2 + 4kcv} \\
&\quad \times \left((2k+1)\sqrt{c^2+4kcv} + ((2k+1)c + 2kv)\right)\Big)^{-1}.
\end{aligned} \tag{41}$$

Therefore $(dp_{\mathrm{ra}}/ds) < 0$ for $v \geq (4k+2)c$, and p_{ra} is decreasing in s for any $s \in (0, c)$. $\lim_{s\to 0^+}(dp_{\mathrm{ra}}/ds) > 0$ for $v < (4k+2)c$. Note that p_{ra} is concave in s. We know that there exists a point s_0 in $(0, c)$ such that $(dp_{\mathrm{ra}}/ds) < 0$ for $s \in (s_0, c)$ and $(dp_{\mathrm{ra}}/ds) > 0$ for $s \in (0, s_0)$. It is easy to see that p_{ra} is decreasing in $s \in (s_0, c)$ and p_{ra} is increasing in $s \in (0, s_0)$.

To obtain the relationship between Q_{ra} and s, we define

$$g(s) = \sqrt{(c-s)^2 + 4k(c-s)(v-s)}. \tag{42}$$

Then we have

$$g'(s) = -\frac{(2k+1)(c-s) + 2k(v-s)}{g(s)}. \tag{43}$$

Notice that

$$\frac{dQ_{\mathrm{ra}}}{ds} = \frac{2A\left(g(s) + g'(s)(c-s)\right)}{\left(g(s) + c - s\right)^2}. \tag{44}$$

We have

$$\begin{aligned}
&\frac{dQ_{\mathrm{ra}}}{ds} \\
&= \frac{2A\left(g^2(s) - (2k+1)(c-s)^2 - 2k(v-s)(c-s)\right)}{g(s)\left(g(s) + c - s\right)^2} \\
&= \frac{4kA\left((c-s)(v-s) - (c-s)^2\right)}{g(s)\left(g(s) + c - s\right)^2} > 0.
\end{aligned} \tag{45}$$

So, Q_{ra} is increasing in s on $(0, c)$. □

Proposition 8 characterizes the impacts of salvage value on equilibrium price and quantity. Chen et al. [26] used conditional value at risk as a risk-averse measure to investigate the combined pricing and ordering problem for newsvendor facing myopic customers. They showed that for the multiplicative demand, the optimal price for a risk-averse retailer is strictly increasing in s under some mild conditions. For the additive demand, they claimed that the optimal price is strictly increasing in s without any condition. On the other hand, Proposition 2 presents that the RE equilibrium price for risk-neutral newsvendor is strictly increasing in s. Contrary to the previous results, our findings show that risk-averse newsvendor is not always setting higher RE equilibrium retail price for larger salvage price. The relationship between them is dependent on other system parameters such as product cost, customer valuation, and risk-averse degree. Specifically, RE equilibrium price p_{ra} is decreasing in salvage price when the product cost c is sufficiently low or when the customer valuation v is sufficiently high. Otherwise, there exists a threshold for s. The RE equilibrium retail price is decreasing in salvage price when the salvage price is higher than the threshold and is increasing in salvage price otherwise. That is, to say that the RE equilibrium retail price for the risk-averse newsvendor in our model is increasing in salvage price only for the situation in which the product cost is sufficiently high, customer valuation is sufficiently low relatively to product

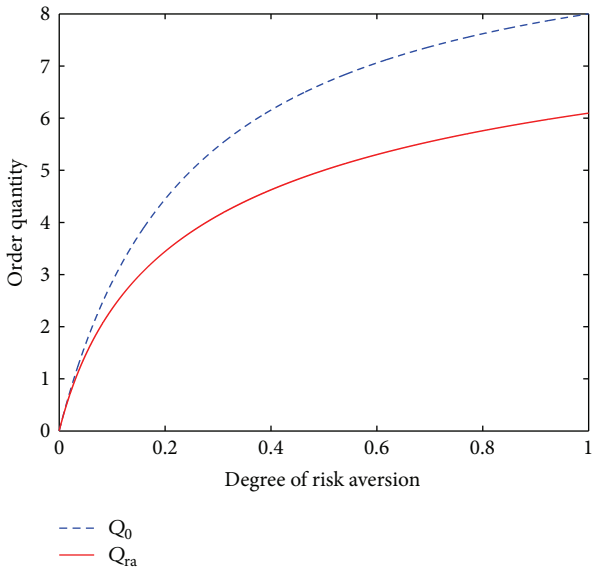

FIGURE 1: The effect of risk-averse level on Q_0 and Q_{ra}.

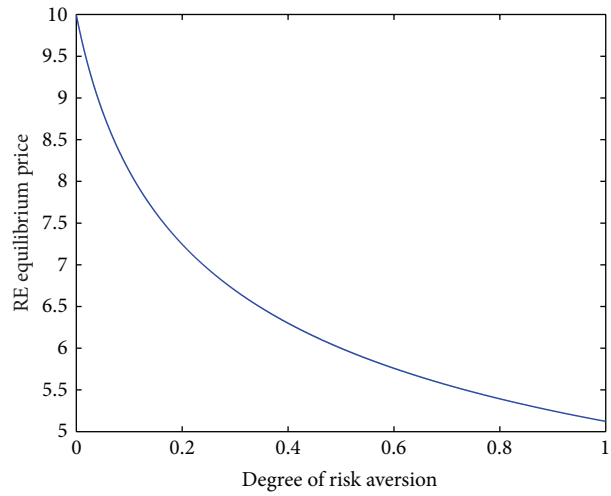

FIGURE 2: The effect of risk-averse level on p_{ra}.

cost, and salvage price is low. In addition, there exists a threshold for risk-averse level k for reasonable fixed v and c. That is, the RE equilibrium retail price is decreasing in salvage price for more risk-averse retailer, and the relation between the RE equilibrium sale price and salvage price for a less risk-averse retailer depends on the threshold for s.

4. Numerical Examples

In this section, we present some numerical examples to illustrate how the RE equilibrium price and quantity are affected by the degree of risk-aversion and salvage price.

Firstly, we analyze the impact of risk-averse level on RE equilibrium quantity and price. Suppose that random demand ξ is uniform distribution on $[0, 10]$, $v = 10$, $c = 4$, and $s = 2$. Then $Q_{rn} = 5$ and $p_{rn} = 6$. According to (17), (20), and (32), we know that

$$p_{ra} = 2 + \frac{\sqrt{1+16k}-1}{k},$$
$$Q_{ra} = \frac{5\left(\sqrt{1+16k}-1\right)^2}{8k}, \tag{46}$$
$$Q_0 = \frac{40k}{4k+1}.$$

From Figures 1 and 2, we find that $Q_0 > Q_{ra}$ and $p_{ra} < v$ for $k \in (0, 1)$ so as to induce the customers to purchase the product by lowering the sale price and generating scarcity and competition among them through restricting the availability of the product and maintaining an image of exclusivity. In addition, Q_0 and Q_{ra} are increasing in k, and p_{ra} is decreasing in k.

Solving $p_{ra} = p_{rn}$, $Q_{ra} = Q_{rn}$, and $Q_0 = Q_{rn}$, we get $k_0 = 0.5$, $k_1 = 0.5$, and $k_2 = 0.25$. So p_{ra} is lower than p_{rn} for $k \in (0.5, 1)$, and Q_{ra} is higher than Q_{rn} for $k \in (0.5, 1)$. Furthermore, Q_{ra} is higher than Q_{rn} for $k \in (0.25, 1)$. These results are contrary to usual results for risk-averse newsvendor relative to risk-neutral retailer when he faces myopic newsvendor.

In what follows, we investigate the influence of salvage price on the RE equilibrium price and quantity with the same parameters as mentioned previously. According to (16) and (17), we know that

$$p_{ra} = s + \frac{\sqrt{(4-s)^2 + 4k(4-s)(10-s)} - (4-s)}{2k},$$
$$Q_{ra} = \frac{10\left(\sqrt{(4-s)^2 + 4k(4-s)(10-s)} - (4-s)\right)}{\sqrt{(4-s)^2 + 4k(4-s)(10-s)} + 4 - s}. \tag{47}$$

Solving the equation $v = (4k+2)c$, we get $k^* = 0.125$. Let $k = 0.1$, $k = 0.05$, and $k = 0.025$ in Figure 3; we find that the RE equilibrium price is decreasing in salvage price. Let $k = 0.2$, $k = 0.6$, and $k = 0.8$ in Figure 4; we find that the RE equilibrium price is firstly increasing in salvage price and then decreasing. In Figure 5, we observe that the RE equilibrium quantity is decreasing in salvage price.

In addition, we notice that s_0 is increasing in k from Figure 4. Letting $(dp_{ra}/ds) = 0$ where $k > k^*$, we can obtain the following equation:

$$1 + \frac{1}{2k} - \frac{(2k+1)(4-s_0) + 2k(10-s_0)}{2k\sqrt{(4-s_0)^2 + 4k(4-s_0)(10-s_0)}} = 0. \tag{48}$$

Its solution is $s_0 = 4 - 6/(4k+1)$. So, the changes of s_0 with k is consistent with our observation. That is to say that the range

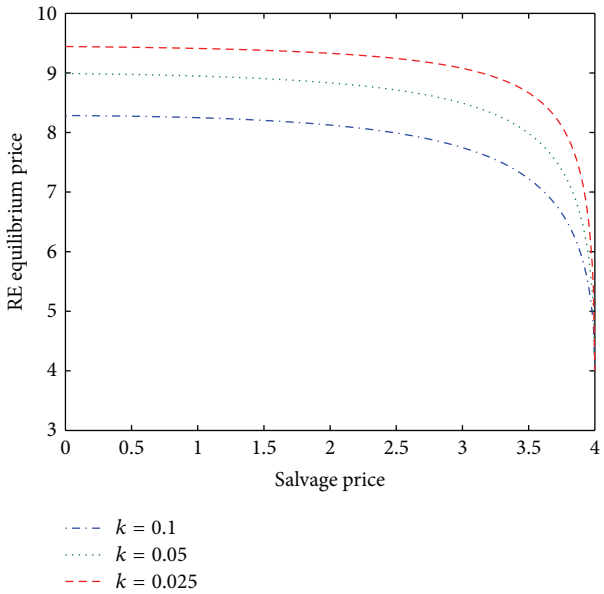

FIGURE 3: The effect of salvage price on p_{ra} when $k \leq k^*$.

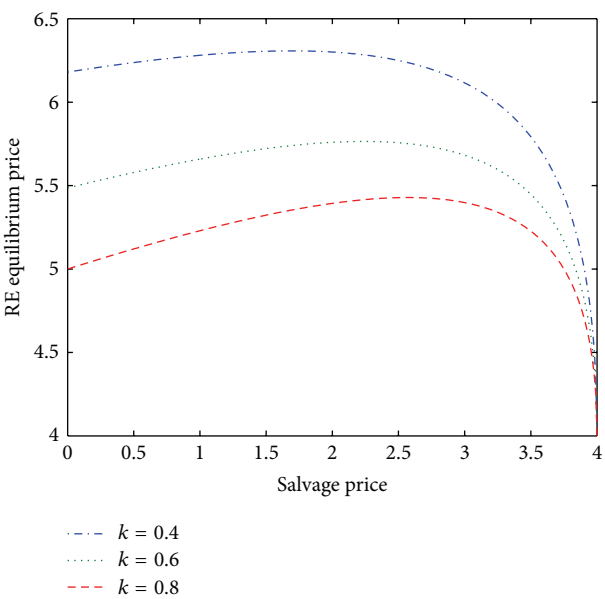

FIGURE 4: The effect of salvage price on p_{ra} when $k > k^*$.

is more narrow for more risk-averse retailer in which the RE equilibrium price is increasing in salvage price.

5. Conclusions

This paper investigates the RE equilibrium decisions for the risk-averse newsvendor facing strategic consumers. One of our basic premises is that the newsvendor is risk averse another is that consumers look ahead and plan purchases with future opportunities in mind. To derive structural results and generate managerial insights into the equilibrium decisions of the risk-averse newsvendor problem with strategic consumers, we present specific results for the situation in which demand is a uniformly distributed and utility is a general power function. We obtain the analytical representations for the RE equilibrium price and quantity and perform the sensitivity analysis to investigate the impacts of ordering cost, salvage price, and degree of risk-aversion on the equilibrium. Our analysis complements the work in the literature and offers managerial insights to the practitioners or managers. We show that the optimal ordering quantity for some risk-averse newsvendor facing myopic consumers is higher than the RE equilibrium ordering quantity for a risk-neutral retailer, and the RE equilibrium retail price for some risk-averse newsvendor is lower than the RE equilibrium price for risk-neutral retailer, and the stocking quantity in RE equilibrium for some risk-averse newsvendor is higher than the RE equilibrium price for risk-neutral retailer. We also find the conditions under which the sale price in RE equilibrium for risk-averse newsvendor is decreasing in salvage price.

The problem studied here can further be researched in several directions. First, it is meaningful to make clear of the value of price commitment and quantity commitment for risk-averse newsvendor facing strategic consumer and the impacts of contracts on decentralized supply chain with risk-averse newsvendor facing forward-looking customers. Second, it is important to investigate the combined effects of risk-averse level of consumer, consumer's forward-looking purchasing behavior, and degree of risk aversion for newsvendor. Finally, investigating the optimal decisions for risk-averse newsvendor facing strategic consumers in competitive setting would be a very interesting topic.

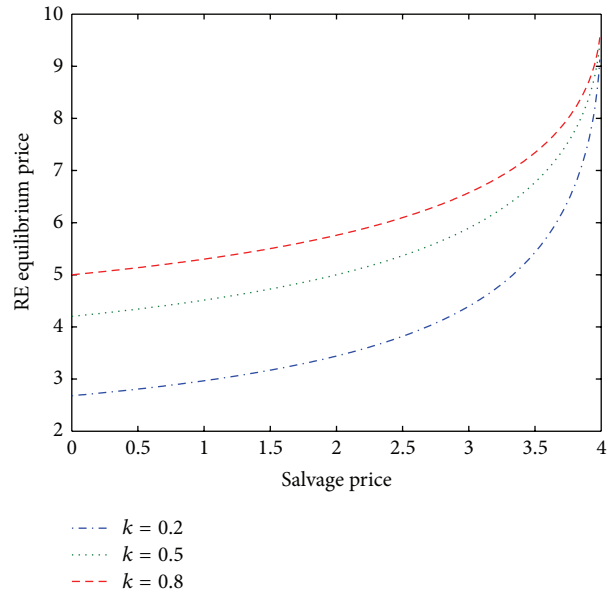

FIGURE 5: The effect of salvage price on Q_{ra}.

Acknowledgment

This work is partially supported by National Natural Science Foundation of China (no. 71271059).

References

[1] M. E. Schweitzer and G. P. Cachon, "Decision bias in the newsvendor problem with a known demand distribution: experimental evidence," *Management Science*, vol. 46, no. 3, pp. 404–421, 2000.

[2] J. F. Muth, "Rational expectations and the theory of price movements," *Econometrica*, vol. 29, pp. 315–335, 1961.

[3] M. Khouja, "The single-period (news-vendor) problem: literature review and suggestions for future research," *Omega*, vol. 27, no. 5, pp. 537–553, 1999.

[4] N. C. Petruzzi and M. Dada, "Pricing and the newsvendor problem: a review with extensions," *Operations Research*, vol. 47, no. 2, pp. 183–194, 1999.

[5] L. Yao, Y. Chen, and H. Yan, "The newsvendor problem with pricing: extensions," *International Journal of Management Science and Engineering Management*, vol. 1, pp. 3–16, 2006.

[6] M. A. Lariviere and E. L. Porteus, "Selling to the newsvendor: an analysis of price-only contracts," *Manufacturing and Service Operations Management*, vol. 3, no. 4, pp. 293–305, 2001.

[7] Y. Gerchak and Y. Wang, "Periodic-review inventory models with inventory-level-dependent demand," *Naval Research Logistics*, vol. 41, no. 1, pp. 99–116, 1994.

[8] A. Balakrishnan, M. S. Pangburn, and E. Stavrulaki, "'Stack them high, let 'em fly': lot-sizing policies when inventories stimulate demand," *Management Science*, vol. 50, no. 5, pp. 630–644, 2004.

[9] J. D. Dana Jr. and N. C. Petruzzi, "Note: the newsvendor model with endogenous demand," *Management Science*, vol. 47, no. 11, pp. 1488–1497, 2001.

[10] A. Balakrishnan, M. S. Pangburn, and E. Stavrulaki, "Integrating the promotional and service roles of retail inventories," *Manufacturing and Service Operations Management*, vol. 10, no. 2, pp. 218–235, 2008.

[11] S. Liu, K. C. So, and F. Zhang, "Effect of supply reliability in a retail setting with joint marketing and inventory decisions," *Manufacturing and Service Operations Management*, vol. 12, no. 1, pp. 19–32, 2010.

[12] T. A. Taylor, "Supply chain coordination under channel rebates with sales effort effects," *Management Science*, vol. 48, no. 8, pp. 992–1007, 2002.

[13] H. Krishnan, R. Kapuscinski, and D. A. Butz, "Coordinating contracts for decentralized supply chains with retailer promotional effort," *Management Science*, vol. 50, no. 1, pp. 48–63, 2004.

[14] Y. Gerchak and M. Parlar, "A Single period inventory problem with partially controllable demand," *Computers and Operations Research*, vol. 14, no. 1, pp. 1–9, 1987.

[15] M. Khouja and S. S. Robbins, "Linking advertising and quantity decisions in the single-period inventory model," *International Journal of Production Economics*, vol. 86, no. 2, pp. 93–105, 2003.

[16] S.-D. Wang and Y.-W. Zhou, "Supply chain coordination models for newsvendor-type products: considering advertising effect and two production modes," *Computers and Industrial Engineering*, vol. 59, pp. 220–231, 2011.

[17] S.-D. Wang, Y.-W. Zhou, and J.-P. Wang, "Supply chain coordination with two production modes and random demand depending on advertising expenditure and selling price," *International Journal of Systems Science*, vol. 41, no. 10, pp. 1257–1272, 2010.

[18] S.-D. Wang, Y.-W. Zhou, and J.-P. Wang, "Coordinating ordering, pricing and advertising policies for a supply chain with random demand and two production modes," *International Journal of Production Economics*, vol. 126, no. 2, pp. 168–180, 2010.

[19] H. S. Lau, "The newsboy problem under alternative optimization objectives," *Operational Research Society Journal*, vol. 31, no. 6, pp. 525–535, 1980.

[20] L. Eeckhoudt, C. Gollier, and H. Schlesinger, "The risk-averse (and prudent) newsboy," *Management Science*, vol. 41, pp. 786–794, 1995.

[21] B. Keren and J. S. Pliskin, "A benchmark solution for the risk-averse newsvendor problem," *European Journal of Operational Research*, vol. 174, no. 3, pp. 1643–1650, 2006.

[22] V. Agrawal and S. Seshadri, "Impact of uncertainty and risk aversion on price and order quantity in the newsvendor problem," *Manufacturing and Service Operations Management*, vol. 2, no. 4, pp. 410–423, 2000.

[23] F. Chen and A. Federgruen, *Mean-Variance Analysis of Basic Inventory Models. Working Paper*, Columbia University, New York, NY, USA, 2000.

[24] T. Choi, D. Li, and H. Yan, "Mean-variance analysis for the newsvendot problem," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 38, pp. 80–96, 2008.

[25] J.-Y. Gotoh and Y. Takano, "Newsvendor solutions via conditional value-at-risk minimization," *European Journal of Operational Research*, vol. 179, no. 1, pp. 80–96, 2007.

[26] Y. Chen, M. Xu, and Z. G. Zhang, "A risk-averse newsvendor model under the CVaR criterion," *Operations Research*, vol. 57, no. 4, pp. 1040–1044, 2009.

[27] S. Ahmed, U. Çakmak, and A. Shapiro, "Coherent risk measures in inventory problems," *European Journal of Operational Research*, vol. 182, no. 1, pp. 226–238, 2007.

[28] F. Gino and G. Pisano, "Toward a theory of behavioral operations," *Manufacturing and Service Operations Management*, vol. 10, no. 4, pp. 676–691, 2008.

[29] E. Bendoly, R. Croson, P. Goncalves, and K. Schultz, "Bodies of knowledge for research in behavioral operations," *Production and Operations Management*, vol. 19, no. 4, pp. 434–452, 2010.

[30] Z.-J. M. Shen and X. Su, "Customer behavior modeling in revenue management and auctions: a review and New Research Opportunities," *Production and Operations Management*, vol. 16, no. 6, pp. 713–728, 2007.

[31] X. Su, "Bounded rationality in newsvendor models," *Manufacturing and Service Operations Management*, vol. 10, no. 4, pp. 566–589, 2008.

[32] X. Su, "Intertemporal pricing with strategic customer behavior," *Management Science*, vol. 53, no. 5, pp. 726–741, 2007.

[33] X. Su and F. Zhang, "Strategic customer behavior, commitment, and supply chain performance," *Management Science*, vol. 54, no. 10, pp. 1759–1773, 2008.

[34] X. Su and F. Zhang, "On the value of commitment and availability guarantees when selling to strategic consumers," *Management Science*, vol. 55, no. 5, pp. 713–726, 2009.

[35] G. Lai, L. G. Debo, and K. Sycara, "Buy now and match later: impact of posterior price matching on profit with strategic consumers," *Manufacturing and Service Operations Management*, vol. 12, no. 1, pp. 33–55, 2010.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 979862, 13 pages
http://dx.doi.org/10.1155/2013/979862

Research Article

Internal Due Date Assignment in a Wafer Fabrication Factory by an Effective Fuzzy-Neural Approach

Toly Chen

Department of Industrial Engineering and Systems Management, Feng Chia University, 100 Wenhwa Road, Seatwen, Taichung City 407, Taiwan

Correspondence should be addressed to Toly Chen; tcchen@fcu.edu.tw

Received 26 November 2012; Revised 26 August 2013; Accepted 8 September 2013

Academic Editor: Hadi Nasseri

Owing to the complexity of the wafer fabrication, the due date assignment of each job presents a challenging problem to the production planning and scheduling people. To tackle this problem, an effective fuzzy-neural approach is proposed in this study to improve the performance of internal due date assignment in a wafer fabrication factory. Some innovative treatments are taken in the proposed methodology. First, principal component analysis (PCA) is applied to construct a series of linear combinations of the original variables to form a new variable, so that these new variables are unrelated to each other as much as possible, and the relationship among them can be reflected in a better way. In addition, the simultaneous application of PCA, fuzzy c-means (FCM), and back propagation network (BPN) further improved the estimation accuracy. Subsequently, the iterative upper bound reduction (IUBR) approach is proposed to determine the allowance that will be added to the estimated job cycle time. An applied case that uses data collected from a wafer fabrication factory illustrates this effective fuzzy-neural approach.

1. Introduction

Internal due date assignment is to quote an attractive but attainable due date for an arriving customer order. However, the completion time of an order is highly uncertain. It is therefore difficult to accurately forecast the completion time. For this reason, an allowance has to be added to the estimated completion time to reduce the risk [1].

Wafer fabrication is the most technologically complex step in semiconductor manufacturing, which exacerbates the difficulties of internal due date assignment [2]. In theory, this problem is NP-hard. That is why wafer fabrication is investigated in this study. Internal due date assignment in a wafer fabrication factory is difficult because of the following reasons.

(1) Shop floor control in a wafer fabrication factory is a nontrivial task owing to the complexity of wafer fabrication. Some wafer fabrication processes are repeated processes. Thus, wafers need to visit a machine multiple times. An average job cycle time is several months with hundreds of hours of standard deviation. Many studies have shown that accurately predicting the cycle/completion times for such large systems is very difficult [1, 3, 4].

(2) In addition, the completion time predicted using existing approaches is generally unbiased. This means that if the internal due date is set to be equal to the mean of the estimated completion time, then the probability of on-time delivery is only about 50% on average. To reduce the risk, an allowance or fudge factor has to be added to the estimated completion time [5]. The due date allowance factor is determined on the basis of the feedback information about the factory status at the time a job arrives at the factory.

(3) Due date assignment, release control, and buffer control affect each other. Make-to-order wafer fabrication factories are confronted with both due date quotation and production scheduling problems at the same time [6]. If due date assignment and factory scheduling are processed separately by two systems, the overall performance is unlikely to be satisfactory because the two tasks are actually interrelated. Therefore, the interaction between due date assignment methods

and scheduling rules in a wafer fabrication factory needs to be investigated.

To tackle these problems, some treatments have been carried out in the literature. First, various research works have been dedicated to estimate the cycle time using hybrid approaches. For example, Gupta and Sivakumar [7] presented look-ahead batch scheduling for the real-time control of due date objectives. Chen [8] proposed the look-ahead self-organization map (SOM)-fuzzy back propagation network (FBPN) approach for this purpose. A set of fuzzy inference rules were also developed to evaluate the achievability of a cycle time forecast. Subsequently, Chen et al. [1] added a selective allowance to the cycle time estimated using the look-ahead SOM-FBPN approach to determine the internal due date. Further, Chen [9] showed that the combination of SOM and FBPN could be improved by a minor adjustment of the classification results with the estimation error. Chen et al. [10, 11] proposed a postclassification fuzzy-neural approach in which a job was not preclassified but rather postclassified after estimating the cycle time. Experimental results showed that the postclassification approach was better than the preclassification approaches in some cases. To balance the influence of the preclassification results with that of the postclassifying results, Chen [12] proposed a bidirectional classifying approach, in which jobs are not only preclassified but also postclassified. Ankenman et al. [13] proposed a metamodeling approach, which integrates discrete-event simulation, adaptive statistical methods, and analytical queueing analysis to quantify the cycle time-throughput relationship. Chien et al. [4] used nonlinear regression equations and then related the forecasting error to some factory conditions and job attributes with a back propagation network (BPN) to improve the forecasting accuracy. The major disadvantage of statistical analysis is the lack of forecasting accuracy [8].

Second, in traditional due date setting rules, the fudge factor is usually equal to a multiple of the standard deviation of the predicted cycle time [14]. Recently, Chen et al. [1] proposed a selective allowance policy in which the allowance was only assigned to some preselected jobs. In this way, the sum of the allowances added to all jobs was controlled. However, even though the probability of on-time delivery in Chen et al.'s study was only 77% for the testing data, showing that improving the probability of on-time delivery while controlling the fudge factor is a real challenge. In addition, the allowances that were assigned to the chosen jobs in this study were equal, leaving room for improvement. Another way of taking this issue into account is to construct a confidence interval containing the actual completion time [3]. The upper confidence limit sets the internal due date. However, the probability of a job delivered on time is only 99.7% for the testing data, under the assumption that residuals follow a normal distribution. From another point of view, Chen and Wang [15] incorporated the fuzzy c-means (FCM)-BPN approach with a nonlinear programming (NLP) model to construct the inclusion interval of the predicted completion time. Similarly, the upper inclusion limit sets the internal due date. An inclusion interval is narrower than a confidence interval, and the probability of a job delivered on time is 100%, at least for the training data. Chen and Lin [16] modified this approach by gathering a group of experts in related fields to set the due date in a collaborative way. Fuzzy intersection is applied to combine the due dates into a representative value.

The existing approaches have the following problems.

(1) Some factors used to forecast the job cycle time are dependent on each other, which may cause problems in classifying jobs and in fitting the relationship between the job cycle time and these factors.

(2) In Chen and Wang [15] and Chen and Lin [16], NLP models are solved to determine the upper bound of the job cycle time. However, the NLP models involve complicated constraints and therefore are difficult to solve. The NLP models will become too huge if many jobs are to be considered.

To tackle these problems, an effective fuzzy-neural approach is proposed in this study to improve the performance of internal due date assignment in a wafer fabrication factory. The literature provides probabilistic (stochastic) and fuzzy methods that can consider the uncertainty or randomness in the completion time. However, the longest average cycle time exceeds three months with a variation of more than 300 hours. Fitting the cycle time within a future month with a distribution function is not easy, implying that a stochastic approach might not be applicable. That is why a fuzzy approach is proposed in this study.

The effective fuzzy-neural approach has the following innovative characteristics.

(1) Variable replacement using principal component analysis (PCA): PCA uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables to reflect information in a better way.

(2) Updating the upper bound of the job cycle time using the iterative upper bound reduction (IUBR) approach: the IUBR approach is proposed to determine the upper bound of the completion time forecast. A tight upper bound means that the allowance assigned to a job is minimized.

Some recent works in this field are relevant. The differences between the proposed methodology and these methods are summarized in Table 1.

The remainder of this paper is organized as follows. Section 2 introduces the proposed methodology which is composed of four steps. A practical example is used to validate the effectiveness of the proposed methodology. The performance of the proposed methodology is evaluated and compared with those of some existing approaches. Finally, the concluding remarks and some directions for future research are given in Section 4.

2. Methodology

The operating procedure of the effective fuzzy-neural approach consists of several steps that will be described in the following sections.

Table 1: The differences between this study and some recent references.

Method	Data preprocessing method	Forecasting method	Upper bound	Optimization method	Computation complexity
Chen et al. (2008) [1]	SOM	FBPN	Yes	Simulation + fuzzy rules	High
Chien et al. (2011) [4]	No	Nonlinear regression	No	BPN	Low
Chen and Lin (2011) [16]	FCM	BPN	Yes	NLP	Very high
Chen and Wang (2013) [23]	PCA + FCM	BPN	No	An iterative process to reduce outliers	Low
The proposed methodology	PCA + FCM	BPN	Yes	IUBR	Low

NLP: nonlinear programming.

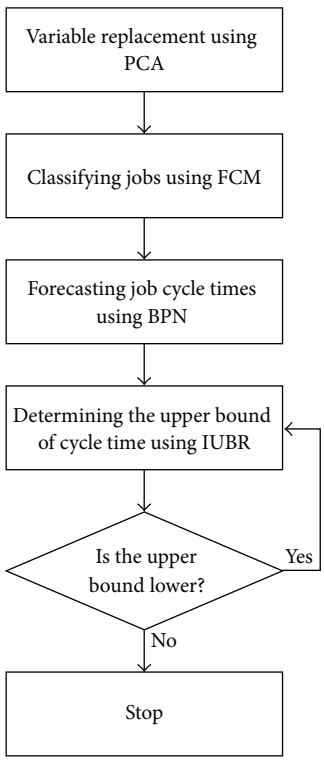

Figure 1: The flowchart of the proposed methodology.

Step 1. Forming new variables by constructing linear combinations of the original variables using PCA.

Step 2. Classifying jobs using fuzzy c-means (FCM).

Step 3. Forecasting the cycle times of jobs in each category using a BPN.

Step 4. Determining the upper bound of the cycle time using the IUBR approach.

A flow chart of the proposed methodology is shown in Figure 1.

2.1. Step 1: Forming New Variables Using PCA. First, PCA is used to replace the inputs to the BPN. PCA was invented by Pearson [18] as an analogue of the principal axes theorem in mechanics; it was later independently developed by Hotelling [19]. In the literature, there are more advanced applications of PCA. For example, Jaiswal et al. [20] used a hybrid of PCA and partial least squares for face recognition. In Mohtasham et al. [21], linear and exponential weighted PCA techniques based on spectral similarity were employed to predict the dye concentration in coloured fabrics. The operating procedure of PCA consists of several steps that are illustrated in Figure 2.

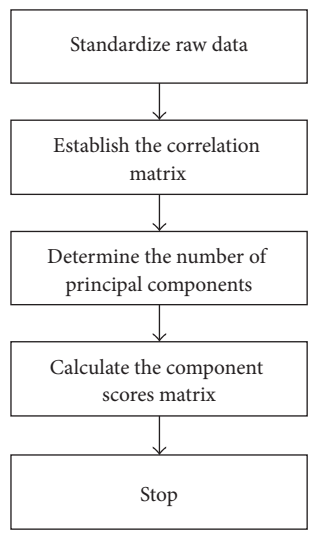

Figure 2: The PCA process.

The references on the combination of PCA, FCM, and BPN are still very limited [17, 22, 23].

2.2. Step 2: Classifying Jobs Using FCM. After employing PCA, examples are then classified using FCM. FCM is one of the most popular fuzzy clustering techniques because it is efficient, straightforward, and easy to implement. However, FCM is sensitive to initialization and is easily trapped in local optima.

The objective function of FCM is to minimize the weighted sum of squared distances such that the jobs in a category will be similar (or related) to one another and

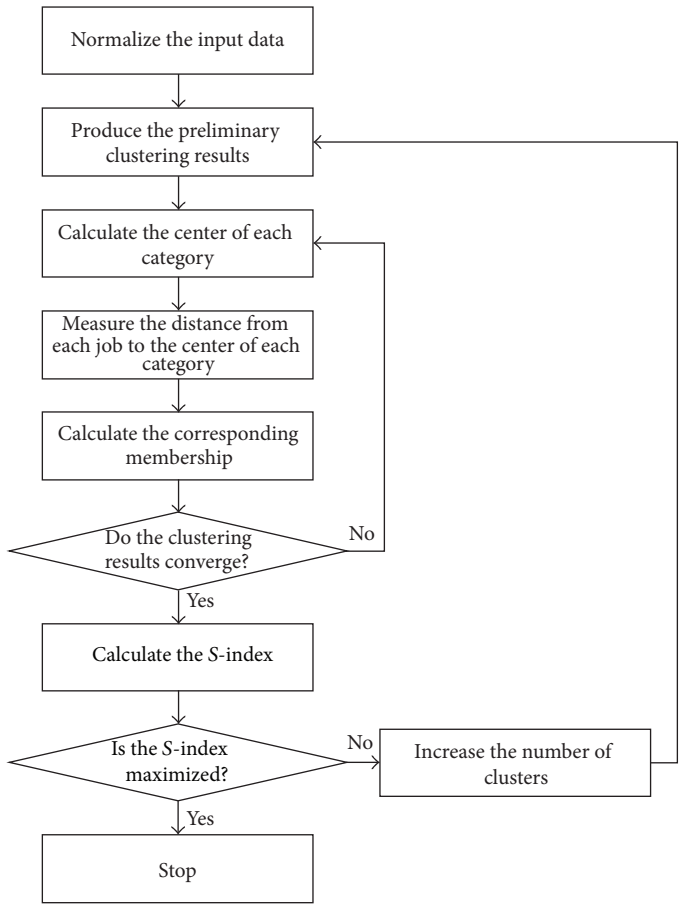

Figure 3: The FCM procedure.

different from (or unrelated to) the jobs in other categories. In FCM, the Euclidean distance between two jobs is measured:

$$e_{j(k)} = \sqrt{\sum_{q=1}^{p} \left(z_{jp} - \overline{z}_{(k)p}\right)^2}, \quad j = 1 \sim n, \ k = 1 \sim K, \tag{1}$$

where

$$\begin{aligned} \overline{z}_{(k)} &= \left\{\overline{z}_{(k)q}\right\}, \quad k = 1 \sim K, \\ \overline{z}_{(k)q} &= \frac{\sum_{j=1}^{n} \mu_{j(k)}^{m} z_{jq}}{\sum_{j=1}^{n} \mu_{j(k)}^{m}}, \quad k = 1 \sim K, \ q = 1 \sim p. \end{aligned} \tag{2}$$

The weight of a job is a function of its membership:

$$\mu_{j(k)} = \frac{1}{\sum_{g=1}^{K} \left(e_{j(k)}/e_{j(g)}\right)^{2/(m-1)}}, \quad j = 1 \sim n, \quad k = 1 \sim K. \tag{3}$$

However, FCM requires prior knowledge about the number of clusters in the data, which may not be known for new data. Then, fuzzy clustering is carried out through an iterative optimization of the objective function (see Figure 3). The clustering process stops when the maximum number of iterations is reached or the improvement in the objective function becomes negligible with more iterations. In addition, the S-index proposed by Xie and Beni [24] is used to give the ideal number of categories automatically:

$$S = \frac{J_m}{n \times e_{\min}^2}, \tag{4}$$

where

$$\begin{aligned} J_m &= \sum_{k=1}^{K} \sum_{j=1}^{n} \mu_{j(k)}^{m} e_{j(k)}^2, \\ e_{\min}^2 &= \min_{k1 \neq k2} \left(\sum_{\text{all } p} \left(\overline{x}_{(k1)p} - \overline{x}_{(k2)p}\right)^2 \right). \end{aligned} \tag{5}$$

Chen and Wang [23] found the empirical relationship between the S-index and the estimation performance.

2.3. Step 3: Forecasting the Cycle Times of Jobs in Each Category with a BPN. Subsequently, the jobs/examples of a category are learned with the same BPN. BPN is a popular tool with applications in a variety of fields. Nevertheless, different problems may require different parameter settings for a given network architecture. In the literature, researchers have used BPNs for estimating cycle times and assigning due dates. The configuration of the BPN is established as follows.

(1) Inputs: the new factors determined by PCA associated with the jth example/job. These factors have to be partially normalized so that their values fall within $[0.1, 0.9]$ [10, 11].

(2) Single hidden layer: generally one or two hidden layers are more beneficial for the convergence property of the BPN [25].

(3) The number of neurons in the hidden layer: 1 to $2K$. An increase in the number of hidden-layer nodes lessens the output errors for the training examples but increases the errors for novel examples. Such a phenomenon is often called "overfitting". There exist many different approaches such as the pruning algorithm, the polynomial time algorithm, the canonical decomposition technique, and the network information criterion for finding the optimal configuration of a BPN [26]. In addition, there has been some research considering the relation among the complexity of a BPN, the performance for the training data, and the number of examples, for example, using Akaike's information criterion (AIC) [27] or the minimum description length (MDL) [28].

(4) Activation/transformation function: there are a number of common activation/transformation functions, such as identity function, binary step function, bipolar step function, sigmoid functions (binary sigmoid function and bipolar sigmoid function), and ramp function. In the proposed methodology, the binary sigmoid function is used:

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{6}$$

Therefore, the output ranges between 0 and 1.

(5) Output (o_j): the (normalized) cycle time forecast of the example. o_j is compared with the normalized cycle time $N(\mathrm{CT}_j)$, for which root mean squared error (RMSE) is calculated:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{n}\left(o_j - N\left(\mathrm{CT}_j\right)\right)^2}{n}}. \tag{7}$$

o_j is derived by transforming the signal transferred to the output layer:

$$o_j = \frac{1}{1 + e^{-n_j^o}}, \tag{8}$$

where

$$n_j^o = I_j^o - \theta^o,$$
$$I_j^o = \sum_{l=1}^{L} w_l^o h_{jl}. \tag{9}$$

Similarly, h_{jl} is derived by transforming the signal transferred to the hidden layer:

$$h_{jl} = \frac{1}{1 + e^{-n_{jl}^h}}, \tag{10}$$

where

$$n_{jl}^h = I_{jl}^h - \theta_l^h,$$
$$I_{jl}^h = \sum_{q=1}^{p} w_{ql}^h z_{jq}. \tag{11}$$

Some algorithms are applicable for training a BPN in the backward phase, such as the gradient descent algorithms, the conjugate gradient algorithms, and the Levenberg-Marquardt algorithm. In this study, the Levenberg-Marquardt algorithm is applied. The Levenberg-Marquardt is the most widely used optimization algorithm. It outperforms simple gradient descent and other conjugate gradient methods in a wide variety of problems. The Levenberg-Marquardt algorithm uses approximation and updates the network parameters in a Newton-like way, as described below.

The network parameters are placed in vector $\boldsymbol{\beta}$. The network output o_j can be represented with $f(\mathbf{x_j}, \boldsymbol{\beta})$. The objective function of the BPN is to minimize RMSE or equivalently the sum of squared error (SSE):

$$\mathrm{SSE}(\boldsymbol{\beta}) = \sum_{j=1}^{n}\left(N\left(\mathrm{CT}_j\right) - f\left(\mathbf{x_j}, \boldsymbol{\beta}\right)\right)^2. \tag{12}$$

The Levenberg-Marquardt algorithm is an iterative procedure. In the beginning, the user should specify the initial values of the network parameters. In each step, the parameter vector is replaced by a new estimate, and the network output by its linearization. When the network converges, the gradient of the objective function will be zero. It should be noted that while the Levenberg-Marquardt method is in no way optimal but is just a heuristic, it works extremely well in practice.

2.4. Step 4: Establishing the Upper Bound for the Job Cycle Time Using the IUBR Approach. In order to apply the BPN obtained at the previous step to determine the internal due date of a job, the parameter values in the BPN must be adjusted. To this end, in Chen and Wang [15] and Chen and Lin [16], the NLP model is constructed to adjust the connection weights and thresholds in the BPN, which is not easy to solve. In the IUBR approach, only the threshold of the output node will be adjusted in an iterative way. This way is much simpler and can also achieve satisfactory results.

Substituting (9) into (8),

$$o_j = \frac{1}{1 + e^{-n_j^o}} = \frac{1}{1 + e^{-(I_j^o - \theta^o)}} = \frac{1}{1 + e^{\theta^o - I_j^o}}. \tag{13}$$

Therefore,

$$\ln\left(\frac{1}{o_j} - 1\right) = \theta^o - I_j^o. \tag{14}$$

So

$$I_j^o = \theta^o - \ln\left(\frac{1}{o_j} - 1\right). \tag{15}$$

Assume that the adjustment made to the threshold of the output node is indicated as $\Delta\theta^o$. After adjustment, the output from the new BPN, o_j', determines the upper bound of the cycle time:

$$o_j' = \frac{1}{1+e^{-n_j^{o\prime}}}, \tag{16}$$

where

$$n_j^{o\prime} = I_j^o - \theta^{o\prime} = I_j^o - \left(\theta^o + \Delta\theta^o\right). \tag{17}$$

Substituting (17) into (16),

$$o_j' = \frac{1}{1+e^{-(I_j^o-\theta^o-\Delta\theta^o)}}. \tag{18}$$

Substituting (15) into (18),

$$\begin{aligned} o_j' &= \frac{1}{1+e^{-(\theta^o-\ln((1/o_j)-1)-\theta^o-\Delta\theta^o)}} = \frac{1}{1+e^{\ln((1/o_j)-1)+\Delta\theta^o}} \\ &= \frac{1}{1+e^{\Delta\theta^o}\left(\left(1/o_j\right)-1\right)}. \end{aligned} \tag{19}$$

Obviously, the maximum of $\Delta\theta^o$ establishes the lowest upper bound.

Since o_j' is the upper bound of the cycle time, $o_j' \geq N(\mathrm{CT}_j)$,

$$\frac{1}{1+e^{\ln((1/o_j)-1)+\Delta\theta^o}} \geq N\left(\mathrm{CT}_j\right), \tag{20}$$

$$\Delta\theta^o \leq \ln\left(\frac{1}{N\left(\mathrm{CT}_j\right)}-1\right) - \ln\left(\frac{1}{o_j}-1\right). \tag{21}$$

Equation (21) holds for all jobs, so

$$\Delta\theta^o \leq \min_j\left(\ln\left(\frac{1}{N\left(\mathrm{CT}_j\right)}-1\right) - \ln\left(\frac{1}{o_j}-1\right)\right). \tag{22}$$

According to (19), the optimal value of $\Delta\theta^o$ should be set to the maximum possible value:

$$\Delta\theta^{o*} = \min_j\left(\ln\left(\frac{1}{N\left(\mathrm{CT}_j\right)}-1\right) - \ln\left(\frac{1}{o_j}-1\right)\right). \tag{23}$$

Then the optimization results of the BPN are sensitive to the initial conditions and may be different for each iteration. Assume that the optimal value of o_j' in the tth iteration is indicated with $o_j'(t)$. After some iterations,

$$o_j'\,(\text{all iterations}) = \min_t o_j'(t). \tag{24}$$

In this way, the upper bound of the cycle time is decreased gradually (see Figure 4). Another merit of the IUBR approach is that it does not rely on the parameters of the BPN.

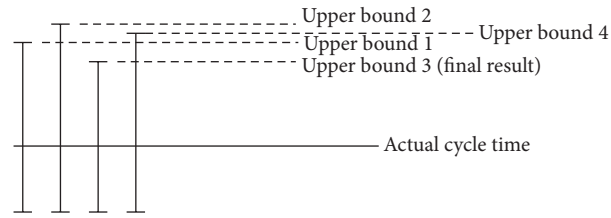

Figure 4: The upper bound is reduced in an iterative manner.

2.5. Ensemble Learning. Ensemble learning is based on the notion of perturbing and combining. An ensemble consists of a collection of ANNs and combines their predictions to obtain a final prediction. In FCM, a job can be classified into several categories to different degrees. In theory, the BPNs of all categories can be applied to predict the cycle time of a job. The forecasts obtained by using the BPNs may not be the same and need to be aggregated. To this end, some treatments have been carried out in the literature.

(1) Linear aggregation [29]:

$$\mathrm{CTE}_j = \frac{\sum_{k=1}^{K}\left(\mu_{j(k)}\cdot\mathrm{CTE}_j(k)\right)}{\sum_{k=1}^{K}\mu_{j(k)}} = \sum_{k=1}^{K}\left(\mu_{j(k)}\cdot\mathrm{CTE}_j(k)\right), \tag{25}$$

where $\sum_{k=1}^{K}\mu_{j(k)} = 1$. $\mathrm{CTE}_j(k)$ is the cycle time of job j estimated by the BPN of category k.

(2) BPN aggregation [29]: the membership and cycle time forecast of a job are fed into another BPN to be aggregated. Consider

$$\mathrm{CTE}_j = \mathrm{BPN}\left(\mu_{j(1)}, \mathrm{CTE}_j(1), \ldots, \mu_{j(K)}, \mathrm{CTE}_j(K)\right). \tag{26}$$

(3) Generalized average method [30]: in FCM, the error is proportional to the distance to the center. For this reason, a natural way to aggregate the forecasts is

$$\mathrm{CTE}_j = \frac{\sum_{k=1}^{K}\sqrt[2/(m-1)]{1/\mu_{j(k)}}\cdot\mathrm{CTE}_j(k)}{\sum_{k=1}^{K}\sqrt[2/(m-1)]{1/\mu_{j(k)}}}. \tag{27}$$

3. Application and Analyses

To demonstrate the application of the proposed methodology, a real case with the data of 40 jobs from a wafer fabrication factory located in Taichung City Scientific Park, Taiwan (see Table 2), was used, where $x_{j1} \sim x_{j6}$ stand for the job size, factory utilization, the queue length on the route, the queue length before the bottleneck, the work in progress (WIP), and the average waiting time. The wafer fabrication factory produces more than ten products and has a monthly capacity of 20,000 wafers. The wafer fabrication processes include photolithography, thermal processes, implantation, chemical vapor deposition, etching, physical vapor deposition, chemical mechanical polishing, process diagnostics and control, and cleaning. The production characteristic of

Table 2: An example.

j	x_{j1}	x_{j2}	x_{j3}	x_{j4}	x_{j5}	x_{j6}	CT_j
1	24	1261	181	781	112	0.92	935
2	24	1263	181	762	127	0.90	958
3	24	1220	176	761	127	0.89	1047
4	23	1282	178	802	127	0.94	1011
5	23	1303	180	780	175	0.93	1068
6	23	1281	183	782	175	0.93	1143
7	23	1242	184	741	163	0.89	1103
8	24	1262	182	681	139	0.86	1250
9	22	1260	182	701	98	0.86	1181
10	22	1260	179	700	257	0.87	1194
11	24	1301	163	722	99	0.84	1260
12	22	1221	184	641	131	0.82	1240
13	23	1323	159	740	247	0.87	1180
14	24	1362	181	782	191	0.95	1227
15	24	1261	181	762	219	0.91	1236
16	23	1321	177	801	219	0.96	1215
17	22	1343	180	822	219	0.97	1228
18	24	1321	177	762	54	0.93	1266
19	25	1343	179	781	54	0.96	1285
20	25	1300	180	740	54	0.92	1272
21	22	1320	181	721	54	0.91	1310
22	24	1321	182	742	49	0.92	1265
23	23	1262	165	680	201	0.80	1308
24	22	1240	161	722	103	0.82	1331
25	23	1183	183	661	53	0.82	1294
26	23	1282	184	701	53	0.88	1314
27	22	1202	177	680	248	0.84	1321
28	23	1202	178	681	248	0.85	1353
29	24	1202	185	701	82	0.86	1226
30	23	1202	158	721	98	0.81	1301
31	24	1343	181	760	67	0.94	1280
32	24	1381	185	801	67	0.97	1286
33	22	1362	156	780	67	0.91	1252
34	23	1282	179	782	223	0.92	1214
35	23	1320	180	782	176	0.93	1251
36	25	1340	176	801	462	0.97	1222
37	23	1320	182	781	168	0.95	1187
38	22	1361	181	781	141	0.94	1205
39	22	1381	179	781	95	0.97	1120
40	23	1363	178	802	179	0.97	1133

"reentry," which is highly relevant to the semiconductor industry, is clearly reflected in this problem. It also shows the difficulties facing production planners and schedulers who attempt to provide an accurate due date for a product with a very complicated routing.

The standard deviations of the six inputs are compared in Figure 5. Note that the variability in x_{j2}, x_{j4}, and x_{j5} is substantially higher than that in the remaining variables.

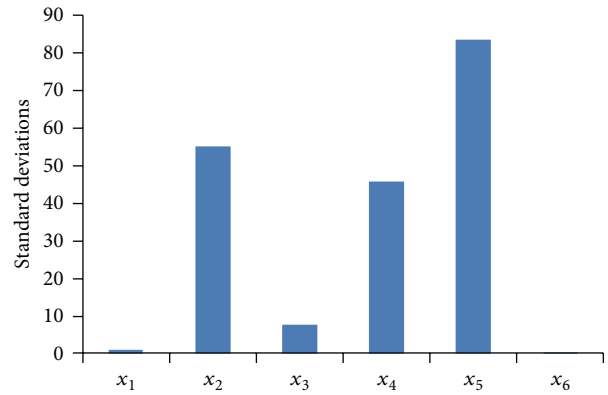

Figure 5: The comparison of the standard deviations of the inputs.

Subsequently, we standardize the data (see Table 3) and obtain the correlation matrix as

$$R = \begin{bmatrix} 0.97 & 0.10 & 0.16 & 0.21 & -0.03 & 0.25 \\ 0.10 & 0.98 & 0.01 & 0.70 & -0.01 & 0.78 \\ 0.16 & 0.01 & 0.98 & 0.05 & -0.07 & 0.37 \\ 0.21 & 0.70 & 0.05 & 0.98 & 0.15 & 0.86 \\ -0.03 & -0.01 & -0.07 & 0.15 & 0.98 & 0.10 \\ 0.25 & 0.78 & 0.37 & 0.86 & 0.10 & 0.98 \end{bmatrix}. \quad (28)$$

The eigenvalues and eigenvectors of R are then calculated. Based on them, the variance contribution rates can be derived as

$$\begin{aligned} &\eta_1 = 46\%, \quad \eta_2 = 20\%, \quad \eta_3 = 16\%, \\ &\eta_4 = 14\%, \quad \eta_5 = 4\%, \quad \eta_6 = 0\%. \end{aligned} \quad (29)$$

Summing up η_q's, we obtain

$$\begin{aligned} &\eta_\Sigma(1) = 46\%, \quad \eta_\Sigma(2) = 65\%, \quad \eta_\Sigma(3) = 81\%, \\ &\eta_\Sigma(4) = 95\%, \quad \eta_\Sigma(5) = 100\%, \quad \eta_\Sigma(6) = 100\%. \end{aligned} \quad (30)$$

After conducting a Pareto analysis, p is chosen as 3 to meet the requirement $\eta_\Sigma(p) \geq 85\% \sim 90\%$. The first three principal components explain roughly 80% of the total variability in the standardized data, so that it might be a reasonable way to reduce the dimensions in order to visualize the data.

Subsequently, the component scores are calculated (see Table 4), which contain the coordinates of the original data in the new coordinate system defined by the principal components, and will be used as the new inputs to the FCM-BPN.

Subsequently, jobs are classified using FCM based on the new variables. The results of the S-test are summarized in Table 5. In this case, the optimal number of job categories was 5. However, there will be some categories with very few jobs. For this reason, the second best solution is used, that is, 4 categories, by setting the threshold of membership to 0.3. The classification results are shown in Table 6.

After preclassification, the three-layer BPN of each category was applied to predict the cycle times of jobs belonging

Table 3: The standardized data.

j	x_{j1}	x_{j2}	x_{j3}	x_{j4}	x_{j5}	x_{j6}
1	0.88	−0.53	0.40	0.74	−0.40	0.37
2	0.88	−0.49	0.48	0.30	−0.22	−0.05
3	0.88	−1.27	−0.17	0.29	−0.22	−0.31
4	−0.22	−0.15	0.07	1.18	−0.22	0.65
5	−0.22	0.24	0.37	0.71	0.35	0.58
6	−0.22	−0.17	0.78	0.74	0.35	0.45
7	−0.22	−0.87	0.90	−0.14	0.21	−0.19
8	0.88	−0.51	0.53	−1.45	−0.08	−0.87
9	−1.32	−0.55	0.60	−1.01	−0.56	−0.81
10	−1.32	−0.54	0.23	−1.03	1.34	−0.55
11	0.88	0.20	−1.87	−0.57	−0.56	−1.19
12	−1.32	−1.25	0.80	−2.33	−0.18	−1.64
13	−0.22	0.60	−2.40	−0.16	1.22	−0.60
14	0.88	1.31	0.47	0.75	0.55	0.94
15	0.88	−0.53	0.40	0.31	0.88	0.19
16	−0.22	0.57	−0.05	1.17	0.88	1.13
17	−1.32	0.97	0.33	1.62	0.88	1.38
18	0.88	0.56	−0.01	0.31	−1.09	0.62
19	1.97	0.96	0.15	0.74	−1.09	1.06
20	1.97	0.19	0.38	−0.16	−1.09	0.36
21	−1.32	0.55	0.51	−0.57	−1.09	0.10
22	0.88	0.55	0.54	−0.13	−1.16	0.32
23	−0.22	−0.52	−1.59	−1.47	0.67	−2.00
24	−1.32	−0.91	−2.11	−0.57	−0.51	−1.54
25	−0.22	−1.95	0.72	−1.89	−1.11	−1.64
26	−0.22	−0.15	0.89	−1.01	−1.11	−0.36
27	−1.32	−1.59	−0.07	−1.47	1.23	−1.26
28	−0.22	−1.60	0.07	−1.47	1.23	−1.11
29	0.88	−1.60	0.93	−1.03	−0.76	−0.87
30	−0.22	−1.59	−2.57	−0.58	−0.56	−1.86
31	0.88	0.97	0.41	0.27	−0.94	0.70
32	0.88	1.65	1.00	1.16	−0.94	1.27
33	−1.32	1.30	−2.73	0.71	−0.94	0.10
34	−0.22	−0.14	0.16	0.74	0.92	0.33
35	−0.22	0.55	0.34	0.75	0.36	0.54
36	1.97	0.91	−0.23	1.17	3.79	1.28
37	−0.22	0.55	0.61	0.73	0.27	0.91
38	−1.32	1.28	0.42	0.72	−0.05	0.81
39	−1.32	1.66	0.14	0.72	−0.60	1.36
40	−0.22	1.33	0.13	1.18	0.40	1.42

Table 4: New inputs to the FCM-BPN.

z_{j1}	z_{j2}	z_{j3}
−0.56	0.91	−0.19
−0.13	0.87	−0.34
0.51	0.57	−0.37
−0.97	−0.10	0.20
−0.87	−0.20	−0.26
−0.75	0.14	−0.51
0.57	0.56	−0.66
1.30	1.18	−0.55
1.55	0.31	0.47
1.37	−0.87	−1.04
1.11	−0.59	0.91
3.04	0.63	−0.20
0.51	−2.44	−0.02
−1.94	0.12	−0.43
−0.30	0.35	−1.29
−1.62	−0.84	−0.48
−2.04	−1.24	−0.17
−0.87	0.77	0.89
−1.92	1.34	0.64
−0.58	1.70	0.34
0.22	0.23	1.29
−0.62	1.31	0.73
2.54	−1.26	−0.16
2.39	−1.64	1.20
3.02	1.57	0.14
0.89	1.21	0.66
2.56	−0.74	−1.19
2.19	−0.13	−1.54
1.61	1.90	−0.42
2.72	−1.23	0.87
−1.27	0.99	0.71
−2.56	1.07	0.78
−0.37	−2.44	2.47
−0.60	−0.51	−0.82
−1.06	−0.27	−0.17
−2.54	−1.36	−3.41
−1.31	−0.02	−0.18
−1.32	−0.63	0.67
−1.77	−0.58	1.32
−2.13	−0.66	0.12

to the category according to the new variables. Different network architectures were evaluated to compare the forecasting performance. The best-fitted network which was selected, and, therefore, the architecture which presented the best forecasting accuracy, is composed of three inputs, six hidden and one output neurons.

The convergence condition in training networks was established as either the improvement in MSE becomes less than 10^{-6} with one more epoch or 1000 epochs have already been run. 3/4 of the adopted examples in each category are fed as "training examples" into the BPN. The remaining 1/4 is left for testing. For example, category 3 has 8 jobs; 6 of them are randomly chosen for training the BPN while the remaining 2 jobs are left for testing. The forecasting accuracy can be evaluated with mean absolute error (MAE), mean absolute percentage error (MAPE), and RMSE. The forecasting performances are summarized in Table 7. The forecasting results are shown in Figure 6. The performance of the proposed methodology is compared with those of statistical analysis (i.e., multiple linear regression), BPN,

Table 5: The results of the S-test.

Number of categories (K)	J_m	$e^2_{\min}$	S
2	1.96	0.14	0.34
3	1.21	0.09	0.34
4	0.86	0.07	0.30
5	0.67	0.06	0.26
6	0.53	0.03	0.43

Table 6: The classifying results ($\mu_L = 0.3$).

Category	Jobs
1	4–6, 14–17
2	10-11, 13, 23-24, 27, 30, 33
3	1, 18–22, 31-32
4	2-3, 7–9, 12, 25-26, 28-29, 34–40

Table 7: The forecasting performances.

Category	MAE (hrs)	MAPE	RMSE (hrs)
1	20	1.7%	44
2	6	0.5%	14
3	5	0.4%	12
4	8	0.6%	18
Total	11	0.9%	29

Table 8: Comparison of the forecasting performances.

Category	MAE (hrs)	MAPE	RMSE (hrs)
Statistical analysis	73	6.1%	99
BPN	30	2.4%	69
FCM-BPN	15	1.2%	38
PCA-BPN	27	2.3%	67
PCA-FCM-BPN	11	0.9%	29

FCM-BPN, and PCA-BPN in Table 8. The nonlinear nature of this problem is obvious since the performance of statistical analysis (a linear approach) is poor. In addition, the simple combination of PCA and BPN does not have much effect. The main effect of PCA is to improve the correctness of job classification, as mentioned in Chen and Wang [23].

Subsequently, the IUBR approach is applied to determine the upper bound of the cycle time. In the first iteration, $\Delta\theta^{o*}(t)$ is -0.808, and the upper bounds of the cycle times are shown in Figure 7.

The process stops after five iterations because the upper bounds remain unchanged after the fifth iteration. The results of the five iterations are summarized in Table 9, from which the allowances which are 25, 33, 54, 48, 56, 57, 58, 44, 54, 53, 42, 46, 55, 48, 47, 50, 48, 41, 36, 39, 30, 41, 31, 24, 34, 29, 27, 13, 37, 24, 47, 15, 34, 50, 44, 49, 53, 50, 53, and 53 added to the cycle times are derived with an average of 42 (hours). The due date of a job is then set to the release time plus the upper bound of the cycle time.

To make a comparison, six other allowance determination policies are also applied to the collected data.

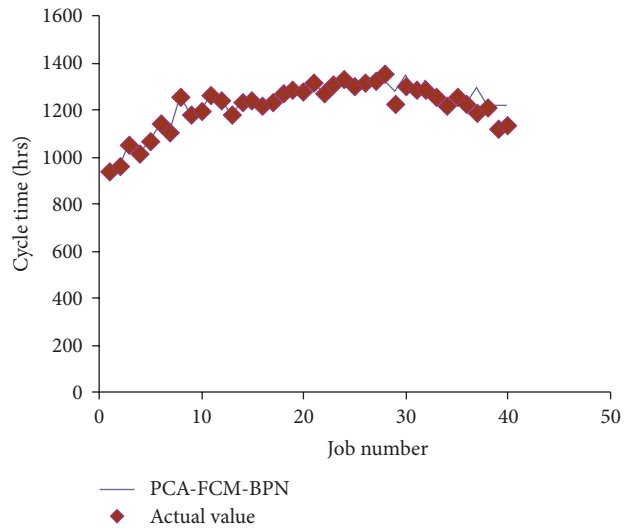

Figure 6: The forecasting results using PCA-FCM-BPN.

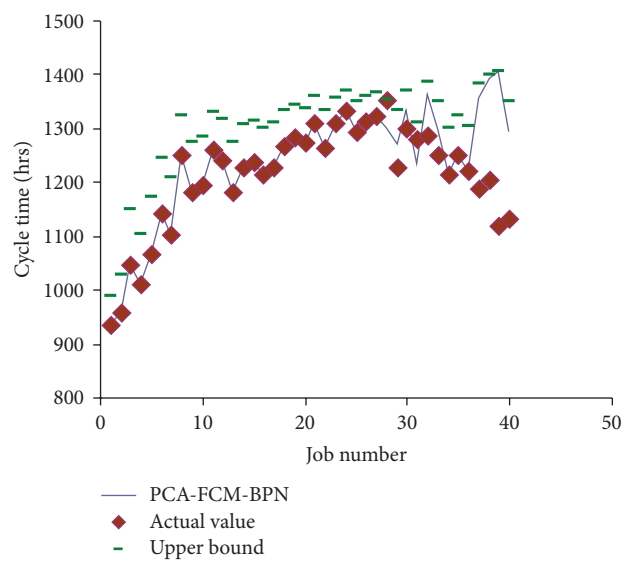

Figure 7: The upper bounds of the cycle times.

(1) Total work content policy (TWK): in TWK, the due date allowance factor is estimated based on historical data by a regression model. There is another product in the wafer fabrication factory with an average cycle time of 1278 hours. The total processing time and cycle time standard deviation of the product are 317 and 87 hours, respectively. The product was adopted as the comparison basis, and in this case the cycle time forecast and allowance are determined as follows:

$$\text{Cycle time forecast} = \frac{1278}{317} * \text{the total processing time},$$
$$\text{Allowance} = 3 * 87 * \left(\frac{\text{the total processing time}}{317}\right). \tag{31}$$

Table 9: The results of the five iterations.

Iteration 1 ($\Delta\theta^{o*}(t) = -0.808$)	Iteration 2 ($\Delta\theta^{o*}(t) = -0.669$)	Iteration 3 ($\Delta\theta^{o*}(t) = -0.446$)	Iteration 4 ($\Delta\theta^{o*}(t) = -0.446$)	Iteration 5 ($\Delta\theta^{o*}(t) = -0.446$)
987	976	960	960	960
1025	1011	991	991	991
1147	1129	1100	1100	1100
1103	1086	1059	1058	1058
1171	1153	1124	1123	1123
1243	1227	1200	1199	1199
1207	1190	1161	1161	1161
1322	1312	1294	1294	1294
1274	1259	1235	1235	1235
1283	1270	1246	1246	1246
1329	1319	1302	1302	1302
1316	1305	1286	1286	1286
1273	1259	1234	1234	1234
1307	1296	1276	1275	1275
1313	1302	1282	1282	1282
1299	1286	1265	1265	1265
1308	1296	1276	1276	1276
1332	1323	1307	1307	1307
1344	1336	1321	1321	1321
1336	1327	1311	1311	1311
1358	1352	1340	1340	1340
1331	1322	1305	1305	1305
1357	1350	1338	1338	1338
1369	1364	1355	1355	1355
1349	1342	1328	1328	1328
1360	1354	1343	1343	1343
1364	1358	1348	1348	1348
1353	1353	1353	1353	1353
1334	1334	1320	1320	1320
1369	1364	1355	1355	1355
1311	1300	1280	1280	1280
1384	1381	1376	1376	1376
1350	1343	1329	1329	1329
1298	1286	1264	1263	1263
1323	1313	1295	1295	1295
1304	1292	1271	1271	1271
1381	1378	1371	1250	1250
1399	1399	1397	1265	1265
1405	1405	1405	1247	1247
1348	1340	1326	1244	1244

(2) Gamma distribution fitting method (Gamma): the waiting time of a job is fitted with a Gamma distribution. For example, the waiting time of a job with 24 pieces of wafers is fitted with a Gamma distribution in Figure 8. The 50% and 95% percentiles are 929 and 1160, respectively, and the total processing time is 251 hours. So the cycle time forecast is 1160 + 251 = 1411 hours, and allowance is 1160 − 929 = 231 hours.

(3) Constant allowance policy (CON, PCA-FCM-BPN + CON): add three times the RMSE of the prediction approach to the completion time forecasts to determine the due date.

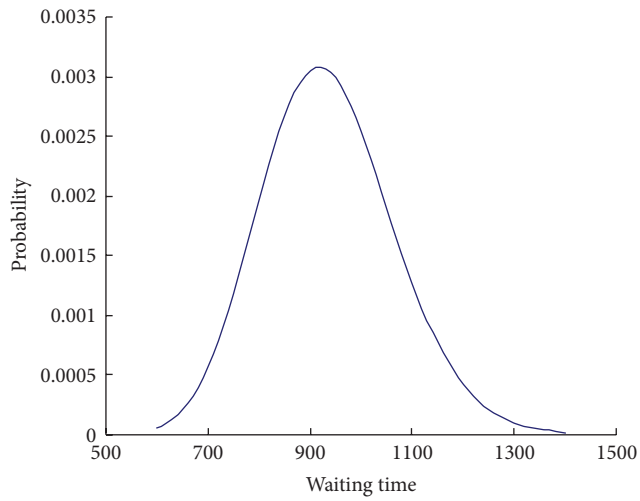

Figure 8: Fitting the waiting time with a Gamma distribution (jobs with 24 wafers).

(4) Selective allowance policy (SAP, PCA-FCM-BPN + SAP): add three times the RMSE of the prediction approach to the completion time forecasts of a small quantity of jobs that might encounter difficulties in keeping the internal due date. Such jobs are chosen in the following way:

$$\sum_{i=1}^{6} N\left(x_{ji}\right) \geq 0.5 \cdot \max_{j} \sum_{i=1}^{6} N\left(x_{ji}\right) + 0.5 \cdot \min_{j} \sum_{i=1}^{6} N\left(x_{ji}\right). \tag{32}$$

In other words, these jobs are among the 50% percentiles.

(5) Random assignment policy (RAP, PCA-FCM-BPN + RAP): add the extra allowance to the completion time forecasts of the same quantity of jobs that are randomly chosen.

(6) No allowance policy (NAP, PCA-FCM-BPN + NAP): no allowance will be assigned to any job.

Due date related performances are impacted by the quality of the due date assignment methods. After applying the seven allowance determination policies, the following performance measures are compared:

(1) number of tardy jobs (N_T);

(2) mean tardiness ($\overline{T}$);

(3) sum of allowances.

The comparison results are summarized in Table 10. The proposed IUBR approach outperforms the other allowance determination policies.

(1) It guarantees the on-time delivery of the jobs. Both N_T and $\overline{T}$ are zeros. Among the other allowance determination policies, only Gamma and CON can achieve that at the expense of adding some extra allowance.

(2) The percentage of reduction in the sum of allowances over CON is 52%. The advantages over TWK, Gamma, SAP, and RAP are 79%, 74%, 12%, and 12%, respectively. The percentage of on-time delivery is not derived from a greater buffer on the completion time prediction.

(3) The performance of SAP is not better than that of RAP, which shows it is not easy to anticipate jobs that may delay.

(4) Compared with TWK and Gamma, the other policies effectively reduce the allowances added to the job cycle times, which is due to the forecasting accuracy of the PCA-FCM-BPN approach.

4. Conclusions and Directions for Future Research

Owing to the complexity of the wafer fabrication, the due date assignment of each job presents a challenging problem to the production planning and scheduling people. The firm has to offer a price reduction if the due date is far away from the expected one. Conversely, the looser the due date is set, the higher the probability that the job will be completed or delivered on time is. That is very important to maintain a good reputation with the customers. This study explores a new application of fuzzy-neural approaches in the due date assignment problem of the wafer fabrication factory. The proposed methodology decomposes internal due date assignment in a wafer fabrication factory into two subproblems: completion time prediction and allowance determination. To overcome the problems with the existing approaches, two innovative treatments are taken in the proposed methodology. First, PCA is applied to construct a series of linear combinations of the original variables to form a new variable, so that these new variables are unrelated to each other as much as possible, and the relationship among them can be reflected in a better way. The combination of PCA and BPN also reduces the space for storing the input variables in the modeling of the wafer fabrication system. In addition, the simultaneous application of PCA, FCM, and BPN further improved the estimation accuracy. Subsequently, the IUBR approach is proposed to determine the allowance that will be added to the estimated job cycle time. Our result is existentially tight.

The validity that the effective fuzzy-neural approach for internal due date assignment is able to improve on-time delivery has been proved by the case study. Based on the above analysis,

(1) the forecasting accuracy (measured with MAE, MAPE, and RMSE) of the PCA-FCM-BPN was significantly better than those of many existing approaches;

(2) it is easier to determine the allowance in the IUBR method than the method based on NLP;

(3) the bound on the job cycle time is tighter than the bounds by TWK, Gamma, and CON and simpler than the bound by Chen and Wang [15], which requires NLP optimization.

Table 10: The performances of various allowance determination policies.

	TWK	Gamma	CON	SAP	RAP	NAP (basis)	IUBR
Sum of allowances	7982	6479	3480	1914	1914	0	1680
N_T	31	0	0	9	7	14	0
$\overline{T}$ (hours)	82	0	0	3.9	2.1	3.5	0

However, there are two limitations that need to be acknowledged and addressed regarding the present study.

(1) The first limitation concerns the experimental nature of this research. The proposed methodology was studied within a short period of time. There is an apparent danger involved whenever conclusions are drawn from such a limited sample and then applied in the highly dynamic semiconductor manufacturing environment.

(2) The BPN part in the methodology is usually regarded as a black box. To exploit the knowledge embedded in the back box, and to facilitate the practical application of the proposed methodology, some association rules have to be extracted from the estimation results.

The IUBR approach only modifies the threshold of the output node. In future studies, other parameters in the BPN can be modified in similar ways. However, it is a challenge to make the modification results independent of the original parameter values. In addition, the concept of customer satisfaction can be incorporated into the proposed methodology; thereby, the due date can achieve a higher level of customer satisfaction. In contrast, the proposed methodology only guarantees a positive level of customer satisfaction.

Acknowledgment

This study was financially supported by the National Science Council of Taiwan.

References

[1] T. Chen, A. Jeang, and Y. Wang, "A hybrid neural network and selective allowance approach for internal due date assignment in a wafer fabrication plant," *International Journal of Advanced Manufacturing Technology*, vol. 36, no. 5-6, pp. 570–581, 2008.

[2] T. Chen, "A flexible way of modelling the long-term cost competitiveness of a semiconductor product," *Robotics & Computer Integrated Manufacturing*, vol. 29, no. 3, pp. 31–40, 2013.

[3] W. L. Pearn, S. H. Chung, and C. M. Lai, "Due-date assignment for wafer fabrication under demand variate environment," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 2, pp. 165–175, 2007.

[4] C. F. Chien, C. Y. Hsu, and C. W. Hsiao, "Manufacturing intelligence to forecast and reduce semiconductor cycle time," *Journal of Intelligent Manufacturing*, vol. 23, no. 6, pp. 2281–2294, 2011.

[5] Y. Wilamowsky, S. Epstein, and B. Dickman, "Optimal common due-date with completion time tolerance," *Computers and Operations Research*, vol. 23, no. 12, pp. 1203–1210, 1996.

[6] Y. C. Wang, T. Chen, and C. W. Lin, "A slack-diversifying nonlinear fluctuation smoothing rule for job dispatching in a wafer fabrication factory," *Robotics & Computer Integrated Manufacturing*, vol. 29, no. 3, pp. 41–47, 2013.

[7] A. K. Gupta and A. I. Sivakumar, "Optimization of due-date objectives in scheduling semiconductor batch manufacturing," *International Journal of Machine Tools and Manufacture*, vol. 46, no. 12-13, pp. 1671–1679, 2006.

[8] T. Chen, "A hybrid look-ahead SOM-FBPN and FIR system for wafer-lot-output time prediction and achievability evaluation," *International Journal of Advanced Manufacturing Technology*, vol. 35, no. 5-6, pp. 575–586, 2007.

[9] T. Chen, "A SOM-FBPN-ensemble approach with error feedback to adjust classification for wafer-lot completion time prediction," *International Journal of Advanced Manufacturing Technology*, vol. 37, no. 7-8, pp. 782–792, 2008.

[10] T. Chen, Y. Wang, and H. Tsai, "Lot cycle time prediction in a ramping-up semiconductor manufacturing factory with a SOM-FBPN-ensemble approach with multiple buckets and partial normalization," *International Journal of Advanced Manufacturing Technology*, vol. 42, no. 11-12, pp. 1206–1216, 2009.

[11] T. Chen, H. Wu, and Y. Wang, "Fuzzy-neural approaches with example post-classification for estimating job cycle time in a wafer fab," *Applied Soft Computing Journal*, vol. 9, no. 4, pp. 1225–1231, 2009.

[12] T. Chen, "Job cycle time estimation in a wafer fabrication factory with a bi-directional classifying fuzzy-neural approach," *International Journal of Advanced Manufacturing Technology*, vol. 56, no. 9–12, pp. 1007–1018, 2011.

[13] B. E. Ankenman, J. M. Bekki, J. Fowler, G. T. Mackulak, B. L. Nelson, and F. Yang, "Simulation in production planning: an overview with emphasis on recent developments in cycle time estimation," *International Series in Operations Research & Management Science*, vol. 151, pp. 565–591, 2011.

[14] V. Gordon, J.-M. Proth, and C. Chu, "A survey of the state-of-the-art of common due date assignment and scheduling research," *European Journal of Operational Research*, vol. 139, no. 1, pp. 1–25, 2002.

[15] T. Chen and Y. Wang, "Incorporating the FCM-BPN approach with nonlinear programming for internal due date assignment in a wafer fabrication plant," *Robotics and Computer-Integrated Manufacturing*, vol. 26, no. 1, pp. 83–91, 2010.

[16] T. Chen and Y. Lin, "A collaborative fuzzy-neural approach for internal due date assignment in a wafer fabrication plant," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 9, pp. 5193–5210, 2011.

[17] T. Chen, "A collaborative and artificial intelligence approach for semiconductor cost forecasting," *Computers & Industrial Engineering*, vol. 66, no. 2, pp. 476–484, 2013.

[18] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 11, pp. 559–572, 1901.

[19] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.

[20] A. Jaiswal, N. Kumar, and R. K. Agrawal, "A hybrid of principal component analysis and partial least squares for face recognition across pose," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 7441 of *Lecture Notes in Computer Science*, pp. 67–73, Springer, Berlin, Germany, 2012.

[21] J. Mohtasham, A. S. Nateri, and H. Khalili, "Textile colour matching using linear and exponential weighted principal component analysis," *Coloration Technology*, vol. 128, no. 3, pp. 199–203, 2012.

[22] T. Chen and Y. C. Wang, "Long-term load forecasting by the collaborative fuzzy-neural approach," *International Journal of Electrical Power and Energy Systems*, vol. 43, pp. 454–464, 2012.

[23] T. Chen and Y. C. Wang, "An iterative procedure for optimizing the performance of the fuzzy-neural job cycle time estimation approach in a wafer fabrication factory," *Mathematical Problems in Engineering*, vol. 2013, Article ID 740478, 15 pages, 2013.

[24] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841–847, 1991.

[25] J. Heaton, *Introduction to Neural Networks with Java*, Heaton Research, St. Louis, Mo, USA, 2013.

[26] M. Khashei, *Forecasting the Esfahan steel company production price in Tehran metals exchange using artificial neural networks (ANNs) [M.S. thesis]*, Isfahan University of Technology, Isfahan, Iran, 2005.

[27] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[28] J. Rissanen, "Stochastic complexity and modeling," *The Annals of Statistics*, vol. 14, no. 3, pp. 1080–1100, 1986.

[29] T. Chen, "Incorporating fuzzy c-means and a back-propagation network ensemble to job completion time prediction in a semiconductor fabrication factory," *Fuzzy Sets and Systems*, vol. 158, no. 19, pp. 2153–2168, 2007.

[30] H. C. Wu and T. Chen, "A fuzzy-neural ensemble and geometric rule fusion approach for scheduling a wafer fabrication factory," *Mathematical Problems in Engineering*, vol. 2013, Article ID 956978, 14 pages, 2013.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 745153, 6 pages
http://dx.doi.org/10.1155/2013/745153

Research Article

Fuzzy Approach to Statistical Control Charts

Shahryar Sorooshian

Faculty of Technology, University Malaysia Pahang, Gambang Kuantan, 26300 Pahang, Malaysia

Correspondence should be addressed to Shahryar Sorooshian; sorooshian@gmail.com

Received 4 February 2013; Accepted 22 August 2013

Academic Editor: Hadi Nasseri

After investigating the advantages and disadvantages of current methods of statistical process control, it becomes important to overcome the disadvantages and then use the advantages to improve a method for monitoring a process with categorical observations. An approach which considers uncertainty and vagueness is tried for this study; and for this purpose, fuzzy set theory is inevitable to use. So, a new approach based on fuzzy set theory is introduced in this research for monitoring attribute quality characteristics. This approach is then compared with the current related approach to see the difference in performance.

1. Introduction

With regard to the continuous improvement in the products and service quality as a main factor for customer satisfaction, improving the tools of monitoring the quality characteristics has become inevitable. Statistical process control (SPC) is a well-known methodology for improving the quality. SPC is a powerful collection of problem-solving tools beneficial in achieving process stability and enhancing capability and quality through the reduction of variability [1]. Control chart is utilized as the most essential tool of SPC that is frequently employed to determine whether a process is in a state of statistical control. According to Montgomery [1], the control chart refers to a graphical display of a quality characteristic that has been measured or computed from a sample versus the sample number or time. Variable control charts are used to monitor continuous characteristics of the products, while attribute control charts are applied to monitor the quality characteristics, which are not possible to express in numerical scale. From the literature, first, it is concluded that there are some advantages and disadvantages for using attribute control charts like p chart by comparing it to the variable control chart like $\overline{X} - R$. Some advantages of using attribute control charts are as follows.

(i) Attribute control charts could monitor more than one quality characteristic simultaneously.

(ii) Attribute control charts need less cost and time for inspection than variable control charts.

Disadvantages of attribute control charts are as follows.

(i) Attribute control charts need larger sample size than variable control charts.

(ii) Attribute information could not determine the reason of being out of control, so correction action is meaningful.

However, the binary classification into conforming and nonconforming used in p chart might not be appropriate in many situations where there might be a number of intermediate levels [2]. In this case, for measuring the quality-related characteristics, it is necessary to use several intermediate levels besides conforming and nonconforming. For example, the quality of the product can be classified by one of the following terms: *perfect*, *good*, *median*, *poor*, and *fair*, depending on deviation from specifications. Data obtained in this way are called categorical data.

Some research has been done for monitoring processes with categorical observations, such as multinomial distribution based and grouped data approach which have several disadvantages as follows:

(i) cannot specify if the change in the quality is a result of quality improvement or not [3],

(ii) control limits do not depend on sample size [3],

(iii) for the trinomial distribution, Cochran [4] rules require all expected frequencies to be at least five. So, a large sample size is required, but collecting such sample size is so hard in real applications,

(iv) however, the majority of our information about the surrounding phenomena is fuzzy and we expressed them by means of linguistic variable. Furthermore, the quality level of each product is determined by the interaction between the linguistic and qualitative variables which are usually vague, and in each organization, operators and experts are the responders of determining the quality level and the estimation of the quality which they have done mentally in uncertain situations. So it is necessary to use an approach that is applicable and capable to register the linguistic variable and estimate them with appropriate approximation. In fact the main problem is vagueness that corresponds to the mental affect [5].

2. Statistical and Fuzzy Control Charts

In general, statistical and fuzzy methodologies exist to deal with the categorical data. Early research on statistical methodologies goes back to Duncan [6] who introduced a chi-square control chart for monitoring a multinomial process with categorical data. Later, this type of control chart is discussed further by Marcucci [7] and Nelson [8]. Marcucci [7] introduced a statistical approach for the case, where the proportion of each category is not known before. In the case of fuzzy methodologies, several approaches are proposed. Bradshaw Jr. [9], for the first time, used fuzzy sets as a basic for explaining the measurement of conformity of each product unit with the specifications. Williams and Zigli [10] showed that quality assurance techniques, especially in service industries, are not without imprecision of human judgments. This imprecision and vagueness can be treated with the help of fuzzy set theory. Raz and Wang [2] and Wang and Raz [11] proposed a probabilistic approach and a membership approach. Kanagawa et al. [12] developed a new control chart for the monitoring of the mean and deviation of attribute variables. Franceschini and Romano [13] proposed an approach based on the use of linguistic quantifiers for constructing control charts. Probabilistic and membership approach are discussed by Laviolette et al. [14], Almond [15], and Kandel et al. [16] and reviewed by Woodall et al. [17] and Taleb and Limam [3]. Gülbay and Kahraman [18–20] proposed α-level fuzzy control chart for attributes in order to reflect the vagueness of data and tightness of inspection. Cheng [21] proposed an approach to deal with the expert's subjective judgments based on the ranking scores assigned by the individual inspectors to the inspected items. Shu and Wu [22] used resolution identity to construct the control limits of fuzzy p chart using fuzzy data. They also proposed a ranking method to determine the process condition in linguistic form such as rather in control or rather out of control. Pandurangan and Varadharajan [23] proposed a control chart for fuzzy multinomial processes with variable sample size. In their approach, control limits for the fuzzy multinomial chart are obtained using multinomial distribution.

3. Highlights

The principle of fuzzy approaches proposed by Raz and Wang [2] and other researchers in this field are like the generalized p-chart, and each product unit is categorized with a linguistic variable, whereas each product unit might belong to several linguistic variables simultaneously in a vague environment. This statement is declared by Wang and Raz [11] themselves as "in a term set consisting of t linguistic values, each sample is completely specified by a t-dimensional vector with elements corresponding to the number of items in the sample describing each linguistic value. This vector is a random variable from a multinomial distribution." Other researchers have also indicated that Raz and Wang do not use fuzzy logic correctly. This disadvantage is also declared by Kandel et al. [16], Dubois and Prade [24], and Laviolette et al. [14]. Unfortunately, all of the recent methods model their approach based on a multinomial distribution without considering the fact that maybe an item could belong to two or even more categories at the same time.

4. Materials and Methods

In this research, for the first time, we try to use a fuzzy inference system to transfer the subjective rating of the quality of the products by the inspectors to a crisp number, so that we can use any variable control chart to monitor the quality of the process. Consider that the attribute characteristics of a specific product would be considered as a linguistic variable in the antecedent of an if-then rule which consists of two terms, *good* and *fair*. The quality of the product is considered as the linguistic variable in the consequent, which consists of two terms, *conforming* and *nonconforming*. Therefore, by considering the number of linguistic variables and their terms, it can be concluded that the fuzzy system used in this approach consists of two if-then rules as below.

Rule 1. If the quality characteristic is "good" then the quality is "conform".

Rule 2. If the quality characteristic is "fair" then the quality is "nonconforming".

Detailed construction procedures appear in the future step by step, followed by an example. In the following, we provide a step by step description of the construction of the fuzzy inference system and monitor the process.

Step 1 (fuzzify input). Before the rules can be evaluated, the inputs must be fuzzified according to each of the linguistic sets. So the second step is to take the inputs (scores), which are crisp integer numbers and determine the degree to which they belong to the appropriate fuzzy sets via membership functions:

$$\alpha_r = \mu^i \left(x^{\text{input}}\right), \tag{1}$$

Table 1: Estimated parameters of the "yellowness" and "blackness" membership function.

	Yellowness			Black	
	β_0	β_1		β_0	β_1
Coefficient value	−3.8672	0.849	Coefficient value	4.0568	−0.716
t-student value	6.01	23.08	t-student value	6.81	25.82

Yellowness: R-adj = 96.8, $F = 219.73$, $a = -\beta_1 = -0.849$, $c = \beta_0/a = 4.556$.
Black: R-adj = 95.2, $F = 321.69$, $a = -\beta_1 = 0.716$, $c = \beta_0/a = 5.667$.

where i = "fair", "good" and $r = 1, 2$ (rules number), then α_r which is a single truth value will be applied to the output function.

Step 2 (apply implication method). A consequent is a fuzzy set represented by a membership function and is reshaped using a function associated with the antecedent (α_r),

$$\mu_r^{\text{conseq}}(u) = \min\left\{\alpha_r, \mu^j(u)\right\}. \tag{2}$$

The input for the implication process is a single number given by the antecedent, and the output is a fuzzy set.

Step 3 (aggregate all outputs). Since decisions are based on the testing of all rules in an FIS, the rules must be combined in some manner in order to make a decision. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set,

$$\mu^{\text{conseq}}(u) = \max_r\left\{\mu_r^{\text{conseq}}(u)\right\}. \tag{3}$$

The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process is one fuzzy set for each output variable.

Step 4 (defuzzify). The input for the defuzzification process is a fuzzy set (the aggregate output fuzzy set), and the output is a single number. As much as fuzziness helps rule evaluation during the intermediate steps, the final desired output for each variable is generally a single number. However, the aggregate of a fuzzy set encompasses a range of output values and so must be defuzzified in order to resolve a single output variable from the set.

There are different ways for defuzzifying, the most popular of which are the center of area (COA) and the mean of maxima (MOM). We use COA method which returns the center of area under the curve,

$$U^{\text{COA}} = \frac{\int_u u \cdot \mu^{\text{conseq}}(u) \cdot du}{\int_u \mu^{\text{conseq}}(u) \cdot du}. \tag{4}$$

Step 5 (monitoring). Finally, in the last step we can monitor the outputs of the fuzzy systems which are crisp continuous data representing the quality of the product unit with traditional control charts.

A numerical example is used to evaluate the proposed approach. After the numerical example, a comparison study is performed based on average run length (ARL) to compare the performance of proposed approach with that of current related approaches.

5. Results and Discussion

In this section, we employ monitoring color problem of boats as an example to illustrate our approach. A boat factory intends to monitor the color of its products as one of the important quality characteristics. The color should be black and does not have any yellowness. So, the rules are formed as below.

Rule 1. If the color is black then the quality is conform.

Rule 2. If the color is yellowness then the quality is nonconforming.

After collecting 30 observations, "a" and "c" are estimated by using a regression model as illustrated in Table 1.

Now, by taking a shift in 25 preliminary samples of 20 rated color of boats by inspectors, the parameters "p" and "q" are determined by using a simulation programming with the goal of minimizing the ARL_1 as 0.1, 0.2. These values of "p" and "q" can be used in the future.

Suppose the color of one boat is rated 8 by an inspector, so we can get the color as "black" with degree of 0.8416 and "yellowness" with degree of 0.2548.

Accordingly, the consequences of the rules are

$$\begin{aligned} \mu_1^{\text{conseq}}(u) &= \min\left\{0.8416, \mu^{\text{black}}(u)\right\}, \\ \mu_2^{\text{conseq}}(u) &= \min\left\{0.2548, \mu^{\text{yellow}}(u)\right\}. \end{aligned} \tag{5}$$

So

$$\mu^{\text{conseq}}(u) = \max\left\{\mu_1^{\text{conseq}}(u), \mu_2^{\text{conseq}}(u)\right\}. \tag{6}$$

And at the end by using COA defuzzification method we have

$$u^{\text{COA}} = 0.761. \tag{7}$$

6. Evaluation Criteria

To compare the performance of different proposed approaches for monitoring the categorical data, average run length (ARL) is suggested as an evaluation criteria.

ARL is the average of the number of samples which should occurr before a sample shows the out-of-control condition. As Montgomery [1] declared, if the observations from the process are not autocorrelated, ARL could be calculated based on the following equation for every type of traditional control chart,

$$\text{ARL} = \frac{1}{p}, \tag{8}$$

where p is the probability of being out of control limits for each points. It should be noted that there are two different ARLs: in control and out of control.

In control average run length is shown by ARL_0. It is the average of the number of samples which should occurr before a sample shows an out-of-control condition when the process is in fact in the state of in-control.

For a traditional type control charts with 3 sigma control limits, the probability of type I error which is the probability of being out-of-control of a point when the process is in fact in the control is equal to 0.0027. So, the ARL when the process is in the control is

$$\mathrm{ARL}_0 = \frac{1}{p} = \frac{1}{0.0027} \cong 3701. \tag{9}$$

It means that, averagely, after each 370 points, a point shows an alarm of out-of-control when the process is in fact in the state of in control.

An average run length when the process is out-of-control is shown by ARL_1. ARL_1 is the average of the number of samples which take place until a point shows an out-of-control condition when the process is in fact out-of-control. ARL_1 could be calculated by the following equation:

$$\mathrm{ARL}_1 = \frac{1}{1-\beta}, \tag{10}$$

where β is the probability of not detecting a shift with the first point after the occurrance of a shift in the process.

7. Comparison Study

Here, by using simulation with MATLAB release R2009a, a comparison study was run to compare the performance of a proposed approach with the current related approach. Proposed approach, probabilistic approach proposed by Raz and Wang [2], generalized p chart proposed by Marcucci [7], and α-cut approach proposed by Gülbay and Kahraman [20] are considered in the comparison study. As Raz and Wang [2] and Taleb and Limam [3] declared that the probabilistic approach has a better performance over the membership approach; however just the probabilistic approach is considered in this comparison study. Base variable in this comparison study consists of four linguistic terms: standard (S), second choice (SC), third choice (TC), and chipped (C). Each linguistic term has its own membership function as below:

$$\mu_{\mathrm{S}}(x) = \begin{cases} 0 & x \le 0 \\ -x+1 & 0 \le x \le 1 \\ 0 & x \ge 1, \end{cases}$$

$$\mu_{\mathrm{SC}}(x) = \begin{cases} 0 & x \le 0 \\ 4x & 0 \le x \le \frac{1}{4} \\ -\frac{4}{3}x + \frac{4}{3} & \frac{1}{4} \le x \le 1 \\ 0 & x \ge 1, \end{cases}$$

$$\mu_{\mathrm{TC}}(x) = \begin{cases} 0 & x \le 0 \\ 2x & 0 \le x \le \frac{1}{2} \\ -2x+2 & \frac{1}{2} \le x \le 1 \\ 0 & x \ge 1, \end{cases}$$

$$\mu_{\mathrm{C}}(x) = \begin{cases} 0 & x \le 0 \\ x & 0 \le x \le 1 \\ 0 & x \ge 1. \end{cases} \tag{11}$$

Raz and Wang [2] showed that there are not any theoretical advantages over the using of different transformation techniques, so in this study fuzzy mode is used as the transformation technique for probabilistic approach.

Table 2 shows the representative values for different membership functions based on fuzzy mode and fuzzy median. It also shows the relationship between the score used for ranking the production items by the inspectors and the linguistic terms used to run other approaches.

As mentioned before, for generating the data and running the simulation, MATLAB release R2009a has been used. For generating the data, first random data was generated based on beta distribution with parameters α and β. Then generated data was multiplied by 10 and at last by using floor function, we could have discrete number from 0 to 10,

$$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}; \quad 0 \le x \le 1. \tag{12}$$

The final observations were used as the input of the fuzzy system. By using Table 1, the data for simulating the other approaches could be used.

Here a beta distribution with parameter $\alpha = 9$ and $\beta = 2$ was used. Figure 1 depicted this distribution.

Tables 3, 4, 5, and 6 show the ARL_1 which is obtained from a 10000 replication of generating data with sample size 5 when there is a shift equal to 0.5σ to 2σ in the process.

Some results could be obtained from this comparison study as below:

(i) proposed approach has a better performance in every cases,

(ii) especially in small shifts and small sample size, the proposed approach could detect the abnormal condition faster than other approaches,

(iii) comparing between generalized p chart and probabilistic approach shows that in every case the generalized p chart has a better performance,

(iv) α-cut approach has the weakest performance among these methods.

8. Conclusion

The first note in this approach is that variable quality characteristics are also better to consider as attribute and categorical quality characteristics. But, control charts for monitoring

Table 2: Representative value for linguistic terms.

Score	Linguistic term	Fuzzy mode	Fuzzy median
1–4	S	0	0.293
5-6	SC	0.25	0.387
7-8	TC	0.5	0.5
9-10	C	1	1

Table 3: ARL_1 for different approach when there is a shift equal to 0.5σ.

Sample size	Proposed method	Probabilistic Wang	Generalized p Marcucci	α-cut Gulbay
5	1.091	14.983	27.911	33.54
10	1.001	5.373	21.276	24.83
15	1.047	3.362	6.573	7.01
20	1.032	2.531	5.217	5.35
50	1.017	1.73	1.525	1.62

Table 4: ARL_1 for different approach when there is a shift equal to 1σ.

Sample size	Proposed method	Probabilistic Wang	Generalized p Marcucci	α-cut Gulbay
5	1.0073	3.663	3.721	5.12
10	1.016	1.284	1.391	2.01
15	1.001	1.059	1.102	1.65
20	1.001	1.013	1.032	1.23
50	1	1	1	1.12

Table 5: ARL_1 for different approach when there is a shift equal to 1.5σ.

Sample size	Proposed method	Probabilistic Wang	Generalized p Marcucci	α-cut Gulbay
5	1.0051	1.2172	1.2262	2.011
10	1.0032	1.0041	1.0079	1.0821
15	1.0011	1.0026	1.0038	1.0439
20	1	1.0013	1.0020	1.0091
50	1	1.0007	1.0011	1.0039

Table 6: ARL_1 for different approach when there is a shift equal to 2σ.

Sample size	Proposed method	Probabilistic Wang	Generalized p Marcucci	α-cut Gulbay
5	1.0009	1.1181	1.1231	1.981
10	1.0008	1.0022	1.0038	1.0616
15	1.0005	1.0015	1.0024	1.0291
20	1	1.0009	1.0011	1.0031
50	1	1.0001	1.0006	1.0021

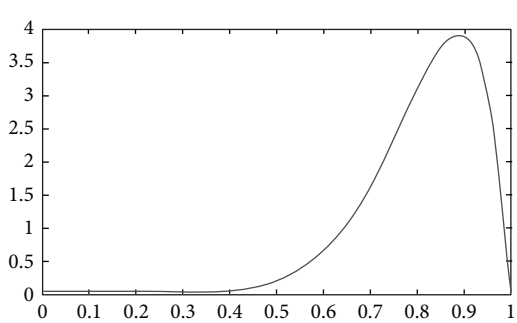

Figure 1: A beta distribution with $\alpha = 9$ and $\beta = 2$.

attribute quality characteristics in comparison to variable control charts have some disadvantages in structure which should be solved first. The second note is for monitoring attribute quality characteristics; which because of mental inspection and human judgments, have some level of vagueness and uncertainty. This research proposed a new approach to quality control, a fuzzy approach for monitoring the process when vagueness and uncertainty arise. The case study and comparison study show the proposed approach has a better performance and could detect abnormal shifts in the process, especially in small shifts and small sample size, faster than current related approaches.

References

[1] D. C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley & Sons, New York, NY, USA, 6th edition, 2009.

[2] T. Raz and J. H. Wang, "Probabilistic and membership approaches in the construction of control charts for linguistic data," *Production Planning and Control*, vol. 1, no. 3, pp. 147–157, 1990.

[3] H. Taleb and M. Limam, "On fuzzy and probabilistic control charts," *International Journal of Production Research*, vol. 40, no. 12, pp. 2849–2863, 2002.

[4] W. G. Cochran, "The chi square test of goodness of fit," *The Annals of Mathematical Statistics*, pp. 315–345, 1952.

[5] R. R. Yager and L. A. Zadeh, *An Introduction to Fuzzy Logic Applications in Intelligent Systems*, Kluwer Academic, 1992.

[6] A. Duncan, "A chi-square chart for controlling a set of percentages," *Industrial Quality Control*, vol. 7, pp. 11–15, 1950.

[7] M. Marcucci, "Monitoring multinomial processes," *Journal of Quality Technology*, vol. 17, no. 2, pp. 86–91, 1985.

[8] L. S. Nelson, "A chi-square control chart for several proportions," *Journal of Quality Technology*, vol. 19, no. 4, pp. 229–231, 1987.

[9] C. W. Bradshaw Jr., "A fuzzy set theoretic interpretation of economic control limits," *European Journal of Operational Research*, vol. 13, no. 4, pp. 403–408, 1983.

[10] R. H. Williams and R. M. Zigli, "Ambiguity impedes quality in the service industries," *Quality Progress*, vol. 20, no. 7, pp. 14–17, 1987.

[11] J. H. Wang and T. Raz, "On the construction of control charts using linguistic variables," *The International Journal of Production Research*, vol. 28, no. 3, pp. 477–487, 1990.

[12] A. Kanagawa, F. Tamaki, and H. Ohta, "Control charts for process average and variability based on linguistic data," *The International Journal of Production Research*, vol. 31, no. 4, pp. 913–922, 1993.

[13] F. Franceschini and D. Romano, "Control chart for linguistic variables: a method based on the use of linguistic quantifiers," *The International Journal of Production Research*, vol. 37, no. 16, pp. 3791–3801, 1999.

[14] M. Laviolette, J. W. Seamanb, J. D. Barrettc, and W. H. Woodallc, "A probabilistic and statistical view of fuzzy methods," *Technometrics*, vol. 37, no. 3, pp. 249–261, 1995.

[15] R. G. Almond, "Discussion: fuzzy logic: better science? Or better engineering?" *Technometrics*, vol. 37, no. 3, pp. 267–270, 1995.

[16] A. Kandel, A. Martins, and R. Pacheco, "Discussion: on the very real distinction between fuzzy and statistical methods," *Technometrics*, vol. 37, no. 3, pp. 276–281, 1995.

[17] W. Woodall, K. Tsui, and G. Tucker, "A review of statistical and fuzzy control charts based on categorical data," *Frontiers in Statistical Quality Control*, vol. 5, pp. 83–89, 1997.

[18] M. Gülbay, C. Kahraman, and D. Ruan, "α-cut fuzzy control charts for linguistic data," *International Journal of Intelligent Systems*, vol. 19, no. 12, pp. 1173–1195, 2004.

[19] M. Gülbay and C. Kahraman, "Development of fuzzy process control charts and fuzzy unnatural pattern analyses," *Computational Statistics and Data Analysis*, vol. 51, no. 1, pp. 434–451, 2006.

[20] M. Gülbay and C. Kahraman, "An alternative approach to fuzzy control charts: direct fuzzy approach," *Information Sciences*, vol. 177, no. 6, pp. 1463–1480, 2007.

[21] C.-B. Cheng, "Fuzzy process control: construction of control charts with fuzzy numbers," *Fuzzy Sets and Systems*, vol. 154, no. 2, pp. 287–303, 2005.

[22] M.-H. Shu and H.-C. Wu, "Monitoring imprecise fraction of nonconforming items using p control charts," *Journal of Applied Statistics*, vol. 37, no. 8, pp. 1283–1297, 2010.

[23] A. Pandurangan and R. Varadharajan, "Fuzzy multinomial control chart with variable sample size," *International Journal of Engineering Science*, vol. 3, 2011.

[24] D. Dubois and H. Prade, "Fuzzy sets—a convenient fiction for modeling vagueness and possibility," *IEEE Transactions on Fuzzy Systems*, vol. 2, no. 1, pp. 16–21, 1994.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 658635, 7 pages
http://dx.doi.org/10.1155/2013/658635

Research Article

An Extension of Cross Redundancy of Interval Scale Outputs and Inputs in DEA

Farhad Hosseinzadeh-Lotfi,[1] Gholam-Reza Jahanshahloo,[2] and Mansour Mohammadpour[1]

[1] *Department of Mathematics, Science and Research Branch, Islamic Azad University, Hesarak, Poonak, Tehran, Iran*
[2] *Faculty of Mathematical Science and Computer Engineering, University for Teacher Education, 599 Taleghani Avenue, Tehran 15618, Iran*

Correspondence should be addressed to Mansour Mohammadpour; mansour_mohammadpour@yahoo.com

Received 19 November 2012; Accepted 22 May 2013

Academic Editor: Hadi Nasseri

It is well known that data envelopment analysis (DEA) models are sensitive to selection of input and output variables. As the number of variables increases, the ability to discriminate between the decision making units (DMUs) decreases. Thus, to preserve the discriminatory power of a DEA model, the number of inputs and outputs should be kept at a reasonable level. There are many cases in which an interval scale output in the sample is derived from the subtraction of nonnegative linear combination of ratio scale outputs and nonnegative linear combination of ratio scale inputs. There are also cases in which an interval scale input is derived from the subtraction of nonnegative linear combination of ratio scale inputs and nonnegative linear combination of ratio scale outputs. Lee and Choi (2010) called such interval scale output and input a cross redundancy. They proved that the addition or deletion of a cross-redundant output variable does not affect the efficiency estimates yielded by the CCR or BCC models. In this paper, we present an extension of cross redundancy of interval scale outputs and inputs in DEA models. We prove that the addition or deletion of a cross-redundant output and input variable does not affect the efficiency estimates yielded by the CCR or BCC models.

1. Introduction

In many DEA applications, such as income, an interval scale output in the sample is derived from the subtraction of nonnegative linear combination of ratio scale outputs and nonnegative linear combination of ratio scale inputs. There are also many cases, like cost, in which an interval scale input is derived from the subtraction of nonnegative linear combination of ratio scale inputs and nonnegative linear combination of ratio scale outputs, although the effect of such dependencies on DEA is not clear. Lee and Choi [1] called such interval scale output and input a *cross redundancy*. They proved that the addition or deletion of a cross-redundant output variable does not affect the efficiency estimates yielded by the CCR or BCC models. Francisco J. López [2] generalized the contributions of Lee and Choi by introducing specific definitions and conducting some additional analysis on the impact of the presence of other types of linear dependencies among the inputs and outputs of a DEA model. In this paper, we deal with cross-redundant output and input variables simultaneously in DEA models. We prove that the addition or deletion of a cross-redundant output and input variable does not affect the efficiency estimates yielded by the CCR or BCC models. The paper is organized as follows. In Section 2, we introduce preliminaries of DEA. In Section 3, we present our main results. In Section 4, we will illustrate that the addition or deletion of cross-redundant output variable and input variable does not affect the efficiency estimates yielded by the CCR or BCC models. Conclusions are summarized in Section 5.

2. Preliminaries

Suppose that we have $n \geq 2$ peer observed DMUs, $\{\text{DMU}_j : j = 1, 2, \ldots, n\}$ which produce multiple outputs $y_{rj}, (r =$

$1,\ldots,s$), by utilizing multiple inputs $x_{ij}, (i = 1,\ldots,m)$. The input and output vectors of DMU_j are denoted by $\mathbf{x}_j$ and $\mathbf{y}_j$, respectively, and we assume that $\mathbf{x}_j$ and $\mathbf{y}_j$ are semipositive, that is, $\mathbf{x}_j \geq 0, \mathbf{x}_j \neq 0$ and $\mathbf{y}_j \geq 0, y_j \neq 0$ for $i = 1,\ldots,n$. We use $(\mathbf{x}_j, \mathbf{y}_j)$ to descript DMU_j and specially use $(\mathbf{x}_o, \mathbf{y}_o)$ (o element of $\{1,2,\ldots,n\}$) as the DMU under evaluation. Throughout this paper, vectors will be denoted by bold letters.

The input-oriented CCR [3] multiplier model evaluates the efficiency of each DMU_o by solving the following linear program:

$$\begin{aligned} \theta^* = \max \quad & \mathbf{u}^t\mathbf{y}_o, \\ & \mathbf{v}^t\mathbf{x}_o = 1 \\ \text{s.t.} \quad & \mathbf{u}^t\mathbf{y}_j \leq \mathbf{v}^t\mathbf{x}_j, \quad j = 1,\ldots,n, \\ & \mathbf{u} \geq o,\ \mathbf{v} \geq o. \end{aligned} \tag{1}$$

Because $\mathbf{x}_j$ and $\mathbf{y}_j$ are semipositive for $j = 1,\ldots,n$, $\theta^* > 0$. Also since $\mathbf{u}^t\mathbf{y}_o \leq \mathbf{v}^t\mathbf{x}_o$ and $\mathbf{v}^t\mathbf{x}_o = 1$, we have $\theta^* \leq 1$. Thus $0 < \theta^* \leq 1$. θ^* represents the input-oriented CCR-efficiency value of DMU_o.

The output-oriented CCR multiplier model evaluates the efficiency of each DMU_o by solving the following linear program:

$$\begin{aligned} \varphi^* = \min \quad & \mathbf{v}^t\mathbf{x}_o, \\ & \mathbf{u}^t\mathbf{y}_o = 1 \\ \text{s.t.} \quad & \mathbf{v}^t\mathbf{x}_j \geq \mathbf{u}^t\mathbf{y}_j, \quad j = 1,\ldots,n, \\ & \mathbf{u} \geq 0,\ \mathbf{v} \geq 0. \end{aligned} \tag{2}$$

Since $\mathbf{u}^t\mathbf{y}_o \leq \mathbf{v}^t\mathbf{x}_o$ and $\mathbf{u}^t\mathbf{y}_o = 1$, we have $\varphi^* \geq 1$. $1/\varphi^*$ represents the output-oriented CCR-efficiency value of DMU_o. Also $\theta^* = 1/\varphi^*$ [4].

The input-oriented BCC [4] multiplier model evaluates the efficiency of each DMU_o by solving the following linear program:

$$\begin{aligned} z^* = \max \ & \mathbf{u}^t\mathbf{y}_o + u_o, \\ & \mathbf{v}^t\mathbf{x}_o = 1 \\ \text{s.t.} \quad & \mathbf{u}^t\mathbf{y}_j + u_o \leq \mathbf{v}^t\mathbf{x}_j, \quad j = 1,\ldots,n, \\ & \mathbf{u} \geq \mathbf{0},\ \mathbf{v} \geq \mathbf{0},\ u_o \text{ is free.} \end{aligned} \tag{3}$$

Let $(\mathbf{u}^*, \mathbf{v}^*)$ be an optimal feasible solution for model (1); then $(\mathbf{u}^*, \mathbf{v}^*, u_o^*)$, where $u_o^* = 0$, will be a feasible solution of model (3). Thus $z^* \geq \theta^*$; therefore, $0 < z^* \leq 1$. z^* represents the input-oriented BCC-efficiency value of DMU_o.

Finally, the output-oriented BCC multiplier model evaluates the efficiency of each DMU_o by solving the following linear program:

$$\begin{aligned} t^* = \min \ & \mathbf{v}^t\mathbf{x}_o - v_o, \\ & \mathbf{u}^t\mathbf{y}_o = 1 \\ \text{s.t.} \quad & \mathbf{v}^t\mathbf{x}_j - v_o \geq \mathbf{u}^t\mathbf{y}_j, \quad j = 1,\ldots,n, \\ & \mathbf{u} \geq \mathbf{0},\ \mathbf{v} \geq \mathbf{0},\ v_o \text{ is free.} \end{aligned} \tag{4}$$

It can be easily confirmed that $t^* \geq 1$. $1/t^*$ represents the output-oriented BCC-efficiency value of DMU_o.

3. Main Results

In this section, we prove that the addition or deletion of a cross-redundant output variable and/or input variable does not affect the efficiency estimates yielded by the BCC multiplier model in input- and output-oriented versions. Similarly, it can be proved that the addition or deletion of cross-redundant variable does not affect efficiency estimates yielded by the CCR multiplier model in input- and output-oriented versions.

Theorem 1. *Let each DMU have $m+1$ inputs and $s+1$ outputs, that is, $\mathbf{x}_j = (x_{1j},\ldots,x_{(m+1)j})$ and $\mathbf{y}_j = (y_{1j},\ldots,y_{(s+1)j})$ for $j = 1,2,\ldots,n$. Let*

$$x_{(m+1)j} = \sum_{i=1}^{m}\beta_i x_{ij} - \sum_{r=1}^{s}\alpha_r y_{rj}; \quad j = 1,\ldots,n, \tag{5}$$

$$y_{(s+1)j} = \sum_{r=1}^{s} a_r y_{rj} - \sum_{i=1}^{m} b_i x_{ij}; \quad j = 1,\ldots,n, \tag{6}$$

where $\beta_i \geq 0$, $b_i \geq 0$, $i = 1,\ldots,m; \alpha_r \geq 0, a_r \geq 0, r = 1,\ldots,s$.

Then the optimal objective function value of the following model:

$$\begin{aligned} \rho^* = \max \quad & \sum_{r=1}^{s+1} p_r y_{ro} + p_o, \\ & \sum_{i=1}^{m+1} q_i x_{io} = 1, \\ \text{s.t} \quad & \sum_{r=1}^{s+1} p_r y_{ro} - \sum_{i=1}^{m+1} q_i x_{io} - p_o \leq 0, \quad j = 1,\ldots,n, \\ & p_r \geq 0, \quad q_i \geq 0, \quad r = 1,\ldots,s+1,\ i = 1,\ldots,m+1 \end{aligned} \tag{7}$$

is equal to the optimal objective function value of the following model (3).

Proof. Let $(p_1^*, \ldots, p_{s+1}^*, q_1^*, \ldots, q_{m+1}^*, p_o^*)$ be an optimal solution for model (7); then we have

$$\rho^* = \sum_{r=1}^{s+1} p_r^* y_{ro} - p_o^*, \tag{8}$$

$$\sum_{i=1}^{m+1} q_i^* x_{io} = 1, \tag{9}$$

$$\sum_{r=1}^{s+1} p_r^* y_{ro} - \sum_{i=1}^{m+1} q_i^* x_{io} - p_o^* \le 0. \tag{10}$$

By (6) and (9), it follows that

$$\rho^* = \sum_{r=1}^{s} \left(p_r^* + p_{s+1}^*\right) y_{ro} - \sum_{i=1}^{m} p_{s+1}^* b_i x_{io} + p_o^*. \tag{11}$$

Also, by (5) and (9) it concludes that

$$\sum_{i=1}^{m} \left(q_i^* + q_{m+1}^* \beta_i\right) x_{io} - \sum_{r=1}^{s} q_{m+1}^* \alpha_r y_{ro} = 1. \tag{12}$$

Now, let

$$\begin{gathered} \overline{v}_i = \frac{\left(q_i^* + q_{m+1}^* \beta_i\right) \rho^*}{AB} + \frac{p_{s+1}^* b_i}{A}, \quad \text{for } i = 1, \ldots, m, \\ \overline{u}_r = \frac{\left(p_r^* + p_{s+1}^* a_r\right) \rho^*}{AB} + \frac{\left(q_{m+1}^* \alpha_r\right) \rho^*}{B}, \quad \text{for } i = 1, \ldots, m, \\ \overline{u}_o = \frac{p_o^* \rho^*}{AB}, \end{gathered} \tag{13}$$

where

$$\begin{gathered} A = \sum_{r=1}^{s} \left(p_r^* + p_{s+1}^* a_r\right) y_{ro} + p_o^* \\ B = 1 + \sum_{r=1}^{s} q_{m+1}^* \alpha_r y_{ro}. \end{gathered} \tag{14}$$

Then, by (7), we have

$$\overline{v}_i \ge 0, \quad i = 1, \ldots, m, \quad \overline{u}_r \ge 0, \quad r = 1, \ldots, s. \tag{15}$$

Also, by (12) and (13), we obtain

$$\begin{aligned} \sum_{i=1}^{m} \overline{v}_i x_{io} &= \frac{\rho^*}{AB} \sum_{i=1}^{m} \left(q_i^* + q_{m+1}^* \beta_i\right) x_{io} \\ &+ \frac{1}{A} \sum_{i=1}^{m} \left(p_{s+1}^* b_i\right) x_{io} = \frac{\rho^* (B)}{AB} + \frac{1}{A}\left(A - \rho^*\right) = 1. \end{aligned} \tag{16}$$

In addition,

$$\begin{aligned} \theta^* &\ge \sum_{r=1}^{s} \overline{u}_r y_{ro} - \overline{u}_o = \frac{\rho^*}{AB} \sum_{r=1}^{s} \left(p_r^* + p_{s+1}^* a_r\right) y_{ro} \\ &\quad + \frac{z_2^*}{B} \sum_{r=1}^{s} \left(q_{m+1}^* \alpha_r\right) - \frac{p_o^* \rho^*}{AB} \\ &= \frac{\rho^*}{AB} \left(\sum_{r=1}^{s} \left(p_r^* + p_{s+1}^* a_r\right) y_{ro} - p_o^* \right) + \frac{\rho^* (B-1)}{B} \\ &= \frac{z_2^*}{AB} (A) + \frac{\rho^* (B-1)}{B} = \rho^*. \end{aligned} \tag{17}$$

Also

$$\begin{aligned} &\sum_{i=1}^{m} \overline{v}_i x_{ij} - \sum_{r=1}^{s} \overline{u}_r y_{rj} + \overline{u}_o \\ &\quad = \frac{1}{AB} \sum_{i=1}^{m} \rho^* \left(q_i^* + q_{m+1}^* \beta_i\right) x_{ij} + \frac{1}{A} \sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} \\ &\qquad - \frac{1}{AB} \sum_{r=1}^{s} \rho^* \left(p_r^* + p_{s+1}^* a_r\right) y_{rj} \\ &\qquad - \frac{1}{B} \sum_{r=1}^{s} \rho^* \left(q_{m+1}^* \alpha_r\right) y_{rj} + \frac{1}{AB} p_o^* \rho^* \\ &\quad = \frac{1}{AB} \left[\rho^* \left(\sum_{i=1}^{m} q_i^* x_{ij} - \sum_{r=1}^{s} p_r^* y_{rj} + p_o^* \right) \right. \\ &\qquad + \rho^* \left(\sum_{i=1}^{m} q_{m+1}^* \beta_i x_{ij} - \sum_{r=1}^{s} p_{s+1}^* a_r y_{rj} \right) \\ &\qquad \left. + B \sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} - A \rho^* \sum_{r=1}^{s} \left(q_{m+1}^* \alpha_r\right) y_{rj} \right]. \end{aligned} \tag{18}$$

So that by (10) we have

$$\begin{aligned} &\sum_{i=1}^{m} \overline{v}_i x_{ij} - \sum_{r=1}^{s} \overline{u}_r y_{rj} + \overline{u}_o \\ &\quad \ge \frac{1}{AB} \left[\rho^* \left(\sum_{r=1}^{s} \left(p_{s+1}^* a_r + q_{m+1}^* \alpha_r\right) y_{rj} \right. \right. \\ &\qquad \left. - \sum_{i=1}^{m} \left(q_{m+1}^* \beta_i + p_{s+1}^* b_i\right) x_{ij} \right) \end{aligned}$$

$$+\rho^*\left(\sum_{i=1}^{m} q_{m+1}^*\beta_i x_{ij}-\sum_{r=1}^{s} p_{s+1}^* a_r y_{rj}\right)$$

$$+B\sum_{i=1}^{m} p_{s+1}^* b_i x_{ij}-A\rho^*\sum_{r=1}^{s}\left(q_{m+1}^*\alpha_r\right) y_{rj}\Bigg]$$

$$=\frac{1}{AB}\Bigg[\rho^*\sum_{r=1}^{s} q_{m+1}^*\alpha_r y_{rj}-\rho^*\sum_{i=1}^{m} p_{s+1}^* b_i x_{ij}$$

$$+B\sum_{i=1}^{m} p_{s+1}^* b_i x_{ij}-A\rho^*\sum_{r=1}^{s}\left(q_{m+1}^*\alpha_r\right) y_{rj}\Bigg]$$

$$=\frac{1}{AB}\left[\left(B-\rho^*\right)\sum_{i=1}^{m} p_{s+1}^* b_i x_{ij}+\rho^*\left(1-A\right)\sum_{r=1}^{s} q_{m+1}^*\alpha_r y_{rj}\right]. \tag{19}$$

Therefore,

$$\sum_{i=1}^{m}\overline{v}_i x_{ij}-\sum_{r=1}^{s}\overline{u}_r y_{rj}+\overline{u}_o$$

$$\geq\frac{1}{AB}\Bigg[\left(B-\rho^*\right)\sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} \tag{20}$$

$$+\rho^*\left(1-A\right)\sum_{r=1}^{s} q_{m+1}^*\alpha_r y_{rj}\Bigg]\geq 0.$$

Consequently, $(\overline{\mathbf{u}},\overline{\mathbf{v}},\overline{u}_o)$, where $\overline{\mathbf{u}}=(\overline{u}_1,\ldots,\overline{u}_s)$ and $\overline{\mathbf{v}}=(\overline{v}_1,\ldots,\overline{v}_m)$, is a feasible solution for model (1), which for.

$$\theta^*\geq\sum_{r=1}^{s}\overline{u}_r y_{ro}-\overline{u}_o=\frac{\rho^*}{AB}\sum_{r=1}^{s}\left(p_r^*+p_{s+1}^* a_r\right) y_{ro}$$

$$+\frac{z_2^*}{B}\sum_{r=1}^{s}\left(q_{m+1}^*\alpha_r\right)-\frac{p_o^*\rho^*}{AB}$$

$$=\frac{\rho^*}{AB}\left(\sum_{r=1}^{s}\left(p_r^*+p_{s+1}^* a_r\right) y_{ro}-p_o^*\right)+\frac{\rho^*\left(B-1\right)}{B} \tag{21}$$

$$=\frac{z_2^*}{AB}\left(A\right)+\frac{\rho^*\left(B-1\right)}{B}=\rho^*.$$

Now, let $(\mathbf{u}^*,\mathbf{v}^*,u_o^*)$ be an optimal solution for model (1); then $(\overline{\mathbf{p}},\overline{\mathbf{q}},\overline{p}_o)$, where $\overline{\mathbf{p}}=(\overline{p}_1,\ldots,\overline{p}_{s+1})$ and $\overline{\mathbf{q}}=(\overline{q}_1,\ldots,\overline{q}_{m+1})$, with $\overline{p}_r=\overline{u}_r$, $r=1,\ldots,s$; $\overline{p}_{s+1}=0$; $\overline{q}_i=\overline{v}_i$, $i=1,\ldots,m$; $\overline{q}_{m+1}=0$; $\overline{p}_o=u_o^*$, is a feasible solution for model (2), which for $\theta^*=\sum_{r=1}^{s}u_r^* y_{ro}-u_o^*=\sum_{r=1}^{s+1}\overline{p}_r y_{ro}-\overline{p}_o\leq\rho^*$. Thus $\theta^*=\rho^*$. □

Theorem 2. *Let each DMU have $m+1$ inputs and $s+1$ outputs with conditions* (5) *and* (6).

Then the optimal objective function value of the following model:

$$\begin{aligned}
w^*=\min\quad&\sum_{r=1}^{s+1} q_i x_{io}-q_o,\\
&\sum_{r=1}^{s+1} p_r y_{ro}=1\\
\text{s.t}\quad&\sum_{i=1}^{m+1} q_i x_{io}-\sum_{r=1}^{s+1} p_r y_{ro}-q_o\geq 0,\quad j=1,\ldots,n,\\
p_r\geq 0,\quad&q_i\geq 0,\quad r=1,\ldots,s+1,\ i=1,\ldots,m+1
\end{aligned} \tag{22}$$

is equal to the optimal objective function value of the following model (4).

Proof. Let $(p_1^*,\ldots,p_{s+1}^*,q_1^*,\ldots,q_{m+1}^*,q_o^*)$ be an optimal solution for model (22); then we have

$$\omega^*=\sum_{i=1}^{m+1} q_i^* x_{io}-q_o^* \tag{23}$$

$$\sum_{r=1}^{s+1} p_r^* y_{ro}=1 \tag{24}$$

$$\sum_{i=1}^{m+1} q_i^* x_{ij}-\sum_{r=1}^{s} p_r^* y_{rj}-q_o^*\geq 0. \tag{25}$$

By (6) and (15), it follows that

$$\omega^*=\sum_{i=1}^{m}\left(q_i^*+q_{m+1}^*\beta_i\right) x_{io}-q_o^*-\sum_{r=1}^{s} q_{m+1}^*\alpha_r y_{ro}. \tag{26}$$

Also, by (5) and (16) it concludes that

$$\sum_{r=1}^{s}\left(p_r^*+p_{s+1}^* a_r\right) y_{ro}-\sum_{i=1}^{m} p_{s+1}^* b_i x_{io}=1. \tag{27}$$

Now, let

$$\begin{aligned}
\overline{v}_i&=\frac{\left(q_i^*+q_{m+1}^*\beta_i\right) w^*}{AB}+\frac{w^* p_{s+1}^* b_i}{B},\qquad\text{for } i=1,\ldots,m,\\
\overline{u}_r&=\frac{\left(p_r^*+p_{s+1}^* a_r\right) w^*}{AB}+\frac{\left(q_{m+1}^*\alpha_r\right)}{A},\qquad\text{for } r=1,\ldots,s,\\
\overline{v}_o&=\frac{q_o^* w^*}{AB},
\end{aligned} \tag{28}$$

where

$$\begin{aligned}
A&=\sum_{i=1}^{m}\left(q_i^*+q_{m+1}^*\beta_i\right) x_{io}-q_o^*,\\
B&=1+\sum_{r=1}^{s} q_{m+1}^*\alpha_r y_{ro}.
\end{aligned} \tag{29}$$

Then, by (7), we have

$$\begin{aligned} \overline{v}_i &\geq 0, \quad i = 1, \ldots, m, \\ \overline{u}_r &\geq 0, \quad r = 1, \ldots, s. \end{aligned} \tag{30}$$

Also, by (26) and (27), we obtain

$$\begin{aligned} \sum_{r=1}^{s} \overline{u}_r y_{ro} &= \frac{w^*}{AB} \sum_{i=1}^{m} \left(p_r^* + p_{s+1}^* \alpha_r\right) y_{ro} \\ &\quad + \frac{w^*}{B} \sum_{i=1}^{m} \left(q_{m+1}^* \alpha_r\right) y_{ro} = \frac{w^* (B)}{AB} + \frac{\left(A - w^*\right)}{A} = 1 \cdot \end{aligned} \tag{31}$$

In addition

$$\begin{aligned} &\sum_{i=1}^{m} \overline{v}_i x_{ij} - \sum_{r=1}^{s} \overline{u}_r y_{rj} - \overline{v}_o \\ &\quad = \frac{1}{AB} \sum_{i=1}^{m} w^* \left(q_i^* + q_{m+1}^* \beta_i\right) x_{ij} \\ &\qquad + \frac{w^*}{B} \sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} \\ &\qquad - \frac{w^*}{AB} \sum_{r=1}^{s} \left(p_r^* + p_{s+1}^* a_r\right) y_{rj} \\ &\qquad - \frac{1}{A} \sum_{r=1}^{s} \left(q_{m+1}^* \alpha_r\right) y_{rj} - \frac{1}{AB} q_o^* w^* \\ &\quad = \frac{1}{AB} \left[w^* \left(\sum_{i=1}^{m} q_i^* x_{ij} - \sum_{r=1}^{s} p_r^* y_{rj} - q_o^* \right) \right. \\ &\qquad + w^* \left(\sum_{i=1}^{m} q_{m+1}^* \beta_i x_{ij} - \sum_{r=1}^{s} p_{s+1}^* a_r y_{rj} \right) \\ &\qquad \left. + A w^* \sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} - B \sum_{r=1}^{s} \left(q_{m+1}^* \alpha_r\right) y_{rj} \right]. \end{aligned} \tag{32}$$

So that by (14), we have

$$\begin{aligned} &\sum_{i=1}^{m} \overline{v}_i x_{ij} - \sum_{r=1}^{s} \overline{u}_r y_{rj} - \overline{v}_o \\ &\quad \geq \frac{1}{AB} \left[w^* \left(\sum_{r=1}^{s} \left(p_{s+1}^* a_r + q_{m+1}^* \alpha_r\right) y_{rj} \right. \right. \\ &\qquad \left. - \sum_{i=1}^{m} \left(q_{m+1}^* \beta_i + p_{s+1}^* b_i\right) x_{ij} \right) \\ &\qquad + w^* \left(\sum_{i=1}^{m} q_{m+1}^* \beta_i x_{ij} - \sum_{r=1}^{s} p_{s+1}^* a_r y_{rj} \right) \\ &\qquad \left. + A w^* \sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} - B \sum_{r=1}^{s} \left(q_{m+1}^* \alpha_r\right) y_{rj} \right] \\ &\quad = \frac{1}{AB} \left[w^* \sum_{r=1}^{s} q_{m+1}^* \alpha_r y_{rj} - w^* \sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} \right. \\ &\qquad \left. + A w^* \sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} - B \sum_{r=1}^{s} \left(q_{m+1}^* \alpha_r\right) y_{rj} \right] \\ &\quad = \frac{1}{AB} \left[w^* (A - 1) \sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} \right. \\ &\qquad \left. + \left(w^* - B\right) \sum_{r=1}^{s} q_{m+1}^* \alpha_r y_{rj} \right]. \end{aligned} \tag{33}$$

Therefore,

$$\begin{aligned} &\sum_{i=1}^{m} \overline{v}_i x_{ij} - \sum_{r=1}^{s} \overline{u}_r y_{rj} - \overline{v}_o \\ &\quad \geq \frac{1}{AB} \left[w^* (A - 1) \sum_{i=1}^{m} p_{s+1}^* b_i x_{ij} \right. \\ &\qquad \left. + \left(w^* - B\right) \sum_{r=1}^{s} q_{m+1}^* \alpha_r y_{rj} \right] \geq 0. \end{aligned} \tag{34}$$

Consequently, $(\overline{\mathbf{u}}, \overline{\mathbf{v}}, \overline{v}_o)$, where $\overline{\mathbf{u}} = (\overline{u}_1, \ldots, \overline{u}_s)$ and $\overline{\mathbf{v}} = (\overline{v}_1, \ldots, \overline{v}_m)$, is a feasible solution for model (4), which for

$$\begin{aligned} z^* &\leq \sum_{i=1}^{m} \overline{v}_i x_{io} - \overline{v}_o \\ &= \frac{w^*}{AB} \sum_{i=1}^{m} \left(q_i^* + q_{m+1}^* \beta_r\right) x_{io} \\ &\quad + \frac{w^*}{B} \sum_{r=1}^{s} \left(p_{s+1}^* b_i\right) x_{io} - \frac{q_o^* w^*}{AB} \\ &= \frac{w^*}{AB} (A) + \frac{w^* (B - 1)}{B} = w^*. \end{aligned} \tag{35}$$

Now let $(\mathbf{u}^*, \mathbf{v}^*, v_o^*)$ be an optimal solution for model (4), and then $(\overline{\mathbf{p}}, \overline{\mathbf{q}}, \overline{p}_o)$, where $\overline{\mathbf{p}} = (\overline{p}_1, \ldots, \overline{p}_{s+1})$ and $\overline{\mathbf{q}} = (\overline{q}_1, \ldots, \overline{q}_{m+1})$, with $\overline{p}_r = \overline{u}_r$, $r = 1, \ldots, s$; $\overline{p}_{s+1} = 0$; $\overline{q}_i = \overline{v}_i$, $i = 1, \ldots, m$; $\overline{q}_{m+1} = 0$; $\overline{p}_o = v_o^*$, is a feasible solution for model (22), which for $\omega^* \geq \sum_{r=1}^{s+1} \overline{p}_r y_{ro} - \overline{p}_o = \sum_{r=1}^{s} u_r^* y_{ro} - u_o^* = z^*$. Thus $z^* = w^*$. □

Theorem 3. *Let each DMU have $m+1$ inputs and $s+1$ outputs with conditions* (5) *and* (6).

Then, the optimal objective function value of the following model:

$$\begin{aligned} \tilde{\rho} = \max \quad & \sum_{r=1}^{s+1} p_r y_{ro}, \\ & \sum_{i=1}^{m+1} q_i x_{io} = 1 \\ \text{s.t} \quad & \sum_{r=1}^{s+1} p_r y_{rj} - \sum_{i=1}^{m+1} q_i x_{ij} \le 0, \quad j = 1, \ldots, n \\ p_r \ge 0, \quad & q_i \ge 0, \quad r = 1, \ldots, s+1, \ i = 1, \ldots, m+1 \end{aligned} \tag{36}$$

is equal to the optimal objective function value of the following model (1).

Proof. This proof is similar to the proof of Theorem 1. □

Theorem 4. *Let each DMU have $m+1$ inputs and $s+1$ outputs with conditions* (5) *and* (6).

Then, the optimal objective function value of the following model:

$$\begin{aligned} \tilde{w} = \min \quad & \sum_{r=1}^{s+1} q_i x_{io}, \\ & \sum_{r=1}^{s+1} p_r y_{ro} = 1 \\ \text{s.t} \quad & \sum_{i=1}^{m+1} q_i x_{ij} - \sum_{r=1}^{s+1} p_r y_{rj} \ge 0, \quad j = 1, \ldots, n, \\ p_r \ge 0, \quad & q_i \ge 0, \quad r = 1, \ldots, s+1, \ i = 1, \ldots, m+1 \end{aligned} \tag{37}$$

is equal to the optimal objective function value of the following model (2).

Proof. This proof is similar to the proof of Theorem 2. □

4. Illustrative Example

In this section, we use the data recorded in Table 1 to illustrate that the addition or deletion of a cross-redundant output variable and input variable does not affect the efficiency estimates yielded by the CCR or BCC models. These correspond to 20 DMUs, whose efficiency is assessed using four inputs and four outputs where

$$\begin{aligned} x_{4j} &= \left(x_{1j} + x_{2j} + 2x_{3j}\right) \\ &\quad - \left(0.5y_{1j} + 0.5y_{2j} + 0.5y_{3j}\right); \\ &\qquad j = 1, \ldots, n, \\ y_{4j} &= \left(0.5y_{1j} + 0.5y_{2j} + 0.5y_{3j}\right) - 0.5x_{3j}; \\ &\qquad j = 1, \ldots, n. \end{aligned} \tag{38}$$

Table 1: Dataset.

	Inp 1	Inp 2	Inp 3	Inp 4	Out 1	Out 1	Out 3	Out 4
Unit 1	7	1	4	12.75	1	2.5	3	1.25
Unit 2	3	7	4	14.75	2.5	1	3	1.25
Unit 3	6	6	3	14.25	2.5	2	3	2.25
Unit 4	3	1	3	1.75	4	5.5	7	6.75
Unit 5	6	0.5	3	5.25	5	3.5	6	5.75
Unit 6	4	0.5	3	3.5	2	6	6	5.5
Unit 7	1.5	2.5	3	1.5	6	4	7	7
Unit 8	0.5	4	4	6.25	1.5	5	6	4.25
Unit 9	2.75	1.75	4	4	8	3	6	6.5
Unit 10	1	3	3	1	8	3	7	7.5
Unit 11	2	2	3	1.25	5.5	5	7	7.25
Unit 12	2.5	1.5	3	2	7	3	6	6.5
Unit 13	4.5	1.5	6	13	4	2	4	2
Unit 14	2	4	7	16.25	1.5	2	4	0.25
Unit 15	4	3	6	12.25	6.5	3.5	3.5	3.75
Unit 16	2	5	4	8.75	5	3.5	4	4.25
Unit 17	1.5	6	4	8.5	4.5	4.5	5	5
Unit 18	0.5	4	3	3	3.5	5.5	6	6
Unit 19	3.5	0.75	3	2.5	7.5	2.5	6	6.5
Unit 20	6	3.5	4	11	3.5	3.5	6	4.5

Table 2: Example results.

	θ^*	$\tilde{\rho}$	z^*	ρ^*
Unit 1	0.3461538	0.3461538	0.7500000	0.7500000
Unit 2	0.3214286	0.3214286	0.7500000	0.7500000
Unit 3	0.4285714	0.4285714	1.0000000	1.0000000
Unit 4	1.0000000	1.0000000	1.0000000	1.0000000
Unit 5	1.0000000	1.0000000	1.0000000	1.0000000
Unit 6	1.0000000	1.0000000	1.0000000	1.0000000
Unit 7	1.0000000	1.0000000	1.0000000	1.0000000
Unit 8	1.0000000	1.0000000	1.0000000	1.0000000
Unit 9	0.9973190	0.9973190	1.0000000	1.0000000
Unit 10	1.0000000	1.0000000	1.0000000	1.0000000
Unit 11	1.0000000	1.0000000	1.0000000	1.0000000
Unit 12	1.0000000	1.0000000	1.0000000	1.0000000
Unit 13	0.4444444	0.4444444	0.6666667	0.6666667
Unit 14	0.3809524	0.3809524	0.6666667	0.6666667
Unit 15	0.5607702	0.5607702	0.5714286	0.5594240
Unit 16	0.6052279	0.6052279	0.7500000	0.7500000
Unit 17	0.6847156	0.6847156	0.7500000	0.7500000
Unit 18	1.0000000	1.0000000	1.0000000	1.0000000
Unit 19	1.0000000	1.0000000	1.0000000	1.0000000
Unit 20	0.6428571	0.6428571	0.7500000	0.7500000

In other words, the forth input and the forth output are cross-redundant variables. In Table 2, θ^*, $\tilde{z}$, ρ^*, and $\tilde{\rho}$, respectively, record the efficiency measure provided by model (1), model (3), model (7), and model (36). It is evident from Table 2 that the addition or deletion of cross-redundant output variable

and/or input variable does not affect the efficiency estimates yielded by the input-oriented CCR or BCC multiplier models.

5. Conclusions

In this paper, we have studied the effect of the cross redundancy between interval scale input and output variables on the efficiency estimates yielded by the CCR multiplier model in input- and output-oriented versions and the BCC multiplier model in input- and output-oriented versions. We proved that the addition or deletion of a cross-redundant output variable and input variable does not affect the efficiency estimates yielded by the input-oriented BCC multiplier model and the output-oriented BCC multiplier model. Similarly, it can be proved that the addition or deletion of cross-redundant variable does not affect efficiency estimates yielded by the CCR multiplier model in input- and output-oriented versions.

References

[1] K. Lee and K. Choi, "Cross redundancy and sensitivity in DEA models," *Journal of Productivity Analysis*, vol. 34, no. 2, pp. 151–165, 2010.

[2] F. J. López, "Generalizing cross redundancy in data envelopment analysis," *European Journal of Operational Research*, vol. 214, no. 3, pp. 716–721, 2011.

[3] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[4] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, no. 9, pp. 1078–1092, 1984.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 138272, 12 pages
http://dx.doi.org/10.1155/2013/138272

Research Article

Structural Credit Risk Models with Subordinated Processes

Martin Gurny,[1,2] Sergio Ortobelli Lozza,[1,3] and Rosella Giacometti[1]

[1] *Department of Management, Economics and Quantitative Methods, University of Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy*
[2] *Department of Applied Finance and Actuarial Studies, Macquarie University, Eastern Road, North Ryde, NSW 2109, Australia*
[3] *Department of Finance, Faculty of Economics, VŠB-Technical University of Ostrava, Sokolská 33, 70121 Ostrava, Czech Republic*

Correspondence should be addressed to Sergio Ortobelli Lozza; sol@unibg.it

Received 31 January 2013; Accepted 15 July 2013

Academic Editor: Xiaoning Zhang

We discuss structural models based on Merton's framework. First, we observe that the classical assumptions of the Merton model are generally rejected. Secondly, we implement a structural credit risk model based on stable non-Gaussian processes as a representative of subordinated models in order to overcome some drawbacks of the Merton one. Finally, following the KMV-Merton estimation methodology, we propose an empirical comparison between the results obtained from the classical KMV-Merton model and the stable Paretian one. In particular, we suggest alternative parameter estimation for subordinated processes, and we optimize the performance for the stable Paretian model.

1. Introduction

Estimating a borrower's risk level, namely, the probability of default (PD), by assigning an appropriate PD is a widely employed strategy by many financial institutions as well as the supervisory authorities. PD indicates a probability that a given counterparty will not be able to meet its obligations. The incorrect estimation of PD leads to, among other things, unreasonable ratings and incorrect pricing of financial instruments, and thereby it is one of the causes of the recent global financial crisis. Undervaluation of the risk caused the collapse of the financial system which has been extended through credit derivatives on the global markets. PD is also a crucial parameter used in the calculation of economic or regulatory capital, under the Basel II and Basel III Accords for banking institutions. These reasons highlight how important the estimation of PD is and why it has been a significant research topic for a long time.

The probability of default, as one of the key risk parameters in the IRB approach, has many methodologies for its estimation. In general, we can classify the existing methodologies into three groups: structural models, reduced-form models, and credit-scoring (statistical) models. We will focus on the first type of models only in this paper. This structural approach was proposed in 1974 by Robert Merton [1] in his seminal paper on the valuation of corporate debt. Largely as a logical extension of the Black-Scholes [2] option pricing framework in 1973, he introduced a model for assessing the credit risk of a company by characterizing a company's equity as a derivative on its assets.

A number of researchers have examined the contribution of the Merton model over the past several years. An overview of structural credit risk models can be found in Bluhm et al. [3] and in Duffie and Singleton [4]. Practitioners employed by either Moody's or KMV were the first ones who analysed Merton model carefully. Moreover, the KMV default probability model is summarized by Crosbie and Bohn [5]. Bohn et al. [6] argue that the KMV-Merton model captures all of the information in traditional agency ratings and well-known accounting variables. The model's predictive power is examined, for instance, by Du and Suo [7] and Hillegeist et al. [8]. Duffie et al. [9] show that the KMV-Merton model probabilities have significant predictive power in modelling default probabilities over time. Farmen et al. [10] investigate default probabilities and their comparative statics in the Merton framework using objective probability measure. The main theoretical models of risky debt valuation built on Merton [1] and Black and Cox [11] are discussed in Bohn

[12]. In the literature on bank deposit insurance, a contingent claim valuation of equity has been used extensively. In this case, the equity call model is reversed to generate estimates of the market value of assets from observed share prices. This approach allows for the calculation of fair deposit insurance premium. Duan [13] proposes another method for estimating asset value and volatility which is based on the maximum likelihood estimation using equity prices.

The Merton model requires a number of simplifying assumptions (the company can default only at debt's maturity time T but not before; the model is not able to distinguish among the different types of debt, constant and flat term structure of interest rates, etc.). Notwithstanding, one of the most important drawbacks is an assumption that company value follows the log-normal distribution. It is well known that log-returns of equities are not Gaussian distributed, and several empirical investigations have shown that log-returns of equities present skew distributions with excess kurtosis which leads to a greater density in the tails, and that the normal distribution with a comparatively thinner tail simply cannot describe this phenomenon (see Mandelbrot [14–16], Fama [17–19], or Rachev and Mittnik [20]).

The main contribution of this paper is twofold. First, we introduce a structural credit risk model based on the stable Paretian distributions as a representative of subordinated models. Secondly, we show that it is possible to use this model in the Merton's framework, and we propose an empirical comparison of the KMV methodology applied to the Merton model and our subordinated one. In particular, we prove that the basic assumption of the Merton model is generally rejected, and thus the log-returns of the company's assets value are not Gaussian distributed. For this reason, we discuss the possibility for using other subordinated processes to approximate the behaviour of the log-returns of the company value. Thus, we propose to use the Hurst et al. [21] option pricing model based on the stable Paretian distributions which generalizes the standard Merton methodology.

The practical and theoretical appeal of the stable non-Gaussian approach is given by its attractive properties that are almost the same as the normal ones. As a matter of fact, the Gaussian law is a particular stable Paretian one, and thus the stable Paretian model is a generalization of the Merton one. The first relevant desirable property of the stable distributional assumption is that stable distributions have domain of attraction. The generalized central limit theorem for the normalized sums of i.i.d. random variables determines the domain of attraction of each stable law. Therefore, any distribution in the domain of attraction of a specified stable distribution will have properties close to those of the stable distribution. Another attractive aspect of the stable Paretian assumption is the stability property; that is, stable distributions are stable with respect to summation of i.i.d. random stable variables. Hence, the stability governs the main properties of the underlying distribution. In addition, in the empirical financial literature, it is well documented that the asset returns have a distribution whose tail is heavier than that of the distributions with finite variance.

The idea of using subordinated stable Paretian processes goes back to the seminal work of Mandelbrot and Taylor [22]. Stable laws have been applied in several financial sectors (see Rachev [23] and Rachev and Mittnik [20]). For these reasons, the stable Paretian law is the first candidate as a subordinated model investigating for credit risk modeling, and in this paper we discuss how to use the Hurst et al. [21] stable subordinated model in the framework of structural credit risk models. In particular, as for the Merton model, we propose two different methodologies for the parameter estimation: the first is to generalize the maximum likelihood parameter estimation proposed by Duan [13]; the second is a generalization of the KMV methodology.

This paper is organized as follows. In Section 2, we firstly review the theory and the distributional assumptions of the Merton model. Subsequently, we introduce the credit risk models with subordinated processes and describe the Mandelbrot-Taylor distributional assumptions. Section 3 is devoted to the parameters estimation for both the Merton and the subordinated models. We characterize empirical data and make a comparison between the obtained results in Section 4. Finally, in the last section, we provide a brief summary.

2. Merton and Subordinated Credit Risk Models

The core concept of the Merton model [1] introduced in 1974 is to treat company's equity and debt as a contingent claim written on company's assets value. In this framework, the company is considered to have a very simple capital structure. It is assumed that the company is financed by one type of equity with a market value E_t at time t and a zero-coupon debt instrument at $t(D_t)$ with a face value of L maturing at time T. (Generally, in a credit risk models framework we assume one-year time horizon for debt maturity and subsequent estimation of PD. One year is perceived as being of sufficient length for a bank to raise additional capital on account of increase in portfolio credit risk (if any).) The exercise price of a call option is defined as the value L. Let A_t be the company's asset value at time t. Naturally, the following accounting identity holds for every time point:

$$A_t = E_t + D_t. \tag{1}$$

In the Merton framework the value of company's equity at maturity time T is given by

$$E_T = \max\left[A_T - L, 0\right]. \tag{2}$$

2.1. The Merton-Black-Scholes Distributional Assumptions. Under the Merton model, the assets value is assumed to follow a geometric Brownian motion (GBM) in the following form:

$$dA_t = \mu A_t dt + \sigma A_t dW_t, \tag{3}$$

where μ is the expected return (drift coefficient), σ is the volatility (diffusion coefficient), both unobserved, and W_t is

the normal variable $N(0,1)$. Using Ito's lemma, we can obtain the solution of (3) as follows:

$$A_T = A_t \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)(T-t) + \sigma\sqrt{(T-t)}W_t\right], \tag{4}$$

where $(T-t)$ is a remaining maturity.

In accordance with the Black-Scholes option pricing theory [2], the Merton model stipulates that the company's equity value satisfies the following equation for pricing the call option within a risk neutral framework:

$$E_t = A_t\Phi(d_1) - Le^{-r(T-t)}\Phi(d_2), \tag{5}$$

where

$$d_1 = \frac{\ln(A_t/L) + \left(r + (1/2)\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}, \tag{6}$$

$$d_2 = d_1 - \sigma\sqrt{(T-t)}, \tag{7}$$

r is the risk-free interest rate and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal variable. (The Treasury bill yields are commonly used as the risk-free interest rate r. Their rates are considered an important benchmark because treasury securities are back by the full faith and credit of the U.S. Treasury. Therefore, they represent the rate at which investment is considered risk-free.) Equation (7) is referred to as the *distance-to-default* (DD) by Moody's KMV. The larger the number in DD is, the less chance the company will default.

We can estimate PD by rearranging (4) as follows:

$$\begin{aligned}\mathrm{PD}_t &= P\left[A_T \le L\right] \\ &= P\left[\ln(A_t) + \left(\mu - \frac{1}{2}\sigma^2\right)(T-t) \right. \\ &\qquad \left. + \sigma\sqrt{(T-t)}\, W_t \le \ln(L)\right] \\ &= \int_{-\infty}^{-(\ln(A_t/L)+(\mu-(1/2)\sigma^2)(T-t))/\sigma\sqrt{(T-t)}} \phi(x)\,dx,\end{aligned} \tag{8}$$

where ϕ is the probability density function of a standard normal variable. Note that unlike (8), (5) is not a function of μ, but it is a function of r (we would get PD under the risk neutral probability measure). When we estimate PD, the risk-free interest rate r has to be replaced with real company drift μ since this step has nothing to do with option pricing. Thereby, the default probability of the company under the objective probability measure is given by

$$\begin{aligned}\mathrm{PD}_t &= \Phi\left(-\hat{d}_2\right) \\ &= \Phi\left(-\frac{\ln(A_t/L) + \left(\mu - (1/2)\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}\right).\end{aligned} \tag{9}$$

Further discussion on this topic can be found in Deliandes and Geske [24] who showed that risk neutral PDs can serve as an upper bound to objective PDs.

2.2. Credit Risk Models with Subordinated Assumptions. Using subordinated processes, we are usually able to capture empirically observed anomalies which are presented in the evolution of return processes over time. That is, we substitute the physical (calendar) time with a so-called intrinsic (operational) time which provides distribution tail effects often observed in the market (see Hurst et al. [21] and Rachev and Mittnik [20]). Thus, if $W = \{W(t), t \ge 0\}$ is a stochastic process and $T = \{T(t), t \ge 0\}$ is a non-negative stochastic process defined on the same probability space and adapted to the same filtration, a new process $Z = \{Z(t) = W(T(t)), t \ge 0\}$ may be formed, and it is defined as subordinated to W by the intrinsic time process T. Next, we will suppose that W is a standard Brownian motion. In this case, if the intrinsic time process T is the deterministic physical time, that is, $T(t) = t$, we obtain the classical log-normal model (see Osborne [25]). Typically, subordinated models with random intrinsic time are leptokurtic with heavier tails compared to the normal distribution. Feller [26] showed that if the intrinsic time process has non-negative stationary independent increments, then the subordinated process Z also has stationary independent increments.

Generally, we assume frictionless markets, where the log-price process Z is subordinated to a standard Brownian motion W by the independent intrinsic time process T. Therefore, we model the assets price process A_t (the company's assets value in our case) by a stochastic equation of the type as follows:

$$\begin{aligned}A(t) = A(t_0)\exp\left\{\int_{t_0}^{t}\mu(s)\,ds + \int_{t_0}^{t}\rho(s)\,dT(s)\right. \\ \left. + \int_{t_0}^{t}\sigma(s)\,dW(T(s))\right\},\end{aligned} \tag{10}$$

where the drift in the physical time scale $\mu(s)$, the drift in the intrinsic time scale $\rho(s)$, and the volatility $\sigma(s)$ are generally assumed to be constant. The appeal of processes subordinated to a standard Brownian motion W by an intrinsic time process T with non-negative stationary independent increments is also due to the option pricing formula which follows from the classical Black-Scholes one in a frictionless complete market and a risk-minimizing strategy in incomplete markets. (In incomplete markets, there exist nonredundant claims carrying an intrinsic risk. In order to evaluate a contingent claim, a risk-minimizing strategy is often applied (see Hofmann et al. [27], Follmer and Sondermann [28], and Follmer and Schweizer [29]).) Hurst et al.'s stable subordinated model [21] uses the unique continuous martingale that makes sense in a discrete setting, but a priori it is not derived from a risk-minimizing strategy even if the markets are incomplete (see Rachev and Mittnik [20]). Following the same notation as in Merton's framework, the value of a European call option at time t (the value of company's equity) with exercise price L (face value of a zero-coupon debt instrument) and time to maturity t(here, we change the notation of maturity time from T (used in the Merton's framework) to t since T denotes the intrinsic time

process in the subordinated option pricing models) is given by

$$E_t = A(t_0) F_+\left(\ln\left(\frac{A(t_0)}{L_{r,t_0,t}}\right)\right) - L_{r,t_0,t} F_-\left(\ln\left(\frac{A(t_0)}{L_{r,t_0,t}}\right)\right), \tag{11}$$

where

$$F_\pm(x) = \int_0^{+\infty} \Phi\left(\frac{x \pm (1/2)\, y}{\sqrt{y}}\right) dF_Y(y), \tag{12}$$

$\Phi(\cdot)$ is the cumulative distribution function of the standard normal variable, F_Y is the cumulative distribution function of a random variable $Y = \int_{t_0}^{t} \sigma^2(s) dT(s)$, and $L_{r,t_0,t} = L\exp(-\int_{t_0}^{t} r(s)ds)$ is the discounted exercise price (the right continuous with left-hand limits (RCLL) time-dependent function $r(t)$ defines the short term interest rate). Considering a continuous distribution of the random variable Y with density function f_Y, then $F_\pm(x)$ can now be numerically integrated over the finite interval $[0,1]$ the transformation $y = u(1-u)^{-3}$ (see Rachev and Mittnik [20]); that is,

$$\begin{aligned} F_\pm(x) &= \int_0^{+\infty} \Phi\left(\frac{x \pm (1/2)\lambda y}{\sqrt{\lambda y}}\right) f_Y(y)\, dy \\ &= \int_0^1 \Phi\left(\frac{x \pm (1/2)\lambda u\,(1-u)^{-3}}{\sqrt{\lambda u\,(1-u)^{-3}}}\right) \\ &\quad \times f_Y\left(u\,(1-u)^{-3}\right) \frac{1+2u}{(1-u)^4} du. \end{aligned} \tag{13}$$

Moreover, as for the classical Black-Scholes model, in the case of subordinated models, we can also monitor the variation in the derivative price with respect to the parameters that enter into the option formula (i.e., the Greeks). For our purposes, it is sufficient to define delta, which is given by

$$\text{delta} = \Delta_E = \frac{\partial E_t}{\partial A} = F_+\left(\ln\left(\frac{A(t_0)}{L_{r,t_0,t}}\right)\right). \tag{14}$$

Analogously to the Merton model, the probability of default can be estimated under the risk neutral probability measure as follows:

$$\begin{aligned} \mathrm{PD}_t &= F_+\left(\ln\left(\frac{L_{r,t_0,t}}{A(t_0)}\right)\right) \\ &= \int_0^{+\infty} \Phi\left(\frac{\ln\left(L_{r,t_0,t}/A(t_0)\right) + (1/2)\, y}{\sqrt{y}}\right) dF_Y(y). \end{aligned} \tag{15}$$

Recall that under the risk neutral measure the stationary increment $Z(t+s) - Z(t)$ has mean $\mu_{Z,s} = 0$ and variance $\sigma^2_{Z,s} = \mu_{T,s}\sigma^2$, where σ and $\mu_{T,s}$ are, respectively, the volatility and the mean of the increment of the stationary process T when they exist (see [21]). The skewness coefficient of this increment is zero (models are symmetric around the zero mean). Kurtosis of the subordinated models is defined as $k_{Z,s} = 3((1+\sigma^2_{T,s})/\mu_{T,s})$, for all $s \geq 0$ (where $\sigma^2_{T,s}$ is the variance of the random variable $T(t+s) - T(t)$ when it exists); that is, subordinated models with intrinsic random time are leptokurtic. Thereby, the model we consider in the following presents heavier tails and higher peaks around the origin than the normal distribution.

2.3. The Mandelbrot-Taylor Distributional Assumptions. Mandelbrot [14–16] and Mandelbrot and Taylor [22] have proposed the stable Paretian distribution to estimate the log-returns. An α-stable distribution $S_\alpha = (\sigma, \beta, \mu)$ depends on four parameters: the index of stability $\alpha \in (0,2]$ ($\alpha = 2$ in the Gaussian case), the skewness parameter $\beta \in [-1,1]$, the scale parameter $\sigma \in (0,+\infty)$, and the location parameter $\mu \in (-\infty,+\infty)$ (see Samorodnitsky and Taqqu [30] for further details on stable distributions). Mandelbrot and Taylor [22] supposed that the intrinsic time process T has stationary independent increments as follows:

$$T(t+s) - T(t) \overset{d}{=} S_{\alpha/2}\left(cs^{2/\alpha}, 1, 0\right), \tag{16}$$

for all $s, t \geq 0$, $\alpha \in (0,2)$, and $c > 0$. Here, the index of stability is $\alpha/2$; the scale parameter is $cs^{\alpha/2}$; the stable skewness is 1; and the location parameter is zero. Under the Mandelbrot-Taylor assumptions; the subordinated process $Z(t) = \ln(A_{th})$ is a symmetric α-stable Lévy motion with stationary independent increments as follows:

$$Z(t+s) - Z(t) = \ln\left(\frac{A_{th}}{A_{(t-s)h}}\right) \overset{d}{=} S_\alpha\left(\nu s^{1/\alpha}, 0, 0\right), \tag{17}$$

for all $s, t > 0$, where

$$\nu = \frac{\sigma\sqrt{c}}{\sqrt{2}(\cos(\pi\alpha/4))^{1/\alpha}}. \tag{18}$$

If we consider the constant scalar parameter σ, then the random variable Y in (11) is as follows:

$$Y = \sigma^2\left(T(t) - T(t_0)\right) = \lambda V, \tag{19}$$

where $\lambda = c\sigma^2(t-t_0)^{2/\alpha}$ and $V = S_{\alpha/2}(1,1,0)$. Hence, with

$$c = 2\left(\cos\left(\frac{\pi\alpha}{4}\right)\right)^{2/\alpha}, \tag{20}$$

it follows that $Z(t) \overset{d}{=} S_\alpha(\sigma t^{1/\alpha}, 0, 0)$. Thus, we can estimate the index of stability α and the scalar parameter σ using the maximum likelihood method (see [20] and the references therein). Moreover, considering the density function f_V of the $\alpha/2$ stable random variable V, we obtain the following expression for $F_\pm(x)$:

$$\begin{aligned} F_\pm(x) &= \int_0^1 \Phi\left(\frac{x \pm (1/2)\lambda u\,(1-u)^{-3}}{\sqrt{\lambda u\,(1-u)^{-3}}}\right) \\ &\quad \times f_V\left(u\,(1-u)^{-3}\right) \frac{1+2u}{(1-u)^4} du. \end{aligned} \tag{21}$$

The probability of default under the risk neutral probability measure is then given by

$$\mathrm{PD}_t = \int_0^1 \Phi\left(\frac{\ln\left(L_{r,t_0,t}/A\left(t_0\right)\right) + (1/2)\,\lambda u\,(1-u)^{-3}}{\sqrt{\lambda u(1-u)^{-3}}}\right) \times f_V\left(u\,(1-u)^{-3}\right)\frac{1+2u}{(1-u)^4}du. \tag{22}$$

3. Estimation Methodology

While for the Merton model there are just three parameters necessary for the estimation of default probabilities—namely, the company's market value A_t at time t, the asset drift μ, and the asset volatility σ—in the case of the subordinated models, we have to estimate the company's market value at time t and the parameters of the subordinated process. Clearly, different distributional hypothesis of the subordinated model could require the estimation of several different parameters. For example, in the α stable Lévy process, once the index of stability α is estimated, the scalar parameter σ is the unique parameter that should be estimated since the skewness parameter and the location parameter have been fixed equal to zero in the model.

3.1. Parameter Estimates for the KMV-Merton Model. The unknown parameters of KMV-Merton model come from (5). Since the market value of assets is a random variable and cannot be observed directly, it is impossible to directly estimate the drift and the volatility in a movement of log-returns on A_t. Therefore, these three parameters have to be estimated in a different way. In fact, we use the observed market value of equity E_t along with (5) to estimate them indirectly.

Generally, the starting point for the two iterative methodologies proposed in literature (the maximum likelihood estimation method and the Moody's KMV method) is based on the so-called *calibration method* (see [3, 5, 31] or [32]), which finds two unknown parameters (A_t and σ) by solving the system of two equations as follows:

$$E_t = A_t\Phi\left(d_1\right) - Le^{-r(T-t)}\Phi\left(d_2\right), \qquad \sigma_E = \frac{A_t}{E_t}\Phi\left(d_1\right)\sigma, \tag{23}$$

where σ_E is the standard deviation of the equity log returns $\ln(E_{th}/E_{(t-1)h})$. Nevertheless, this method does not estimate asset drift μ; it determines the risk neutral probability of default PD using the risk free asset r. As a consequence, Jovan [33] showed that this method provides different estimates of probability of defaults for the same obligors compared to the two following iterative methodologies: the maximum likelihood estimation method and the Moody KMV method.

3.1.1. Maximum Likelihood Estimation Method. This methodology was initially proposed by Duan [13] and enhanced by Duan et al. [34] later. The time series of daily market value of equity E_t is equal to n days, where $t = (0,\ldots,n)$. In Duan et al. [34] the time step h is introduced. Typically, the value of this coefficient for daily data would be $h = 1/250$. The methodology is iterative. Then, the following log-likelihood function for the estimation of μ and σ of model (3), where $th = (0,\ldots,nh)$, is defined on the basis of observed values of E_t as follows:

$$L\left(\widehat{\theta};\widehat{A}_{th} \mid E_{th}\right) = -\frac{n}{2}\ln\left(2\pi\widehat{\sigma}^2 h\right) - \frac{1}{2}\sum_{t=1}^{n}\frac{\left(\widehat{R}_t - \left(\widehat{\mu} - (1/2)\,\widehat{\sigma}^2\right)h\right)^2}{\widehat{\sigma}^2 h} - \sum_{t=1}^{n}\ln\left(\widehat{A}_{th}\right) - \sum_{t=1}^{n}\ln\left(\Phi\left(d_1\right)\right), \tag{24}$$

where

$$\widehat{R}_t = \ln\left(\frac{\widehat{A}_{th}}{\widehat{A}_{(t-1)h}}\right) \tag{25}$$

and where $\widehat{\theta} \equiv (\widehat{\mu},\widehat{\sigma})$ and $\widehat{A}_{th}$ is estimated from (5). To launch the iteration process we could insert as initial values entered into the iteration process the values obtained by solving the system (23). Despite the fact that these estimates are not the best ones from a solution point of view, they can be good enough as the initial values for different kinds of iterative procedures. Each iteration produces a time series of daily values $\widehat{A}_{th}^{(i)}$, where the debt maturity ranges over $1 \le (T - th) \le T$. We maximize (24) to obtain estimates of the unobserved asset drift and volatility $\widehat{\theta}^{(i)}$. Since this is an iterative procedure, we use the new estimates obtained from (24) and the new market value of assets obtained from (5) for maximizing (24) once again. The procedure is repeated until the differences in $\widehat{\mu}^{(i)}$ and $\widehat{\sigma}^{(i)}$ between the successive iterations are sufficiently small (i.e., until $|\widehat{\mu}^{(i+1)} - \widehat{\mu}^{(i)}| + |\widehat{\sigma}^{(i+1)} - \widehat{\sigma}^{(i)}| \le \varepsilon$ for a given small ε).

Duan et al. [34] found that the Moody's KMV method provides the same estimates as the MLE method, even though they state that the latter method is preferable for inference statistics.

3.1.2. Moody's KMV Methodology. This iterative procedure follows a disclosed part of Moody's KMV methodology for a calculation of expected default frequency (see Duan et al. [34], Duffie et al. [9], Crosbie and Bohn [5], or Vassalou and Xing [35]). This method is quite similar to the MLE method. The unique difference is that in order to obtain estimates of the asset drift and volatility, instead of maximizing the log-likelihood function, we have explicit formulas.

The first step is exactly the same calculation of the daily value of $\widehat{A}_{th}^{(i)}$, $th = (0,\ldots,nh)$ from (5). As the initial values can be used again, the estimates can be obtained by solving

the system (23). Then, the arithmetic mean of the sample is given by

$$\overline{R}^{(i)} = \frac{1}{n}\sum_{t=1}^{n}\widehat{R}_t^{(i)}, \tag{26}$$

where $\widehat{R}_t$ is defined in (25). Another step is the calculation of estimates of the asset volatility $\widehat{\sigma}$ and the drift $\widehat{\mu}$ of model (3) which are defined as follows:

$$\begin{aligned}\widehat{\sigma}^{(i+1)} &= \sqrt{\frac{1}{nh}\sum_{t=1}^{n}\left(\widehat{R}_t^{(i)} - \overline{R}^{(i)}\right)^2},\\ \widehat{\mu}^{(i+1)} &= \overline{R}^{(i)}\frac{1}{h} + \frac{1}{2}\widehat{\sigma}^{2(i+1)}.\end{aligned} \tag{27}$$

Since this is again an iterative procedure, we use the new estimates obtained from (27) to calculate $A_{th}^{(i+1)}$. The procedure is repeated until the differences in $\widehat{\mu}$ and $\widehat{\sigma}$ among successive iterations are sufficiently small.

It is worth to mention that the Merton model with parameters estimated according to the methodology described above differs from the one actually employed by Moody's KMV. How well the Merton model performs substantially relies on the simplifying assumptions facilitating its implementation. These simplifying assumptions are not really realistic in practice, though. That is why Moody's KMV does not rely solely on these assumptions. (In 2002, Moody's Corporation completed acquisition of KMV. KMV Corporation is now renamed as Moody's KMV.) Indeed, the founders of KMV, Oldrich Vasicek and Stephen Kealhofer, developed a new model called Vasicek-Kealhofer (VK) (see Arora et al. [36]) to estimate the distance-to-default of an individual company. One of the most important differences is that while we use the cumulative normal distribution to convert distances-to-default into "real" (non risk-neutral) default probabilities in classical Merton model, Moody's KMV uses its large historical database to estimate the real empirical distribution of distances-to-default, and it calculates default probabilities based on that distribution.

3.2. Parameter Estimates for Subordinated Models. We can extend the estimation methodologies proposed for the KMV-Merton model in order to estimate the parameters of a subordinated model.

3.2.1. Maximum Likelihood Estimation Method. Obviously, in order to use this method, we have to revise (24). Actually, (24) can be derived from the more general formula which can be used for the derivation of log-likelihood functions for any subordinated model. This formula is defined in the following way:

$$\begin{aligned}&L\left(\widehat{\theta}; \widehat{A}_{th} \mid E_{th}\right)\\ &\quad= \sum_{t=1}^{n}\ln\left(f_Z\left(\widehat{R}_t\right)\right) - \sum_{t=1}^{n}\ln\left(\widehat{A}_{th}\right) - \sum_{t=1}^{n}\ln\left(\Delta_E\right),\end{aligned} \tag{28}$$

where $\widehat{\theta}$ represents the set of the parameters in the density function $f_Z(\widehat{R}_t)$ of the stationary increment $\ln(A_{th}/A_{(t-1)h}) = Z(t+1) - Z(t)$, $\widehat{A}_{th}$ is estimated from (11), $\widehat{R}_t$ is defined in (25), and Δ_E is given by (14). The initial values $\widehat{A}_{th}^{(1)}$ of the iteration process could be the ones obtained by solving the system (23). The procedure continues iteratively till the distance $\|\widehat{\theta}^{(i+1)} - \widehat{\theta}^{(i)}\|$ is sufficiently small. Typically, there are two problems regarding this maximum likelihood method. The first difficulty is related to computation time. This method generally presents more local optima, and it can be very time consuming to reach a global optimum. Secondly, it is often very problematic to implement this methodology since many subordinated models do not have close form equation for the density function f_Z.

3.2.2. An Extended KMV Methodology. As for Moody's KMV iterative methodology, we have first to compute the daily value of $\widehat{A}_{th}^{(i)}$, $th = (0, \ldots, nh)$ solving (11), then the other parameters of the subordinated process $\widehat{\theta}^{(i+1)}$ are estimated on the series $\widehat{R}_t^{(i)} = \ln(\widehat{A}_{th}^{(i)}/\widehat{A}_{(t-1)h}^{(i)})$ considering the distributional assumption of the subordinated model. The procedure continues iteratively till the distance $\|\widehat{\theta}^{(i+1)} - \widehat{\theta}^{(i)}\|$ is sufficiently small. In particular, for the α stable Lévy model, we first suggest to determine the index of stability α. Secondly, the unique parameter that must be estimated is the scalar parameter σ since the skewness parameter and the location parameter are fixed equal to zero. Clearly, even in this case, we need to insert some initial values $\widehat{A}_{th}^{(1)}$ of the iteration process that could be the ones obtained by solving the system (23). Moreover (as for the Merton model—see Duan et al. [34]), the extended KMV methodology provides the same estimates as the MLE method when the parameter estimates $\widehat{\theta}^{(i+1)}$ are the MLE on the series $\widehat{R}_t^{(i)}$.

4. Application and Results

In this section, we first describe the data used in the computational analysis and apply the Merton model. Secondly, we test the distributional assumption of this model. Finally, we apply the Stable Lévy model and compare it with the Merton one. In the application of the models, we use the extended KMV methodology.

To apply the previous models to a particular company, we need the market value of equity E_t, the face value of the zero-coupon debt instrument L, and the risk-free interest rate r. For risk-free interest rate we used 13-week Treasury bill and Thomson Reuters Datastream dataset to obtain the market value of equity and the face value of the zero-coupon debt instrument of 24 US companies with strong capitalization in US market. (The companies are (1) Boeing, (2) Cisco Systems, (3) Chevron, (4) E. I. du Pont de Nemours, (5) Walt Disney, (6) Home Depot, (7) Hewlett-Packard, (8) IBM, (9) Intel, (10) Johnson & Johnson, (11) Coca Cola, (12) McDonalds, (13) 3M, (14) Merck & Co., (15) Microsoft, (16) Pfizer, (17) Procter & Gamble, (18) AT & T, (19) UnitedHealth Group, (20) United Technologies, (21) Verizon Communications, (22) WalMart

Stores, (23) Exxon Mobil, and (24) Travelers Companies.) For a sample period we used data from January 3, 2000, to December 30, 2011. As market value of equity, we used consolidated market value of a company which is defined as a share price multiplied by the number of ordinary shares in issue. Finally, for the face value of the zero-coupon debt instrument, we used the sum of the short-term debt, current portion of the long-term debt, and half of the long-term debt. (There need to be chosen an amount of the debt that is relevant to a potential default during a one year period. Total debt is inadequate when not all of it is due in one year (it is assumed one-year time horizon for debt maturity and subsequent estimation of PD), as the firm may remain solvent even when the value of assets falls below its total liabilities. Using the short-term debt for the default barrier would be often wrong, for instance, when there are covenants that force the company to serve other debts when its financial situation deteriorates. Prior studies generally follow KMV methodology and choose the short-term debt plus half of the long-term debt for the default barrier (see Bharath and Shumway [37], Vassalou and Xing [35] or Duffie et al. [9]).) While the short-term debt and current portion of the long-term debt represent that portion of the debt payable within one year including current portion of the long-term debt and sinking fund requirements of preferred stock or debentures, the long-term debt represents all interest bearing financial obligations excluding amounts due within one year.

4.1. Analysis of the Distributional Assumption of the Company Value Log-Returns. The Merton model distributional assumption implies that the unobservable company value log-returns are Gaussian distributed. In order to test this assumption, we use the daily log-returns of the companies' assets value obtained from both the KMV-Merton model and the alpha stable Lévy model, from January 3, 2000, to December 30, 2011 (for a total of 3157 daily values).

First of all, we test the Gaussian and the stable non-Gaussian hypotheses on the company value log-returns obtained from the KMV-Merton model. Thus, we compute different statistics every day on the last 250 daily company values (1 year of daily values). Table 1 reports the average among all the firms and for all the expost period of different statistics applied to company value log-returns to test the Gaussian hypothesis and the stable non-Gaussian hypothesis. In particular we consider the average of the follwing: the mean, the standard deviation, the skewness $E((X-E(X))^3)/E((X-E(X))^2)^{1.5}$, the kurtosis $E((X-E(X))^4)/E((X-E(X))^2)^2$, the percentage of rejection of the Gaussian hypothesis using the Jarque-Bera (JB) test (at the 5% significance level) (see [38]), the stable index of stability "alpha," the stable index of skewness "beta," the stable scalar parameter "sigma," the stable location parameter "mu," and the percentage of rejection of the stable Paretian hypothesis using the Kolmogorov-Smirnov (K-S) test (at the 5% significance level).

In particular, the results reported in Table 1 suggest that (1) the returns exhibit heavy tails since the average of the stability parameters alpha is always less than 2 and the average of kurtosis is much higher than 3; (2) the returns are slightly asymmetric since the average of the skewness parameter and the average of the stable parameter beta are always different from zero; and (3) the Gaussian hypothesis is almost always rejected for all companies while the stable Paretian hypothesis is generally rejected for four companies of the considered sample.

Table 1: Average of some statistics for the daily log-returns of the companies' assets value obtained from the KMV-Merton model.

Mean	0.00002
St.dev.	0.0196
Skewness	−0.6140
Kurtosis	33.4351
JB test (95%)	96.77%
Alpha	1.7089
Beta	0.0062
Sigma	0.0106
Mu	0.0001
K-S test (95%)	16.56%

Secondly, we test the different distributional hypothesis of the companies value log-returns obtained by the stable Lévy model using a Kolmogorov-Smirnov (K-S) test (at the 5% significance level). From this test, we observe almost the same percentage of rejection (16.55%) we get using the companies value log-returns as that obtained from the KMV-Merton model (16.56%). Similarly, we get 98.44% of rejection of the Gaussian hypothesis applying the Jarque-Bera test to the companies value log-returns obtained by stable Lévy model (compared to 96.77% obtained from the KMV-Merton model).

From this preliminary analysis, we deduce that the classical distributional hypothesis of the Merton model is almost never verified. Moreover, the stable non-Gaussian hypothesis appears more realistic than the Gaussian one. Therefore, it is appropriate to apply a Stable Lévy model which is able to capture empirically observed anomalies that contradict the classical normality assumption. The results we get here are not a real surprise since the stable Paretian laws generalize the Gaussian one.

4.2. Estimate of Default Probabilities with KMV-Merton Model. We used Moody's KMV methodology (We perform our analysis using MATLAB.) for the estimation of the parameters for the Merton model used for the computation of the probability of default of any company. The results of the empirical analysis are reported in Figure 1 and Table 2. In Table 2, there are listed average values of ratio between the debt and the company's assets value and average values of risk neutral PD and distance-to-default obtained from the KMV-Merton model for any company. In particular, we observe that generally when the average ratio between debt and company value is high, we observe an analogous higher probability of default and a lower distance to default. This aspect could be a problem when the KMV-Merton model is used to compute the risk neutral and real probabilities of default of a bank since

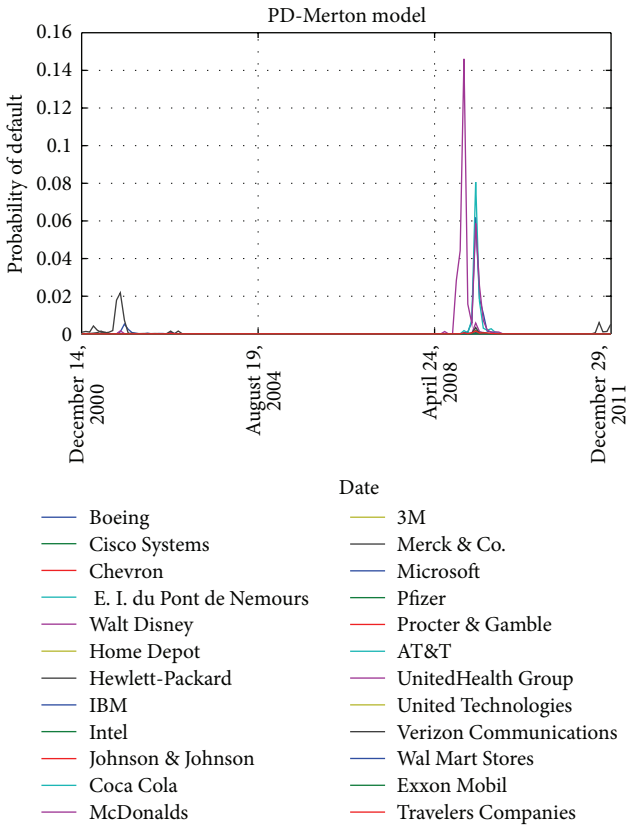

Figure 1: PDs-KMV-Merton model.

financial institutions have significantly greater debt compared to other companies. Therefore, the Merton model is not plausible for the estimation of PDs of financial institutions unless some adjustments are made. (For example, Byström [39] shows that one of the main implications of his simplified "spread sheet" version of the Merton model is the fact that the default probability's insensitivity to the leverage ratio at high levels of debt makes it possible to apply his model to banks and other highly leveraged firms.)

Moreover, Figure 1 describes the evolution of the risk neutral PDs on the monthly basis. These probabilities are almost null during all the decade. However, we can distinguish three periods of increased PDs for some companies which are as follows: at the beginning of the century after the high tech crisis and September 11, during the subprime crisis, and during the country credit risk crisis. During the first period and the country credit risk crisis, the most evident grown of PD is due to the Hewlett-Packard firm (its PD increased up to 2.1% in the first period and to 1% in the last one). The period with more significant growth in PDs is dated from September 2008. This might be easily explained by the subprime mortgage crisis that reached a critical stage during the first week of September 2008 and was characterized by severely contracted liquidity in the global credit markets and insolvency threats to investment banks and other institutions. Beginning with bankruptcy of Lehman Brothers on September 14, 2008, the financial crisis entered an acute phase marked by the failures of prominent American and European banks and efforts by the American and European governments to rescue distressed financial institutions. Among the companies from our sample which were affected the most by this crisis belong UnitedHealth Group, E. I. du Pont de Nemours, and Boeing. UnitedHealth Group is a care company which offers a spectrum of products and services. This company suffered a jump in PD from 0% in May 2008 up to 14.6% in November 2008. E. I. du Pont de Nemours is a chemical company and was the world's third largest chemical company based on market capitalization in 2009. This company's PD increased from 0% in October 2008 to 8.1% in February 2009. Finally, Boeing as a representative of aerospace industry suffered an increase in PD from 0% in October 2008 to 6.2% in February 2009. This phase of financial crisis lasted approximately one year, and in October 2009 the values of PD of observed companies went back to zero.

Table 2: Average values of ratio (L/A) and risk neutral PD and DD obtained from the KMV-Merton model.

Company	Average ratio (L/A)	Average PD	Average DD
(1) Boeing	0.1326	0.000830	8.9020
(2) Cisco Systems	0.0262	0.000000	20.6010
(3) Chevron	0.0613	0.000000	13.8524
(4) E. I. du Pont de Nemours	0.1169	0.000845	9.9706
(5) Walt Disney	0.1312	0.000083	8.5109
(6) Home Depot	0.0600	0.000002	11.8297
(7) Hewlett-Packard	0.0909	0.000511	8.3242
(8) IBM	0.1037	0.000000	11.4799
(9) Intel	0.0099	0.000000	14.2761
(10) Johnson & Johnson	0.0331	0.000000	22.8225
(11) Coca Cola	0.0615	0.000000	17.5142
(12) McDonalds	0.1031	0.000015	12.2037
(13) 3M	0.0493	0.000000	14.9342
(14) Merck & Co.	0.0611	0.000037	11.1672
(15) Microsoft	0.0068	0.000000	21.4008
(16) Pfizer	0.0815	0.000019	11.0915
(17) Procter & Gamble	0.1010	0.000000	13.9819
(18) AT&T	0.1619	0.000013	8.4346
(19) UnitedHealth Group	0.0924	0.002424	10.2912
(20) United Technologies	0.0800	0.000001	12.1045
(21) Verizon Communications	0.2117	0.000106	8.8750
(22) Wal Mart Stores	0.0957	0.000000	12.4895
(23) Exxon Mobil	0.0208	0.000000	18.0516
(24) Travelers Companies	0.1298	0.000035	8.9095

Table 3: Indexes of stability alpha average values of ratio (L/A), and risk neutral PD and DD obtained from the stable Lévy model.

Company	Alpha	Average ratio (L/A)	Average PD	Average DD
(1) Boeing	1.6619	0.1308	0.0149	8.9153
(2) Cisco Systems	1.5756	0.0262	0.0116	20.4104
(3) Chevron	1.6671	0.0606	0.0067	13.7868
(4) E. I. du Pont de Nemours	1.6575	0.1169	0.0137	10.0480
(5) Walt Disney	1.5680	0.1305	0.0265	8.5155
(6) Home Depot	1.6101	0.0599	0.0173	11.9741
(7) Hewlett-Packard	1.5850	0.0914	0.0253	8.3069
(8) IBM	1.6110	0.1032	0.0120	11.5404
(9) Intel	1.6411	0.0098	0.0131	14.3321
(10) Johnson & Johnson	1.5803	0.0330	0.0068	22.9854
(11) Coca Cola	1.5505	0.0614	0.0120	17.6094
(12) McDonalds	1.7570	0.1012	0.0032	12.3247
(13) 3M	1.5590	0.0494	0.0136	14.9028
(14) Merck & Co.	1.5909	0.0610	0.0150	11.1738
(15) Microsoft	1.5459	0.0068	0.0082	21.1204
(16) Pfizer	1.6691	0.0813	0.0085	11.2040
(17) Procter & Gamble	1.4745	0.1010	0.0204	13.9846
(18) AT&T	1.5985	0.1607	0.0176	8.5163
(19) UnitedHealth Group	1.5839	0.0925	0.0256	10.3436
(20) United Technologies	1.6064	0.0798	0.0138	12.0951
(21) Verizon Communications	1.6645	0.2106	0.0114	8.9470
(22) Wal Mart Stores	1.6398	0.0955	0.0080	12.5641
(23) Exxon Mobil	1.6494	0.0207	0.0060	18.1822
(24) Travelers Companies	1.4659	0.1291	0.0464	8.9419

4.3. Estimate of Default Probabilities with the Stable Lévy Model. In order to evaluate the stable Lévy model, we estimate the parameters using the extended KMV methodology. First of all, we compute the indices of stability on the daily log-returns of the companies' asset values, obtained by the stable Lévy model, which are reported in Table 3. To evaluate the stable parameters and the distributions of subordinator f_V in (21), we perform a maximum likelihood estimator that uses the fast Fourier transform (see [1, 20, 40]). The estimated index of stability is maintained constant for each firm and for all the period of analysis. Clearly, we could adapt more dynamically the model requiring that the index of stability changes periodically with the scalar and location stable parameters. However, this should require the knowledge of the subordinator density distribution f_V that changes with the index of stability. Since this distribution is obtained by inverting the Fourier transform, the iterating procedure of the KMV methodology would require too long computational time in that case. In Table 3, there are also listed the average values of ratio between the debt and the company's assets value and average values of risk neutral PD and distance-to-default obtained from the stable Lévy model for any company.

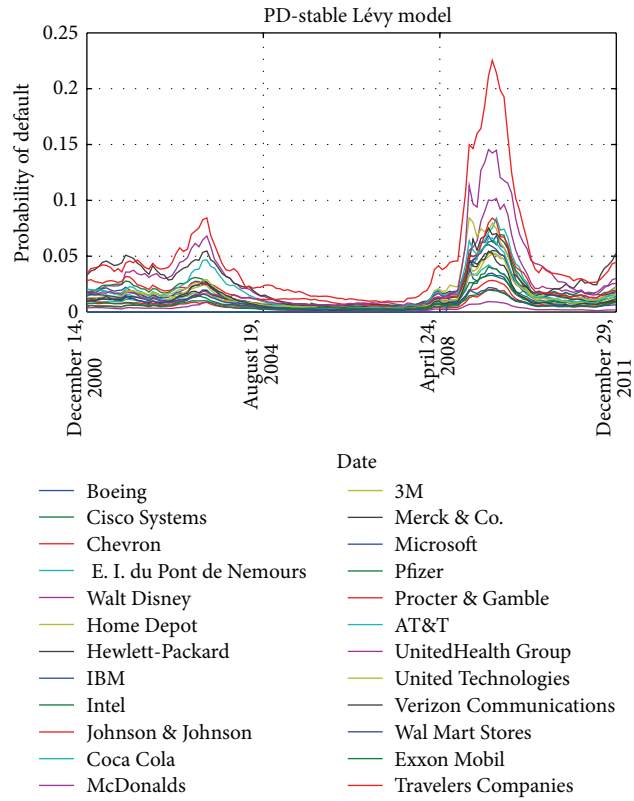

Figure 2: PDs-stable Lévy model.

4.4. Comparison of the Two Models. In particular, we observe that there are not very large differences between the company values obtained by the stable Lévy model and the company values obtained by the Merton model. This aspect is important since we couldn't expect strong differences in the company values that represent an unobservable objective variable whose big differences could not be easily justifiable. This observation implies that there are not large differences between the two models with respect to the follwing: (1) the average ratio between debt and company value; (2) the average distance-to-default.

Figure 3 reports the main differences between the two models for those companies which present the highest peaks in default probabilities (E. I. du Pont de Nemours, Walt Disney, Hewlett-Packard, UnitedHealth Group, and Travelers Companies). In particular, Figures 3(a) and 3(b) show that the main differences in the ratio between the debt and the company value and between the distances-to-default are concentrated during the high volatility period after September 11, 2001. However, this difference (as remarked previously) is almost null during the big crisis following the Lehman Brothers bankruptcy. Figures 3(c) and 3(d) show default

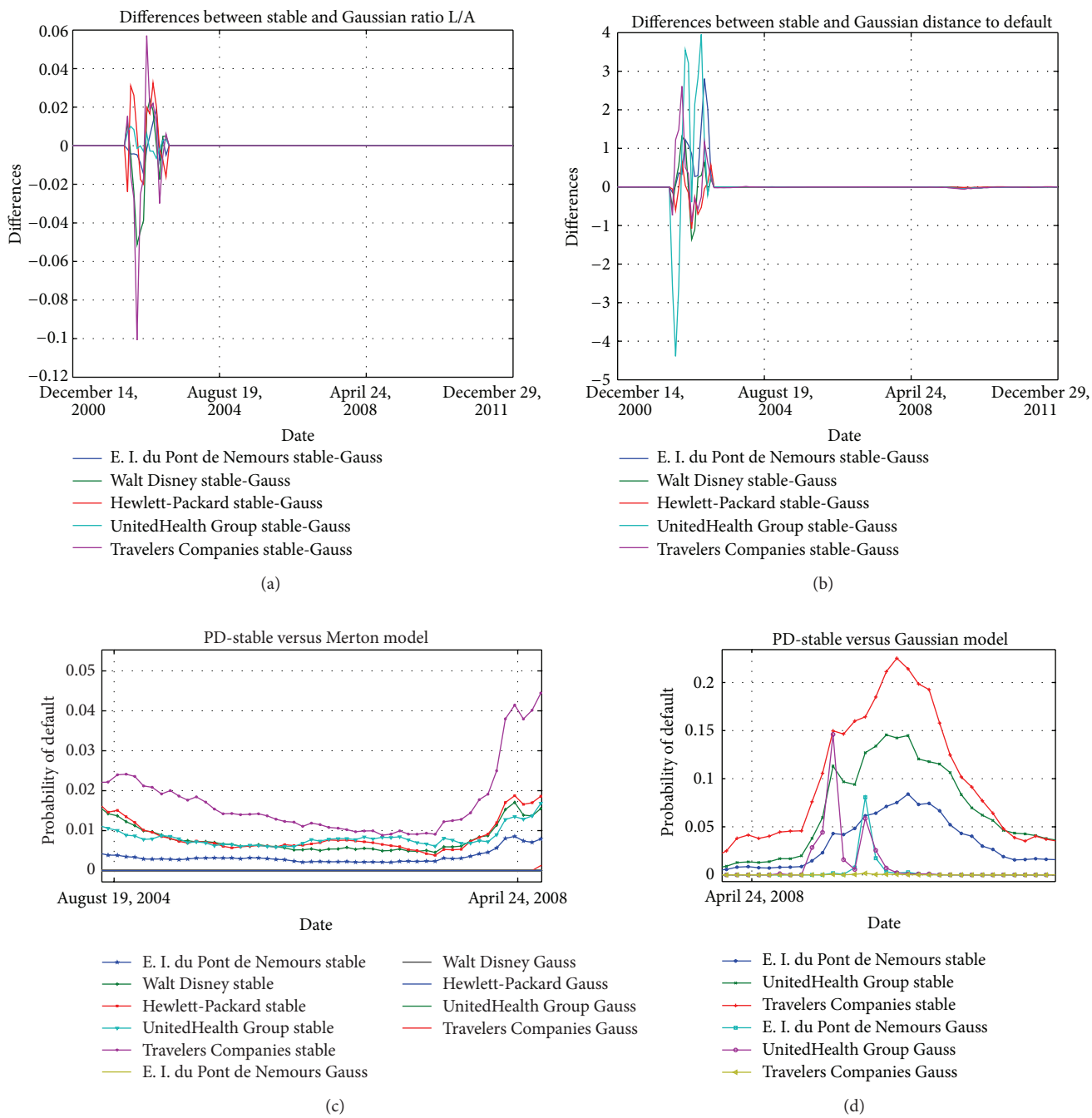

Figure 3: Differences between Stable Lévy model and KMV-Merton model. (a) Difference between stable ratio L/A and Gaussian ratio. (b) Difference between the stable and Gaussian distance-to-default. (c) Probabilities of default during "calm periods." (d) Probabilities of default during the crisis.

probabilities of chosen companies during "calm" periods and during periods of the crisis. In this case, we observe very big differences among PDs. On one hand, the probabilities of default computed by the Merton model are almost null during the "calm" periods and increase during one or two months of the crisis. On the other hand, the probabilities of default computed by the Lévy stable model are never null during the "calm" periods and become very high during the months of the crisis and in the close subsequent periods.

In particular, we observe the biggest difference for the Travelers Companies for which the Merton model does not register any significant difference in the default probabilities while the stable Lévy model shows the highest values. The reason of this difference is essentially caused by the combination of two aspects. First, the index of stability of the Travelers Companies is very small, and that means very fat tails with high probability of losses. Secondly, the ratio between the debt and the Travelers Companies value is high. This analysis

confirms the previous one that shows the average default probabilities obtained by the stable Lévy model are much higher than those obtained by the Merton model. This is not a real surprise because while the probability tails of the Gaussian distribution tend to zero exponentially, the probability tails of stable non-Gaussian distribution tend to zero in polynomial order. Thus the probability of losses computed by the stable Lévy model is much higher than the one computed by the Merton model. This effect is also emphasized in Figure 2 that reports the evolution of default probabilities during the decade 2001–2011. Figure 2 shows a much higher sensitivity of these probabilities of all companies with respect to the periods of crises. Moreover, since all the tests have shown that the stable non-Gaussian hypothesis appears more realistic than the Gaussian one, we deduce that the KMV-Merton model underestimates the probability of default.

5. Conclusion

In this paper, we propose alternative structural credit risk models, and we discuss how to evaluate the probability of default of a given firm under different distributional hypotheses. Finally, we apply and compare the stable Lévy credit risk model with the Merton one. The empirical analysis suggests that the probability of default is generally underestimated by the Merton model. Clearly, these first results should be further discussed and compared with other distributional models in a future research.

Conflict of Interests

According to the editorial policy of the journal the authors declare that they have no conflict of interests with the firms recalled in the paper.

Acknowledgments

The authors thank the two anonymous referees for their comments and helpful suggestions that improved the paper greatly. The research was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence Project (CZ.1.05/1.1.00/02.0070), including the access to the supercomputing capacity, the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296, and the state budget of the Czech Republic. The paper has been also supported by the Italian funds ex MURST 60% 2011 and MIUR PRIN MISURA Project, 2013–2015.

References

[1] R. C. Merton, "On the pricing of corporate debt: the risk structure of interest rates," *Journal of Finance*, vol. 29, pp. 449–470, 1974.

[2] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, pp. 637–659, 1973.

[3] C. Bluhm, L. Overbeck, and C. Wagner, *An Introduction to Credit Risk Modeling*, CRC Press, Boca Raton, Fla, USA, 1st edition, 2003.

[4] D. Duffie and K. J. Singleton, *Credit Risk: Pricing, Measurement, and Management*, Princeton University Press, Princeton, NJ, USA, 1st edition, 2003.

[5] P. J. Crosbie and J. R. Bohn, "Modelling default risk," Working Paper, KMV Corporation, 2003.

[6] J. R. Bohn, N. Arora, and I. Korablev, "Power and level validation of the EDF credit measure in the U.S. market," Moody's KMV Working Paper, 2005.

[7] Y. Du and W. Suo, "Assessing credit quality from equity markets: is structural model a better approach," Working Paper, Queen's University, 2003.

[8] S. A. Hillegeist, E. K. Keating, D. P. Cram, and K. G. Lundstedt, "Assessing the probability of Bankruptcy," Working Paper, Kellog School of Management, 2002.

[9] D. Duffie, L. Saita, and K. Wang, "Multi-period corporate default prediction with stochastic covariates," *Journal of Financial Economics*, vol. 83, no. 3, pp. 635–665, 2007.

[10] E. S. Farmen, S. Westgaard, S. Fleten, and N. Wijst, "Default risk and its Greeks under an objective probability measure," in *Proceedings of the Stockholm School of Economics—Department of Finance Seminar Series*, November 2003.

[11] F. Black and C. Cox, "Valuing corporate securities: some effects of bond indenture provisions," *Journal of Finance*, vol. 31, no. 2, pp. 351–367, 1976.

[12] J. R. Bohn, "A survey of contingent-claims approaches to risky debt valuation," *The Journal of Risk Finance*, vol. 1, no. 3, pp. 53–70, 2000.

[13] J. C. Duan, "Maximum likelihood estimation using price data of the derivative contract," *Mathematical Finance*, vol. 4, no. 2, pp. 155–157, 1994.

[14] B. Mandelbrot, "New methods in statistical economics," *Journal of Political Economy*, vol. 71, pp. 421–440, 1963.

[15] B. Mandelbrot, "The variation of certain speculative prices," *Journal of Business*, vol. 26, pp. 394–419, 1963.

[16] B. Mandelbrot, "The variation of some other speculative prices," *Journal of Business*, vol. 40, pp. 393–413, 1967.

[17] E. Fama, "Mandelbrot and the stable Paretian hypothesis," *Journal of Business*, vol. 36, pp. 420–429, 1963.

[18] E. Fama, "The behavior of stock market prices," *Journal of Business*, vol. 38, pp. 34–105, 1965.

[19] E. Fama, "Portfolio analysis in a stable Paretian market," *Management Science*, vol. 11, pp. 404–419, 1965.

[20] S. Rachev and S. Mittnik, *Stable Paretian Model in Finance*, John Wiley & Sons, Chichester, UK, 2000.

[21] S. R. Hurst, E. Platen, and S. T. Rachev, "Option pricing for a logstable asset price model," *Mathematical and Computer Modelling*, vol. 29, no. 10–12, pp. 105–119, 1999.

[22] T. Mandelbrot and M. Taylor, "On the distribution of stock price differences," *Operations Research*, vol. 15, pp. 1057–1062, 1967.

[23] S. T. Rachev, *Handbook of Heavy Tailed Distributions in Finance*, Elsevier, Amsterdam, The Netherlands, 2003.

[24] G. Deliandes and R. Geske, "Credit risk and risk neutral default probabilities—information about rating migrations and defaults," Working Paper, UCLA, 2003.

[25] M. F. Osborne, "Brownian motion in the stock market," *Operations Research*, vol. 7, pp. 145–173, 1959.

[26] W. Feller, *An Introduction to Probability Theory and Its Applications II*, John Wiley & Sons, New York, NY, USA, 1966.

[27] N. Hofmann, E. Platen, and M. Schweizer, "Option pricing under incompleteness and stochastic volatility," Tech. Rep., Department of Mathematics, University of Bonn, 1982.

[28] H. Follmer and D. Sondermann, "Hedging of non-redundant contingent claims," in *Contributions to Mathematical Economics*, W. Hildenbrand and A. M. Aollell, Eds., pp. 205–223, North-Holland, Amsterdam, The Netherlands, 1986.

[29] H. Follmer and M. Schweizer, "Microeconomic approach to diffusion models for stock prices," *Mathematical Finance*, vol. 3, pp. 1–23, 1989.

[30] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, New York, NY, USA, 1994.

[31] M. Bruche, "Estimating structural bond pricing models via simulated maximum likelihood," 2005, http://www.defaultrisk.com/pp_model_96.htm.

[32] J. Ericsson and J. Reneby, "Estimating structural bond pricing models," *Journal of Business*, vol. 78, no. 2, pp. 707–736, 2005.

[33] M. Jovan, "The Merton structural model and IRB compliance," *Metodološki Zvezki*, vol. 7, no. 1, pp. 39–57, 2010.

[34] J. C. Duan, G. Gauthier, and J. G. Simonato, "On the equivalence of the KMV and maximum likelihood methods for structural credit risk models," 2004, http://www.defaultrisk.com/pp_model_97.htm.

[35] M. Vassalou and Y. Xing, "Default risk in equity returns," *Journal of Finance*, vol. 59, no. 2, pp. 831–868, 2004.

[36] N. Arora, J. R. Bohn, and F. Zhu, "Reduced form versus structural models of credit risk: a case study of three models," Moody's KMV Working Paper, 2005.

[37] S. T. Bharath and T. Shumway, "Forecasting default with the Merton distance to default model," *Review of Financial Studies*, vol. 21, no. 3, pp. 1339–1369, 2008.

[38] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review*, vol. 55, no. 2, pp. 163–172, 1987.

[39] H. N. Byström, "Merton for dummies: a flexible way of modelling default risk," Research Paper 112, University of Technology, Sydney, Australia, 2003.

[40] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Communications in Statistics C*, vol. 13, no. 4, pp. 759–774, 1997.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 315752, 6 pages
http://dx.doi.org/10.1155/2013/315752

Research Article

Joint Implementation of Signal Control and Congestion Pricing in Transportation Network

Wei Mao,[1] Feifei Qin,[1] Yihong Hu,[1] and Zhijia Tan[2]

[1] *School of Economics and Management, Tongji University, Siping Road 1239, Shanghai 200092, China*
[2] *College of Management, Huazhong University of Science & Technology, 1037 Luoyu Road, Wuhan 430074, China*

Correspondence should be addressed to Wei Mao; maoweitj@yahoo.cn

Received 1 February 2013; Revised 25 May 2013; Accepted 13 July 2013

Academic Editor: Xiaoning Zhang

The policy of jointly implementing signal control and congestion pricing in the transportation network is investigated. Bilevel programs are developed to model the simultaneous optimization of signal setting and congestion toll. The upper level aims to maximize the network reserve capacity or minimize the total travel time, subject to signal setting and toll constraints. The lower level is a deterministic user equilibrium problem given a plan of signal setting and congestion charge. Then the bilevel programs are transferred into the equivalent single level programs, and the solution methods are discussed. Finally, a numerical example is presented to illustrate the concepts and methods, and it is shown that the joint implementation policy can achieve promising results.

1. Introduction

As a result of urbanization and industrialization, almost all big cities in the world face serious problems of traffic congestion. Therefore, it becomes more and more important to mitigate traffic congestion and to enhance the potential reserve capacity of road networks. Over the past forty years, many researchers investigated various management methods, such as signal control [1–4], congestion pricing [5–13], route guidance [14], and credit management [15, 16] to improve the network performance.

In the previous studies of traffic management, some researchers set their objective functions to be the minimization of the total travel time over the whole network [8, 15]. Furthermore, enhancing network capacity is also often used as an alternative objective function in traffic management [10, 17]. In those cases, network capacity is defined as the maximal demand that can be accommodated in the network, without violating capacity constraints of the links.

In the analysis of traffic management, scholars often use single approach for traffic management, and the combined methods are seldom utilized. With the rapid development of intelligent computing and control technologies [18, 19], it becomes more feasible to jointly implement various traffic management approaches. Due to lack of analytical research on the joint implementation of signal control and congestion pricing, the purpose of this paper is to present a policy of simultaneously optimizing traffic signal setting and congestion toll. Furthermore, both reserve capacity maximization and travel time minimization are used as the objectives of the policy planner.

The proposed problem is formulated as a Stackelberg game with the bilevel optimization structure. The upper level either minimizes total travel time or maximizes network reserve capacity with both signal setting parameters and congestion tolls as control variables. The lower level is a deterministic user equilibrium (DUE) problem given the signal setting and tolls assigned by the upper level. After replacing the lower level traffic assignment problem with its first order conditions, the proposed bilevel problem can be transferred into its equivalent single level formulation. By transferring the objective functions and link cost functions into piecewise linear functions, the whole problem becomes a linear program that can be solved by using commercial computing package, such as CPLEX.

The remainder of the paper is organised as follows. The bilevel programs that combine traffic problems are formulated in Section 2. In Section 3, the bilevel models are then

transferred into single level models, and their linearized formulations are also discussed. Section 4 presents the numerical example and discussed results. Conclusions are provided in Section 5.

2. Bilevel Formulation of Improving Network Capacity with Simultaneous Implementation of Signal Control and Congestion Pricing

Let $G = (N, A)$ be a directed transportation network defined by a set N of nodes and a set A of directed links. Each link ($a \in A$) has an associated flow-dependent travel time, $t_a(v_a)$, which presents the travel time per unit flow or average travel time on each link. The travel time function, $t_a(v_a)$, is assumed to be differentiable and monotonically increasing with the traffic flow, v_a. Let W denote the set of origin-destination (O-D) pairs and let P_w be the set of all paths between O-D pair. Each feasible path, $p \in P_w$, between O-D pair has a travel time $t_p^w = \sum_{a\in A} t_a(v_a)\delta_{ap}^w$. Herein δ_{ap}^w equals 1 if the path p between O-D pair uses link a, and 0 otherwise. The existing demand between O-D pair is denoted as q^wsuch that $\mathbf{q}$ is the vector of q^w.

The set of signal-controlled intersections is denoted as $I(I \subset N)$. Meanwhile, let A_i be the set of links entering the signalized intersection and let $\overline{A}$ be the set of all signal-controlled links, $\overline{A} = \{A_i, i \in I\}$.

The signal timing variables for links approaching a given signalized intersection should satisfy some linear constraints, which include cycle time, clearance time, and minimum and maximum green times. These constraints can be mathematically described in the following form:

$$\mathbf{G}_i\lambda_i \geq \mathbf{b}_i, \quad i \in I, \tag{1}$$

where λ_i is a vector of timing variables associated with signalized intersection. Both the matrix $\mathbf{G}_i$ and vector $\mathbf{b}_i$ depend on the particular timing specification for intersection, whether it is stage based or group based. For more detailed descriptions, the reader may refer to Allsop [20].

In this paper, we allow for link based tolls, which means a certain amount of toll (τ_a) is imposed on link a. Let $\boldsymbol{\tau}$ be the vector of link-based toll, τ_a. After transferring the monetary cost τ_a into the equivalent cost in time unit according to the value of time (VOT), ρ, the generalized travel cost (GTC) (including both travel time and toll charge) of passing link a is $t_a(v_a, \boldsymbol{\lambda}) + (\tau_a/\rho)$. Here, we assume $J(J \subseteq A)$ to be the set of toll links.

2.1. Bilevel Model of Minimizing Total Travel Time. The travel time minimization problem over the whole network can be modelled by a bilevel program, or a Stackelberg game. In such a leader-follower game, the leader cannot directly control the decision of the follower, but it can affect the behaviour of the follower by making its own decisions and anticipating the results. However, the follower can only react according to the decisions of the leader. In this study, the transport system planner is viewed as the leader, and its decision variables are signal setting and congestion tolls. The followers are travellers, and their route choice behaviours can be characterized by a deterministic traffic assignment, given the decisions made by the leader.

The behaviour of the leader, namely, the upper level problem, is given below

$$\begin{aligned} \min_{\mathbf{v},\boldsymbol{\tau},\boldsymbol{\lambda}} \quad & \sum_{a\in A} t_a(v_a, \boldsymbol{\lambda})\, v_a(\boldsymbol{\lambda}, \boldsymbol{\tau}) \\ \text{s.t.} \quad & \mathbf{G}_i\lambda_i \geq \mathbf{b}_i, \quad i \in I \\ & \tau_j^u \geq \tau_j \geq \tau_j^l, \quad j \in J, \end{aligned} \tag{2}$$

where $\mathbf{v}(\boldsymbol{\lambda}, \boldsymbol{\tau})$ is the vector of link traffic volume. $v_a(\boldsymbol{\lambda}, \boldsymbol{\tau})$ represents an equilibrium traffic flow which obtains from the following lower-level program [5]:

$$\min_{\mathbf{v}} \quad \sum_{a\in A} \int_0^{v_a} t_a(\omega, \boldsymbol{\lambda})\, d\omega + \sum_{j\in J} \frac{v_j\tau_j}{\rho} \tag{3}$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{p\in P_w} f_p^w = q^w, \quad w \in W \\ & v_a = \sum_{w\in W}\sum_{p\in P_w} f_p^w \delta_{ap}^w, \quad a \in A \\ & f_p^w \geq 0, \quad p \in P_w, \quad w \in W. \end{aligned} \tag{4}$$

2.2. Bilevel Model of Maximizing Network Reserve Capacity. In this subsection, we introduce the bilevel program of network reserve capacity maximization, which is jointly implementing signal control and congestion pricing. If the current OD matrix is multiplied by a factor μ, then it becomes $\mu\mathbf{q}$. Given the new demand matrix $\mu\mathbf{q}$, link flow $\mathbf{v}$ can be obtained by solving a traffic assignment based on the vector of the signal timing variables $\boldsymbol{\lambda}$ for all signalized intersections and the vector $\boldsymbol{\tau}$ of congestion tolls. If the degree of saturation on any link does not exceed a prescribed benchmark value of that link at the equilibrium condition, the congestion and emission in the network are acceptable. Namely, the following condition has to be satisfied:

$$v_a(\mu, \boldsymbol{\lambda}, \boldsymbol{\tau}) \leq p_a C_a(\boldsymbol{\lambda}), \quad a \in A, \tag{5}$$

where $C_a(\boldsymbol{\lambda})$ is the capacity of link which is dependent on the signal timings ($\boldsymbol{\lambda}$). Furthermore, $v_a(\mu, \boldsymbol{\lambda}, \boldsymbol{\tau})$ is the equilibrium traffic flow of link a that depends on the demand multiplier, signal settings, and congestion tolls. Parameter pa is the maximum acceptable degree of saturation for link a. The above constraint should be fulfilled for links at closely spaced intersections, since queues block neighbour intersections and congestion would spread over the whole network.

The largest multiplier of the O-D matrix that can be accommodated without violating the capacity constraints can be obtained by maximizing μ within the feasible region defined by the constraints for all links. Let the maximum acceptable value of OD multiplier μ be μ^*. Therefore, if $\mu^* > 1$ the network has reserve capacity of $100(\mu^* - 1)\mathbf{q}$, and if $\mu^* < 1$ the network is overloaded by $100(1 - \mu^*)\mathbf{q}$.

The combined signal control and congestion pricing optimization problem is formulated as a bilevel model or a Stackelberg game. The leader, namely, the system planner aims to maximize the network reserve capacity, by setting appropriate signals and tolls, whereas the follower, namely, travellers follow deterministic user equilibrium in terms of the generalized travel cost (GTC), which describes the minimum-cost path finding behavior of drivers in the transportation networks. Consequently, the upper level program is given by

$$\underset{\mu,\boldsymbol{\lambda},\boldsymbol{\tau}}{\text{Maximise}} \quad \mu \tag{6}$$

$$\begin{aligned} \text{subject to} \quad & v_a(\mu,\boldsymbol{\lambda},\boldsymbol{\tau}) \le p_a C_a(\boldsymbol{\lambda}), \quad a \in A \\ & \mathbf{G}_i \lambda_i \ge \mathbf{b}_i, \quad i \in I \\ & \tau_j^u \ge \tau_j \ge \tau_j^l, \quad j \in J, \end{aligned} \tag{7}$$

where the equilibrium flow $v_a(\mu,\boldsymbol{\lambda},\boldsymbol{\tau})$ is obtained by solving the following lower-level network equilibrium problem:

$$\min_{\mathbf{v}} \quad \sum_{a\in A}\int_0^{v_a} t_a(\omega,\boldsymbol{\lambda})\,d\omega + \sum_{j\in J}\frac{v_j\tau_j}{\rho} \tag{8}$$

$$\begin{aligned} \text{subject to} \quad & \sum_{p\epsilon P_w} \mu q^w, \quad w\epsilon W \\ & v_a = \sum_{w\epsilon W}\sum_{p\epsilon P_w} f_p^w \delta_{ap}, \quad a\epsilon A \\ & f_p^w \ge 0, \quad p\epsilon P_w, \quad w\epsilon W. \end{aligned} \tag{9}$$

3. Transformation to an Equivalent Single Level Formulation

Obviously, the proposed models are very difficult to solve due to their bilevel structure. In this section, we transfer the bilevel models into the single-level program and approximate them into a set of mixed integer linear programs. Therefore, they can be solved by commercial software, such as CPLEX.

3.1. Equivalent Single-Level Model of Minimizing Total Travel Time. First, we replace the user equilibrium traffic assignment problem with its first order condition. Accordingly, the total travel time minimization problem becomes a single-level formulation:

$$\min_{\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\tau}} \quad Z = \sum_{a\in A} t_a(x_a,\boldsymbol{\lambda})\,x_a \tag{10}$$

$$\text{s.t.} \quad \mathbf{G}_i\lambda_i \ge \mathbf{b}_i, \quad i \in I \tag{11}$$

$$\tau_j^u \ge \tau_j \ge \tau_j^l, \quad j \in J \tag{12}$$

$$\sum_{p\in P_w} f_p^w = q^w, \quad w \in W, \tag{13}$$

$$v_a = \sum_{w\in W}\sum_{p\in P_w} f_p^w \delta_{ap}^w, \quad a \in A, \tag{14}$$

$$c_p^w = \sum_{a\in A} t_a(v_a,\boldsymbol{\lambda})\,\delta_{ap}^w + \sum_{j\in J}\frac{\tau_j\delta_{jp}^w}{\rho}, \tag{15}$$

$$\begin{aligned} & f_p^w\left(c_p^w - \pi^w\right) = 0, \\ & p \in P_w, \quad w \in W, \end{aligned} \tag{16}$$

$$c_p^w - \pi^w \ge 0, \quad p \in P_w, \quad w \in W, \tag{17}$$

$$f_p^w \ge 0, \quad p \in P_w, \tag{18}$$

where π^w represents the least travel cost, including both travel time and toll of OD pair w. The symbol c_p^w denotes the generalized travel cost in the path p.

In the above proposed model, constraints (16)–(18) represent the deterministic route choice behaviour. Clearly this complementary condition is nonlinear and nonconvex, so it cannot be put into a linear program. Fortunately, Wang and Lo [21] formulated it into a set of mixed-integer constraints, as below

$$\begin{aligned} & L\varphi_p^w + \varepsilon \le f_p^w \le U\left(1-\varphi_p^w\right), \quad p \in P_w, \quad w \in W, \\ & \varphi_p^w \in \{0,1\}, \quad p \in P_w, \quad w \in W, \\ & L\varphi_p^w \le c_p^w - \pi^w \le U\varphi_p^w, \quad p \in P_w, \quad w \in W, \\ & c_p^w - \pi^w \ge 0, \quad p \in P_w, \quad w \in W, \end{aligned} \tag{19}$$

where L represents a negative constant with a very large absolute value, U is viewed as a very large positive constant, while ε is treated a very small positive value. And φ_p^w is a binary variable. Specifically, if $\varphi_p^w = 0$, one has $f_p^w > 0$ and $c_p^w = \pi^w$. If $\varphi_p^w = 1$, one has $f_p^w = 0$ and $c_p^w > \pi^w$. These two cases of φ_p^w are exactly equivalent to the above complementary condition. Therefore, conditions (16)–(18) can be replaced by conditions (19).

If we let $T_a(x_a,\boldsymbol{\lambda}) = t_a(x_a,\boldsymbol{\lambda})x_a$, the objective function becomes $\sum_{a\in A} T_a(x_a,\boldsymbol{\lambda})$. If $t_a(x_a,\boldsymbol{\lambda})$ and $T_a(x_a,\boldsymbol{\lambda})$ are transferred into linear functions, the above problem is a linear program. Fortunately, $t_a(x_a,\boldsymbol{\lambda})$ and $T_a(x_a,\boldsymbol{\lambda})$ can be approximated by a piecewise linear function with multiple segments. Let λ_k be the kth variable in the vector $\boldsymbol{\lambda}$, and K is the cardinality of $\boldsymbol{\lambda}$. The feasible domain of v_a, $[\underline{v}_a, \overline{v}_a]$ is partitioned into N segments, and the feasible domain of λ_k, $[\underline{\lambda}_k, \overline{\lambda}_k]$ is partitioned into M segments, respectively. Theoretically, the accuracy of linearizing $t_a(x_a,\boldsymbol{\lambda})$ can be guaranteed by setting sufficiently large N and M. In this study, for each link a, a series of values of $V_{a,n}$ are used to partition the feasible domain of v_a into many small segments, where $\underline{v}_a < V_{a,n} < V_{a,n+1} < \overline{v}_a$, $(n = 1,\ldots N-1)$. We denote $[V_{a,n}, V_{a,n+1}]$ as the region n of v_a. Similarly a series of values of $Z_{k,m}$ are used to partition the feasible domain of λ_k into many small segments, where $\underline{\lambda}_k < Z_{k,m} < Z_{k,m+1} < \overline{\lambda}_k$, $(m = 1,\ldots,M-1)$. We denote $[Z_{k,m}, V_{k,m+1}]$ as region m of λ_k. For each region $(n, m_1,\ldots,m_k,\ldots,m_K)$, the following linear function

is specified to approximate the nonlinear travel time function, $t_a(x_a, \boldsymbol{\lambda})$

$$t_a(v_a, \lambda) = E_n^a v_a + \sum_{k=1}^{K} F_m^k \lambda_k + G_{m_1,\ldots,m_k,\ldots,m_K}^{a,n} \quad \text{if } U_{a,n} \le v_a \le U_{a,n+1},\ Z_{k,m} \le \lambda_k \le Z_{k,m+1}, \tag{20}$$

where E_n^a and F_m^k are coefficients. The first-order Taylor series is applied to approximate the travel time function $t_a(x_a, \boldsymbol{\lambda})$. Therefore, the coefficients E_n^a and F_m^k are determined by the derivatives of the travel time function with respect to v_a and λ_k that are evaluated at $V_{a,n}$ and $Z_{k,m}$, namely,

$$E_n^a = \left.\frac{\partial t_a}{\partial v_a}\right|_{(V_{a,n}, Z_{1,m_1},\ldots,Z_{k,m_k},\ldots Z_{K,m_K})}. \qquad F_m^k = \left.\frac{\partial t_a}{\partial \lambda_i}\right|_{(V_{a,n}, Z_{1,m_1},\ldots,Z_{k,m_k},\ldots Z_{K,m_K})}. \tag{21}$$

And the coefficient $G_{m_1,\ldots,m_k,\ldots,m_K}^a$ can be evaluated by equating the values of the original function and the piecewise linear approximated function at $U_{a,n}$ and V_{λ_k,m_k}, and thus given by

$$\begin{aligned} G_{m_1,\ldots,m_k,\ldots,m_K}^{a,n} &= t_a\left(V_{a,n}, Z_{1,m_1}, \ldots, Z_{k,m_k}, \ldots Z_{K,m_K}\right) \\ &\quad - V_{a,n} \left.\frac{\partial t_a}{\partial v_a}\right|_{(V_{a,n}, Z_{1,m_1},\ldots,Z_{k,m_k},\ldots Z_{K,m_K})} \\ &\quad - \sum_{k=1}^{K} Z_{k,m_k} \left.\frac{\partial t_a}{\partial \lambda_k}\right|_{(V_{a,n}, Z_{1,m_1},\ldots,Z_{k,m_k},\ldots Z_{K,m_K})}. \end{aligned} \tag{22}$$

Subsequently, the piecewise linear travel time function of each link is transferred into the following equivalent mixed-integer linear constraints:

$$\begin{aligned} &L \cdot \xi_{a,n} \le v_a - V_{a,n} \le U \cdot (1 - \xi_{a,n}) - \varepsilon, \\ &\theta_{a,n} = \xi_{a,n+1} - \xi_{a,n}, \\ &L \cdot \varsigma_{m_k} \le \lambda_i - Z_{k,m_k} \le U \cdot \left(1 - \varsigma_{m_k}\right) - \varepsilon, \\ &\vartheta_{m_k} = \varsigma_{m_k+1} - \xi_{m_k}, \\ &\psi_{n,m_1,\ldots,m_k,\ldots,m_K}^{a} = \theta_{a,n} + \sum_{k=1}^{K} \vartheta_{m_k}, \\ &L \cdot \left(K + 1 - \psi_{m_1,\ldots,m_k,\ldots,m_K}^{a,n}\right) \le t_a - \left(E_n^a v_a + \sum_{k=1}^{K} F_m^k \lambda_k\right) \\ &\qquad \le U \cdot \left(K + 1 - \psi_{m_1,\ldots,m_k,\cdots,m_K}^{a,n}\right), \\ &\xi_{a,n} \in \{0,1\}, \quad \varsigma_{m_k} \in \{0,1\}, \quad n = 1,2,\ldots,N, \\ &m_k = 1,2,\ldots,M, \quad \forall a \in A, \end{aligned} \tag{23}$$

where L and U still represent very large negative and positive constants, respectively. And ε is still a very small positive constant. The binary variable $\xi_{a,n}$ indicates the comparison between v_a and $V_{a,n}$. Specifically, $\xi_{a,n} = 0$ indicates $v_a \ge V_{a,n}$, $\xi_{a,n} = 1$ indicates $v_a < U_{a,n}$. Thus $\theta_{a,n}$ indicates whether v_a falls in segment n or not. If $\theta_{a,n} = 1$, v_a is in segment n. Similarly, $\vartheta_{m_k} = 1$ means λ_k falls in segment m_k. If $\psi_{m_1,\ldots,m_k,\ldots,m_K}^{a,n} = K + 1$, then the corresponding approximated linear function in the region $(n, m_1, \ldots, m_k, \ldots, m_K)$ is utilized, namely, $t_a = E_n^a v_a + \sum_{k=1}^{K} F_m^k \lambda_k + G_{m_1,\ldots,m_k,\ldots,m_K}^{a,n}$. In this way, all the nonlinear constraints of the single-level formulation have been transferred into linear ones.

Tracing the same way, $T_a(x_a, \boldsymbol{\lambda})$ can be approximated into linear functions. Thus the whole problem becomes a mixed integer linear program, which can be solved by commercial software, such as CPLEX.

3.2. Equivalent Single-Level Model of Maximizing Network Reserve Capacity. If the lower level problem is replaced with its first order condition, the combined signal control and pricing problem with the objective of maximizing reserve capacity can be transferred into a single-level formulation, as below

$$\underset{\mu,\boldsymbol{\lambda},\boldsymbol{\tau}}{\text{Maximise}} \quad \mu \tag{24}$$

$$\begin{aligned} \text{s.t.} \quad & v_a(\mu, \boldsymbol{\lambda}, \boldsymbol{\tau}) \le p_a C_a(\boldsymbol{\lambda}), \quad a \in \overline{A} \\ & \mathbf{G}_i \lambda_i \ge \mathbf{b}_i, \quad i \in I, \\ & \tau_j^u \ge \tau_j \ge \tau_j^l, \quad j \in J, \\ & \sum_{p \in P_w} f_p^w = \mu q^w, \quad w \in W, \\ & v_a = \sum_{w \in W} \sum_{p \in P_w} f_p^w \delta_{ap}^w, \quad a \in A, \\ & c_p^w = \sum_{a \in A} t_a(v_a, \boldsymbol{\lambda}) \delta_{ap}^w + \sum_{j \in J} \frac{\tau_j \delta_{jp}^w}{\rho}, \\ & L\varphi_p^w + \varepsilon \le f_p^w \le U\left(1 - \varphi_p^w\right), \\ & p \in P_w, \quad w \in W, \\ & \varphi_p^w \in \{0,1\}, \quad p \in P_w, \quad w \in W, \\ & L\varphi_p^w \le c_p^w - \pi^w \le U\varphi_p^w, \\ & p \in P_w, \quad w \in W, \\ & c_p^w - \pi^w \ge 0, \quad p \in P_w, \quad w \in W, \\ & f_p^w \ge 0, \quad p \in P_w, \end{aligned} \tag{25}$$

where π^w represents the minimum generalized travel cost of OD pair w. Notation c_p^w denotes the generalized travel cost of path p. Furthermore, $\mathbf{f}$, $\mathbf{c}$, and $\boldsymbol{\pi}$ are vectors of f_p^w, c_p^w, and π^w, respectively.

In this program, the objective function is obviously a linear function. And $t_a(x_a, \boldsymbol{\lambda})$ can also be linearized as done in Section 3.1.

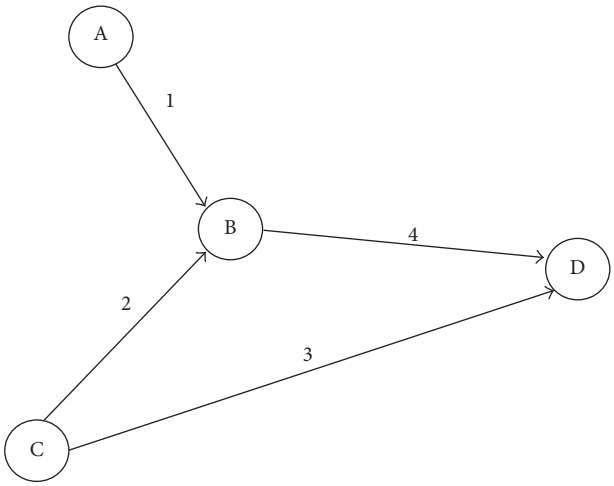

Figure 1: A signal-controlled road network.

Table 1: Input data to the example network.

Link number a	1	2	3	4
Free-flow time t_a^0 (min)	4.6	5.2	5.1	3.9
Saturation flow s_a (veh/min)	52	50	20	80

Table 2: Solution of total travel time minimization with signal control only.

Link number a	1	2	3	4
Traffic volume (veh/min)	10	0	20	10
Travel time (min)	4.95	4.83	9.00	4.18
Volume to capacity ratio	0.19	NA	1.00	0.13

Table 3: Solution of total travel time minimization with both signal and pricing.

Link number a	1	2	3	4
Traffic volume (veh/min)	10	4.70	15.30	14.70
Travel time (min)	5.12	5.78	8.08	4.30
Volume to capacity ratio	0.29	0.28	0.77	0.18

Table 4: Solution of reserve capacity maximization with both signal and pricing.

Link number a	1	2	3	4
Traffic volume (veh/min)	23.92	27.0	20.0	50.28
Travel time (min)	6.36	7.29	8.98	5.28
Volume to capacity ratio	1.0	1.0	1.0	0.63

4. A Numerical Example

Consider an example network, shown in Figure 1, with 7 links and 6 nodes, of which nodes E and F are signal-controlled intersections. The current O-D demand from node A to node D is 10 veh/min and that from C to D is 20 veh/min. There is only one path ABD for the O-D pair (A, B),while there are two paths CBD and CD for O-D pair (C, D). The delay formula of links takes the following form:

$$t_a(v_a, \lambda_a) = t_a^0 + \theta_a \times \left(\frac{v_a}{\lambda_a s_a}\right), \tag{26}$$

where λ_a is the proportion of a cycle that is effectively green for link α, and $\lambda_a = 1.0$ for any link that does not enter into a signal-controlled junction. The values of t_a^0, θ_a, and s_a are given in Table 1.

For the signalized intersections B, signal control is represented by two split parameters (proportions of green times) λ_1 and λ_2. The proportion of green time allocated to link 1 is λ_1, and the proportion allocated to link 2 is λ_2. Loss time of phase transition is ignored, namely, $\lambda_1 + \lambda_2 = 1$. Therefore the capacity of link 1 is $\lambda_1 s_1$, and the capacity of link 2 is $\lambda_2 s_2$. The lower and upper bounds of the proportion of green time are $0.05 \le \lambda_1, \lambda_2 \le 0.95$.

With only signal control, the total travel time in the network is minimized at $\lambda_1 = 1.00$ and $\lambda_2 = 0$, which means green time is fully assigned to link 1. The minimal total travel time is 271.21 min. The corresponding link volume, link travel time, and volume to capacity ratio are listed in Table 2.

Now we consider the policy of the joint implementation of signal control and congestion pricing, wherein a toll τ_3 is charged on link 3. Assume that the value of time is 1.0 \$/min for travellers. Using the model developed in Section 2.1, after piecewisely linearizing the objective function and the delay formula, the solution is obtained at $\lambda_1 = 0.66$, $\lambda_2 = 0.34$, and $\tau_3 = \$2.0$. At the optimum, total travel time is 265.36 min. By introducing congestion pricing, the total travel time decreases by 5.85 min. The corresponding link volume, link travel time, and volume to capacity ratio are listed in Table 3.

Assume that the maximum acceptable level of link volume is exactly its capacity, and then we investigate the reserve capacity maximization. If signal control is the only policy, the solution is given by $\lambda_1 = 1.00$ and $\lambda_2 = 0$, which is the same as the case of the total travel time minimization. The maximal demand multiplier is 1.0, and the network has no reserve capacity. The link volumes are also shown in Table 2. Clearly, link 3 is the critical link since its volume reaches the capacity.

When signal control and congestion pricing are simultaneously implemented, the model developed in Section 2.2 is used to maximize the reserve capacity. After piecewisely linearizing the delay formula, the solution is obtained at $\lambda_1 = 0.46$, $\lambda_2 = 0.54$, and $\tau_3 = \$3.59$. At the optimum, the reserve capacity is 2.34. By introducing congestion pricing, the network demand multiplier increases from 1.0 to 2.34. In other words, the network reserve capacity increases from 0 to 1.34. In this case, the link volume, link travel time, and volume to capacity ratio are listed in Table 4. Clearly links 1, 2, and 3 are critical links, because they will be operated at their full capacities when the network serves 2.34 times the existing demand levels.

This numerical example shows that the network performance (in terms of both reserve capacity maximization and system time minimization) can be significantly improved by further introducing congestion pricing, besides implementing signal control.

5. Conclusions

This paper proposed a joint implementation policy of signal control and congestion toll optimization in the transportation network. The objective of the system planner is either to minimize the total travel time of the whole network or to maximize the reserve capacity of the network. The reserve capacity of a network is defined as a multiplier that raises the demand of each OD pair by the same proportion without violating capacity constraints of all links. The policy is formulated as two bilevel models, depending on which objective is chosen. The objective in the upper level is to minimize system travel time or to maximize reserve capacity. The problem in the lower level is a traffic assignment problem, considering both signal setting and congestion pricing. By reformulating the lower level problem with its first order conditions, we then transfer the bilevel programs into the equivalent single level programs. After transferring the objective function of the system travel time and the link cost formula into the piecewise linear functions, the whole problem can be characterized as a mixed integer program. The numerical example indicates that the network performance can be significantly improved by further introducing congestion pricing, besides implementing signal control.

Acknowledgments

This study has been substantially supported by two grants from the National Natural Science Foundation Council of China (Projects nos. 71125004, 71361130014).

References

[1] R. E. Allsop, "Estimating the traffic capacity of a signalized road junction," *Transportation Research*, vol. 6, pp. 245–255, 1972.

[2] S. Yagar, "Capacity of a signalized road junction: critique and Extensions," *Transportation Research*, vol. 8, pp. 137–147, 1974.

[3] S. Yagar, "Minimizing delays for transient demands with application to signalized road junctions," *Transportation Research*, vol. 11, pp. 53–62, 1977.

[4] H. Yang and S. Yagar, "Traffic assignment and signal control in saturated road networks," *Transportation Research A*, vol. 29, pp. 125–139, 1995.

[5] H. Yang and X. N. Zhang, "Multiclass network toll design problem with social and spatial equity constraints," *Journal of Transportation Engineering-ASCE*, vol. 128, no. 5, pp. 420–428, 2002.

[6] H. Yang and X. N. Zhang, "Optimal toll design in second-best link-based congestion pricing," *Transportation Research Record*, no. 1857, pp. 85–92, 2003.

[7] H. Yang, X. N. Zhang, and Q. Meng, "Modeling private highways in networks with entry-exit based toll charges," *Transportation Research B*, vol. 38, no. 3, pp. 191–213, 2004.

[8] X. N. Zhang and H. Yang, "The optimal cordon-based network congestion pricing problem," *Transportation Research B*, vol. 38, no. 6, pp. 517–537, 2004.

[9] X. N. Zhang and B. van Wee, "Efficiency comparison of various parking charge schemes considering daily travel cost in a linear city," *European Journal of Transport and Infrastructure Research*, vol. 11, no. 2, pp. 234–255, 2011.

[10] X. N. Zhang and B. van Wee, "Enhancing transportation network capacity by congestion pricing with simultaneous toll location and toll level optimization," *Engineering Optimization*, vol. 44, no. 4, pp. 477–488, 2012.

[11] X. N. Zhang, H.-J. Huang, and H. M. Zhang, "Integrated daily commuting patterns and optimal road tolls and parking fees in a linear city," *Transportation Research B*, vol. 42, no. 1, pp. 38–56, 2008.

[12] X. N. Zhang, H. Yang, and H.-J. Huang, "Multiclass multicriteria mixed equilibrium on networks and uniform link tolls for system optimum," *European Journal of Operational Research*, vol. 189, no. 1, pp. 146–158, 2008.

[13] X. N. Zhang, H. M. Zhang, H. J. Huang, L. J. Sun, and T. Q. Tang, "Competitive, cooperative and Stackelberg congestion pricing for multiple regions in transportation networks," *Transportmetrica*, vol. 7, no. 4, pp. 297–320, 2011.

[14] H. Yang and X. Zhang, "Modeling competitive transit and road traffic information services with heterogeneous endogenous demand," *Journal of Transportation Research Board, Transportation Research Record*, vol. 1783, pp. 7–16, 2002.

[15] H. Yang and X. L. Wang, "Managing network mobility with tradable credits," *Transportation Research B*, vol. 45, no. 3, pp. 580–594, 2011.

[16] X. N. Zhang, H. Yang, and H.-J. Huang, "Improving travel efficiency by parking permits distribution and trading," *Transportation Research B*, vol. 45, no. 7, pp. 1018–1034, 2011.

[17] S. C. Wong and H. Yang, "Reserve capacity of a signal-controlled road network," *Transportation Research B*, vol. 31, no. 5, pp. 397–402, 1997.

[18] B. He, Z. P. Wang, C. Q. Liu, Y. G. Li, and R. J. Shen, "Swimming behavior analysis based on bacterial chemotaxis in solution," *Journal of Bionic Engineering*, vol. 9, no. 3, pp. 315–321, 2012.

[19] B. He, Z. P. Wang, Q. Li, H. Xie, and R. J. Shen, "Analytic method for kinematics and dynamics of multiple-backbone continuum robot," *International Journal of Advanced Robotic Systems*, vol. 10, no. 84, pp. 1–13, 2013.

[20] R. E. Allsop, "Evolving applications of mathematical optimization in design and operation of individual signal-controlled road junctions," in *Mathematics in Transport Planning and Control*, J. D. Griffiths, Ed., pp. 1–24, Clarendon Press, Oxford, UK, 1989.

[21] D. Z. W. Wang and H. K. Lo, "Global optimum of the linearized network design problem with equilibrium flows," *Transportation Research Part B*, vol. 44, no. 4, pp. 482–492, 2010.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 492421, 20 pages
http://dx.doi.org/10.1155/2013/492421

Review Article

A Review of Ranking Models in Data Envelopment Analysis

F. Hosseinzadeh Lotfi,[1] G. R. Jahanshahloo,[1] M. Khodabakhshi,[2] M. Rostamy-Malkhlifeh,[1] Z. Moghaddas,[1] and M. Vaez-Ghasemi[1]

[1] *Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran*
[2] *Department of Mathematics, Faculty of Science, Lorestan University, Khorramabad, Iran*

Correspondence should be addressed to F. Hosseinzadeh Lotfi; farhad@hosseinzadeh.ir

Received 26 January 2013; Accepted 27 May 2013

Academic Editor: Hadi Nasseri

In the course of improving various abilities of data envelopment analysis (DEA) models, many investigations have been carried out for ranking decision-making units (DMUs). This is an important issue both in theory and practice. There exist a variety of papers which apply different ranking methods to a real data set. Here the ranking methods are divided into seven groups. As each of the existing methods can be viewed from different aspects, it is possible that somewhat these groups have an overlapping with the others. The first group conducts the evaluation by a cross-efficiency matrix where the units are self- and peer-evaluated. In the second one, the ranking units are based on the optimal weights obtained from multiplier model of DEA technique. In the third group, super-efficiency methods are dealt with which are based on the idea of excluding the unit under evaluation and analyzing the changes of frontier. The fourth group involves methods based on benchmarking, which adopts the idea of being a useful target for the inefficient units. The fourth group uses the multivariate statistical techniques, usually applied after conducting the DEA classification. The fifth research area ranks inefficient units through proportional measures of inefficiency. The sixth approach involves multiple-criteria decision methodologies with the DEA technique. In the last group, some different methods of ranking units are mentioned.

1. Introduction

Data envelopment analysis as a mathematical tool was initiated by Farrell [1] and Charnes et al. [2]. They formulated a linear programming problem with which it is possible to evaluate decision-making units (DMUs) with multiple inputs and outputs. Note that in this technique it is not necessary to know the production function. In this technique, an LP problem is solved for each DMU, and the relative efficiency of each unit obtained as a linear combination of corresponding optimal weights. In this problem, weights are free to get their value to show the under evaluation unit in optimistic viewpoint. Those units with optimal objective function equal to one are called "best practice." These units are located onto the efficient frontier, and those far away from this frontier are called inefficient. As proved in DEA literature, at least one of the units is located onto this frontier. Note that units located unto this frontier can be considered as benchmarks for inefficient units. Based on what Charnes et al. [2] provided, many extensions to DEA Models are presented in the literature. As the example of the most important one Banker et al. [3] can be mentioned. Also, multiplicative and additive models developed in the literature by Charnes et al. [4–6]. In that time Thrall [7] provided a complete comparison of all classic DEA models.

One of the important issues discussed in DEA literature is ranking efficient units since the efficient units obtained in the efficiency score of one cannot be compared with each other on the basis of this criterion any more. Therefore, it seems necessary to provide models for further discrimination among these units. Many papers are presented in the literature review for ranking the efficient units. Note that Adler and Golany [8] used principle component analysis for improving the discrimination of DEA. But this attempt was not sufficient

and just made a reduction in number of efficient units not rank units completely. Many papers presented in the literature for ranking efficient units; one of the first papers is Young and Hamer [9]. One important field in ranking is cross-efficiency; to name a few, consider Sexton et al. [10], Rödder and Reucher [11], Örkcü and Bal [12], Wu et al. [13], Jahanshahloo et al. [14], Wang et al. [15], Ramón et al. [16], Guo and Wu [17], Contreras [18], Wu et al. [19], Zerafat Angiz et al. [20], and Washio and Yamada [21].

In the literature there exist other methods based on finding optimal weights in DEA analysis as Jahanshahloo et al. [22], Wang et al. [23], Alirezaee and Afsharian [24], Liu and Hsuan Peng [25], Wang et al. [26], Hatefi and Torabi [27], Hosseinzadeh Lotfi et al. [28], Wang et al. [29], and Ramón et al. [30].

One of the important fields in ranking is super efficiency presented by Andersen and Petersen [31], Mehrabian et al. [32], Tone [33], Jahanshahloo et al. [34], Jahanshahloo et al. [35], Chen and Sherman [36], Amirteimoori et al. [37], Jahanshahloo et al. [38], Li et al. [39], Sadjadi et al. [40], Gholam Abri et al. [41], Jahanshahloo et al. [42], Noura et al. [43], Ashrafi et al. [44], Chen et al. [45], Rezai Balf et al. [46], and Chen et al. [47].

Another important field in ranking is benchmarking methods such as Torgersen et al. [48], Sueyoshi [49], Jahanshahloo et al. [50], Lu and Lo [51], and Chen and Deng [52].

One important field is using statistical tools for ranking units first suggested by Friedman and Sinuany-Stern [53] and Mecit and Alp [54].

One of the significant fields in ranking is unseeing multicriteria decision-making (MCDM) methodologies and DEA analysis. To mention a few, consider Joro et al. [55], Li and Reeves [56], Belton and Stewart [57], Sinuany-Stern et al. [58], Strassert and Prato [59], Chen [60], Jablonsky [61], Wang and Jiang [62], and Hosseinzadeh Lotfi et al. [63].

Also there exist some other ranking methods not much developed and extended in the literature, Seiford and Zhu [64], Jahanshahloo [65], Jahanshahloo et al. [34], Jahanshahloo et al. [66], Jahanshahloo and Afzalinejad [67], Amirteimoori [68], Kao [69], Khodabakhshi and Aryavash [70], and Zerafat Angiz et al. [71].

In addition to the theoretical papers presented in ranking literature there exist a variety of papers which used these new models in applications such as Charnes et al. [72], Cook and Kress [73], Cook et al. [74], Martić and Savić [75], De Leeneer and Pastijn [76], Lins et al. [77], Paralikas and Lygeros [78], Ali and Nakosteen [79], Martin and Roman [80], Raab and Feroz [81], Wang et al. [26], Williams and Van Dyke [82], Jürges and Schneider [83], Giokas and Pentzaropoulos [84], Darvish et al. [85], Lu and Lo [51], Feroz et al. [86], Sadjadi et al. [40], Ramón et al. [30], and Sitarz [87].

There exist some papers which reviewed ranking methods, as Adler et al. [88]. In this paper, most of the ranking methods, specially the new ones described in the literature, are reviewed. Here, the different ranking methods are classified into seven groups after reviewing the basic DEA method in Section 2. In Section 3, the cross-efficiency technique will be discussed. In this method first suggested by Sexton et al. [10], the DMUs are self- and peer-assessed. In Section 4, some of the ranking methods based on optimal weights obtained from DEA models, common set of weights are briefly reviewed. Super-efficiency methods first introduced by Andersen and Petersen [31] will be reviewed in Section 5. The basic idea is based on the idea of leaving out one unit and assessing by the remaining units. Section 6 discusses the evaluation of DMUs through benchmarking, an approach originating in Torgersen et al. [48]. Section 7 will review the papers which use the statistical tools for ranking units first suggested by Friedman and Sinuany-Stern [53] such as canonical correlation analysis and discriminant analysis. Section 8 discusses the ranking of units based on multi-criteria decision-making (MCDM) methodologies and DEA. Section 9 discusses some different ranking methods existing in the DEA literature. Section 10 presents the results of the various methodologies applied to an example.

2. Data Envelopment Analysis

Data envelopment analysis is a mathematical programming technique for performance evaluation of a set of decision-making units.

Let a set consists of n homogeneous decision-making units to be evaluated. Assume that each of these units uses m inputs x_{ij} $(i = 1, \ldots, m)$ to produce s outputs y_{rj} $(r = 1, \ldots, s)$. Moreover, $X_j \in R^m$ and $Y_j \in R^s$ are considered to be nonnegative vectors. We define the set of production possibility as $T = \{(X, Y) \mid X \text{ can produce } Y\}$.

When variable returns to scale form of technology is assumed we have $T = T_{\mathrm{BCC}}$ and

$$T_{\mathrm{BCC}} = \left\{ (x, y) \mid x \geq \sum_{j=1}^{n} \lambda_j x_j, \; y \leq \sum_{j=1}^{n} \lambda_j y_j, \sum_{j=1}^{n} \lambda_j = 1, \; \lambda_j \geq 0, \; j = 1, \ldots, n \right\}, \tag{1}$$

and when constant returns to scale form of technology is assumed we have $T = T_{\mathrm{CCR}}$ and

$$T_{\mathrm{CCR}} = \left\{ (x, y) \mid x \geq \sum_{j=1}^{n} \lambda_j x_j, y \leq \sum_{j=1}^{n} \lambda_j y_j, \; \lambda_j \geq 0, \; j = 1, \ldots, n \right\}. \tag{2}$$

The two-phase enveloping problem with constant returns to scale form of technology, first provided by Charnes et al. [2], is as follows:

$$\begin{aligned} \min \quad & \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \\ \text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{io}, \quad i = 1, \ldots, m, \\ & \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{ro}, \quad r = 1, \ldots, s, \\ & \lambda_j \geq, \quad j = 1, \ldots, n. \end{aligned} \tag{3}$$

The two-phase enveloping problem with variable returns to scale form of technology, first provided by Banker et al. [3], is as follows:

$$\begin{aligned} \min \quad & \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \\ \text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{io}, \quad i = 1, \ldots, m, \\ & \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{ro}, \quad r = 1, \ldots, s, \\ & \sum_{j=1}^{n} \lambda_j = 1, \\ & \lambda_j \geq 0, \quad j = 1, \ldots, n. \end{aligned} \tag{4}$$

As regards the above-mentioned problems and due to corresponding feasible region it is evident that $\theta^{*\mathrm{CCR}} \leq \theta^{*\mathrm{BCC}}$.

According to the definition of T_{CCR} and T_{BCC}, an envelop constructed through units called best practice or efficient. Considering mentioned problems if a DMU_o is not CCR (BCC) efficient, it is possible to project this DMU onto the CCR (BCC) efficiency frontier considering the following formulas:

$$\begin{aligned} \hat{x}_{io} &= \theta^* x_{io} - s_i^{-*} = \sum_{j=1}^{n} \lambda_j^* x_{ij}, \quad i = 1, \ldots, m, \\ \hat{y}_{ro} &= y_{ro} + s_r^{+*} = \sum_{j=1}^{n} \lambda_j^* y_{rj}, \quad r = 1, \ldots, s. \end{aligned} \tag{5}$$

The dual model corresponding to the following model is as follows:

$$\begin{aligned} \max \quad & \sum_{r=1}^{s} u_r y_{ro} \\ \text{s.t.} \quad & \sum_{i=1}^{m} v_i x_{io} = 1, \\ & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad j = 1, \ldots, n, \\ & U \geq 0, \quad V \geq 0. \end{aligned} \tag{6}$$

For overcoming the problem of zero weights, and variability of weights use of assurance region is suggested by Thompson et al. [89–91].

As mentioned in the literature usually there exist more than one efficient units, and these units cannot be further compared to each other on basis of efficiency scores. Thus, it felt necessary to provide new models for ranking these units. There exist a variety of ranking models in context of data envelopment analysis. In the remaining of this paper we review some of these models.

3. Cross-Efficiency Ranking Techniques

Sexton et al. [10] provided a method for ranking units based on this idea that units are self- and peer-evaluated. For deriving the cross-efficiency of any DMU_j using weights chosen by DMU_o, they proposed the following equation:

$$\theta_{oj} = \frac{U_o^* Y_j}{V_o^* X_j}, \tag{7}$$

where U^*, V^* are optimal weights obtained from the following model for DMU_o under assessment

$$\begin{aligned} \min \quad & V^t X_o \\ \text{s.t.} \quad & U^t Y_o = 1, \\ & U^t Y_j - V^t X_j \leq 0, \quad j = 1, \ldots, n, \\ & U \geq 0, \quad V \geq 0. \end{aligned} \tag{8}$$

Now DMU_o received the average cross-efficiency score as $\overline{\theta}_o = \sum_{j=1}^{n} \theta_{oj}/n$; for further details about this averaging see also Green et al. [92]. Doyle and Green [93] also used cross-efficiency matrix for ranking units. According to this method for ranking DMUs, many investigations have been done as reviewed in Adler et al. [88].

Rödder and Reucher [11] presented a consensual peer-based DEA model for ranking units. As the authors said, this method is generalized twofold. The first is an optimal efficiency improving input allocation; the second aim is the choice of a peer DMU whose corresponding price is acceptable for the other units. Consider

$$\begin{aligned} \max \quad & V_k^{*T} W_l \\ \text{s.t.} \quad & U_k^{*T} Y_l - V_k^{*T} W_l \leq 0, \\ & W_l - \sum_j \mu_{lj} X_j \geq 0, \\ & \sum_j \mu_{lj} Y_j \geq Y_l, \\ & \mu_{lj} \geq 0, \quad \forall j. \end{aligned} \tag{9}$$

The higher the degree of input variation is, the better the chance to be efficient will be.

Örkcü and Bal [12] provided a goal programming technique to be used in the second stage of the cross-evaluation. Their modified model is as follows:

$$\begin{aligned}
\min\ & a = \left\{ \sum_{j=1}^{n} \eta_j, \sum_{j=1}^{n} \alpha_j \right\} \\
\text{s.t.}\ & \sum_{i=1}^{m} v_{ip} x_{ip} = 1, \\
& \sum_{r=1}^{s} u_{rp} y_{rp} - \theta_{pp}^{*} \sum_{i=1}^{m} v_{ip} x_{ip} = 0, \\
& \sum_{r=1}^{s} u_{rp} y_{rj} - \sum_{i=1}^{m} v_{ip} x_{ij} + \alpha_j = 0, \quad j = 1, \ldots, n, \\
& M - \alpha_j + \eta_j - p_j = 0, \quad j = 1, \ldots, n, \\
& u_{rp} \geq 0, \quad v_{ip} \geq 0, \quad r = 1, \ldots, s, \quad i = 1, \ldots, m, \\
& \alpha_j \geq 0, \quad \eta_j \geq 0, \quad p_j \geq 0, \quad j = 1, \ldots, n.
\end{aligned} \tag{10}$$

As the authors noted, there exist alternative optimal solutions.

Wu et al. [13] described the main suffering of cross-efficiency when the ultimate average cross-efficiency utilized for ranking units. For removing this shortcoming they eliminated the assumption of average and utilized the Shannon entropy in order to obtain the weights for ultimate cross-efficiency scores. Jahanshahloo et al. [14] provided a method for selecting symmetric weights to be used in DEA cross-efficacy.

Step 1. Efficiency of DMUs needs to be computed.

Step 2. Choose the solutions, in accordance with the secondary goal for each DMU, as follows:

$$\begin{aligned}
\min\ & e^T Z_o e \\
\text{s.t.}\ & u_o y_o = 1, \\
& v_o X_o = \theta_o, \\
& u_o Y - v_o X \leq 0, \\
& u_{oi} y_{oi} - u_{oj} y_{oj} \leq z_{oij}, \quad \forall i, j, \\
& u_{oj} y_{oj} - u_{oi} y_{oi} \leq z_{oij}, \quad \forall i, j, \\
& u_o, v_o \geq d.
\end{aligned} \tag{11}$$

Step 3. The cross-efficiency for any DMU_j, using the weights that DMU_o has chosen in the previous model, is then used as follows:

$$\theta_{oj} = \frac{u_o^* Y_j}{v_o^* X_j}. \tag{12}$$

Wang et al. [15] provided a cross-efficiency evaluation based on ideal and anti-ideal units for ranking. As the authors mentioned, a DMU could choose a unique set of input and output weights to make its distance from ideal DMU as small as possible, or the distance from anti-ideal DMU as large as possible, or the both. Thus, according to this idea they proposed the following procedure for cross-efficiency evaluation.

Model 1. Minimization of the distance from ideal DMU.

Model 2. Maximization of the distance from anti-ideal DMU.

Model 3. Maximization of the distance between ideal DMU and anti-ideal DMU.

Model 4. Maximization of the relative closeness.

The authors mentioned that the bigger the relative closeness of a DMU is the better performance it will have.

In a paper, Ramón et al. [16] selected the profiles of weights used in cross-efficiency assessment. As the authors said they tried to prevent unrealistic weighting. They have discussed the zero weights as they excluded variables from the evaluation. In the calculation of cross-efficiency scores, they proposed to ignore the profiles of those weights of the unit under evaluation that among their alternate optimal solutions cannot choose nonzero weights. They also proposed the "peer-restricted" cross-efficiency evaluation where the units assessed in a peer evaluation which means profiles of weights of some inefficient units are not considered. Finally, the presented approach extended to derive a common set of weights. Guo and Wu [17] provided a complete ranking of DMUs with undesirable outputs using restriction in DEA. As the author mentioned this model is presented to realize a unique ranking of units by "maximal balanced index" according to the obtained optimal shadow prices

$$\begin{aligned}
\max\ & \sum_{i=1}^{m} v_i w_i + \sum_{t=1}^{k} \eta_t h_t - \sum_{r=1}^{s} u_r q_r \\
\text{s.t.}\ & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - \sum_{t=1}^{k} \eta_r b_{tj} \leq 0, \quad j = 1, \ldots, n, \\
& \sum_{i=1}^{m} v_i x_{ip} + \sum_{t=1}^{k} \eta_r b_{tp} = 1, \\
& \sum_{r=1}^{s} u_r y_{rp} = \mathrm{EEF}_p, \\
& U, V, \eta \geq 0,
\end{aligned} \tag{13}$$

where EEF_p is the optimal objective function of multiplier model.

Contreras [18] used cross-evaluation for ranking units in DEA methodology. The idea is based upon introducing a model for optimizing the rank position of DMUs

$$\begin{aligned}
&\min \quad r_{kk} \\
&\text{s.t.} \quad \theta_{kk} = \theta_{kk}^*, \\
&\qquad \theta_{lk} - \theta_{jk} + \delta_{lj}^k \beta \geq 0, \quad l \neq j, \\
&\qquad \theta_{lk} - \theta_{jk} + \gamma_{lj}^k \beta \geq \varepsilon, \quad l \neq j, \\
&\qquad \delta_{lj}^k + \delta_{jl}^k \leq 1, \quad l \neq j, \\
&\qquad \delta_{lj}^k + \gamma_{jl}^k = 1, \quad l \neq j, \\
&\qquad r_{lk} = \sum_{\substack{j=1 \\ j \neq l}}^{n} \frac{\delta_{lj}^k + \gamma_{jl}^k}{2} + 1, \quad l = 1, \ldots, n, \\
&\qquad \gamma_{lj}^k, \delta_{lj}^k \in \{0, 1\}, \quad l \neq j.
\end{aligned} \tag{14}$$

Consider $\theta_{jk} = U_k^* Y_j / V_k^* X_j$ and solve the following model:

$$\begin{aligned}
&\min \quad r_{kk} \\
&\text{s.t.} \quad \theta_{kk}^* \cdot \sum_{h=1}^{m} v_{hk} x_{hj} - \sum_{r=1}^{s} u_{rk} y_{rj} + \delta_{kj} \beta \geq 0, \quad j \neq k, \\
&\qquad -\theta_{kk}^* \cdot \sum_{h=1}^{m} v_{hk} x_{hj} + \sum_{r=1}^{s} u_{rk} y_{rj} + \delta_{jk} \beta \geq 0, \quad j \neq k, \\
&\qquad \theta_{kk}^* \cdot \sum_{h=1}^{m} v_{hk} x_{hj} - \sum_{r=1}^{s} u_{rk} y_{rj} + \gamma_{kj} \beta \geq \varepsilon, \quad j \neq k, \\
&\qquad -\theta_{kk}^* \cdot \sum_{h=1}^{m} v_{hk} x_{hj} + \sum_{r=1}^{s} u_{rk} y_{rj} + \gamma_{jk} \beta \geq \varepsilon, \quad j \neq k, \\
&\qquad \delta_{lj}^k + \delta_{jl}^k \leq 1, \quad l \neq j, \\
&\qquad \delta_{lj}^k + \gamma_{jl}^k = 1, \quad l \neq j, \\
&\qquad r_{kk} = 1 + \frac{1}{2} \sum_{\substack{j=1 \\ j \neq k}}^{n} \delta_{kj} + \gamma_{kj}, \\
&\qquad \gamma_{kj}, \gamma_{jk}, \delta_{kj}, \delta_{jk} \in \{0, 1\}, \quad j \neq k.
\end{aligned} \tag{15}$$

As it is obvious, nonuniqueness of optimal weight may occur.

Wu et al. [19] proposed a weight balanced DEA model to reduce differences in weights data and zero weights

$$\begin{aligned}
&\min \quad \sum_{r=1}^{s} \left|\alpha_r^d\right| + \sum_{i=1}^{m} \left|\beta_i^d\right| \\
&\text{s.t.} \quad \sum_{i=1}^{m} w_{id} x_{ij} - \sum_{r=1}^{s} \mu_{rd} y_{rj} \geq 0, \quad j = 1, \ldots, n, \\
&\qquad \sum_{i=1}^{m} w_{id} x_{id} = 1, \\
&\qquad \sum_{i=1}^{m} \mu_{rd} y_{rd} = E_{dd}, \\
&\qquad \mu_{rd} y_{rd} + \alpha_r^d = \frac{E_{dd}}{s}, \quad r = 1, \ldots, s, \\
&\qquad w_{id} x_{id} + \beta_i^d = \frac{1}{m}, \quad i = 1, \ldots, m, \\
&\qquad w \geq 0, \quad \mu \geq 0, \ \beta^d, \alpha^d \text{ free.}
\end{aligned} \tag{16}$$

Therefore, the cross-efficiency score of DMU_j is the average of these cross-efficiencies:

$$E_j = \frac{1}{n} \sum_{d=1}^{n} E_{dj}, \quad j = 1, \ldots, n, \tag{17}$$

where

$$E_{dj} = \frac{\sum_{r=1}^{s} \mu_{rd}^* y_{rj}}{\sum_{i=1}^{m} w_{id}^* x_{ij}}. \tag{18}$$

As it is obvious, nonuniqueness of optimal weight may occur.

Zerafat Angiz et al. [20] introduced a cross-efficiency matrix based on this idea that ranking order is much more significant than individual efficiency score. Thus, they have provided the following procedure.

Step 1. Considering CCR model, calculate the efficiency score of all DMUs and consider Z_{pp}^* as the efficiency score of DMU_p.

Step 2. Now the cross-efficiency matrix Z can be constructed by $(z_{jp}^*)_{n \times n}$. Note that Z_{pp}^* is used as the diagonal elements of Z.

Step 3. Convert the cross-efficiency matrix into a cross-ranking matrix R as $(r_{jp})_{n \times n}$, in which r_{jp} is the ranking order of z_{jp}^* in column p of matrix Z.

Step 4. Construct the preference matrix W as $(w_{jk})_{n \times n}$ considering matrix R where w_{jk} is the number of time that DMU_j is placed in rank k.

Step 5. Construct matrix Ω as $(\widehat{\theta}_{jp})_{n \times n}$ in which $\widehat{\theta}_{jk}$ is calculated by summing the efficiency scores in matrix Z, corresponds to DMU_j, being placed in rank k.

Step 6. Obtain a common set of weight for final ranking of DMUs using the following modified method:

$$\begin{aligned} \max\ & \beta = \frac{\sum_{k=1}^{n} \mu_k \widehat{\theta}_{jk}}{\beta_j^*} \\ \text{s.t.}\ & \sum_{k=1}^{n} \mu_k \widehat{\theta}_{jk} \le 1, \quad j = 1, \ldots, n, \\ & \mu_k - \mu_{k+1} \ge d(k, \varepsilon), \quad k = 1, \ldots, n-1, \\ & \mu_n \ge d(n, \varepsilon), \end{aligned} \tag{19}$$

where $\widehat{\theta}_{jk}$ obtained from Step 5 and β_j^* is the optimal solution of the following model. Finally, the DMUs are ranked based on their $z_j^* = \sum_{k=1}^{n} \mu_k^* \widehat{\theta}_{jk}$ values.

Washio and Yamada [21] discussed that in real cases finding the best ranking is more significant than acquiring the most advantage weight and maximizing the efficiency. Thus they presented a model called rank-based measure (RBM) for evaluating units from different viewpoint. Thus, they suggested a method for acquiring those weight resulted from the best ranking as long as calculating those weight that maximizes the efficiency score. Finally they applied the presented model to the cross-efficiency assessment.

4. Ranking Techniques Based on Finding Optimal Weights in DEA Analysis

Jahanshahloo et al. [22] gave a note on some of the DEA models for complete ranking using common set of weights. They proved that by solving only one problem, it is possible to determine the common set of weights

$$\begin{aligned} \max\ & z \\ \text{s.t.}\ & \sum_{r=1}^{s} u_r y_{rj} + u_0 - z \sum_{i=1}^{m} v_i x_{ij} \ge 0, \quad j \in A, \\ & \sum_{r=1}^{s} u_r y_{rj} + u_0 - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1, \ldots, n,\ j \notin A, \\ & U, V, \eta \ge \varepsilon,\ u_0 \text{ free}, \end{aligned} \tag{20}$$

where A is the set of efficient units of the following model:

$$\begin{aligned} \max\ & \left\{ \frac{\sum_{r=1}^{s} u_r y_{r1} + u_0}{\sum_{i=1}^{m} v_i x_{i1}} \cdots \frac{\sum_{r=1}^{s} u_r y_{rn} + u_0}{\sum_{i=1}^{m} v_i x_{in}} \right\} \\ \text{s.t.}\ & \frac{\sum_{r=1}^{s} u_r y_{rj} + u_0}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \quad j = 1, \ldots, n, \\ & U, V, \eta \ge \varepsilon,\ u_0 \text{ free}. \end{aligned} \tag{21}$$

Note that DMUs can be ranked based on the evaluation of their efficiencies.

Wang et al. [23] provided an aggregating preference ranking. In this paper use of ordered weighted averaging (OWA) operator is proposed for aggregating preference rankings. Let w_j be the relative importance weight given to the jth ranking place and v_{ij} the vote candidate i receives in the jth ranking place. The total score of each candidate is defined as

$$z_i = \sum_{i=1}^{m} v_{ij} W_j, \quad i = 1, \ldots, m. \tag{22}$$

Alirezaee and Afsharian [24] discussed multiplier model, in which the variables are considered as shadow prices; note that $\sum_{r=1}^{s} u_r y_{rj}$ and $\sum_{i=1}^{m} v_i x_{ij}$ are total revenue and cost of DMU_j which are considered in optimization problem. The authors claimed that $\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0$, $j = 1, \ldots, n$ is the profit restriction for DMU_j. If $F(x, y) = 0$ is considered to be the efficient production function, then

$$\sum_i \frac{\partial F}{\partial x_i} x_i + \sum_j \frac{\partial F}{\partial y_j} y_j = 0. \tag{23}$$

They mentioned that the connected profit of the DMU is zero when shadow prices are derived from the technology and called this situation as balance situation. As the authors mentioned therefore in the case that DMU_1 is efficient but DMU_2 is inefficient or the efficiency score of both DMUs is the same, and it obtains more negative quantity in balance index, it can be concluded that DMU_1 has a better rank than DMU_2.

Liu and Hsuan Peng [25] in their paper proposed a method for determining the common set of weights for ranking units. In common weights analysis methodology they provided the following model:

$$\begin{aligned} \Delta^* = \min\ & \sum_{j \in E} \Delta_j^o + \Delta_j^i \\ \text{s.t.}\ & \frac{\sum_{r=1}^{s} y_{rj} U_r + \Delta_j^o}{\sum_{i=1}^{m} x_{ij} V_i - \Delta_j^i} = 1, \quad j \in E, \\ & \Delta_j^o, \Delta_j^i \ge 0, \quad \forall j \in E, \\ & U_r \ge \varepsilon, \quad r = 1, \ldots, s, \\ & V_i \ge \varepsilon, \quad i = 1, \ldots, m. \end{aligned} \tag{24}$$

The mentioned ratio form the linear equations.

Wang et al. [26] proposed a paper for ranking decision-making units by imposing a minimum weight restriction in DEA. The authors noted that using data envelopment analysis, it is not possible to distinguish between DEA efficient units. Thus, they presented a method for ranking units using imposing minimum weight restriction for the input-output data. As they mentioned, these weights restrictions are decided by a decision maker (DM) or an assessor as regards the solutions to a series of LP models considered for determining a maximin weight for each efficient DMU.

Hatefi and Torabi [27] proposed a common weight multi criteria decision analysis (MCDA)-data envelopment analysis (DEA) for constructing composite indicators (CIs). As the authors proved the presented model can discriminate between efficient units. The obtained common weights have discriminating power more than those obtained from

previous models. Finally they studied the robustness and discriminating power of the proposed method by Spearman's rank correlation coefficient.

Hosseinzadeh Lotfi et al. [28] proposed one DEA ranking model based on applying aggregate units. In doing so artificial units called aggregate units are defined as follows. The aggregate unit is shown by $\mathrm{DMU}_{\overline{a}}$

$$x_{i\overline{a}}^{p} = \sum_{k\in R_p} x_{ik}, \qquad y_{r\overline{a}}^{p} = \sum_{k\in R_p} y_{rk}, \qquad i = 1,\ldots,m,\ r = 1,\ldots,s, \tag{25}$$

where $R_p = \{j \mid \mathrm{DMU}_j \in E_p\}, E_p = E\{\mathrm{DMU}_p\}$. Note that E is the set of efficient units.

First, it is tried to maximize the efficiency score of the $\mathrm{DMU}_{\overline{a}}$ and then to maximize the efficiency score of the $\mathrm{DMU}_{\overline{a}}^{p}$. For resolving the existence of alternative solutions the authors presented an approach comprising $(m+s)$ simple linear problems to achieve the most appropriate optimal solutions among all alternative optimal solutions. Finally, they proposed the RI index for ranking all efficient DMUs. Let $U_{\overline{a}}, V_P\overline{a}$ and $U_{\overline{a}}^{p}, V_{\overline{a}}^{p}$ be the optimal solutions of the multiplier model for assent $\mathrm{DMU}_{\overline{a}}$ and $\mathrm{DMU}_{\overline{a}}^{p}$. Consider $\eta_{\overline{a}} = \sum_{r=1}^{s} u_{r\overline{a}} - \sum_{i=1}^{m} v_{i\overline{a}}$, $\eta_{\overline{a}}^{p} = \sum_{r=1}^{s} u_{r\overline{a}}^{p} - \sum_{i=1}^{m} v_{i\overline{a}}^{p}$, thus $\mathrm{RI}_p = \eta_{\overline{a}}^{o} - \eta_{\overline{a}}$.

Wang et al. [29] presented two nonlinear regression models for deriving common set of weights for fully rank units

$$\begin{aligned} &\min\ z = \sum_{j=1}^{n}\left(\theta_j^* - \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}}\right)^2 \\ &\text{s.t.}\quad U, V \geq 0, \end{aligned}$$

$$\begin{aligned} &\min\ \sum_{j=1}^{n}\left(\sum_{r=1}^{s} u_r y_{rj} - \theta_j^* \sum_{i=1}^{m} v_i x_{ij}\right)^2 \\ &\text{s.t.}\quad \sum_{r=1}^{s} u_r\left(\sum_{j=1}^{n} y_{rj}\right) + \sum_{i=1}^{m} v_i\left(\sum_{j=1}^{n} x_{ij}\right) = n, \\ &\qquad U, V \geq 0. \end{aligned} \tag{26}$$

Ramón et al. [30] aimed at deriving a common set of weights for ranking units. As the authors mentioned the idea is based upon minimization of the deviations of the common weights from the nonzero weights obtained from DEA. Furthermore, several norms are used for measuring such differences.

5. Super-Efficiency Ranking Techniques

Super efficiency models introduced in DEA technique are based upon the idea of leave one out and assessing this unit trough the remanding units.

Andersen and Petersen [31] introduced a model for ranking efficient units. The proposed model is as follows:

$$\begin{aligned} &\min\ \theta \\ &\text{s.t.}\quad \sum_{\substack{j=1\\ j\neq o}}^{n} \lambda_j x_{ij} \leq \theta x_{io}, \quad i = 1,\ldots,m, \\ &\qquad \sum_{\substack{j=1\\ j\neq o}}^{n} \lambda_j y_{rj} \geq y_{ro}, \quad r = 1,\ldots,s, \\ &\qquad \lambda_j \geq 0, \quad j = 1,\ldots,n,\ j\neq o. \end{aligned} \tag{27}$$

Although this idea is useful for further discriminating efficient units, it has been shown in the literature that it may be infeasible and nonstable. Thrall [7] mentioned the infeasibility of super-efficiency CCR model. Also a condition under which infeasibility occurred in super-efficiency DEA models is mentioned by Zhu [94], Seiford and Zhu [95], and Dulá and Hickman [96].

Hashimoto [97] provided a model based on the idea of one leave out and assurance region for ranking units.

Mehrabian et al. [32] presented a complete ranking for efficiency units in DEA context. As the authors mentioned this model does not have difficulties of A.P. model

$$\begin{aligned} &\min\ w_p + 1 \\ &\text{s.t.}\quad \sum_{j=1\ j\neq p}^{n} \lambda_j x_{ij} \leq x_{ip} + w_p 1, \quad i = 1,\ldots,m, \\ &\qquad \sum_{j=1\ j\neq p}^{n} \lambda_j y_{rj} \geq y_{rp}, \quad r = 1,\ldots,s, \\ &\qquad \lambda_j \geq 0, \quad j\neq p. \end{aligned} \tag{28}$$

With this method it is not possible to rank nonextreme efficient units.

Tone [33] presented super efficiency of SBM model. This model has the advantages of nonradial models, and it is always feasible and stable

$$\begin{aligned} &\min\ \frac{\sum_{i=1}^{m} \overline{x}_i / x_{io}}{\sum_{r=1}^{s} \overline{y}_r / y_{ro}} \\ &\text{s.t.}\quad \sum_{j=1\,j\neq o}^{n} \lambda_j x_{ij} \leq \overline{x}_i, \quad i = 1,\ldots,m, \\ &\qquad \sum_{j=1\,j\neq o}^{n} \lambda_j y_{rj} \geq \overline{y}_r, \quad r = 1,\ldots,s, \\ &\qquad \overline{x}_i \geq x_{io}, \quad 0 \leq \overline{y}_r \leq y_{ro}, \quad i = 1,\ldots,m, \quad r = 1,\ldots,s, \\ &\qquad \lambda_j \geq 0, \quad j\neq o. \end{aligned} \tag{29}$$

Jahanshahloo et al. [34] added some ratio constraints to the multiplier form of A.P. model and introduced a new method for ranking DMUs

$$\begin{aligned}
\min\ & \sum_{r=1}^{s} u_r y_{ro} \\
\text{s.t.}\ & \sum_{i=1}^{m} v_r x_{io} = 1, \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_r x_{ij} \le 0, \quad j = 1,\dots,n,\ j \ne o, \\
& \tilde{t}^{pq} \le \frac{v_p}{v_q} \le \bar{t}^{pq}, \quad p,q = 1,\dots,m,\ p < q, \\
& \tilde{t}^{kw} \le \frac{v_k}{v_w} \le \bar{t}^{kw}, \quad k,w = 1,\dots,s,\ k < w, \\
& U, V \ge \varepsilon.
\end{aligned} \tag{30}$$

DMU_o is efficient if the optimal objective function of the previous model is greater than or equal one.

Jahanshahloo et al. [35] presented a method for ranking efficient units on basis of the idea of one leave out and L_1 norm. As the authors proved this model is always feasible and stable

$$\begin{aligned}
\min\ & \sum_{i=1}^{m} x_i - \sum_{r=1}^{s} y_r + \alpha \\
\text{s.t.}\ & \sum_{j=1, j\ne o}^{s} \lambda_j x_{ij} \le x_i, \quad i = 1,\dots,m, \\
& \sum_{j=1, j\ne o}^{s} \lambda_j y_{rj} \ge y_r, \quad r = 1,\dots,s, \\
& x_i \ge x_{io}, \quad i = 1,\dots,m, \\
& 0 \le y_r \le y_{ro}, \quad r = 1,\dots,s, \\
& \lambda_j \ge 0, \quad j = 1,\dots n,\ j \ne o,
\end{aligned} \tag{31}$$

where $\alpha = \sum_{r=1}^{s} y_{ro} - \sum_{i=1}^{m} x_{io}$.

In their paper Chen and Sherman [36] presented a non-radial super-efficiency method and discussed the advantage of it. They verified that this model is invariant to units of input/output measurement. Let $J^o = J/\{\mathrm{DMU}_o\}$.

Step 1. Solve the following model to find the extreme efficient units in J^o:

$$\begin{aligned}
\min\ & \theta_k^{\text{super}} \\
\text{s.t.}\ & \sum_{j\ne o,k} \lambda_j x_{ij} \le \theta_k^{\text{super}} x_{io}, \quad i = 1,\dots,m, \\
& \sum_{j\ne o,k} \lambda_j y_{rj} \ge y_r, \quad r = 1,\dots,s, \\
& \lambda_j \ge 0, \quad j = 1,\dots n,\ j \ne o, k.
\end{aligned} \tag{32}$$

Consider E^o as the set of efficient units of J^o.

Step 2. Solve the following model:

$$\begin{aligned}
\min\ & \theta_k^{\text{super}} \\
\text{s.t.}\ & \sum_{j\in E^o} \lambda_j x_{ij} + s_i^{o-} = \theta x_{io}, \quad i = 1,\dots,m, \\
& \sum_{j\in E^o} \lambda_j y_{rj} - s_r^{o+} = y_{ro}, \quad r = 1,\dots,s, \\
& \lambda_j \ge 0, \quad j \in E^o.
\end{aligned} \tag{33}$$

Consider θ^* as the optimal solution of the previous model and solve the following model for $p \in \{1,\dots,m\}$:

$$\begin{aligned}
\max\ & s_p^{o-} \\
\text{s.t.}\ & \sum_{j\in E^o} \lambda_j x_{pj} + s_p^{o-} = \theta^* x_{po}, \\
& \sum_{j\in E^o} \lambda_j x_{ij} + s_i^{o-} = \theta^* x_{io}, \quad i \ne p, \\
& \sum_{j\in E^o} \lambda_j y_{rj} - s_r^{o+} = y_{ro}, \quad r = 1,\dots,s, \\
& \lambda_j \ge 0, \quad j \in E^o.
\end{aligned} \tag{34}$$

According to the obtained optimal solution of the previous model for $i = 1,\dots,m$ let $x_{io}^{(1)}, s_i^{o-*}(0) = s_i^{o-*}$, and $\theta^*(0) = \theta^*$, $I(t) = \{i : s_i^{o-*}(t-1) \ne 0\}$.

Step 3. Solve the following model:

$$\begin{aligned}
\min\ & \theta(t) \\
\text{s.t.}\ & \sum_{j\in E^o} \lambda_j x_{ij} \le \theta(t) x_{io}^{(t)}, \quad i \in I(t) \\
& \sum_{j\in E^o} \lambda_j x_{ij} = x_{io}^{(t)}, \quad i \notin I(t), \\
& \sum_{j\in E^o} \lambda_j y_{rj} - s_r^{o+} = y_{ro}, \quad r = 1,\dots,s, \\
& \lambda_j \ge 0, \quad j \in E^o.
\end{aligned} \tag{35}$$

Now according to the optimal solution of the previous model, solve the following model for each $p \in I(t)$:

$$\begin{aligned}
\min\ & s_p^{o-}(t) \\
\text{s.t.}\ & \sum_{j\in E^o} \lambda_j x_{ij} + s_p^{o-}(t) = \theta^*(t) x_{po}^{(t)}, \quad p \in I(t), \\
& \sum_{j\in E^o} \lambda_j x_{ij} + s_p^{o-}(t) = \theta^*(t) x_{io}^{(t)}, \quad i \ne p \in I(t), \\
& \sum_{j\in E^o} \lambda_j x_{ij} = x_{io}^{(t)}, \quad i \notin I(t), \\
& \sum_{j\in E^o} \lambda_j y_{rj} - s_r^{o+}(t) = y_{ro}, \quad r = 1,\dots,s, \\
& \lambda_j \ge 0, \quad j \in E^o.
\end{aligned} \tag{36}$$

Step 4. Let $x_{io}^{(t+1)} = \theta^*(t) x^{(t)}$ for $i \in I(t)$ and $x_{io}^{(t+1)} = x^{(t)}$ for $i \in I(t)$. If $I(t+1) = \emptyset$, then stop otherwise; if $I(t+1) \ne \emptyset$,

let $t = t + 1$ and go to Step 3. Now define θ^o as the average RNSE-DEA index

$$\theta^o = \frac{\sum_{t=1}^{T} n_I(T)\,\theta_t^o + \tilde{n}_I(T)\,\theta_{t+1}^o}{\sum_{t=1}^{T} n_I(T) + \tilde{n}_I(T)}. \tag{37}$$

Amirteimoori et al. [37] provided a distance-based approach for ranking efficient units. The presented method is a new method utilized L_2 norm. As noted in their paper, this new approach does not have difficulties of other methods

$$\begin{aligned}
\max\ & \beta^T Y_p - \alpha^T X_p \\
\text{s.t.}\ & \beta^T Y_j - \alpha^T X_j + s_j = 0, \quad j \in E,\ j \neq p, \\
& \alpha^T 1_m + \beta^T 1_s = 1, \\
& s_j \le \left(1 - \gamma_j\right) M, \quad j \in E,\ j \neq p, \\
& \sum_{j \in E, j \neq p} \gamma_j \ge m + s + 1, \\
& \alpha \ge \varepsilon \cdot 1_m, \qquad \beta \ge \varepsilon \cdot 1_s, \\
& \alpha_j \in \{0, 1\}, \quad j \in E,\ j \neq p.
\end{aligned} \tag{38}$$

This method cannot rank nonextreme efficient units.

Jahanshahloo et al. [38] presented modified MAJ model for ranking efficiency units in DEA technique

$$\begin{aligned}
\min\ & w_p + 1 \\
\text{s.t.}\ & \sum_{j=1 j \neq p}^{n} \lambda_j \frac{x_{ij}}{M_i} \le \frac{x_{ip}}{M_i} + w_p 1, \quad i = 1, \ldots, m, \\
& \sum_{j=1 j \neq p}^{n} \lambda_j y_{rj} \ge y_{rp}, \quad r = 1, \ldots, s, \\
& \lambda_j \ge 0, \quad j \neq p,
\end{aligned} \tag{39}$$

where $M_i = \text{Max}\{x_{ij} \mid \text{DMU}_j \text{ is efficient}\}$. It cannot rank nonextreme efficient units.

Li et al. [39] presented a new method for ranking which does not have difficulties of earlier methods. The presented model is always feasible and stable

$$\begin{aligned}
\min\ & 1 + \frac{1}{m} \sum_{i=1}^{m} \frac{s_{i2}^+}{R_i^-} \\
\text{s.t.}\ & \sum_{j=1 j \neq p}^{n} \lambda_j x_{ij} + s_{i1}^- - s_{i2}^+ = x_{ip}, \quad i = 1, \ldots, m, \\
& \sum_{j=1 j \neq p}^{n} \lambda_j y_{rj} - s_r^+ = y_{rp}, \quad r = 1, \ldots, s, \\
& s_{i2}^- \ge 0,\ s_{i1}^- \ge 0,\ s_r^+ \ge 0, \quad i = 1, \ldots, m, \\
& \qquad r = 1, \ldots, s, \\
& \lambda_j \ge 0, \quad j \neq p.
\end{aligned} \tag{40}$$

In this case extreme efficient units cannot be ranked.

In a paper Khodabakhshi [98] addresses super efficiency on improved outputs. He mentioned that as A.P. model may be infeasible under variable returns to scale technology, using the presented model gives a complete ranking when getting an input combination for improving outputs is suitable.

Sadjadi et al. [40] presented a robust super-efficiency DEA for ranking efficient units. They noted that as in most of the times exact data do not exist and the stochastic super-efficiency model presented in their paper incorporates the robust counterpart of super-efficiency DEA

$$\begin{aligned}
\min\ & \theta_o^{\text{RS}} \\
\text{s.t.}\ & \sum_{j=1 j \neq o}^{n} \lambda_j \overline{x}_{ij} - \theta_o^{\text{RS}} \overline{x}_{io} \\
& + \varepsilon \Omega \left(\sum_{\substack{j=1 \\ j \neq o, j \in J_i}}^{n} \lambda_j^2 \overline{x}_{ij}^2 + \left(\theta_o^{\text{RS}} \overline{x}_{io}\right)^2 \right)^{(1/2)} \le 0, \\
& \qquad i = 1, \ldots, m \\
& \sum_{\substack{j=1 \\ j \neq o}}^{n} \lambda_j \overline{y}_{rj} - \varepsilon \Omega \left(\overline{y}_{ro}^2 + \sum_{\substack{j=1 \\ j \neq o, j \in J_i}}^{n} \lambda_j^2 \overline{y}_{rj}^2 \right)^{(1/2)} \ge \overline{y}_{ro}, \\
& \qquad r = 1, \ldots, s, \\
& \lambda_j \ge 0, \quad j = 1, \ldots, n,
\end{aligned} \tag{41}$$

where $\overline{X}, \overline{Y}$ are input-output data. This model does not rank nonextreme efficient units. It may be unstable and infeasible.

Gholam Abri et al. [41] proposed a model for ranking efficient units. They used representation theory and represented the DMU under assessment as a convex combination of extreme efficient units. As the authors noted it is expected that the performance of DMU_o is the same as the performance of convex combination of extreme efficient units. Thus, it is possible to represent DMU_o as follows:

$$\left(X_o, Y_o\right) = \sum_{j=1}^{s} \lambda_j \left(X_j, Y_j\right), \quad \sum_{j=1}^{n} \lambda_j = 1,\ j = 1, \ldots, s. \tag{42}$$

As regards representation theorem, this system has $m + s - 1$ constraints and s variables, $(\lambda_1, \ldots, \lambda_s)$. If this system has a

unique solution we will have $\theta_o^* = \sum_{j=1}^{n} = 1, \lambda_j^* \theta_j$, otherwise two models should be considered

$$\begin{aligned}
&\min \sum_{j=1}^{s} = 1,\ \lambda_j \theta_j \\
&\text{s.t.} \quad (X_o, Y_o) = \sum_{j=1}^{s} \lambda_j \left(X_j, Y_j\right), \\
&\qquad \sum_{j=1}^{s} \lambda_j = 1, \\
&\qquad \lambda_j \geq 0, \quad j = 1, \ldots, s, \\
&\max \sum_{j=1}^{s} = 1,\ \lambda_j \theta_j \\
&\text{s.t.} \quad (X_o, Y_o) = \sum_{j=1}^{s} \lambda_j \left(X_j, Y_j\right), \\
&\qquad \sum_{j=1}^{s} \lambda_j = 1, \\
&\qquad \lambda_j \geq 0, \quad j = 1, \ldots, s.
\end{aligned} \tag{43}$$

Consider θ_1 and θ_2 as the optimal solution of the above-mentioned models, respectively. If $\theta_1 = \theta_2$, this is the same as what has been mentioned previously. If $\theta_1 < \theta_2$, then mentioned models provide an interval which helps rank units from the worst to the best. Sometimes it will obtain the ranking score with a bounded interval $[\theta_1, \theta_2]$.

Jahanshahloo et al. [42] presented models for ranking efficient units. The presented models are somehow a modification of cross-efficiency model that overcomes the difficulty of alternative optimal weights. The authors in their paper with regard to the changes and also utilizing TOPSIS technique presented a new super-efficient method for ranking units. The presented model is as follows:

$$\theta_{ij} = \left\{ \frac{U^t Y_i}{V^t X_i} \mid \frac{U^t Y_l}{V^t X_l} \leq 1, \frac{U^t Y_j}{V^t X_j} = \theta_{jj}, \right. \\ \left. l = 1, \ldots, n,\ l \neq i,\ U \geq 0,\ V \geq 0 \right\}, \tag{44}$$

where θ_{jj} is the efficiency score of DMU_J using corresponding weights. Also θ_{ij} is the efficiency of DMU_i using optimal weights of DMU_j. Noura et al. [43] provided a method for ranking efficient units based on this idea that more effective and useful units in society should have better rank.

Step 1. For each efficient unit choose the lower and upper limit for each inputs and outputs. Let E be the set of efficient units

$$\begin{aligned}
&x_i^{*u} = \mathrm{Max}_{j \in E} \left|x_{ij}\right|, \quad x_i^{*l} = \mathrm{Min}_{j \in E} \left|x_{ij}\right|, \quad i = 1, \ldots, m, \\
&y_r^{*u} = \mathrm{Max}_{j \in E} \left|y_{rj}\right|, \quad y_r^{*l} = \mathrm{Min}_{j \in E} \left|y_{rj}\right|, \quad r = 1, \ldots, s.
\end{aligned} \tag{45}$$

Step 2. In accordance with the previous step, here, the utility inputs and outputs are as follows:

$$\begin{aligned}
&\underline{x} = x_i^{*l}, \quad \forall i \left(i \in D_i^-\right), \qquad \overline{x} = x_i^{*u}, \quad \forall i \left(i \in D_i^+\right), \\
&\underline{y} = y_r^{*l}, \quad \forall r \left(r \in D_o^-\right), \qquad \overline{y} = y_r^{*u}, \quad \forall r \left(r \in D_o^+\right).
\end{aligned} \tag{46}$$

Step 3. Consider dimensionless (d_i, d_r) introduced as follows for each efficient unit belonging to E:

$$\begin{aligned}
&\forall i \left(i \in D_i^+\right) \quad d_{ij} = \frac{x_{ij}}{x_{ij} + \xi}, \\
&\forall i \left(i \in D_i^-\right) \quad d_{ij} = \frac{\overline{x}_i}{x_i + \xi}, \\
&\forall r \left(r \in D_r^+\right) \quad d_{rj} = \frac{y_{rj}}{y_r + \xi}, \\
&\forall r \left(r \in D_r^-\right) \quad d_{rj} = \frac{\overline{y}_r}{y_{rj} + \xi},
\end{aligned} \tag{47}$$

where ξ is representative of a small and nonzero number used for not dividing by zero. Now consider $D - j$ as follows which shows that more successful DMU_j will be if the larger value of the D_j is

$$D_j = \sum_{i \in I} d_{ij} + \sum_{r \in R} d_{rj}, \tag{48}$$

where $I = D_i^+ \cup D_i^-,\ R = D_r^+ \cup D_r^-$.

Ashrafi et al. [44] introduced an enhanced Russell measure of super efficiency for ranking efficient units in DEA. The linear counterpart of the proposed model is as follows:

$$\begin{aligned}
&\max \ \frac{1}{m} \sum_{i=1}^{m} u_i \\
&\text{s.t.} \quad \sum_{r=1}^{s} v_r = s, \\
&\qquad \sum_{\substack{j=1 \\ j \neq o}}^{n} \alpha_j x_{ij} \leq u_i x_{io}, \quad i = 1, \ldots, m, \\
&\qquad \sum_{\substack{j=1 \\ j \neq o}}^{n} \alpha_j y_{rj} \leq v_i y_{ro}, \quad r = 1, \ldots, s, \\
&\qquad \alpha_j \geq 0,\ u_i \geq \beta,\ j = 1, \ldots, n,\ j \neq k,\ i = 1, \ldots, m, \\
&\qquad o \leq \beta \leq,\ v_r \geq \beta,\ r = 1, \ldots s.
\end{aligned} \tag{49}$$

It cannot rank nonextreme efficient units.

Chen et al. [45] proposed a modified super-efficiency method for ranking units based on simultaneous input-output projection. The presented model overcomes the infeasibility problem

$$\begin{aligned} p_1 = \min\ & \frac{\theta_o^{sr}}{\phi_o^{sr}} \\ \text{s.t.}\ & \sum_{\substack{j=1\\ j\neq o}} \lambda_j x_{ij} \le \theta_o^{sr} x_{io}, \quad i=1,\dots,m, \\ & \sum_{\substack{j=1\\ j\neq o}} \lambda_j y_{rj} \ge \phi_o^{sr} y_{ro}, \quad r=1,\dots,s, \\ & \sum_{\substack{j=1\\ j\neq o}} \lambda_j = 1, \\ & 0 < \theta_o^{sr} \le 1, \quad \phi_o^{sr} \ge 1, \quad i=1,\dots,m, \\ & \qquad r=1,\dots,s, \\ & \lambda_j \ge 0, \quad j=1,\dots,n. \end{aligned} \tag{50}$$

Rezai Balf et al. [46] provide a model for ranking units based on Tchebycheff norm. As proved that this model is always feasible and stable, it seems to have superiority over other models

$$\begin{aligned} \max\ & V_p \\ \text{s.t.}\ & V_p \ge \sum_{\substack{j=1\\ j\neq o}}^{n} \lambda_j x_{ij} - x_{ip}, \quad i=1,\dots,m, \\ & V_p \ge y_{rp} - \sum_{\substack{j=1\\ j\neq o}}^{n} \lambda_j y_{rj}, \quad r=1,\dots,s, \\ & \lambda_j \ge 0, \quad j=1,\dots,n, \end{aligned} \tag{51}$$

where

$$\begin{aligned} V_p = \ \mathrm{Max} \Bigg\{ & \left(\sum_{\substack{j=1\\ j\neq o}}^{n} \lambda_j x_{ij} - x_{ip} \right) i=1,\dots,m, \\ & \left(y_{rp} - \sum_{\substack{j=1\\ j\neq p}}^{n} \lambda_j y_{rj} \right) r=1,\dots,s \Bigg\}. \end{aligned} \tag{52}$$

Chen et al. [47] for overcoming the infeasibility problem that occurred in variable returns to scale super-efficiency DEA model according to a directional distance function developed Nerlove-Luenberger (N-L) measure of super-efficiency.

6. Benchmarking Ranking Techniques

Sueyoshi et al. [49] proposes a "benchmark approach" for baseball evaluation. This method is the combination of DEA and (Offensive earned-run average) OERA. As the authors noted, using this method it is possible to select best units and also their ranking orders. They mentioned that using only DEA may result in a shortcoming in assessment as many efficient units can be identified. Thus, the authors used slack-adjusted DEA model and OERA to overcome this difficulty.

Jahanshahloo et al. [50] presented a new model for ranking DMUs based on alteration in reference set. The idea is based on this fact that efficient units can be the target unit for inefficient units

$$\begin{aligned} \min\ & \partial_{a,b} = \theta - \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right) \\ \text{s.t.}\ & \sum_{j\in J-\{b\}} \lambda_j x_{ij} + s_i^- - \theta x_{ia}, \quad i=1,\dots,m, \\ & \sum_{j\in J-\{b\}} \lambda_j y_{rj} - s_r^+ = y_{ra}, \quad r=1,\dots,s, \\ & \theta \text{ free}, \ s_i^- \ge 0, \ s_r^+ \ge 0, \ i=1,\dots,m, \ r=1,\dots,s, \\ & \lambda_j \ge 0, \quad j \in J-\{b\}. \end{aligned} \tag{53}$$

Finally the ranking order for each efficient unit b can be computing by

$$\mho_b = \sum_{a\in J_n} \frac{\partial_{a,b}}{\tilde{n}}. \tag{54}$$

Note that this method cannot rank nonextreme efficient units.

Lu and Lo [51] provided an interactive benchmark model for ranking units. The idea is based upon considering a fixed unit as a benchmark and calculating the efficiency of other units, pair by pair, to this unit. This procedure continues when all units are accounted for as a benchmark unit. Consider DMU_o as a unit under evaluation and DMU_b as a benchmark. Assume

$$\overline{x}_i = x_{io}(1+\phi_i), \qquad \overline{y}_r = y_{ro}(1-\varphi_r).$$

$$\begin{aligned} \min\ & \theta_o^{*b} = \frac{1+(1/m)\sum_{i=1}^{m}\phi_i}{1-(1/s)\sum_{r=1}^{s}\varphi_r} \\ \text{s.t.}\ & \lambda_b x_{ib} - x_{io}\phi_i \le x_{io}, \quad i=1,\dots,m, \\ & \lambda_b y_{rb} - y_{ro}\varphi_r \ge y_{ro}, \quad r=1,\dots,s, \\ & \phi_i \ge o, \quad \varphi_r \le y_{ro}, \quad i=1,\dots,m, \quad r=1,\dots,s, \\ & \lambda_j \ge 0, \quad j=1,\dots,n. \end{aligned} \tag{55}$$

Then using $(\sum_{b=1}^{n}\theta_o^{b*})/n, o = 1,\dots,n$, the efficiency of benchmark DMU_o can be obtained. Using the following index which indicates the increment in efficiency of a unit by moving from peer appraisal to self-appraisal, it is now possible to rank units. Note that the less the magnitude of

this index is the better rank for corresponding unit will be obtained

$$\mathrm{FPI}_K^{\mathrm{IBM}} = \frac{\left(\mathrm{TE}_k^{\mathrm{BCC}} - \mathrm{STD}_k^{\mathrm{IBM}}\right)}{\left(\mathrm{STD}_k^{\mathrm{IBM}}\right)}, \tag{56}$$

where $\mathrm{TE}_k^{\mathrm{BCC}}$ and $\mathrm{STD}_k^{\mathrm{IBM}}$ are, respectively, efficiency in BCC model and normalization of $\mathrm{TE}^{\mathrm{IBM}}$ of DMU_k. Chen [52] provided a paper for ranking efficient and inefficient units in DEA. They noted that the evaluation of efficient units is based upon the alterations in efficiency of all inefficient units by omitting in reference set. At first, solve the following model which measures the efficiency of DMU_a when DMU_b is eliminated from the reference set

$$\begin{aligned}
\min\ & \rho_{a,b} = \theta_a^s - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \\
\text{s.t.}\ & \sum_{j=1\, j\neq b}^{n} \lambda_j x_{ij} + s_i^- = \theta_a^s x_{ia}, \quad i = 1,\ldots,m, \\
& \sum_{j=1\, j\neq b}^{n} \lambda_j y_{rj} - s_r^+ = y_{ra}, \quad r = 1,\ldots,s, \\
& \sum_{j=1\, j\neq b}^{n} = 1, \\
& \theta_a^s \geq 0,\ s_i^- \geq 0,\ s_r^+ \geq 0,\ i = 1,\ldots,m,\ r = 1,\ldots,s, \\
& \lambda_j \geq 0, \quad j \neq b.
\end{aligned} \tag{57}$$

Then, again, solve the previous model, this time for calculating the efficiency score of each inefficient unit when each of efficient units, in turn, is eliminated from the reference set η_a^* and calculate the efficiency change

$$\tau_{a,b} = \rho_{a,b}^* - \eta_a^*. \tag{58}$$

Then, calculate the following index called as "MCDE" index for each efficient and inefficient unit:

$$E_a = \sum_{b\in V_E} W_b^* \rho_{a,b}^*, \qquad E_b = \sum_{b\in V_I} W_b^* \rho_{a,b}^*. \tag{59}$$

Now, in accordance with the magnitude of the acquired index it is possible to rank units. Those units with higher score have better ranking order.

7. Ranking Techniques by Multivariate Statistics in the DEA

As described in DEA literature in DEA technique frontier is taken into consideration rather than central tendency considered in regression analysis. DEA technique considers that an envelope encompasses through all the observations as tight as possible and does not try to fit regression planes in center of data. In DEA methodology each unit is considered initially and compared to the efficient frontier, but in regression analysis a procedure is considered in which a single function fits to the data. DEA uses different weights for different units but does not let the units use weights of other units. Canonical correlation analysis, as Friedman and Sinuany-Stern [53] noted, can be used for ranking units. This method is somehow the extension of regression analysis, Adler et al. [88]. The aim in canonical correlation analysis is to find a single vector common weight for the inputs and outputs of all units. Consider Z_j, W_j as the composite input and output and $\acute{V}, \acute{U}$ as corresponding weights, respectively. The presented model by Tatsuoka and Lohnes [99] is as follows:

$$\begin{aligned}
\max\ & r_{z,w} = \frac{\acute{V} S_{xy} U}{\left(\acute{V} S_{xx} V\right)\left(\acute{U} S_{yy} U\right)^{(1/2)}} \\
\text{s.t.}\ & \acute{V}_{xx} V = 1, \\
& \acute{U}_{yy} U = 1,
\end{aligned} \tag{60}$$

where S_{xx}, S_{yy}, and S_{xy} are respectively defined as the matrices of the sums of squares and sums of products of the variables.

Friedman and Sinuany-Stern [53] while defining the ratio of linear combinations of the inputs and outputs, $T_j = W_j/Z_j$ used canonical correlation analysis. As they noted that scaling ratio T_j of the canonical correlation analysis/DEA is unbounded.

Sinuany-Stern et al. [100] used linear discriminant analysis for ranking units. They defined

$$D_j = \sum_{r=1}^{s} u_r y_{rj} + \sum_{i=1}^{m} v_r\left(-x_{ij}\right). \tag{61}$$

DMU_j is said to be efficient if $D_j > D_c$ where D_c is a critical value based on the midpoint of the means of the discriminant function value of the two groups, Morrison [101]. The larger the amount of D_j is the better rank DMU_j will have.

Friedman and Sinuany-Stern [53] noted that as cross-efficiency/DEA, canonical correlation analysis/DEA and discriminant analysis/DEA, ranking orders may vary from each other, thus it seems necessary to introduce the combined ranking (CO/DEA). Combined ranking, for each unit, considered all the ranks obtained from the above-mentioned rankings. Moreover, statistical tests are; one for goodness of fit between DEA and a specific ranking and the other for testing correlation between variety of ranking orders, Siegel and Castellan [102].

8. Ranking with Multicriteria Decision-Making (MCDM) Methodologies and DEA

Li and Reeves [56] for increasing discrimination power of DEA presented a multiple-objective linear program (MOLP). As the authors mentioned using minimax and minsum efficiency in addition to the standard DEA objective function help to increase discrimination power of DEA.

Strassert and Prato [59] presented the balancing and ranking method which uses a three-step procedure for deriving an overall complete or partial final order of options. In the first step, derive an outranking matrix for all options

from the criteria values. Considering this matrix it is possible to show the frequency with which one option is ranked higher than the other options. In the second step, by triangularizing the outranking matrix establish an implicit preordering or provisional ordering of options. The outranking matrix shows the degree to which there is a complete overall order of options. In the third step, based on information given in an advantages-disadvantages table, the provisional ordering is subjected to different screening and balancing operations.

Chen [60] utilized a nonparametric approach, DEA, to estimate and rank the efficiency of association rules with multiple criteria in following steps. Proposed postprocessing approach is as follows.

Step 1. Input data for association rule mining.

Step 2. Mine association rules by using the a priori algorithm with minimum support and minimum confidence.

Step 3. Determine subjective interestingness measures by further considering the domain related knowledge.

Step 4. Calculate the preference scores of association rules discovered in Step 2 by using Cook and Kress's DEA model.

Step 5. Discriminate the efficient association rules found in Step 3 by using Obata and Ishii's [103] discriminate model.

Step 6. Select rules for implementation by considering the reference scores generated in Step 5 and domain related knowledge.

Jablonsky [61] presented two original models, super SBM and AHP, for ranking of efficient units in DEA. As the author mentioned these models are based on multiple criteria decision-making techniques-goal programming and analytic hierarchy process. Super SBM model for ranking units is as follows:

$$\begin{aligned}
\min\ & \theta_q^G = 1 + t\gamma + (1-t)\left(\frac{\sum_{i=1}^m s_{1i}^+}{x_{iq}} + \frac{\sum_{i=1}^m s_{2k}^-}{y_{kq}}\right) \\
\text{s.t.}\ & \sum_{j=1\, j\neq q}^{n} \lambda_j x_{ij} + s_{1i}^- - s_{1i}^+ = x_{iq}, \quad \mathrm{i} = 1,\dots,m, \\
& \sum_{j=1\, j\neq q}^{n} \lambda_j y_{kj} + s_{2k}^- - s_{2k}^+ = y_{kq}, \quad k = 1,\dots,r, \\
& s_{1i}^+ \le \gamma, \quad i = 1,\dots,m, \\
& s_{2k}^- \le \gamma, \quad k = 1,\dots,r, \\
& t \in \{0,1\}, \quad \lambda_j \ge 0, \quad S_1^+, S_2^+ \ge 0, \quad S_1^-, S_2^- \ge 0, \\
& j = 1,\dots,n.
\end{aligned} \tag{62}$$

Wang and Jiang [62] presented an alternative mixed integer linear programming models in order to identify the most efficient units in DEA technique. As the authors mentioned presented models can make full use of input-output information with no need to specify any assurance regions for input and output weights to avoid zero weights

$$\begin{aligned}
\min\ & \sum_{i=1}^m v_i \left(\sum_{j=1}^n x_{ij}\right) - \sum_{r=1}^s u_r \left(\sum_{j=1}^n y_{rj}\right) \\
\text{s.t.}\ & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \le I_j, \quad j = 1,\dots,n, \\
& \sum_{j=1}^n I_j = 1, \\
& I_j \in \{0,1\}, \quad j = 1,\dots,n, \\
& u_r \ge \frac{1}{(m+s)\max_j\{y_{rj}\}}, \quad r = 1,\dots,s, \\
& v_i \ge \frac{1}{(m+s)\max_j\{x_{ij}\}}, \quad i = 1,\dots,m.
\end{aligned} \tag{63}$$

Hosseinzadeh Lotfi et al. [63] provided an improved three-stage method for ranking alternatives in multiple criteria decision analysis. In the first stage, based on the best and worst weights in the optimistic and pessimistic cases, obtain the rank position of each alternative, respectively. The obtained weight in the first stage is not unique; thus it seems necessary to introduce a secondary goal that is used in the second stage. Finally, in the third stage, the ranks of the alternatives compute in the optimistic or pessimistic case. It is mentionable that the model proposed in the third stage is a multicriteria decision-making (MCDM) model, and it is solved by mixed integer programming. As the authors mentioned and provided in their paper this model can be converted into an LP problem

$$\begin{aligned}
\min\ & 2nr_o^o + \sum_{i=1\, i\neq o}^{n} \left(n + 1 - r_i^{i*}\right) r_i^o \\
\text{s.t.}\ & \sum_{j=1}^k w_j^o v_{ij} - \sum_{j=1}^k w_j^o v_{hj} + \delta_{ih}^o M \ge 0, \quad i = 1,\dots,n,\ i \neq h, \\
& \delta_{ih}^o + \delta_{hi}^o = 1, \quad i = 1,\dots,n,\ i \neq h, \\
& \delta_{ih}^o + \delta_{hk}^o + \delta_{ki}^o \ge 1, \quad i = 1,\dots,n,\ i \neq h \neq k, \\
& r_i^o = 1 + \sum_{h\neq i}^n \delta_{ih}^o, \quad i = 1,\dots,n, \\
& w^o \in \phi, \\
& \delta_{ih}^o \in \{0,1\}, \quad i = 1,\dots,n,\ i \neq h.
\end{aligned} \tag{64}$$

In the previous model, the rank vector R^o for each alternative x_o is computed by the ideal rank.

9. Some Other Ranking Techniques

Seiford and Zhu [64] presented the context-dependent DEA method for ranking units. Let $J^1 = \{\mathrm{DMU}_j, j = 1,\dots,n\}$ be

the set of decision-making units. Consider $J^{l+1} = J^l - E^l$ in which $E^l = \{\text{DMU}_k \in J^l \mid \phi^*(l,k) = 1\}$ where $\phi^*(l,k) = 1$ is the optimal value of following model:

$$\begin{aligned} \max_{\lambda_j, \phi(l,k)} \quad & \phi^*(l,k) = \phi(l,k) \\ \text{s.t.} \quad & \sum_{j \in F(J^l)} \lambda_j y_j \geq \phi(l,k)\, y_k, \\ \text{s.t.} \quad & \sum_{j \in F(J^l)} \lambda_j x_j \leq \phi(l,k)\, x_k, \\ \text{s.t.} \quad & \lambda_j \geq 0, \quad j \in F\left(J^l\right). \end{aligned} \tag{65}$$

The previous model with $l = 1$ is the original CCR model. Note that DMUs in E^1 show the first level efficient frontier. $l = 2$ indicates the second level efficient frontier when the first level efficient frontier is omitted. The following algorithm finds these efficient frontiers.

Step 1. Set $l = 1$ and evaluate the entire set of DMUs, J^1. In this way the first level efficient DMU, E^1, is identified.

Step 2. Exclude the efficient DMUs from future DEA runs and set $J^{l+1} = J^l - E^l$. (If $J^{l+1} = \emptyset$ stop).

Step 3. Evaluate the new subset of "inefficient" DMUs, J^{l+1}, to obtain a new set of efficient DMUs, E^{l+1}.

Step 4. Let $l = l + 1$. Go to Step 2.

When $J^{l+1} = \emptyset$, the algorithm stops.

As the authors proved in this way it is possible to rank the DMUs in the first efficient frontier based upon their attractiveness scores and identify the best one.

Jahanshahloo et al. [65] provided a paper for ranking units using gradient line. As the authors mentioned the advantage of this model is stability and robustness

$$\begin{aligned} \max \quad & H_o = -V^T X_o + U^T Y_o \\ \text{s.t.} \quad & -V^T X_j + U^T Y_j \leq 0, \quad j = 1, \ldots, n, \ j \neq o, \\ & V^T e + U^T e = 1, \\ & V, U \geq \varepsilon 1. \end{aligned} \tag{66}$$

Note that $(U^*, -V^*)$ is the gradient of hyperplane which supports on $\acute{T}_c$, the obtained PPS by omitting the DMU under assessment in T_c. As the authors proved a unit is efficient iff the optimal objective function of the following model is greater than zero. Consider

$$\begin{aligned} P_0 &= \{(X,Y) : X = \alpha X_o,\ Y = \beta Y_o\} \\ S_0 &= \{(X,Y) \in P_0 : X = \alpha X_o,\ Y = \beta Y_o,\ \alpha \geq 0,\ \beta \geq 0\}. \end{aligned} \tag{67}$$

Intersection of S_0 and efficient surface of $\acute{T}_c c$ is a half line where its equation is $(-V^{*T} X_o)\alpha + (U^{*T} Y_o)\beta = 0$. To rank DMU_o the length of connecting arc DMU_o with intersection point of line and previous ellipse is calculated in (α, β) space. This intersection is as follows:

$$\begin{aligned} \alpha^* &= \left(\frac{K_\alpha^2 K_\beta^2}{K_\beta^2 + \left(V^{*T} X_o / U^{*T} Y_o\right)^2 K_\alpha^2} \right)^{1/2}, \\ \beta^* &= \left(\frac{V^{*T} X_o}{U^{*T} Y_o} \right) \alpha^*. \end{aligned} \tag{68}$$

Now, the length of connecting arc DMU_o to the point corresponding to α^*, β^* in (α, β) plan is calculated as follows: $I = \int_1^{\alpha^*} (1+\beta)^{1/2} d\alpha$, where

$$\begin{aligned} K_\alpha &= \left(\frac{\sum_{j=1}^n x_{io}^2 + \sum_{r=1}^s y_{io}^2}{\sum_{j=1}^n x_{io}^2} \right)^{1/2}, \\ K_\beta &= \left(\frac{\sum_{j=1}^n x_{io}^2 + \sum_{r=1}^s y_{io}^2}{\sum_{r=1}^s y_{ro}^2} \right)^{1/2}. \end{aligned} \tag{69}$$

Jahanshahloo et al. [34] also provided a paper with the concept of advantage in data envelopment analysis.

In their paper Jahanshahloo et al. [66] considering Monte Carlo method presented a new method of ranking.

Step 1. Generate a uniformly distributed sequence of $\{U_j\}_{j=1}^{2n}$ on$(0,1)$.

Step 2. Random numbers should be classified into N pairs like $(U_1, U_1'), \ldots, (U_N, U_N')$ in a way that each number is used just one time.

Step 3. Compute $X_i = a + U_i(b-a)$ and $f(X_i) > cU_i'$.

Step 4. Estimate the integral I by $\theta_I = c(b-a)(N_H/N)$.

Now consider DMU_o as an efficient unit measure those units that are dominated DMU_o. As for dome DMUs this would be unbounded, thus the authors, for each unit, bounded the region. Then, for DMU_o if $(-X_o, Y_o) \geq (-\overline{X}, \overline{Y})$, then $(-\overline{X}, \overline{Y})$ is in the RED of DMU_o. Now by using $V_p = V^*(N_H/N)$ all the hits (the above condition) can be counted, where V^* is the measure of the whole region. Jahanshahloo and Afzalinejad [67] presented a method based on distance of

the unit under evaluation to the full inefficient frontier. They presented two models, radian and nonradial

$$\begin{aligned}
&\min \ \phi \\
&\text{s.t.} \quad \sum_{j=1}^{n} \lambda_j x_{ij} = \phi x_{io}, \quad i = 1, \ldots, m, \\
&\qquad \sum_{j=1}^{n} \lambda_j y_{rj} = y_{ro}, \quad r = 1, \ldots, s, \\
&\qquad \sum_{j=1}^{n} = 1, \\
&\qquad \lambda_j \geq 0, \quad j = 1, \ldots, m,
\end{aligned}$$

$$\begin{aligned}
&\max \ \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \\
&\text{s.t.} \quad \sum_{j \in J-\{b\}} \lambda_j x_{ij} - s_i^- = x_{io}, \quad i = 1, \ldots, m, \\
&\qquad \sum_{j \in J-\{b\}} \lambda_j y_{rj} + s_r^+ = y_{ro}, \quad r = 1, \ldots, s, \\
&\qquad \sum_{j=1}^{n} = 1, \\
&\qquad s_i^- \geq 0, \quad s_r^+ \geq 0, \quad i = 1, \ldots, m, \quad r = 1, \ldots, s, \\
&\qquad \lambda_j \geq 0, \quad j = 1, \ldots, n.
\end{aligned} \tag{70}$$

Amirteimoori [68] based on the same idea considered both efficient and antiefficient frontiers for efficiency analysis and ranking units.

Kao [69] mentioned that determining the weights of individual criteria in multiple criteria decision analysis in a way that all alternatives can be compared according to the aggregate performance of all criteria is of great importance. As Kao noted this problem relates to search for alternatives with a shorter and longer distance, respectively, to the ideal and anti-ideal units. He proposed a measure that considered the calculation of the relative position of an alternative between the ideal and anti-ideal for finding an appropriate rankings.

Khodabakhshi and Aryavash [70] presented a method for ranking all units using DEA concept.

First, Compute the minimum and maximum efficiency values of each DMU in regard to this assumption that the sum of efficiency values of all DMUs equals to 1.

Second, determine the rank of each DMU in relation to a combination of its minimum and maximum efficiency values.

Zerafat Angiz et al. [71] proposed a technique in order to aggregate the opinions of experts in voting system. As the authors mentioned the presented method uses fuzzy concept, and it is computationally efficient and can fully rank alternatives. At first, number of votes given to a rank position was grouped to construct fuzzy numbers, and then the artificial ideal alternative introduced. Furthermore, by performing DEA the efficiency measure of alternatives was obtained considering artificial ideal alternative compared by each of the alternatives pair by pair. Thus alternatives are ranked in accordance with their efficiency scores. If this method cannot completely rank alternatives, weight restrictions based on fuzzy concept are imposed into the analysis.

10. Different Applications in Ranking Units

As discussed formerly there exist a variety of ranking methods and applications in theliterature. Nowadays DEA models are widely used in different areas for efficiency evaluation, benchmarking and target setting, ranking entities, and so forth. Ranking units, as one of the important issues in DEA, has been performed in different areas. Jahanshahloo et al. [42] ranked cities in Iran to find the best place for creating a data factory. Hosseinzadeh Lotfi et al. [28] used a new method for ranking in order to find the best place for power plant location. Ali and Nakosteen [79], Amirteimoori et al. [37] and Alirezaee and Afsharian [24], Soltanifar and Hosseinzadeh Lotfi [104], Zerafat Angiz et al. [71], Hosseinzadeh Lotfi et al. [28], Jahanshahloo [50], Chen and Deng [52], and Jablonsky [61] used different ranking methods in banking system. Sadjadi [40] ranked provincial gas companies in Iran. Mehrabian et al. [32], Li et al. [39], Örkcü and Bal [12], and Wu et al. [19] ranked different departments of universities. Jahanshahloo et al. [35], Jahanshahloo and Afzalinejad [67] utilized the presented models in ranking 28 Chinese cities. Jahanshahloo at al. [35] provided an application to burden sharing amongst NATO member nations. Jahanshahloo et al. [14] utilized the presented model for ranking nursery homes. Contreras [18] and Wang et al. [23] used the provided ranking techniques in ranking candidates. Lu and Lo [51] applied their method on an application to financial holding companies. Hosseinzadeh Lotfi et al. [63] consider an empirical example in which voters are asked to rank two out of seven alternatives. Wang and Jiang [62] utilized the provided method in facility layout design in manufacturing systems and performance evaluation of 30 OECD countries. Chen [60] used an example of market basket data in order to illustrate the provided approach.

11. Application

In this section, some of the reviewed models are applied on the example used in Soltanifar and Hosseinzadeh Lotfi [104]. Consider twenty commercial banks of Iran with input-output data tabulated in Table 1 and summarized as follows. Also the results of CCR model are listed in this table. As it can be seen seven units are efficient, DMUS 1,4, 7, 12, 15, 17, and 20. *Inputs* are staff, computer terminal, and space. *Outputs* are deposits, loans granted, and charge.

Consider some of the important ranking methods in literature as follows:

R.M1: A.P. model [31] is based upon the idea of leave unit evaluation out and measuring the distance of the unit under evaluation from the production possibility set constructed by the remaining DMUs.

Table 1: Inputs, outputs, and efficiency scores.

DMU_p	I_1	I_2	I_3	O_1	O_2	O_3	CCR efficiency
DMU_1	0.950	0.700	0.155	0.190	0.521	0.293	1.0000
DMU_2	0.796	0.600	1.000	0.227	0.627	0.462	0.8333
DMU_3	0.798	0.750	0.513	0.228	0.970	0.261	0.9911
DMU_4	0.865	0.550	0.210	0.193	0.632	1.000	1.0000
DMU_5	0.815	0.850	0.268	0.233	0.722	0.246	0.8974
DMU_6	0.842	0.650	0.500	0.207	0.603	0.569	0.7483
DMU_7	0.719	0.600	0.350	0.182	0.900	0.716	1.0000
DMU_8	0.785	0.750	0.120	0.125	0.234	0.298	0.7978
DMU_9	0.476	0.600	0.135	0.080	0.364	0.244	0.7877
DMU_{10}	0.678	0.550	0.510	0.082	0.184	0.049	0.290
DMU_{11}	0.711	1.000	0.305	0.212	0.318	0.403	0.6045
DMU_{12}	0.811	0.650	0.255	0.123	0.923	0.628	1.0000
DMU_{13}	0.659	0.850	0.340	0.176	0.645	0.261	0.8166
DMU_{14}	0.976	0.800	0.540	0.144	0.514	0.243	0.4693
DMU_{15}	0.685	0.950	0.450	1.000	0.262	0.098	1.0000
DMU_{16}	0.613	0.900	0.525	0.115	0.402	0.464	0.6390
DMU_{17}	1.000	0.600	0.205	0.090	1.000	0.161	1.0000
DMU_{18}	0.634	0.650	0.235	0.059	0.349	0.068	0.4727
DMU_{19}	0.372	0.700	0.238	0.039	0.190	0.111	0.4088
DMU_{20}	0.583	0.550	0.500	0.110	0.615	0.764	1.0000

Table 2: Matrix of properties.

	p_1	p_2	p_3	p_4	p_5	p_6	p_7
R.M1	—	—	√	—	√	√	—
R.M2	√	—	√	√	√	√	√
R.M3	√	—	√	√	√	√	√
R.M4	√	—	—	√	√	√	√
R.M5	√	—	√	√	√	—	√
R.M6	√	—	—	√	√	—	√
R.M7	√	—	—	√	√	—	√
R.M8	√	—	—	√	√	√	√
R.M9	√	√	√	√	—	—	—
R.M10	√	√	√	√	—	—	—
R.M11	√	—	√	√	√	—	—
R.M12	√	—	√	√	—	—	√
R.M13	√	—	—	√	√	—	√
R.M14	√	—	—	—	—	√	—
R.M15	√	—	—	√	√	—	√

Table 3: Ranking orders.

E.D.	R.M1	R.M2	R.M3	R.M4	R.M5	R.M6	R.M7	R.M8
DMU_1	7	7	7	6	7	7	7	6
DMU_4	2	2	2	2	2	2	2	2
DMU_7	5	3	4	3	3	4	3	4
DMU_{12}	6	6	6	5	6	5	5	5
DMU_{15}	1	1	1	1	1	1	1	1
DMU_{17}	3	5	3	5	4	3	4	7
DMU_{20}	4	4	5	4	5	6	6	3

Table 4: Ranking orders.

E.D.	R.M9	R.M10	R.M11	R.M12	R.M13	R.M14	R.M15
DMU_1	7	7	7	7	7	5	5
DMU_4	2	1	2	2	2	1	2
DMU_7	1	2	3	5	5	2	6
DMU_{12}	4	3	5	6	6	3	7
DMU_{15}	3	6	1	1	1	6	1
DMU_{17}	6	5	4	3	3	6	3
DMU_{20}	5	4	6	4	4	4	4

R.M2: MAJ model [32] presented for ranking efficient which is always stable but might be infeasible in some cases.

R.M3: Modified MAJ model [38] overcomes the problem which might occurr in MAJ model.

R.M4: A new model based on the idea of alterations in the reference set of the inefficient units [50].

R.M5: A model presented by Li et al. [39] which is a super-efficiency method that does not have the suffering in previous methods.

R.M6: Slack-based model [33] is based upon the input and output variables at the same time.

R.M7: SA DEA model [49] overcomes the problem of infeasibility which existed in A.P. model.

R.M8: Cross-efficiency [10] is provided based on using weights of each unit under evaluation in optimality for other units.

R.M9: A model based on finding common set of weights [25] which determine the common set of weights for DMUs and ranked DMUs based this idea.

R.M10: A model based on finding common set of weights [22, 66] for ranking efficient units.

R.M11: L1-norm model [34, 35, 65], the idea is based upon the leave-one-out efficient unit and l1-norm which is always feasible and stable.

R.M12: L_∞-norm model Rezai Balf et al. [46] provided a method with more ability over other existing methods, based on Tchebycheff norm.

R.M13: An enhanced Russel measure of super-efficiency model for ranking units [44].

R.M14: A ranking model which considers the distance of unit from the full inefficient frontier [38]

R.M15: A modified super-efficiency model [45] which overcomes the infeasibility that may happen in problem. This model is based on simultaneous projection of input output.

In accordance with properties of different ranking models in order to rank efficient units, consider Table 2.

p_1: Feasibility

p_2: Ranking extreme efficient units

p_3: Complexity in computation

p_4: Instability

p_5: Absence of multiple optimal solution

p_6: Dependency to θ and slacks

p_7: Dependency to the number of efficient and inefficient units.

In Tables 3 and 4 the rank of efficient units considering the above mentioned methods is listed. Note that E.D. shows efficient DMUs (E.D.), and $\mathrm{R.M}_j, (j = 1, \ldots, 15)$ are those explained previously.

12. Conclusion

In this paper, the DEA ranking was reviewed and classified into seven general groups. In the first group, those papers based on a cross-efficiency matrix were reviewed. In this field, DMUs have been evaluated by self- and peer pressure. The second group of papers is based on those papers looking for optimal weights in DEA analysis. The third one is the super-efficiency method. By omitting the under evaluation unit and constructing a new frontier by the remaining units, the unit under evaluation can get an equal or greater score. The fourth group is based on benchmarking idea. In this class, the effect of an efficient unit considered as a target for inefficient units is investigated. This idea is very useful for the managers in decision making. Another class, the fifth one, involves the application of multivariate statistical tools. The sixth section discusses the ranking methods based on multicriteria decision-making (MCDM) methodologies and DEA. The final section includes some various ranking methods presented in the literature. Finally, in an application, the result of some of the above-mentioned ranking methods is presented.

References

[1] M. J. Farrell, "The measurement of productive efficiency," *Journal of the Royal Statistical Society A*, vol. 120, pp. 253–281, 1957.

[2] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[3] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, no. 9, pp. 1078–1092, 1984.

[4] A. Charnes, W. W. Cooper, L. Seiford, and J. Stutz, "A multiplicative model for efficiency analysis," *Socio-Economic Planning Sciences*, vol. 16, no. 5, pp. 223–224, 1982.

[5] A. Charnes, C. T. Clark, W. W. Cooper, and B. Golany, "A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. air forces," *Annals of Operations Research*, vol. 2, no. 1, pp. 95–112, 1984.

[6] A. Charnes, W. W. Cooper, B. Golany, L. Seiford, and J. Stutz, "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions," *Journal of Econometrics*, vol. 30, no. 1-2, pp. 91–107, 1985.

[7] R. M. Thrall, "Duality, classification and slacks in DEA," *Annals of Operations Research*, vol. 66, pp. 109–138, 1996.

[8] N. Adler and B. Golany, "Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe," *European Journal of Operational Research*, vol. 132, no. 2, pp. 260–273, 2001.

[9] F. W. Young and R. M. Hamer, *Multidimensional Scaling, History, Theory and Applications*, Lawrence Erlbaum, London, UK, 1987.

[10] T. R. Sexton, R. H. Silkman, and A. J. Hogan, "Data envelopment analysis: critique and extensions," in *Measuring Efficiency: An Assessment of Data Envelopment Analysis*, R. H. Silkman, Ed., pp. 73–105, Jossey-Bass, San Francisco, Calif, USA, 1986.

[11] W. Rödder and E. Reucher, "A consensual peer-based DEA-model with optimized cross-efficiencies: input allocation instead of radial reduction," *European Journal of Operational Research*, vol. 212, no. 1, pp. 148–154, 2011.

[12] H. H. Örkcü and H. Bal, "Goal programming approaches for data envelopment analysis cross efficiency evaluation," *Applied Mathematics and Computation*, vol. 218, no. 2, pp. 346–356, 2011.

[13] J. Wu, J. Sun, L. Liang, and Y. Zha, "Determination of weights for ultimate cross efficiency using Shannon entropy," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5162–5165, 2011.

[14] G. R. Jahanshahloo, F. H. Lotfi, Y. Jafari, and R. Maddahi, "Selecting symmetric weights as a secondary goal in DEA cross-efficiency evaluation," *Applied Mathematical Modelling*, vol. 35, no. 1, pp. 544–549, 2011.

[15] Y.-M. Wang, K.-S. Chin, and Y. Luo, "Cross-efficiency evaluation based on ideal and anti-ideal decision making units," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10312–10319, 2011.

[16] N. Ramón, J. L. Ruiz, and I. Sirvent, "Reducing differences between profiles of weights: a "peer-restricted" cross-efficiency evaluation," *Omega*, vol. 39, no. 6, pp. 634–641, 2011.

[17] D. Guo and J. Wu, "A complete ranking of DMUs with undesirable outputs using restrictions in DEA models," *Mathematical and Computer Modelling*, 2012.

[18] I. Contreras, "Optimizing the rank position of the DMU as secondary goal in DEA cross-evaluation," *Applied Mathematical Modelling*, vol. 36, no. 6, pp. 2642–2648, 2012.

[19] J. Wu, J. Sun, and L. Liang, "Cross efficiency evaluation method based on weight-balanced data envelopment analysis model," *Computers and Industrial Engineering*, vol. 63, pp. 513–519, 2012.

[20] M. Zerafat Angiz, A. Mustafa, and M. J. Kamali, "Cross-ranking of decision making units in data envelopment analysis," *Applied Mathematical Modelling*, vol. 37, no. 1-2, pp. 398–405, 2013.

[21] S. Washio and S. Yamada, "Evaluation method based on ranking in data envelopment analysis," *Expert Systems with Applications*, vol. 40, no. 1, pp. 257–262, 2013.

[22] G. R. Jahanshahloo, A. Memariani, F. H. Lotfi, and H. Z. Rezai, "A note on some of DEA models and finding efficiency and complete ranking using common set of weights," *Applied Mathematics and Computation*, vol. 166, no. 2, pp. 265–281, 2005.

[23] Y.-M. Wang, Y. Luo, and Z. Hua, "Aggregating preference rankings using OWA operator weights," *Information Sciences*, vol. 177, no. 16, pp. 3356–3363, 2007.

[24] M. R. Alirezaee and M. Afsharian, "A complete ranking of DMUs using restrictions in DEA models," *Applied Mathematics and Computation*, vol. 189, no. 2, pp. 1550–1559, 2007.

[25] F.-H. F. Liu and H. Hsuan Peng, "Ranking of units on the DEA frontier with common weights," *Computers and Operations Research*, vol. 35, no. 5, pp. 1624–1637, 2008.

[26] Y.-M. Wang, Y. Luo, and L. Liang, "Ranking decision making units by imposing a minimum weight restriction in the data envelopment analysis," *Journal of Computational and Applied Mathematics*, vol. 223, no. 1, pp. 469–484, 2009.

[27] S. M. Hatefi and S. A. Torabi, "A common weight MCDA-DEA approach to construct composite indicators," *Ecological Economics*, vol. 70, no. 1, pp. 114–120, 2010.

[28] F. Hosseinzadeh Lotfi, A. A. Noora, G. R. Jahanshahloo, and M. Reshadi, "One DEA ranking method based on applying aggregate units," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13468–13471, 2011.

[29] Y.-M. Wang, Y. Luo, and Y.-X. Lan, "Common weights for fully ranking decision making units by regression analysis," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9122–9128, 2011.

[30] N. Ramón, J. L. Ruiz, and I. Sirvent, "Common sets of weights as summaries of DEA profiles of weights: with an application to the ranking of professional tennis players," *Expert Systems with Applications*, vol. 39, no. 5, pp. 4882–4889, 2012.

[31] P. Andersen and N. C. Petersen, "A procedure for ranking efficient units in data envelopment analysis," *Management Science*, vol. 39, no. 10, pp. 1261–1294, 1993.

[32] S. Mehrabian, M. R. Alirezaee, and G. R. Jahanshahloo, "A complete efficiency ranking of decision making units in data envelopment analysis," *Computational Optimization and Applications*, vol. 14, no. 2, pp. 261–266, 1999.

[33] K. Tone, "A slacks-based measure of super-efficiency in data envelopment analysis," *European Journal of Operational Research*, vol. 143, no. 1, pp. 32–41, 2002.

[34] G. R. Jahanshahloo, M. Sanei, and N. Shoja, "Modified ranking models, using the concept of advantage in data envelopment analysis," Working paper, 2004.

[35] G. R. Jahanshahloo, F. Hosseinzadeh Lotfi, N. Shoja, G. Tohidi, and S. Razavyan, "Ranking using l_1-norm in data envelopment analysis," *Applied Mathematics and Computation*, vol. 153, no. 1, pp. 215–224, 2004.

[36] Y. Chen and H. D. Sherman, "The benefits of non-radial vs. radial super-efficiency DEA: an application to burden-sharing amongst NATO member nations," *Socio-Economic Planning Sciences*, vol. 38, no. 4, pp. 307–320, 2004.

[37] A. Amirteimoori, G. Jahanshahloo, and S. Kordrostami, "Ranking of decision making units in data envelopment analysis: a distance-based approach," *Applied Mathematics and Computation*, vol. 171, no. 1, pp. 122–135, 2005.

[38] G. R. Jahanshahloo, L. Pourkarimi, and M. Zarepisheh, "Modified MAJ model for ranking decision making units in data envelopment analysis," *Applied Mathematics and Computation*, vol. 174, no. 2, pp. 1054–1059, 2006.

[39] S. Li, G. R. Jahanshahloo, and M. Khodabakhshi, "A super-efficiency model for ranking efficient units in data envelopment analysis," *Applied Mathematics and Computation*, vol. 184, no. 2, pp. 638–648, 2007.

[40] S. J. Sadjadi, H. Omrani, S. Abdollahzadeh, M. Alinaghian, and H. Mohammadi, "A robust super-efficiency data envelopment analysis model for ranking of provincial gas companies in Iran," *Expert Systems with Applications*, vol. 38, no. 9, pp. 10875–10881, 2011.

[41] A. Gholam Abri, G. R. Jahanshahloo, F. Hosseinzadeh Lotfi, N. Shoja, and M. Fallah Jelodar, "A new method for ranking non-extreme efficient units in data envelopment analysis," *Optimization Letters*, vol. 7, no. 2, pp. 309–324, 2011.

[42] G. R. Jahanshahloo, M. Khodabakhshi, F. Hosseinzadeh Lotfi, and M. R. Moazami Goudarzi, "A cross-efficiency model based on super-efficiency for ranking units through the TOPSIS approach and its extension to the interval case," *Mathematical and Computer Modelling*, vol. 53, no. 9-10, pp. 1946–1955, 2011.

[43] A. A. Noura, F. Hosseinzadeh Lotfi, G. R. Jahanshahloo, and S. Fanati Rashidi, "Super-efficiency in DEA by effectiveness of each unit in society," *Applied Mathematics Letters*, vol. 24, no. 5, pp. 623–626, 2011.

[44] A. Ashrafi, A. B. Jaafar, L. S. Lee, and M. R. A. Bakar, "An enhanced russell measure of super-efficiency for ranking efficient units in data envelopment analysis," *The American Journal of Applied Sciences*, vol. 8, no. 1, pp. 92–96, 2011.

[45] J.-X. Chen, M. Deng, and S. Gingras, "A modified super-efficiency measure based on simultaneous input-output projection in data envelopment analysis," *Computers & Operations Research*, vol. 38, no. 2, pp. 496–504, 2011.

[46] F. Rezai Balf, H. Zhiani Rezai, G. R. Jahanshahloo, and F. Hosseinzadeh Lotfi, "Ranking efficient DMUs using the Tchebycheff norm," *Applied Mathematical Modelling*, vol. 36, no. 1, pp. 46–56, 2012.

[47] Y. Chen, J. Du, and J. Huo, "Super-efficiency based on a modified directional distance function," *Omega*, vol. 41, no. 3, pp. 621–625, 2013.

[48] A. M. Torgersen, F. R. Foørsund, and S. A. C. Kittelsen, "Slack-adjusted efficiency measures and ranking of efficient units," *Journal of Productivity Analysis*, vol. 7, no. 4, pp. 379–398, 1996.

[49] T. Sueyoshi, "DEA non-parametric ranking test and index measurement: slack-adjusted DEA and an application to Japanese agriculture cooperatives," *Omega*, vol. 27, no. 3, pp. 315–326, 1999.

[50] G. R. Jahanshahloo, H. V. Junior, F. H. Lotfi, and D. Akbarian, "A new DEA ranking system based on changing the reference set," *European Journal of Operational Research*, vol. 181, no. 1, pp. 331–337, 2007.

[51] W.-M. Lu and S.-F. Lo, "An interactive benchmark model ranking performers: application to financial holding companies," *Mathematical and Computer Modelling*, vol. 49, no. 1-2, pp. 172–179, 2009.

[52] J.-X. Chen and M. Deng, "A cross-dependence based ranking system for efficient and inefficient units in DEA," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9648–9655, 2011.

[53] L. Friedman and Z. Sinuany-Stern, "Scaling units via the canonical correlation analysis in the DEA context," *European Journal of Operational Research*, vol. 100, no. 3, pp. 629–637, 1997.

[54] E. D. Mecit and I. Alp, "A new proposed model of restricted data envelopment analysis by correlation coefficients," *Applied Mathematical Modeling*, vol. 3, no. 5, pp. 3407–3425, 2013.

[55] T. Joro, P. Korhonen, and J. Wallenius, "Structural comparison of data envelopment analysis and multiple objective linear programming," *Management Science*, vol. 44, no. 7, pp. 962–970, 1998.

[56] X.-B. Li and G. R. Reeves, "Multiple criteria approach to data envelopment analysis," *European Journal of Operational Research*, vol. 115, no. 3, pp. 507–517, 1999.

[57] V. Belton and T. J. Stewart, "DEA and MCDA: Competing or complementary approaches?" in *Advances in Decision Analysis*, N. Meskens and M. Roubens, Eds., Kluwer Academic, Norwell, Mass, USA, 1999.

[58] Z. Sinuany-Stern, A. Mehrez, and Y. Hadad, "An AHP/DEA methodology for ranking decision making units," *International Transactions in Operational Research*, vol. 7, no. 2, pp. 109–124, 2000.

[59] G. Strassert and T. Prato, "Selecting farming systems using a new multiple criteria decision model: the balancing and ranking method," *Ecological Economics*, vol. 40, no. 2, pp. 269–277, 2002.

[60] M.-C. Chen, "Ranking discovered rules from data mining with multiple criteria by data envelopment analysis," *Expert Systems with Applications*, vol. 33, no. 4, pp. 1110–1116, 2007.

[61] J. Jablonsky, "Multicriteria approaches for ranking of efficient units in DEA models," *Central European Journal of Operations Research*, vol. 20, no. 3, pp. 435–449, 2012.

[62] Y.-M. Wang and P. Jiang, "Alternative mixed integer linear programming models for identifying the most efficient decision making unit in data envelopment analysis," *Computers and Industrial Engineering*, vol. 62, no. 2, pp. 546–553, 2012.

[63] F. Hosseinzadeh Lotfi, M. Rostamy-Malkhalifeh, N. Aghayi, Z. Ghelej Beigi, and K. Gholami, "An improved method for ranking alternatives in multiple criteria decision analysis," *Applied Mathematical Modelling*, vol. 37, no. 1-2, pp. 25–33, 2013.

[64] L. M. Seiford and J. Zhu, "Context-dependent data envelopment analysis: measuring attractiveness and progress," *Omega*, vol. 31, no. 5, pp. 397–408, 2003.

[65] G. R. Jahanshahloo, M. Sanei, F. Hosseinzadeh Lotfi, and N. Shoja, "Using the gradient line for ranking DMUs in DEA," *Applied Mathematics and Computation*, vol. 151, no. 1, pp. 209–219, 2004.

[66] G. R. Jahanshahloo, F. Hosseinzadeh Lotfi, H. Zhiani Rezai, and F. Rezai Balf, "Using Monte Carlo method for ranking efficient DMUs," *Applied Mathematics and Computation*, vol. 162, no. 1, pp. 371–379, 2005.

[67] G. R. Jahanshahloo and M. Afzalinejad, "A ranking method based on a full-inefficient frontier," *Applied Mathematical Modelling*, vol. 30, no. 3, pp. 248–260, 2006.

[68] A. Amirteimoori, "DEA efficiency analysis: efficient and anti-efficient frontier," *Applied Mathematics and Computation*, vol. 186, no. 1, pp. 10–16, 2007.

[69] C. Kao, "Weight determination for consistently ranking alternatives in multiple criteria decision analysis," *Applied Mathematical Modelling*, vol. 34, no. 7, pp. 1779–1787, 2010.

[70] M. Khodabakhshi and K. Aryavash, "Ranking all units in data envelopment analysis," *Applied Mathematics Letters*, vol. 25, no. 12, pp. 2066–2070, 2012.

[71] M. Zerafat Angiz, A. Tajaddini, A. Mustafa, and M. Jalal Kamali, "Ranking alternatives in a preferential voting system using fuzzy concepts and data envelopment analysis," *Computers and Industrial Engineering*, vol. 63, no. 4, pp. 784–790, 2012.

[72] A. Charnes, W. W. Cooper, and S. Li, "Using data envelopment analysis to evaluate efficiency in the economic performance of Chinese cities," *Socio-Economic Planning Sciences*, vol. 23, no. 6, pp. 325–344, 1989.

[73] W. D. Cook and M. Kress, "An mth generation model for weak ranging of players in a tournament," *Journal of the Operational Research Society*, vol. 41, no. 12, pp. 1111–1119, 1990.

[74] W. D. Cook, J. Doyle, R. Green, and M. Kress, "Ranking players in multiple tournaments," *Computers & Operations Research*, vol. 23, no. 9, pp. 869–880, 1996.

[75] M. Martić and G. Savić, "An application of DEA for comparative analysis and ranking of regions in Serbia with regards to social-economic development," *European Journal of Operational Research*, vol. 132, no. 2, pp. 343–356, 2001.

[76] I. De Leeneer and H. Pastijn, "Selecting land mine detection strategies by means of outranking MCDM techniques," *European Journal of Operational Research*, vol. 139, no. 2, pp. 327–338, 2002.

[77] M. P. E. Lins, E. G. Gomes, J. C. C. B. Soares de Mello, and A. J. R. Soares de Mello, "Olympic ranking based on a zero sum gains DEA model," *European Journal of Operational Research*, vol. 148, no. 2, pp. 312–322, 2003.

[78] A. N. Paralikas and A. I. Lygeros, "A multi-criteria and fuzzy logic based methodology for the relative ranking of the fire hazard of chemical substances and installations," *Process Safety and Environmental Protection*, vol. 83, no. 2 B, pp. 122–134, 2005.

[79] A. I. Ali and R. Nakosteen, "Ranking industry performance in the US," *Socio-Economic Planning Sciences*, vol. 39, no. 1, pp. 11–24, 2005.

[80] J. C. Martin and C. Roman, "A benchmarking analysis of spanish commercial airports. A comparison between SMOP and DEA ranking methods," *Networks and Spatial Economics*, vol. 6, no. 2, pp. 111–134, 2006.

[81] R. L. Raab and E. H. Feroz, "A productivity growth accounting approach to the ranking of developing and developed nations," *International Journal of Accounting*, vol. 42, no. 4, pp. 396–415, 2007.

[82] R. Williams and N. Van Dyke, "Measuring the international standing of universities with an application to Australian universities," *Higher Education*, vol. 53, no. 6, pp. 819–841, 2007.

[83] H. Jürges and K. Schneider, "Fair ranking of teachers," *Empirical Economics*, vol. 32, no. 2-3, pp. 411–431, 2007.

[84] D. I. Giokas and G. C. Pentzaropoulos, "Efficiency ranking of the OECD member states in the area of telecommunications: a composite AHP/DEA study," *Telecommunications Policy*, vol. 32, no. 9-10, pp. 672–685, 2008.

[85] M. Darvish, M. Yasaei, and A. Saeedi, "Application of the graph theory and matrix methods to contractor ranking," *International Journal of Project Management*, vol. 27, no. 6, pp. 610–619, 2009.

[86] E. H. Feroz, R. L. Raab, G. T. Ulleberg, and K. Alsharif, "Global warming and environmental production efficiency ranking of the Kyoto Protocol nations," *Journal of Environmental Management*, vol. 90, no. 2, pp. 1178–1183, 2009.

[87] S. Sitarz, "The medal points' incenter for rankings in sport," *Applied Mathematics Letters*, vol. 26, no. 4, pp. 408–412, 2013.

[88] N. Adler, L. Friedman, and Z. Sinuany-Stern, "Review of ranking methods in the data envelopment analysis context," *European Journal of Operational Research*, vol. 140, no. 2, pp. 249–265, 2002.

[89] R. G. Thompson, F. D. Singleton, R. M. Thrall, and B. A. Smith, "Comparative site evaluations for locating a high energy physics lab in Texas," *Interfaces*, vol. 16, pp. 35–49, 1986.

[90] R. G. Thompson, L. N. Langemeier, C.-T. Lee, E. Lee, and R. M. Thrall, "The role of multiplier bounds in efficiency analysis with application to Kansas farming," *Journal of Econometrics*, vol. 46, no. 1-2, pp. 93–108, 1990.

[91] R. G. Thompson, E. Lee, and R. M. Thrall, "DEA/AR-efficiency of U.S. independent oil/gas producers over time," *Computers and Operations Research*, vol. 19, no. 5, pp. 377–391, 1992.

[92] R. H. Green, J. R. Doyle, and W. D. Cook, "Preference voting and project ranking using DEA and cross-evaluation," *European Journal of Operational Research*, vol. 90, no. 3, pp. 461–472, 1996.

[93] J. Doyle and R. Green, "Efficiency and cross-efficiency in DEA: derivations, meanings and uses," *Journal of the Operational Research Society*, vol. 45, no. 5, pp. 567–578, 1994.

[94] J. Zhu, "Robustness of the efficient decision-making units in data envelopment analysis," *European Journal of Operational Research*, vol. 90, pp. 451–460, 1996.

[95] L. M. Seiford and J. Zhu, "Infeasibility of super-efficiency data envelopment analysis models," *INFOR Journal*, vol. 37, no. 2, pp. 174–187, 1999.

[96] J. H. Dulá and B. L. Hickman, "Effects of excluding the column being scored from the DEA envelopment LP technology matrix," *Journal of the Operational Research Society*, vol. 48, no. 10, pp. 1001–1012, 1997.

[97] A. Hashimoto, "A ranked voting system using a DEA/AR exclusion model: a note," *European Journal of Operational Research*, vol. 97, no. 3, pp. 600–604, 1997.

[98] M. Khodabakhshi, "A super-efficiency model based on improved outputs in data envelopment analysis," *Applied Mathematics and Computation*, vol. 184, no. 2, pp. 695–703, 2007.

[99] M. M. Tatsuoka and P. R. Lohnes, *Multivariant Analysis: Techniques For Educational and Psycologial Research*, Macmillan Publishing Company, New York, NY, USA, 2nd edition, 1988.

[100] Z. Sinuany-Stern, A. Mehrez, and A. Barboy, "Academic departments efficiency via DEA," *Computers and Operations Research*, vol. 21, no. 5, pp. 543–556, 1994.

[101] D. F. Morrison, *Multivariate Statistical Methods*, McGraw-Hill, New York, NY, USA, 2nd edition, 1976.

[102] S. Siegel and N. J. Castellan, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York, NY, USA, 1998.

[103] T. Obata and H. Ishii, "A method for discriminating efficient candidates with ranked voting data," *European Journal of Operational Research*, vol. 151, no. 1, pp. 233–237, 2003.

[104] M. Soltanifar and F. Hosseinzadeh Lotfi, "The voting analytic hierarchy process method for discriminating among efficient decision making units in data envelopment analysis," *Computers and Industrial Engineering*, vol. 60, no. 4, pp. 585–592, 2011.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 678783, 11 pages
http://dx.doi.org/10.1155/2013/678783

Research Article

A Hybrid Approach of Bundle and Benders Applied Large Mixed Linear Integer Problem

Placido Rogerio Pinheiro[1] and Paulo Roberto Oliveira[2]

[1] *University of Fortaleza (UNIFOR), Graduate Program in Applied Informatics, Avenida Washington Soares 1321, 60811-905 Fortaleza, CE, Brazil*
[2] *Federal University of Rio de Janeiro, Computing and Systems Engineering Department, 21945-970 Rio de Janeiro, RJ, Brazil*

Correspondence should be addressed to Placido Rogerio Pinheiro; placidrp@uol.com.br

Received 2 February 2013; Accepted 19 May 2013

Academic Editor: Farhad Hosseinzadeh Lotfi

Consider a large mixed integer linear problem where structure of the constraint matrix is sparse, with independent blocks, and coupling constraints and variables. There is one of the groups of constraints to make difficult the application of Benders scheme decomposition. In this work, we propose the following algorithm; a Lagrangian relaxation is made on the mentioned set of constraints; we presented a process heuristic for the calculation of the multiplier through the resolution of the dual problem, structured starting from the bundle methods. According to the methodology proposed, for each iteration of the algorithm, we propose Benders decomposition where quotas are provided for the value function and ε-subgradient.

1. Introduction

The main objective of this work is to develop a methodology that combines the Benders decomposition and a heuristic for calculating the multiplier applied to a relaxed problem to solve a linear problem of large integer

$$\begin{aligned} \min \quad & c^t x + d^t y, \\ \text{s.t.} \quad & Ax + By \le b, \\ & x \in X_R,\ y \in Y, \end{aligned} \tag{1}$$

where $X_R = \{x : Dx \le d, x \ge 0\}$ and $Y = \{y : Fy \le f, y \ge 0, y \text{ integer}\}$.

Relaxing a part of the restrictions has been updated by multiplying the respective heuristic process, solving a local model of dual relaxed. As the algorithm presented, for each iteration with the multiplier obtained, apply iterations of Benders decomposition on the relaxed problem, obtaining an ε-subgradient quotas and lower and upper optimal solution. The motivation of this work arises from an integer linear problem from the large expansion planning of the transmission of a digital telecommunication system in an urban area equivalent to a city the size of Rio de Janeiro/Brazil [1]. Section 2 gives an outline about decomposition methods for large scale, with all the relevant features for this work. Moreover, Section 3 shows the integer linear model in this work. More details of the methodology regularization and Benders decomposition are given in Section 4. Section 5 gives an improvement the Approximate Algorithm Bundle. In Section 6 is presented some conclusion and future works obtained from this work.

2. Decomposition Methods for Large Scale

The large mixed integer linear programming problem has highlighted the difficulty to be solved directly through commercial software. In such cases the Lagrangian, combined with subgradient optimization, is often used to lower levels to find the optimal value of the objective function. These quotes can be used, for example, the method of Branch-and-Bound [2], or just to measure the quality of feasible

solutions. These properties are currently incorporated in commercial software [3]. Other strategies are also considered: obtaining upper bounds [4], more efficient routines on the generation of cuts and the use of parallel processing [5]. The Lagrangian was used by [6, 7] with its work on the traveling salesman problems, and methods of Branch-and-Bound and implicit enumeration had considerable gain in [8] with the Lagrangian, in [9]. There are several questions directed to the Lagrangian in integer linear problems, among them how to calculate the Lagrange multipliers, how to choose among the various relaxations of the problem, and how to obtain viable solutions to the primal problem. Techniques for solving the Lagrangian dual relaxation of combinatorial optimization problems in polynomial time by applying the algorithm as a subroutine of ellipsoids [10] have been presented with [11]. Other methodologies use heuristics decomposition lagrange combining the solution of the Lagrangian Dual by the method of subgradient, also considering the feasible solutions primal heuristics [12]. These techniques were applied to flow problems in networks "multicommodity" in [13], the capacitated location problems [14]. The decomposition Benders [15] is an exact method, finite, effective when the number of integer variables is much smaller than the number of continuous variables in which case the master problem has dimension much smaller than the original problem. However, for large problems, the Benders master problem can be difficult to solve because of the large size. It joins the convergence speed generally slow, making this method inefficient in many cases. Moreover, computational experiments have shown that a general code of Branch-and-Bound applied to solve the problem Benders master produces a tree often much larger than for solving the original problem. Thus, the disadvantage of this decomposition is often the difficulty in solving the master problem, making it inefficient.

Several papers were presented with the objective of solving the problem about master with a higher overall efficiency. Among these, [2, 9, 13, 21, 56, 62] implement the Benders decomposition with Lagrangian applied in cuts master problem. [22–25]. This transfers the difficulty of the master problem in solving iteratively the maximum dual function. In [26, 27] this method is denied due to lack of controllability (the optimal solution in the Benders master problem can never achieve the optimum in relaxed master problem) in the solution of the relaxed master problem. There are also suggestions on how to get a good initial set of cuts for the Benders master problem [28, 29]. In [30] suggests the use of linear relaxation for the Benders master problem in a number of initial iterations.

Motivated by these failures, [31, 32] developed the "Cross Decomposition" while exploiting the structures of primal and dual problems, combining the advantages of Dantzig-Wolfe decomposition [33, 34] and Benders [34, 35]. Reference [27] carried out a comparative study of several approaches to the problem Benders master, presenting an efficient method for solving a linear problem, the whole "Cross Decomposition" [31, 32, 36–39]. Theoretical aspects of Benders decomposition together with the "Cross Decomposition" are also discussed in [36, 40]. Changes in "Cross Decomposition" for integer linear programming problems were made by [13, 37]. These changes are made through the generalization of the method of Kornai and Liptak [41], which eliminates the need to use the master primal and dual problems. The dynamics of this decomposition is the subproblems, which iterates the primal and dual subproblems. Instead of using the last subproblem solution as input to another, using an average of all previous subproblem solutions. The convergence proof of this methodology is found in [42]. For a certain class of location problems are presented structured exact solution methods from the "Cross Decomposition" in [36, 40, 43]. A comparison of techniques Kornai and Liptak for Decomposition and Cross-linear problems with block-angular structures and computational results is discussed in [16, 44]. Both methodologies have also been applied to problems of organizational planning [45]. Reference [46] presented a simplified algorithm of "Cross Decomposition" for multiple-choice right-side constraints. Applications involving stochastic transport problems were addressed in [47], involving comparative study with other methods.

The update of the multipliers can be made by various methods. If formulated as a linear problem, the simplex is traditionally used. Moreover, in general, a dual nondifferentiable and classical approach is the method of subgradient [1, 23–25, 48], which is known not to be a method of lowering. Although more complex, the techniques originally developed for bundle [49–51] are being increasingly used. The method exploits bundle data from previous iterations, vectors iterated, objective function, and subgradients, the bundle information, to produce new iteration. The method of ε-descend [52] considers the method of programming differentiable subgradient conjugates [49, 50]. Kiwiel in [53–55] provides new insight into the method of Bundle based on classical methods of cutting planes developed by [56, 57]. The basic idea of generalization of cutting planes is to add a quadratic regularization to the linear approximation by convex parts to the objective function; this linearization is generated by using the subgradient. To avoid a large bundle, it is necessary to limit it. Reference [58], for example, presented a selection strategy based on the subgradient multipliers associated with the local model, where the bundle that remains in subgradient $n + 2$, n being the size of a variable of the problem, considered three approaches to specify the quadratic stabilization process, which are essentially equivalent. The first technique uses the confidence regions; see [59, 60]. The Moreau-Yosida regularization generates the proximal method used by [61]. A modern synthesis technique using bundle and metric variable is made from the concept of Moreau-Yosida regularization in [62, 63]. Applications in control problems involving the method of bundle can be found in [64], and other applications using Lagrangian decomposition, networks, and comparative tests with other algorithms are developed in [60] and decompositions of large and parallel optimization in [65]. Lemarechal according to [66] "is not an exaggeration to say that 90 percent of the applications of nondifferentiability appear in decompositions of one form or another, while the remaining 10 percent are shown via the calculation of eigenvalues." We mention also [67] when C. Lemaréchal says, "the nondifferentiable optimization has the biggest deficiency of the speed of convergence."

3. Model

Consider the integer linear problem (P), motivated by an application in a telecommunications system [1], as follows:

$$\begin{aligned} \nu_P = \min \quad & \sum_{k=1}^{4} c_k^t x_k + \sum_{j=1}^{2} e_j^t y_j, \\ \text{s.t.} \quad & D_k x_k = d_k, \quad k = 1, \ldots, 4, \\ & F_j y_j \le f_j, \quad j = 1, 2 \\ & A_k x_k + B_k y_1 = K_k, \quad k = 1, \ldots, 4, \\ & C_3 x_3 + C_4 x_4 + C_2 y_2 = K_5, \\ & x_k \ge 0,\ y_j \ge 0 \text{ integer}, k = 1, \ldots, 4\ j = 1, 2, \end{aligned} \tag{P}$$

where the matrices $A_k, B_k, C_2, C_3, C_4, D_k$ and F_j have appropriate dimensions with the vectors $c_k, d_k, e_j, f_j, K_k, K_5, x_k$ and y_j.

On the other hand, consider $X = \prod_{k=1}^{4} X_k$ where $X_k = \{x_k; D_k x_k - d_k = 0 \wedge x_k \ge 0, x_k \text{ integers}\}$ and $Y = \prod_{j=1}^{2} Y_j$ where $Y_j = \{y_j; F_j y_j - f_j \le 0 \wedge y_j \ge 0, y_j \text{ integers}\}$, supposed nonempty and limited, that is finite.

Consider the integer linear programming problem (P):

$$\begin{aligned} \nu_{PI} = \min \quad & \sum_{k=1}^{4} c_k^t x_k + \sum_{j=1}^{2} e_j^t y_j, \\ \text{s.t.} \quad & A_k x_k + B_k y_1 = K_k, \quad k = 1, \ldots, 4, \\ & C_3 x_3 + C_4 x_4 + C_2 y_2 = K_5, \\ & x \in X,\ y \in Y. \end{aligned} \tag{ILP}$$

A relaxation of the continuous variable x (ILP) generates $(\mathrm{ILP_x})$:

$$\begin{aligned} \nu_P = \min \quad & \sum_{k=1}^{4} c_k^t x_k + \sum_{j=1}^{2} e_j^t y_j, \\ \text{s.t.} \quad & A_k x_k + B_k y_1 = K_k, \quad k = 1, \ldots, 4, \\ & C_3 x_3 + C_4 x_4 + C_2 y_2 = K_5, \\ & x \in X_R,\ y \in Y, \end{aligned} \tag{$\mathrm{ILP_x}$}$$

where $X_R = \prod_{k=1}^{4} X_{(k)R}$, and $X_{(k)R} = \{x_k; D_k x_k - d_k = 0 \wedge x_k \ge 0\}$.

Also to relax the variable y $(\mathrm{ILP_y})$ is obtained is follows:

$$\begin{aligned} \nu_{PR} = \min \quad & \sum_{k=1}^{4} c_k^t x_k + \sum_{j=1}^{2} e_j^t y_j, \\ \text{s.t.} \quad & A_k x_k + B_k y_1 = K_k, \quad k = 1, \ldots, 4, \\ & C_3 x_3 + C_4 x_4 + C_2 y_2 = K_5, \\ & x \in X_R, y \in Y_R, \end{aligned} \tag{$\mathrm{ILP_y}$}$$

where $Y_{(j)} = \{y_j; F_j y_j - f_j \le 0 \wedge y_j \ge 0\}$.

Relaxing the last block of constraints (ILP), we have the dual

$$\nu_{\mathrm{D_I}} = \max_{\lambda} \varphi(\lambda), \tag{$\mathrm{D_I}$}$$

where for all λ defines the dual function.

$$\begin{aligned} \varphi(\lambda) = & \min_{(x,y)\in W_I} \quad \ell(x, y, \lambda) \\ = & \min \quad \sum_{k=1}^{4} c_k^t x_k + \sum_{j=1}^{2} e_j^t y_j \\ & + \lambda^t (C_3 x_3 + C_4 x_4 + C_2 y_2 - K_5), \\ & \text{s.t.} \quad (x, y) \in W_I, \end{aligned} \tag{φ}$$

where $W_I = \{(x, y); x \in X_R, y \in Y, A_k x_k + B_k y_1 = K_k, k = 1, \ldots, 4\}$.

The purpose of this relaxation is to ensure separability of blocks of variables x_3 and x_4, y_2 over in order, then applying the Benders decomposition.

4. Methodology Regularization and Benders Decomposition

The slow convergence of the algorithms in structured Benders decomposition applied large-scale integer linear programming problems motivated the development of the methodology to accelerate the classical method. Applies to Benders decomposition to the large-scale integer linear programming problem with a Lagrangian relaxation, to update the Lagrange multipliers by a Bundle Methods.

4.1. Benders Decomposition for the Relaxed Problem. The Benders decomposition applied to the relaxed problem (φ) is to reformulate this problem in an equivalent containing only y-integer variables and a continuous variable. Without loss of generality, assume that the problem has a finite optimal solution for all λ.

For each λ, (φ) can be rewritten as:

$$\begin{aligned} \min_{y\in Q} \Bigg\{ & \sum_{j=1}^{2} e_j^t y_j + \lambda^t C_2 y_2 \\ & + \min \Bigg\{ \sum_{k=1}^{4} c_k^t x_k + \lambda^t (C_3 x_3 + C_4 x_4), \\ & A_k x_k = K_k - B_k y_1, k = 1, \ldots, 4, x \in X_R \Bigg\} \Bigg\}, \end{aligned} \tag{φ'}$$

where $Q = \{y \in Y; \exists x \in X_R$ such that $A_k x_k = K_k - B_k y_1 k = 1, \ldots, 4\}$, nonempty.

For $y \in Q$ with y_1 fixed, the inner minimization subproblem (with explicit X_R)

$$\begin{aligned} v_L = \min_x \quad & \sum_{k=1}^{4} c_k^t x_k + \lambda^t \left(C_3 x_3 + C_4 x_4\right), \\ \text{s.t.} \quad & D_k x_k = d_k, \quad k = 1, \ldots, 4, \\ & A_k x_k = K_k - B_k y_1, \quad k = 1, \ldots, 4, \\ & x_k \geq 0, \quad k = 1, \ldots, 4 \end{aligned} \tag{L}$$

has given its dual

$$\begin{aligned} v_D = \max_{(v,u)} \quad & \sum_{i=1}^{4} d_i^t v_i + \sum_{i=1}^{4} \left(K_i - B_i y_1\right)^t u_i, \\ \text{s.t.} \quad & D_i^t v_i + A_i^t u_i \leq \tilde{c}_i \quad \text{where } \tilde{c}_i = \begin{cases} c_i, & i = 1, 2, \\ c_i + C_i^t \lambda, & i = 3, 4, \end{cases} \end{aligned} \tag{D}$$

where $u = (u_1, \ldots, u_4)^t$ and $v = (v_1, \ldots, v_4)^t$.

We assume that the polyhedral

$$U(\lambda) = \left\{(v, u)\,;\, D_i^t v_i + A_i^t u_i \leq \tilde{c}_i \text{ where } \tilde{c}_i = \begin{cases} c_i, & i = 1, 2, \\ c_i + C_i^t \lambda, & i = 3, 4 \end{cases}\right\} \tag{2}$$

are uniformly bounded, if necessary adding dimensions to variables (v, u) and λ.

Thus we can define the set $\{(v^q, u^q)\lambda\}$ for all $q \in P_{U(\lambda)}$ (finite) of extreme points of $U(\lambda)$. In this case, $\left(\varphi'\right)$ is equal to

$$\min_{y \in Q} \left\{ \sum_{j=1}^{2} e_j^t y_j + \lambda^t C_2 y_2 + \max_{q \in P_{U(\lambda)}} \left\{ \sum_{i=1}^{4} d_i^t v_i^q + \sum_{i=1}^{4} \left(K_i - B_i y_1\right)^t u_i^q \right\} \right\}. \tag{φ_1}$$

Consider that $z_L(\lambda)$, the argument of the minimum, has for any subset $P'_{U(\lambda)} \subseteq P_{U(\lambda)}$, the relaxed Benders master problem:

$$\begin{aligned} z_L(\lambda) = \min_{(z,y)} \quad & z, \\ \text{s.t.} \quad & z \geq \sum_{j=1}^{2} e_j^t y_j + \sum_{i=1}^{4} d_i^t v_i^q + \sum_{i=1}^{4} \left(K_i - B_i y_1\right)^t u_i^q, \\ & z \in \mathbb{R}, \ y \in Y, \ \forall q \in P'_{U(\lambda)}. \end{aligned} \tag{MB}$$

4.2. Quotas. For λ fixed, consider $z_U(\lambda) = \sum_{j=1}^{2} e_j^t y_j + \lambda^t C_2 y_2 + v_D$ upper limit, where v_D was obtained in relaxed primal dual subproblem (D) and $z_L(\lambda)$ was obtained in relaxed primal dual subproblem (MB).

Then

$$z_L(\lambda) \leq \varphi(\lambda) = \min_{(x,y) \in W_I} \ell(x, y, \lambda) \leq z_U(\lambda). \tag{3}$$

For λ variable, if we assume that, as done in the algorithm, the restrictions relaxed Benders master problem will remain the same from one to another iteration in λ, then $z_L(\lambda^{p+1}) \geq z_L(\lambda^p)$.

4.3. Regularization Dual Quadratic Problem. The iterative solution of the dual problem of maximizing into $\varphi(\mathrm{D_I})$, that updates the multiplier λ, is done using a local regulated model, such as the Bundle. However we do not know for each λ, the value of $\varphi(\lambda)$, we only have lower quotas $z_L(\lambda)$ and upper $z_U(\lambda)$.

Suppose that we are in the pth iteration λ^p. Consider the model

$$\begin{aligned} w(\lambda^p) = \max_{(w,\rho)} \quad & w - \frac{1}{2t_p} \left\|\rho - \lambda^p\right\|^2, \\ \text{s.t.} \quad & w \leq \left(g^r\right)^t \left(\rho - \lambda^p\right) + z_U\left(\rho^r\right), \quad r \geq 1, \end{aligned} \tag{$\mathrm{F_I}$}$$

where $t_p > 0$, which determines the size of direction $\rho - \lambda^P$.

For application of the Bundle method, consider the following.

(a) The value of $g^r := C_3 x_3^r + C_4 x_4^r + C_2 y_2^r - K_5$ corresponds to some ε_r-subgradient of φ in λ^r.

Indeed, for $(x^r, y^r) \in W_I$ and any λ^p,

$$\begin{aligned} \varphi(\lambda^p) &= \min_{(x,y) \in W_I} \ell(x, y, \lambda^p) \\ &= \ell\left(x^r, y^r, \lambda^p\right) - \varepsilon_r, \quad \text{for some } \varepsilon_r \geq 0 \end{aligned} \tag{4}$$

Defining ℓ,

$$\begin{aligned} \varphi(\lambda^p) &= \left(\lambda^p - \lambda\right)^t g^r + \ell\left(x^r, y^r, \lambda\right) - \varepsilon_r \\ &\geq \left(\lambda^p - \lambda\right)^t g^r + \varphi(\lambda) - \varepsilon_r, \end{aligned} \tag{5}$$

that is, $g^r \in \partial_{\varepsilon_r} \varphi(\lambda^p)$ (subdifferencial of $\varphi(\lambda^p)$).

(b) The linear cuts, corresponding to the local polyhedral model, and also $\varphi(\lambda^r)$. This value is replaced by the upper bound $z_U(\rho^r)$, provided by the primal relaxed dual subproblem (D). To get g^r, $z_U(\rho^r)$, $z_L(\rho^r)$ may require a few iterations of the Benders algorithm.

Indeed, consider acceptable $(g^r, z_U(\rho^r), z_L(\rho^r))$ if test quality of the approximation of $\varphi(\rho^r)$ is verified as follows:

$$\begin{aligned} & z_U\left(\rho^r\right) - z_L\left(\rho^r\right) \\ & \quad \leq \alpha\left(z_U\left(\lambda^{p-1}\right) - z_L\left(\lambda^{p-1}\right)\right), \quad \text{for some } 0 < \alpha < 1. \end{aligned} \tag{6}$$

It is noted that the convergence Benders method ensures that the test will be checked in a finite number of iterations [15]. With this test it is ensured that the maximum error in calculating φ, from one iteration to another in λ, decreases. Indirectly we also expect $\varepsilon_r \to 0$.

With this set of information, we have the model "approximate"

$$\Omega_p(\rho) := \min_r \left\{ (g^r)^t (\rho - \lambda^p) + z_U(\rho^r) \right\}. \tag{7}$$

Thus, equivalently, $(\mathrm{F_I})$ is as follows:

$$\max_\rho \left\{ \Omega_p(\rho) - \frac{1}{2t_p} \|\rho - \lambda^p\|^2 \right\}. \tag{8}$$

The regularized model $(\mathrm{F_I})$ has embedded in it the process (decomposed) plans secants and intends to determine a direction of ascent through the accumulated residue, with the approximate calculation of the dual function $\varphi(\lambda)$ in $(\mathrm{D_I})$, through Benders decomposition. The lemma and the proposition that follow seek to justify the existence and uniqueness of the solution of the quadratic subproblem, like it's the aggregate subgradient.

Lemma 1 (see Lemma XV.3.1.1 in [69]). *The problem* (8) *has a unique solution* ρ^{p+1} *characterized by*

$$\rho^{p+1} = \lambda^p + t_p \widehat{g}_p, \quad \widehat{g}_p \in \partial\Omega_p\left(\rho^{p+1}\right). \tag{9}$$

Furthermore

$$\Omega(\lambda) \le z_L(\lambda_p) + \left(\widehat{g}_p\right)^t (\lambda - \lambda_p) + \widehat{e}_p, \quad \forall\lambda, \tag{10}$$

where

$$\widehat{e}_p = \Omega\left(\rho^{p+1}\right) - z_L(\lambda_p) - t_p \left\|\widehat{g}_p\right\|^2. \tag{11}$$

Proof. Suppose nonempty set generated by linear constraints, the existence and uniqueness of the solution ρ^{p+1} following the definition of a positive quadratic. The optimality condition for this solution is

$$0 \in \partial\Omega_p\left(\rho^{p+1}\right) - \frac{1}{t_p}\left(\rho^{p+1} + \lambda^p\right). \tag{12}$$

Then

$$\Omega(\lambda) \le \Omega\left(\rho^{p+1}\right) + \left(\widehat{g}_p\right)^t \left(\lambda - \rho^{p+1}\right) \tag{13}$$

which is equivalent to

$$\begin{aligned} \Omega(\lambda) \le{}& z_L(\lambda_p) + \left(\widehat{g}_p\right)^t (\lambda - \lambda_p) - z_L(\lambda_p) \\ &+ \Omega\left(\rho^{p+1}\right) + \left(\widehat{g}_p\right)^t \left(\lambda_p - \rho^{p+1}\right). \end{aligned} \tag{14}$$

Considering (9) recognizes the expression (11) of $\widehat{e}_p$. □

Proposition 2 (see Lemma XV.3.1.2 in [69]). *With the notation of Lemma 1, consider a quadratic function* $\Psi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *satisfying*

$$\Psi(\lambda) \le z_L(\lambda_p) + \left(\widehat{g}_p\right)^t (\lambda - \lambda_p) + \widehat{e}_p =: \widetilde{\varphi}(\lambda), \quad \forall\lambda, \tag{15}$$

where equality in $\lambda = \rho^{p+1}$. *Then* ρ^{p+1} *maximizes the function*

$$\widetilde{\Psi}(\lambda) := \Psi(\lambda) + \frac{1}{2t_p}\|\lambda - \lambda^p\|^2. \tag{16}$$

Proof. Applying (9) and (11) and defining relations Ψ can be written as follows:

$$\begin{aligned} \Psi(\lambda) \le{}& \Omega\left(\rho^{p+1}\right) + \left(\widehat{g}_p\right)^t \left(\lambda - \rho^{p+1}\right) - \Omega\left(\rho^{p+1}\right) \\ &+ \left(\widehat{g}_p\right)^t \left(\rho^{p+1} - \lambda_p\right) + z_L(\lambda_p) + \widehat{e}_p, \\ \therefore\ \Psi(\lambda) \le{}& \Omega\left(\rho^{p+1}\right) + \left(\widehat{g}_p\right)^t \left(\lambda - \rho^{p+1}\right) \end{aligned} \tag{17}$$

with equality in $\lambda = \rho^{p+1}$. Subtracting the term $(1/t_p)(\lambda - \lambda_p) = 0$ both sides,

$$\widetilde{\Psi}(\lambda) \le \Omega\left(\rho^{p+1}\right) + \left(\widehat{g}_p\right)^t \left(\lambda - \rho^{p+1}\right) - \frac{1}{2t_p}\|\lambda - \lambda^p\|^2, \tag{18}$$

even with equality $\lambda = \rho^{p+1}$. Now note that the function of the right side is maximized when

$$\widehat{g}_p - \frac{1}{t_p}(\lambda - \lambda_p) = 0, \tag{19}$$

corresponding to ρ^{p+1}, given by (9). □

The function $\widetilde{\varphi}(\lambda)$ is known to aggregate linearization approximation of φ, where the limit on the model Ω, as described in Lemma 1.

A more convenient way to solve $(\mathrm{F_I})$ is to be made through the dual problem. Define the Lagrangian and their optimality conditions as follows:

For $d := \rho - \lambda^p$,

$$\begin{aligned} \ell^*(w, d, \eta) ={}& w - \frac{1}{2t_p}\|d\|^2 \\ &- \sum_{r\in\{1,\dots,\ell)} \eta^r \left(w - (g^r)^t d - z_U(\rho^r)\right), \quad \eta \ge 0 \\ \ell^*_w = 0 \ \therefore{}& \sum_{r\in\{1,\dots,\ell)} \eta^r = 1, \\ \ell^*_d = 0 \ \therefore{}& -\frac{1}{2t_p} d + \sum_{r\in\{1,\dots,\ell)} \eta^r g^r = 0. \end{aligned} \tag{20}$$

Complementarity,

$$\eta_r \left[w - (g^r)^t d - z_U(\rho^r)\right] = 0, \quad r \ge 1. \tag{21}$$

Substituting these equations into ℓ^*, it has the following quadratic linear problem:

$$\begin{aligned} \min \quad & \wp(\eta), \\ \text{s.t.} \quad & \sum_{r\in\{1,\dots,\ell)} \eta^r = 1, \\ & \eta^r \ge 0, \end{aligned} \tag{$\mathrm{DF_I}$}$$

where

$$\wp(\eta) = \frac{t_p}{2}\left\|\sum_{r\in\{1,\dots,\ell)} \eta^r g^r\right\|^2 + \sum_{r\in\{1,\dots,\ell)} \eta^r z_U(\rho^r). \quad (22)$$

Consider that the optimality conditions have also to update the multiplier

$$\rho = \lambda^p + t_p\left(\sum_{r\in\{1,\dots,\ell)} \eta^r g^r\right), \quad (23)$$

where η is single solution of $(\mathrm{DF_I})$.

5. The Approximate Algorithm of Bundle

5.1. Algorithm Partial Benders. At each iteration, the multiplier λ is used in subproblem (D), what, resolved, provides an upper bound z_U and generates a new Benders cut to be included in the relaxed master problem (MB). Solving this provides a lower limit z_L and a variable y for subproblem (L), which in turn is resolved in x. With λ fixed, this process is repeated and accumulated up all the cuts in the master problem Benders (MB), until the test (6) is satisfied. At the end of this process the values of x, y, z_U, and z_L are taken to the regularized model, a new update to the multiplier.

Note. We chose to include the quadratic model only the cut that corresponds to the test (6). However, we can include all cuts, leaving for future work the selection policy and proper disposal.

5.2. Approximate Test Armijo. One approach test Armijo [8] here will determine that the direction of the approximation increases φ. Thereby,

$$\delta_p := \Omega\left(\rho^{p+1}\right) - z_L\left(\rho^p\right) - \frac{1}{2t_p}\left\|\rho^{p+1} - \lambda^p\right\|^2, \quad (24)$$

where $\rho^{p+1} = \lambda^p + t_p\widehat{g}$ and Ω is given by (7).

Approaching the values of φ by lower quotas $(z_L(\lambda))$ and upper $(z_U(\rho))$ has

$$\varphi\left(\rho^{p+1}\right) - \varphi\left(\lambda^p\right) \le z_U\left(\rho^{p+1}\right) - z_L\left(\lambda^p\right). \quad (25)$$

For $0 < m_1 < 1$ provided, an approximation of the test Armijo will be satisfied in ρ^{p+1} if

$$z_U\left(\rho^{p+1}\right) - z_L\left(\lambda^p\right) \ge m_1\delta_p, \quad (26)$$

where the left side is positive because

$$z_U\left(\rho^{p+1}\right) - z_L\left(\lambda^p\right) \ge z_U\left(\rho^{p+1}\right) - z_L\left(\rho^{p+1}\right) \ge 0. \quad (27)$$

If we compare this to the test that corresponded to the exact calculation of the function φ, observed that the difference between the current and the candidate has been replaced by an increase as much as δ_p is an increase of the exact value. This expects that the test stop approximate Bundle method will not be anticipated, since also ensures a good approximation to the function φ.

5.3. Regularization Algorithm for Updating the Multipliers with Relaxation. Before we present the algorithm and in order to keep the notation, replace the model $(\mathrm{F_I})$ that is equivalent to

$$\begin{aligned} w\left(\lambda^p\right) = \max_{(w,\rho)} \quad & w - \frac{1}{2t_p}\left\|\rho - \lambda^p\right\|^2, \\ \text{s.t.} \quad & w \le \left(g^r\right)^t\left(\rho - \rho^r\right) + e_r + z_U\left(\lambda^p\right), \quad r \ge 1, \end{aligned} \quad (6')$$

where

$$e_r := e\left(\lambda^p, \rho^r, g^r\right) := z_U\left(\rho^r\right) - z_U\left(\lambda^p\right) + \left(g^r\right)^t\left(\rho^r - \lambda^p\right). \quad (28)$$

We used without distinctions $g(\rho^r)$ and g^r.

Algorithm 3. Initialization: They are given tolerance stop $\underline{\delta} \ge 0$ and $\theta > 0$. Consider $\overline{\ell} > 0$ the maximum size of the Bundle, $t_1 > 0$. Get an initial dual feasible solution λ^1, $y^0 \in Y$ and x^0 initial feasible solution (L); that is, for $y = y^0$, x^0 is solution of

$$\begin{aligned} A_k x_k &= K_k - B_k y_1, \quad k = 1,\dots,4, \\ x_k &\in X_k, \quad k = 1,\dots,4. \end{aligned} \quad (29)$$

Calculate $g^1 = g(\lambda^1)$. Make $z_U(\lambda^1) := \ell(x^0, y^0, \lambda^1)$. Estimate $z_L(\lambda^1)$, for example, through an iteration of the Benders method. Choice $m_1 \in (0,1)$ reducing the test Armijo, $\alpha \in (0,1)$ is the reduction in quality test approximation φ. Initialize the set of ascent $P = \phi$, the accountant of iterations $p = 1$, and the size of the bundle $\ell = 1$. For $e_1 = 0$, corresponding to the initial bundle (g^1, e_1), and the initial model

$$\rho \longrightarrow \Omega_1(\rho) := z_U\left(\lambda^1\right) + \left(g^1\right)^t\left(\rho - \lambda^1\right). \quad (30)$$

Step 1 (principal calculation and test stop). Whether ρ^{p+1} is the unique solution of the quadratic problem such that

$$\rho^{p+1} = \lambda^p + t_p\widehat{g}^p, \quad \text{where } \widehat{g}^p \in \partial\Omega_p\left(\rho^{p+1}\right). \quad (31)$$

Make

$$\begin{aligned} \widehat{e}_p &:= \Omega_p\left(\rho^{p+1}\right) - z_L\left(\lambda^p\right) - t_p\left\|\widehat{g}^p\right\|^2, \\ \delta_p &:= \Omega_p\left(\rho^{p+1}\right) - z_L\left(\lambda^p\right) - \frac{t_p}{2}\left\|\widehat{g}^p\right\|^2. \end{aligned} \quad (32)$$

Calculate through of the algorithm

$$z_U\left(\rho^{p+1}\right), \qquad z_L\left(\rho^{p+1}\right), \qquad g\left(\rho^{p+1}\right). \quad (33)$$

If $\delta_p \le \delta$ and $z_U(\rho^{p+1}) - z_L(\rho^{p+1}) > \theta$, stop.

Step 2 (approximation test Armijo). If $z_U(\rho^{p+1}) - z_L(\lambda^p) \ge m_1\delta_p$, $m_1 \in (0,1)$ is "serious step"; otherwise, it is "null step," check to Step 4.

Step 3 (serious step).

Make $\lambda^{p+1} = \rho^{p+1}$.

Add p the set P; for $r = 1, \ldots, \ell$.

Permute e_r and $\widehat{e}_p$ by, respectively,

$$\begin{aligned} &e_r + z_U(\lambda^p) - z_U(\lambda^{p+1}) + (g^r)^t(\lambda^p - \lambda^{p+1}), \\ &\widehat{e}_r + z_U(\lambda^p) - z_U(\lambda^{p+1}) + (\widehat{g}^r)^t(\lambda^p - \lambda^{p+1}). \end{aligned} \tag{34}$$

Step 4 (control of bundle size). If $\ell = \overline{\ell}$, then eliminate at least two of the bundle elements and insert the element $(\widehat{g}^p, \widehat{e}_p)$.

Consider $(g^t, e_t)_{t=1,\ldots,\ell}$ the new bundle obtained (with $\ell = \overline{\ell}$).

Step 5. Insert $(g^{\ell+1}, e_{\ell+1})$ to the bundle, where $e_{\ell+1} = 0$ in the case of serious step, and in the case of null step,

$$e_{\ell+1} = z_U(\rho^{p+1}) - z_U(\lambda^p) + (g^{\ell+1})^t(\rho^{p+1} - \lambda^p). \tag{35}$$

Replace ℓ by $\ell + 1$, and update the model $\rho \to \Omega_{p+1}(\rho) := \min_r\{(g^r)^t(\rho - \lambda^{p+1}) + z_U(\rho^r)\}$.

Step 6. Make $p = p + 1$, and return to Step 1.

5.4. Partial Benders Algorithm for Integer Linear Problem with Relaxation. Initialization: make $q = 1$.

Step 1. Solve

$$\begin{aligned} v_D = \max_{(v_i,u_i)} \quad & \sum_{i=1}^{4} d_i^t V_i + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i, \\ \text{s.t.} \quad & D_i^t v_i + A_i^t u_i \le \widetilde{c}_i, \\ & \text{where } \widetilde{c}_i = \begin{cases} c_i, & i = 1, 2, \\ c_i + C_i^t \rho^{p+1}, & i = 3, 4. \end{cases} \end{aligned} \tag{36}$$

If there is no solution, stop: (φ_1) has no feasible solution. Otherwise, $(v^{(p,q)}, u^{(p,q)})$ solution, and then

$$z_U(\rho^{p+1}) = \sum_{j=1}^{2} e_i^t y_j + \rho^t C_2 y_2 + v_D. \tag{37}$$

Generate a new constraint (cut) from $(\rho^{p+1}, v^{(p,q)}$, and $u^{(p,q)})$. Continue with Step 2.

Step 2. Solve

$$\begin{aligned} \min \quad & z, \\ \text{s.t.} \quad & z \ge \sum_{j=1}^{2} e_j^t y_j + (\rho^{p+1})^t C_2 y_2 + \sum_{i=1}^{4} d_i^t v_i^q \\ & \quad + \sum_{i=1}^{4} (K_i - B_i y)^t, \quad z \in \mathbb{R},\ y \in Y,\ \forall q. \end{aligned} \tag{38}$$

Consider $(z_L(\rho^{p+1}), y^{p+1})$ the optimal solution. Continue with Step 3.

Step 3. Solve

$$\begin{aligned} \min \quad & \sum_{k=1}^{4} c_k^t x_k + (\rho^{p+1})^t (C_3 x_3 + C_4 x_4). \\ \text{s.t.} \quad & A_k x_k = k_k - B_k y_1^{p+1}, \quad k = 1, \ldots, 4. \\ & x_k \in X_R, \quad k = 1, \ldots, 4. \end{aligned} \tag{39}$$

Whether x^{p+1} solution; continue with Step 4.

Step 4 (test quality approach φ).

If

$$z_U(\rho^{p+1}) - z_L(\rho^{p+1}) \le \alpha(z_U(\lambda^p) - z_L(\lambda^p)) \tag{40}$$

end.

Otherwise, make $p+1 = p$, $q = q+1$, and return to Step 1.

Remarks

(1) The test for stopping the algorithm adds to the usual tolerance $\underline{\delta}$ of bundles, the requirement that the function approximation is reasonable. In fact, for crude approximations of φ, it is possible to have false serious steps with error δ false small, hence the need for θ-approximation.

(2) The Benders relaxed master problem (MB) should have some heuristics for selecting cuts, given that the accumulations of all inequalities explode the subproblem.

We present Figure 1 of the algorithm for the problem with integer linear relaxation.

5.5. About Convergence. We opted to observe that θ small enough for the results cited correspond to guarantee the stability of the algorithm of bundle. This can be observed by adding a positive parameter $\theta \to 0$, the expression of errors linearization, and the gains predicted by the model (see, Lemma 3.2.1 in [69]. Thus if only guarantee the local convergence. Furthermore, the quality test approximation (φ_1) should be sufficient for obtaining convergence of the overall strength because the iterative process to arrive at the usual formulation of the bundle, with $\theta = 0$. Undoubtedly, with the risk of being a high cost computational algorithm, as already noted. It presents the known result that guarantees no cycling algorithm Benders.

Theorem 4. *The vectors composed of the vertices and their multipliers (v^p, u^p, and λ^p) generated at each iteration by the algorithm are different.*

Proof. Suppose that the first ($p \ge 1$) extreme points, say $(v^1, u^1), (v^2, u^2), \ldots, (v^p, u^p)$ the problem generated (D) (λ) and $\lambda^1, \ldots, \lambda^p$ obtained problem regularized $(\mathrm{F_I})$.

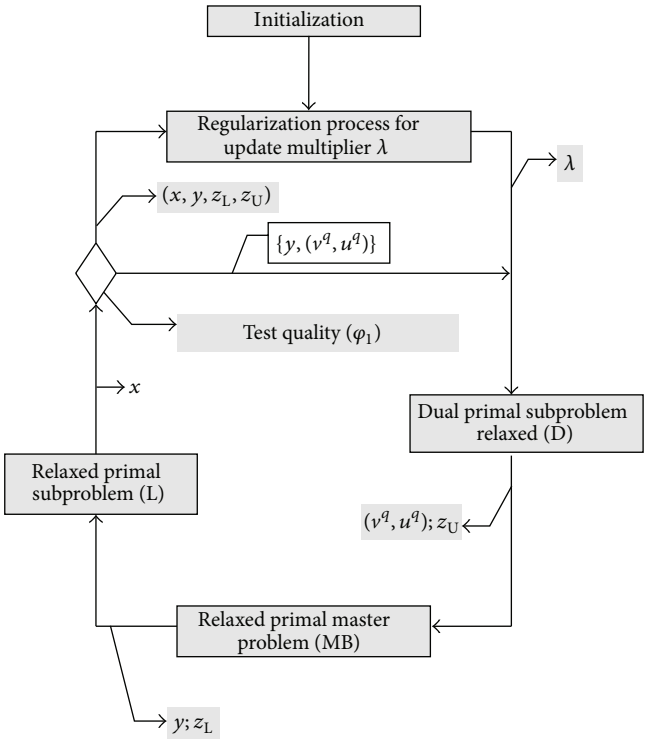

Figure 1

Then with the Step 2, it has

$$\begin{aligned} &\min \quad z, \\ &\text{s.t.} \quad z \geq \sum_{j=1}^{2} e_j^t y_j + (\lambda^s)^t C_2 y_2 + \sum_{i=1}^{4} b_i^t v_i^p + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^p, \\ &\qquad z \in \mathbb{R}, \ \ y \in Y, \ \ s \leq p. \end{aligned} \tag{41}$$

The optimal solution of this problem z, y, that is, for some k $(1 \leq k \leq p)$, $s = 1, \ldots, p$

$$\begin{aligned} z &= \sum_{j=1}^{2} e_j^t y_j (\lambda^k)^t C_2 y_2 + \sum_{i=1}^{4} b_i^t v_i^k + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^k \\ &\geq \sum_{j=1}^{2} e_j^t y_j + (\lambda^s)^t C_2 y_2 + \sum_{i=1}^{4} b_i^t v_i^s + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^s. \end{aligned} \tag{42}$$

As z is a lower bound on the optimal cost primal relaxed φ, $\varphi \geq z$, and with (42),

$$\sum_{i=1}^{4} b_i^t v_i^k + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^k \leq \varphi - \left(\sum_{j=1}^{2} e_j^t y_j + (\lambda^k)^t C_2 y_2 \right). \tag{43}$$

On the other hand, in the next iteration of (D) (λ^{p+1}) the solution (v^{p+1}, u^{p+1}) is a vertex $U(\lambda)$. Then

$$\begin{aligned} &\sum_{i=1}^{4} b_i^t v_i^{p+1} + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^{p+1} \\ &\qquad = \sum_{K=1}^{4} c_k^t x_k + \left(\lambda^{p+1}\right)^t (C_3 x_3 + C_4 x_4), \end{aligned} \tag{44}$$

where x is a solution of (L) (λ).

As (x, y) is a viable solution φ it has

$$\begin{aligned} &\sum_{j=1}^{2} e_j^t y_j + \left(\lambda^{p+1}\right)^t C_2 y_2 + \sum_{K=1}^{4} c_k^t x_k + \left(\lambda^{p+1}\right)^t (C_3 x_3 + C_4 x_4) \\ &\qquad \geq \varphi. \end{aligned} \tag{45}$$

Equivalently

$$\begin{aligned} \varphi - \sum_{j=1}^{2} e_j^t y_j \leq \ & \sum_{K=1}^{4} c_k^t x_k + \left(\lambda^{p+1}\right)^t (C_3 x_3 + C_4 x_4) \\ & + \left(\lambda^{p+1}\right)^t C_2 y_2. \end{aligned} \tag{46}$$

Combining (43) and (44), it has

$$\begin{aligned} &\sum_{i=1}^{4} b_i^t v_i^k + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^k + \left(\lambda^k\right)^t C_2 y_2 \\ &\qquad \leq \varphi - \sum_{j=1}^{2} e_j^t y_j \\ &\qquad \leq \sum_{k=1}^{4} c_k^t x_k + \left(\lambda^{p+1}\right)^t (C_3 x_3 + C_4 x_4) + \left(\lambda^{p+1}\right)^t (C_2 y_2) \\ &\qquad = \sum_{i=1}^{4} b_i^t v_i^{p+1} + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^{p+1} + \left(\lambda^{p+1}\right)^t C_2 y_2. \end{aligned} \tag{47}$$

If $\varphi - (\sum_{k=1}^{4} c_k^t x_k + (\lambda^{p+1})^t (C_3 x_3 + C_4 x_4)) = \sum_{j=1}^{2} e_j^t y_j + (\lambda^{p+1})^t C_2 y_2$ then (x, y) solves the relaxed integer linear problem (φ).

Otherwise,

$$\begin{aligned} &\sum_{i=1}^{4} b_i^t v_i^k + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^k + \left(\lambda^k\right)^t C_2 y_2 \\ &\qquad < \sum_{i=1}^{4} b_i^t v_i^{p+1} + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^{p+1} + \left(\lambda^{p+1}\right)^t C_2 y_2 \end{aligned} \tag{48}$$

in which case $(v^k, u^k, \lambda^k) \neq (v^{p+1}, u^{p+1}, \lambda^{p+1})$.

But inequality (42) is as follows:

$$\sum_{i=1}^{4} b_i^t v_i^k + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^k + \left(\lambda^k\right)^t C_2 y_2 \geq \sum_{i=1}^{4} b_i^t v_i^s + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^s + \left(\lambda^{p+1}\right)^t C_2 y_2, \quad s = 1, \ldots, p. \tag{49}$$

On the other hand, of (48),

$$\sum_{i=1}^{4} b_i^t v_i^s + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^s + (\lambda^s)^t C_2 y_2 < \sum_{i=1}^{4} b_i^t v_i^{p+1} + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^{P+1} + \left(\lambda^{p+1}\right)^t C_2 y_2, \quad s = 1, \ldots, p, \tag{50}$$

and therefore $(v^{p+1}, u^{p+1}, \lambda^{p+1}) \neq (v^s, u^s, \lambda^s)$ $s = 1, \ldots, p$. □

Corollary 5. *If $q > 1$, the mth iteration internal, then $y^{m+1} \neq y^m$.*

Proof. Suppose the contrary, that to solve (MB) with m cuts, the solution y is repeated. In this case, to solve the problem (D), we would obtain a vector (v^{m+1}, u^{m+1}) satisfying

$$\sum_{i=1}^{4} d_i^t v_i^\ell + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^\ell = \sum_{i=1}^{4} d_i^t v_i^{m+1} + \sum_{i=1}^{4} (K_i - B_i y_1)^t u_i^{m+1} \tag{51}$$

for some $\ell = \{1, \ldots, m\}$. However, this only occurs when the optimality criterion is reached. □

6. Conclusion and Future Works

Our main goal was to present an alternative technique using Lagrangian relaxation in solving a problem in integer linear programming. The work introduced a new algorithm structured from a block of relaxation of constraints that the problem presents difficulties when approached by traditional techniques Benders. We hope to take advantage of the computational process smoothing over other heuristic algorithms (Dantzig-Wolfe, subgradient) because its search direction is determined by processes similar to bundle method, which has shown proven results superior to those in many large problems [60]. It seems also unlikely that the technique of "Cross Decomposition" would be adaptable. As future works, we investigate other applications in order to verify the efficiency of the methods on structured problems and extend the decomposition to nonlinear problems and integer nonlinear, using the Lagrangian heuristic process along with the regularization. It is expected that other hybrid methodologies [71–74] can be applied in the solution of the problem (P) [1].

Acknowledgment

The authors are thankful to National Counsel of Technological and Scientific Development (CNPq).

References

[1] P. R. Pinheiro, *Uma Metodologia de Feixe e Benders Aplicada a um Problema Linear Inteiro de Grande Porte [Tese de Doutorado]*, Universidade Federal do Rio de Janeiro, 1998.

[2] H. M. Salkin and K. Mathur, *Foundations of Integer Programming*, North-Holland, 1989.

[3] ILOG, "ILOG CPLEX 10.0 Users Manual," United States of America, 2006.

[4] G. L. Nemhauser, M. W. P. Savelsbergh, and G. C. Sigismondi, "MINTO, a mixed INTeger optimizer," *Operations Research Letters*, vol. 15, no. 1, pp. 47–58, 1994.

[5] E. Kwork and Y. Lee, "Parallelism in mixed integer programming," in *Proceedings of the International Symposium on Mathematical Programming*, Ann Arbor, Mich, USA, 1994.

[6] M. Help and R. M. Karp, "The traveling salesman problem and minimum spanning trees," *Operations Research*, vol. 18, pp. 1138–11162, 1970.

[7] M. Held and R. M. Karp, "The traveling-salesman problem and minimum spanning trees: part II," *Mathematical Programming*, vol. 1, no. 1, pp. 6–25, 1971.

[8] A. M. Geoffrion, "Lagrangian relaxation for integer programming," *Mathematical Programming Study*, vol. 2, pp. 82–114, 1974.

[9] M. L. Fisher, "The lagrangian relaxation method for solving integer programming problems," *Management Science*, vol. 27, no. 1, pp. 1–18, 1981.

[10] L. G. Khachian, "A polynomial algorithm for linear programming," *Soviet Mathematics Doklady*, vol. 20, pp. 191–194, 1979.

[11] D. Bertsimas and J. B. Orlin, "A technique for speeding up the solution of the Lagrangean dual," *Mathematical Programming*, vol. 63, no. 1–3, pp. 23–45, 1994.

[12] K. Aardal and T. Larsson, "A benders decomposition based heuristic for the hierarchical production planning problem," *European Journal of Operational Research*, vol. 45, no. 1, pp. 4–14, 1990.

[13] K. Holmberg, "Mean value cross decomposition applied to integer programming problems," *European Journal of Operational Research*, vol. 97, no. 1, pp. 124–138, 1997.

[14] K. Holmberg and J. Ling, "A Lagrangean heuristic for the facility location problem with staircase costs," *European Journal of Operational Research*, vol. 97, no. 1, pp. 63–74, 1997.

[15] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische Mathematik*, vol. 4, no. 1, pp. 238–252, 1962.

[16] K. Aardal and A. Ari, "On the resemblance between the Kornai-Liptak and cross decomposition techniques for block-angular linear programs," *European Journal of Operational Research*, vol. 46, no. 3, pp. 393–398, 1990.

[17] G. Côté and M. A. Laughton, "Large-scale mixed integer programming: benders-type heuristics," *European Journal of Operational Research*, vol. 16, no. 3, pp. 327–333, 1984.

[18] M. L. Fisher and R. A. Jaikumar, "Decomposition algorithm for large scale vehicle routing," Working Paper 78-11-05, Department of Decision Sciences, University of Pennsylvania.

[19] H. H. Hoc, "Topological optimization of networks: a nonlinear mixed integer model employing generalized benders decomposition," *IEEE Transactions on Automatic Control*, vol. 27, no. 1, pp. 164–169, 1982.

[20] G. G. Paula Jr. and N. Maculan, "A p-median location algorithm based on the convex lagrangean relation of the benders master problem," in *Proceedings of the International Symposium on Mathematical Programming*, Tokyo, Japan, 1988.

[21] K. Rana and R. G. Vickson, "A model and solution algorithm for optimal routing of a time—chartered containership," *Transportation Science*, vol. 22, no. 2, pp. 83–95, 1988.

[22] M. Held, P. Wolfe, and H. P. Crowder, "Validation of subgradient optimization," *Mathematical Programming*, vol. 6, no. 1, pp. 62–88, 1974.

[23] B. T. Poljak, "A general method of solving extremum problems," *Soviet Mathematics*, vol. 8, pp. 593–597, 1967.

[24] B. T. Polyak, "Minimization of unsmooth functionals," *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 3, pp. 14–29, 1969.

[25] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, 1985.

[26] J. A. Ferland and M. Florian, "A sub-optimal algorithm to solve a large scale 0-1 programming problem," in *Survey of Mathematical Programming*, A. Prekopa, Ed., pp. 461–469, North-Holland, Amsterdam, 1979.

[27] K. Holmberg, "On using approximations of the Benders master problem," *European Journal of Operational Research*, vol. 77, no. 1, pp. 111–125, 1994.

[28] M. Minoux, "Subgradient optimization and benders decomposition for large scale programming," in *Mathematical Programming*, R. W. Cottle, M. L. Kelmanson, and B. Korte, Eds., 1984.

[29] M. Minoux, *Mathematical Programming*, Theory and Algorithms, John Wiley and Sons, 1986.

[30] D. McDaniel and M. Devine, "A modified benders' partitioning algorithm for mixed integer programming," *Management Science*, vol. 24, pp. 312–319, 1977.

[31] T. J. Van Roy, "Cross decomposition for mixed integer programming," *Mathematical Programming*, vol. 25, no. 1, pp. 46–63, 1983.

[32] T. J. Van Roy, "A cross decomposition algorithm for capacited facility location," *Operations Research*, vol. 34, no. 1, pp. 145–163, 1986.

[33] G. B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ, USA, 1963.

[34] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations Research*, vol. 8, pp. 101–111, 1960.

[35] A. M. Geoffrion, "Generalized Benders Decomposition," *Journal of Optimization Theory and Applications*, vol. 10, no. 4, pp. 237–260, 1972.

[36] K. Holmberg, "On the convergence of cross decomposition," *Mathematical Programming*, vol. 47, no. 1–3, pp. 269–296, 1990.

[37] K. Holmberg, "Generalized cross decomposition applied to nonlinear integer programming problems: duality gaps and convexification in parts," *Optimization*, vol. 23, pp. 341–3356, 1992.

[38] K. Holmberg, "Cross decomposition applied to integer programming problems: duality gaps and convexification in parts," *Operations Research*, vol. 42, no. 4, pp. 657–668, 1994.

[39] K. Holmberg, "Solving the staircase cost facility location problem with decomposition and piecewise linearization," *European Journal of Operational Research*, vol. 75, no. 1, pp. 41–61, 1994.

[40] K. Homberg, *Decomposition in large scale mathematical programming [Ph.D. dissertation]*, Institute of Technology, Linköping, Sweden, 1985.

[41] J. Kornai and T. Liptak, "Two-level planning," *Econometrica*, vol. 33, pp. 141–191, 1965.

[42] K. Holmberg, "A convergence proof for linear mean value cross decomposition," *ZOR Zeitschrift fü Operations Research Methods and Models of Operations Research*, vol. 39, no. 2, pp. 157–186, 1994.

[43] C. Y. Lee, "A cross decomposition algorithm for a multiproduct-multitype facility location problem," *Computers and Operations Research*, vol. 20, no. 5, pp. 527–540, 1993.

[44] K. Holmberg, "Linear mean value cross decomposition: a generalization of the Kornai-Liptak method," *European Journal of Operational Research*, vol. 62, no. 1, pp. 55–73, 1992.

[45] K. Holmberg, "Primal and dual decomposition as organizational design: price and/or resource directive decomposition," Tecnical Report LiTH-MAT-R-94-04, Linköping Institute of Technology, Linköping, Sweden.

[46] S. Kim, S. C. Cho, and B. S. Um, "A simplified cross decomposition algorithm for multiple right hand choice linear programming," *Journal of the Operations Research Society of Japan*, vol. 32, no. 4, pp. 441–449, 1989.

[47] K. Holmberg and K. O. Jörnsten, "Cross decomposition applied to the stochastic transportation problem," *European Journal of Operational Research*, vol. 17, no. 3, pp. 361–368, 1984.

[48] B. Pchénitchny and Y. Daniline, *Methodes Numériques dans les Problèmes d'Extrémum*, Mir, Moscow, Russia, 1977.

[49] C. Lemaréchal, "An extension of davidon methods to non differentiable problems," *Mathematical Programming Study*, vol. 3, pp. 95–109, 1975.

[50] P. Wolfe, "A method of conjugate subgradient for minimizing nondifferentiable functions," in *Nondifferentiable Optimization, Mathematical Programming Study 3*, M. L. Balinski and P. Wolfe, Eds., pp. 145–1173, 1975.

[51] R. T. Wong, *Accelerating benders decomposition for network design [Ph.D. thesis]*, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1978.

[52] C. Lemaréchal, "Nondifferenciable optimization, subgradient and ε-subgradient methods," *Lecture Notes in Economics and Mathematical Systems*, vol. 117, pp. 191–1199, 1976.

[53] K. C. Kiwiel, *Methods of Descend for Nondifferentiable Optimization*, vol. 1133 of *Lecture Notes in Mathematics*, 1985.

[54] K. C. Kiwiel, "A method for solving certain quadratic programming problems arising in nonsmooth optimization," *IMA Journal of Numerical Analysis*, vol. 6, no. 2, pp. 137–152, 1986.

[55] K. C. Kiwiel and A. Stachurski, *NOA—A Fortran Package of Nondifferentiable Optimization Algorithms*, vol. 6, Systems Research Institute, Polish Academy of Sciences, Newelska, Poland, 1988.

[56] E. W. Cheney and A. A. Goldstein, "Newton's method for convex programming and Tchebycheff approximation," *Numerische Mathematik*, vol. 1, no. 1, pp. 253–268, 1959.

[57] J. E. Kelley, "The cutting plane method for solving convex problems," *SIAM Journal*, vol. 8, pp. 703–712, 1960.

[58] K. C. Kiwiel, "Approximations in proximal bundle methods and decomposition of convex programs," *Journal of Optimization Theory and Applications*, vol. 84, no. 3, pp. 529–548, 1995.

[59] H. Schramm, *Eine Kombination von Bundle-und Trust-Region-Verfahren zur Lösung Nichtdifferenzierbarer Optimierungs Probleme*, Helf 30, Bayreuther Mathematische Schriften, Bayreuth, Germany, 1989.

[60] H. Schramm and J. Zowe, "A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results," *SIAM Journal on Optimization*, vol. 2, no. 1, pp. 121–152, 1992.

[61] K. C. Kiwiel, "Proximity control in bundle methods for convex nondifferentiable minimization," *Mathematical Programming*, vol. 46, no. 1–3, pp. 105–122, 1990.

[62] C. Lemaréchal and C. Sagastizábal, "Application of bundle methods to the unit-commitment problem," Rapport Tecnique 0184, Institu Recherche d'Informatique et d'Automatique, Le Chesnay, 1995.

[63] L. Qi and X. Chen, "A preconditioning proximal Newton method for nondifferentiable convex optimization," *Mathematical Programming B*, vol. 76, no. 3, pp. 411–429, 1997.

[64] M. M. Mäkelä and P. Neittaanmäki, *Nonsmooth Optimization*, World Scientific, 1992.

[65] D. Medhi, *Decomposition of structured large scale optimization problems and parallel optimization [Ph.D. thesis]*, University of Wisconsin, 1987.

[66] SIAG/OPT Views-and-News, "A Forum for the SIAM Activity Group on Optimization," Larry Nazareth Editor, 4, 1994.

[67] C. Lemaréchal, "Interview with Claude Lemaréchal," Optima, 1996.

[68] C. Lemaréchal, "Lagrangian decomposition and nonsmooth optimization: bundle algorithm, prox iteration, augmented lagrangian," in *Nonsmooth Optimization: Methods and Applications*, F. Giannessi, Ed., pp. 201–216, Gordon and Breach, Philadelphia, Pa, USA, 1992.

[69] J. B. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms*, vol. 1-2, 1993.

[70] L. Armijo, "Minimization of function having continuous parcial derivatives," *Pacific Journal of Mathematics*, vol. 16, pp. 1–3, 1966.

[71] P. R. Pinheiro, A. L. V. Coelho, A. B. De Aguiar, and T. O. Bonates, "On the concept of density control and its application to a hybrid optimization framework: investigation into cutting problems," *Computers and Industrial Engineering*, vol. 61, no. 3, pp. 463–472, 2011.

[72] P. R. Pinheiro, A. L. V. Coelho, A. B. de Aguiar, and A. de Meneses Sobreira Neto, "Towards aid by generate and solve methodology: application in the problem of coverage and connectivity in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 790459, 11 pages, 2012.

[73] N. Nepomuceno, P. Pinheiro, and A. L. V. Coelho, "Tackling the container loading problem: a hybrid approach based on integer linear programming and genetic algorithms," *Lecture Notes in Computer Science*, vol. 4446, pp. 154–165, 2007.

[74] R. D. Saraiva, N. V. Nepomuceno, and P. R. Pinheiro, "The generate-and-solve framework revisited: generating by simulated annealing," *Lecture Notes in Computer Science*, vol. 7832, pp. 262–273, 2013.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 795098, 9 pages
http://dx.doi.org/10.1155/2013/795098

Research Article

Interior Point Method for Solving Fuzzy Number Linear Programming Problems Using Linear Ranking Function

Yi-hua Zhong, Yan-lin Jia, Dandan Chen, and Yan Yang

School of Science, Southwest Petroleum University, Chengdu, Sichuan 610500, China

Correspondence should be addressed to Yan-lin Jia; yelinjyl@126.com

Received 29 January 2013; Accepted 16 June 2013

Academic Editor: Hadi Nasseri

Recently, various methods have been developed for solving linear programming problems with fuzzy number, such as simplex method and dual simplex method. But their computational complexities are exponential, which is not satisfactory for solving large-scale fuzzy linear programming problems, especially in the engineering field. A new method which can solve large-scale fuzzy number linear programming problems is presented in this paper, which is named a revised interior point method. Its idea is similar to that of interior point method used for solving linear programming problems in crisp environment before, but its feasible direction and step size are chosen by using trapezoidal fuzzy numbers, linear ranking function, fuzzy vector, and their operations, and its end condition is involved in linear ranking function. Their correctness and rationality are proved. Moreover, choice of the initial interior point and some factors influencing the results of this method are also discussed and analyzed. The result of algorithm analysis and example study that shows proper safety factor parameter, accuracy parameter, and initial interior point of this method may reduce iterations and they can be selected easily according to the actual needs. Finally, the method proposed in this paper is an alternative method for solving fuzzy number linear programming problems.

1. Introduction

Linear programming is one of the most widely used decision-making tools for solving real-word problems. However, real word situations are characterized by imprecision rather than exactness. Then, fuzzy linear programming (FLP) has been developed to treat uncertainty of optimization problems, such as fuzzy data envelopment analysis and fuzzy network optimization [1–3]. Since 1970, various attempts have been made to study FLP problem [4–32]. The concept of FLP was first proposed by Tanaka et al. [4] in the framework of the fuzzy decision of Bellman and Zadeh [5]. For solving FLP, defuzzification methods have been widely studied for some years and applied to fuzzy control and fuzzy expert systems. The most common transforming method is ranking fuzzy numbers method, which is to establish a one-to-one correspondence between fuzzy numbers and real numbers according to the definite rule. Then, every fuzzy number is mapped to a point on the real line. Ranking is a viable approach for ordering fuzzy numbers. A special version of ranking function was first proposed by Yager [33].

Then, many researchers have considered various kinds of FLP problems and have proposed some approaches for solving these problems [8–28]. Maleki et al., Ganesan and Veeramani, and Nasseri et al. [8–12] presented simplex methods for solving fuzzy number linear programming (FNLP) and linear programming with fuzzy variables (FVLP) using the concept of comparison of fuzzy numbers and linear ranking function. This method is similar to the simplex method that was used for solving linear programming problems in crisp environment. Nasseri and Khabiri [13] proposed a revised simplex algorithm for FVLP, which is useful for sensitivity analysis on FVLP. Furthermore, there is a revised simplex algorithm for FNLP problems using linear ranking function proposed [14], which is useful for sensitivity analysis on FNLP. Nasseri et al. [15] considered a kind of linear programming which includes the triangular fuzzy numbers in its parameters and proposed a revised simplex

algorithm for an extended linear programming problem which is equivalent to the original fuzzy linear programming problem. Ebrahimnejad [16] obtained some new results in FLP and gave a new method to obtain an initial fuzzy basic feasible solution for solving FLP problems. Nasseri and Alizadeh [17] thought that finding a basic feasible solution (BFS) is not straightforward and some works to make the simplex algorithm start might be needed, so they proposed a penalty method to solve FVLP problems in which the BFS is not readily available. Ebrahimnejad et al. [18] proposed a new method for bounded linear programming with fuzzy cost coefficients called the bounded fuzzy primal simplex algorithm. Some scholars [19–25] studied duality in FLP. Mahdavi-Amiri, Nasseri and Ebrahimnejad presented the dual simplex algorithm for solving FNLP problem [19, 20] and the dual simplex algorithm for FVLP problem [21]. Ebrahimnejad et al. [22] introduced another efficient method, primal-dual simplex algorithm, to obtain a fuzzy solution of FVLP problem. Ebrahimnejad and Nasseri [23] studied dual simplex algorithm for bounded linear programming with fuzzy numbers. Ebrahimnejad and Nasseri [24] defined a new dual problem for the linear programming problem with trapezoidal fuzzy variables as a linear programming problem with trapezoidal fuzzy variables and deduced the duality results such as weak duality, strong duality, and complementary slackness theorems. Nasseri et al. [25] established the dual of a linear programming problem with symmetric trapezoidal fuzzy numbers, where the coefficients and variables are symmetric trapezoidal fuzzy numbers, and developed some duality results for the fuzzy primal and fuzzy dual problems. Ebrahimnejad and Nasseri [26] used the complementary slackness to solve FNLP and FVLP problems without the need of a simplex tableau. Sigarpich et al. [27] gave a new method for solving the degeneracy in linear programming problems with fuzzy variables by a definite linear function for ranking symmetric triangular fuzzy numbers. Chanas [28] presented the possibility of the identification of a complete fuzzy decision in fuzzy linear programming by use of the parametric programming technique.

Sensitivity analysis is a basic tool for studying perturbations in optimization problems. There is considerable research on sensitivity analysis for some models of operations research and management science such as linear programming and investment analysis. So, many scholars studied the sensitivity analysis for FVLP [29–31] and FNLP [32]. They considered the following variations: change in the cost vector, change in the right-hand side vector, change in the constraint matrix, addition of a new activity (trapezoidal fuzzy variable), and addition of a new constraint.

In a word, existing methods solving FNLP problems are mainly using the concept of comparison of fuzzy numbers and linear ranking function to change the fuzzy number into crisp number, using simplex method and its revised method to solve these FNLP problems. Because the time complexity of simplex methods [10, 11] or revised simplex algorithm [14] is exponential, its iterations will increase rapidly with increasing the number of decision-making variables and constraint conditions. This paper wants to propose a new interior point method to improve the efficiency of solving large-scale FNLP problems, which will revise the feasible direction and step size as well as terminate condition in common interior point method by using trapezoidal fuzzy numbers, linear ranking function, fuzzy vector, and their operations.

This paper is organized as follows. We demonstrate some preliminaries of fuzzy set theory and the concept of ranking functions in Section 2. The simplex method for solving FNLP will be reviewed in Section 3. A new interior point method for solving FNLP will be proposed in Section 4. Example study and algorithm analysis will be shown in Section 5. Finally, we will allocate the Section 6 to conclusions.

2. Preliminaries

In this section, we review some necessary concepts of fuzzy set theory and the ranking function and then present some definition about fuzzy vectors.

Definition 1 (see [5, 19]). A convex fuzzy set $\widetilde{A}$ on R is a *fuzzy number* if the following conditions hold.

(i) Its membership function is piecewise continuous.

(ii) There exist three intervals $[a, b]$, $[b, c]$, and $[c, d]$ such that $\mu_{\widetilde{A}}$ is increasing on $[a, b]$, equal to 1 on $[b, c]$, decreasing on $[c, d]$, and equal to 0 elsewhere.

Definition 2 (see [5, 19]). Let $\widetilde{A} = (a^L, a^U, \alpha, \beta)$ denote the *trapezoidal fuzzy number*, where $(a^L - \alpha, a^U + \beta)$ is the support of $\widetilde{A}$ and $[a^L, a^U]$ is its core.

Remark 3. We denote the set of all trapezoidal fuzzy numbers by $F(R)$.

Theorem 4 (see [5, 19]). *If $\widetilde{a} = (a^L, a^U, \alpha, \beta)$ and $\widetilde{b} = (b^L, b^U, \gamma, \theta)$ are two trapezoidal fuzzy numbers, then*

(i) *for any $x > 0$, $x \in R$, $x\widetilde{a} = (xa^L, xa^U, x\alpha, x\beta)$,*

(ii) *for any $x < 0$, $x \in R$, $x\widetilde{a} = (xa^U, xa^L, -x\beta, -x\alpha)$,*

(iii) $\widetilde{a} + \widetilde{b} = (a^L + b^L, a^U + b^U, \alpha + \gamma, \beta + \theta)$.

Definition 5 (see [34]). The function $\mathscr{R} : F(R) \to R$ which maps each fuzzy number into the real line is called a *ranking function*, where a natural order exists.

Theorem 6 (see [34]). *If $\widetilde{a}, \widetilde{b} \in F(R)$, then*

(i) *$\widetilde{a} \geq_{\mathscr{R}} \widetilde{b}$ if and only if $\mathscr{R}(\widetilde{a}) \geq \mathscr{R}(\widetilde{b})$;*

(ii) *$\widetilde{a} >_{\mathscr{R}} \widetilde{b}$ if and only if $\mathscr{R}(\widetilde{a}) > \mathscr{R}(\widetilde{b})$;*

(iii) $\tilde{a} =_{\mathcal{R}} \tilde{b}$ *if and only if* $\mathcal{R}(\tilde{a}) = \mathcal{R}(\tilde{b})$;

(iv) $\tilde{a} \leq_{\mathcal{R}} \tilde{b}$ *if and only if* $\mathcal{R}(\tilde{a}) \leq \mathcal{R}(\tilde{b})$.

Definition 7 (see [34]). If a ranking function $\mathcal{R}$ such that

$$\mathcal{R}\left(k\tilde{a} + \tilde{b}\right) = k\mathcal{R}(\tilde{a}) + \mathcal{R}\left(\tilde{b}\right) \tag{1}$$

for any $\tilde{a}, \tilde{b} \in F(R)$, $k \in R$, then $\mathcal{R}$ is a *linear ranking function* on $F(R)$.

Theorem 8 (see [33]). *The forms of linear ranking functions on $F(R)$ are often given as follows:*

(i) $\mathcal{R}(\tilde{a}) = c_L a^L + c_U a^U + c_\alpha \alpha + c_\beta \beta$, *where* $\tilde{a} = (a^L, a^U, \alpha, \beta)$ *and* c_L, c_U, c_α, c_β *are constants, at least one of which is nonzero;*

(ii) $\mathcal{R}(\tilde{a}) = (1/2)\int_0^1 (\inf \tilde{a}_\lambda + \sup \tilde{a}_\lambda) d\lambda$, *that is, reduced to*

$$\mathcal{R}(\tilde{a}) = \frac{a^L + a^U}{2} + \frac{1}{4}(\beta - \alpha). \tag{2}$$

Corollary 9 (see [34]). *For any trapezoidal fuzzy number $\tilde{a}$, the relation $\tilde{a} \geq \tilde{0}$ holds if there exist $\varepsilon \geq 0$ and $\alpha \geq 0$ such that $\tilde{a} \geq_{\mathcal{R}} (-\varepsilon, \varepsilon, \alpha, \alpha)$. One realizes that $\mathcal{R}(-\varepsilon, \varepsilon, \alpha, \alpha) = 0$ (one also consider that $\tilde{a} =_{\mathcal{R}} \tilde{0}$ if and only if $\mathcal{R}(\tilde{a}) = 0$). Thus, without loss of generality, throughout the paper one lets $\tilde{0} = (0, 0, 0, 0)$ as the zero trapezoidal fuzzy number.*

Corollary 10 (see [34]). *For any two trapezoidal fuzzy numbers $\tilde{a} = (a^L, a^U, \alpha, \beta)$ and $\tilde{b} = (b^L, b^U, \gamma, \theta)$, $\tilde{a} \geq_{\mathcal{R}} \tilde{b}$ if and only if $a^L + a^U + (1/2)(\beta - \alpha) \geq b^L + b^U + (1/2)(\theta - \gamma)$.*

Definition 11. A *fuzzy vector* of n dimension on $F(R)$ is an n-tuple on $F(R)$: $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n)$, where the fuzzy number $\tilde{c}_i$ is called the ith component of it, $1 \leq i \leq n$.

Definition 12. Let $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n)$ and $\tilde{d} = (\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_n)$ be two fuzzy vectors whose sum is defined as

$$\tilde{c} + \tilde{d} = \left(\tilde{c}_1 + \tilde{d}_1, \tilde{c}_2 + \tilde{d}_2, \ldots, \tilde{c}_n + \tilde{d}_n\right). \tag{3}$$

Remark 13. It is quite easy to get the following rules:

(i) commutativity: $\tilde{c} + \tilde{d} = \tilde{d} + \tilde{c}$;

(ii) associativity: $(\tilde{c} + \tilde{d}) + \tilde{e} = \tilde{c} + (\tilde{d} + \tilde{e})$;

(iii) neutral Element: $\tilde{0} + \tilde{c} = \tilde{c} + \tilde{0} = \tilde{c}$.

Definition 14. Let $a \in R$, $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n)$ be a fuzzy vector; scalar multiplication of $\tilde{c}$ by a is defined as

$$a\tilde{c} = \left(a\tilde{c}_1, a\tilde{c}_2, \ldots, a\tilde{c}_n\right). \tag{4}$$

Remark 15. It is quite easy to get the following rules:

(i) distributivity over fuzzy vectors: $a(\tilde{c} + \tilde{d}) = a\tilde{c} + a\tilde{d}$;

(ii) distributivity over number: $(a + b)\tilde{c} =_{\mathcal{R}} a\tilde{c} + b\tilde{c}$.

Definition 16. Let $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n)$; ranking function operation of $\tilde{c}$ is defined as

$$\mathcal{R}(\tilde{c}) = \left(\mathcal{R}(\tilde{c}_1), \mathcal{R}(\tilde{c}_2), \ldots, \mathcal{R}(\tilde{c}_n)\right). \tag{5}$$

Remark 17. It is quite easy to obtain

$$\mathcal{R}\left(k\tilde{c} + \tilde{d}\right) = k\mathcal{R}(\tilde{c}) + \mathcal{R}\left(\tilde{d}\right), \tag{6}$$

where $\tilde{c}, \tilde{d} \in (F(R))^n$ and $k \in R$.

Definition 18. Let $a = (a_1, a_2, \ldots, a_n)$, $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n)$; vector multiplication of $\tilde{c}$ by a is defined as

$$a\tilde{c}^T = (a_1, a_2, \ldots, a_n)\left(\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n\right)^T = \sum_{i=1}^{n} a_i \tilde{c}_i. \tag{7}$$

3. Simplex Method for Solving Fuzzy Number Linear Programming

In this section, we recall the definition of FNLP and the fuzzy primal simplex algorithm to FNLP.

Definition 19 (see [11, 19]). An *FNLP problem* is defined as follows

$$\begin{aligned} \max \quad & \tilde{z} =_{\mathcal{R}} \tilde{c}^T x \\ \text{s.t.} \quad & Ax \leq b, \\ & x \geq 0, \end{aligned} \tag{8}$$

where $b \in R^m$, $x \in R^n$, $A \in R^{m \times n}$, $\tilde{c} \in (F(R))^n$, and $\mathcal{R}$ is a linear ranking function.

Remark 20. There is another equivalent form of (8) as follows:

$$\begin{aligned} \max \quad & \tilde{z} =_{\mathcal{R}} \tilde{c}^T x \\ \text{s.t.} \quad & Ax \leq b, \end{aligned} \tag{9}$$

where $A \in R^{(m+n) \times n}$ and the other symbols are the same as in (8).

Algorithm 21 (see [11]). The fuzzy primal simplex algorithm.

Assumption. A basic feasible solution with basis B and the corresponding simplex tableau is at hand.

(i) The basic feasible solution is given by $x_B = B^{-1}b$ and $x_N = 0$. The fuzzy objective value is $\tilde{z} =_{\mathcal{R}} (\tilde{c}_B x_B)$.

(ii) Let $\tilde{z}_k - \tilde{c}_k = \min_j\{\tilde{z}_j - \tilde{c}_j\}$, $j = 1, \ldots, n$, $j \neq B_i$, $i = 1, \ldots m$. If $\tilde{z}_k - \tilde{c}_k \geq_{\mathscr{R}} 0$; then stop. The current solution is optimal; else go to step (iii).

(iii) If $y_k \leq 0$, then stop; the problem is unbounded. Otherwise determine the index of the variable x_{B_r} leaving the basis as follows:

$$\frac{b_r}{y_{rk}} = \min_{1 \leq i \leq m} \left\{ \frac{b_i}{y_{ik}} \mid y_{ik} > 0 \right\}. \tag{10}$$

(iv) Pivot on y_{rk} and update the simplex tableau. Go to step (ii).

Remark 22. The idea of this algorithm is to start from a vertex; each step of its iteration is moving to a better vertex until the optimal solution is found or infeasible solution is proved.

In Algorithm 21, searching adjacent vertexes is just only along the edge, and each iteration calculation is very small. But simplex method should go a long way to reach the optimal solution along the feasible region boundary through almost each vertex. For the feasible region of the large-scale application, a problem may have a lot of vertexes, this "boundary method" will encounter the problem of huge calculation generating by iteration. In order to reduce the iterations, alternative method is moving along the "short path" in internal of the feasible region. However, the usual interior point method always needs to consider all the feasible directions in each step of iteration in order to find the best one.

Fortunately, we know that Karmarkar's interior point method [35] is not searching forward along the surface of the feasible region but directly approaching to the optimal solution along search directions in the internal of the feasible region. But this method cannot be used directly to solve FNLP problems. So, in the next section, we will propose a revised interior point method, which can be used directly to solve FNLP problem.

4. A Revised Interior Point Method for Solving Fuzzy Number Linear Programming

In this section, we propose a revised interior-point method to solve FNLP problem.

4.1. The Idea of Revised Interior Point Method. The basic idea of revised interior point is first starting from an interior point x^0 and getting a subsequent point to increase objective function value along the feasible direction, then starting from this interior point, and getting a new subsequent point to make objective function value increase along other feasible direction. Repeating the previous steps will produce a sequence of point $\{x^k\}$ which is subject to $\tilde{c}^T x^{k+1} \geq_{\mathscr{R}} \tilde{c}^T x^k$, where $\tilde{c}^T x^{k+1} \geq_{\mathscr{R}} \tilde{c}^T x^k$ are the operations of ranking function and fuzzy vector. When the iteration is subjected to termination criterion, it will stop. The key of this method is choosing a feasible direction to improve objective function value.

4.2. The Derivation of Computational Formula. Combined with the slack variable v, the problem (9) is converted into the following form:

$$\begin{aligned} \max \quad & \tilde{z} =_{\mathscr{R}} \tilde{c}^T x \\ \text{s.t.} \quad & Ax + v = b, \\ & v \geq 0. \end{aligned} \tag{11}$$

In the kth iteration, define $v^k \geq 0$, $v^k \in R^m$, subject to $v^k = b - Ax^k$. Then, define the diagonal matrix $D_k = \operatorname{diag}(1/V_1^k, 1/V_2^k, \ldots, 1/V_m^k)$.

Let $w = D_k v$, problem (11) is changed as follows:

$$\begin{aligned} \max \quad & \tilde{z} =_{\mathscr{R}} \tilde{c}^T x \\ \text{s.t.} \quad & Ax + D_k^{-1} w = b, \\ & w \geq 0. \end{aligned} \tag{12}$$

Choose the search direction $d = [d_x \;\; d_w]^T$; then it must be one solution of the following equation:

$$D_k A d_x + d_w = 0, \tag{13}$$

$$\begin{aligned} A^T D_k \left(D_k A d_x + d_w\right) &= 0, \\ d_x &= -\left(A^T D_k^2 A\right)^{-1} A^T D_k d_w. \end{aligned} \tag{14}$$

Then,

$$\begin{aligned} \tilde{c}^T d_x &=_{\mathscr{R}} \tilde{c}^T \left[-\left(A^T D_k^2 A\right)^{-1} A^T D_k d_w\right] \\ &=_{\mathscr{R}} -\left[D_k A \left(A^T D_k^2 A\right)^{-1} \tilde{c}\right]^T d_w. \end{aligned} \tag{15}$$

To maximize $\tilde{c}^T d_x$, that is to say, maximize $\mathscr{R}(\tilde{c}^T d_x)$, combined with (6) and (7), then

$$d_w = -\mathscr{R}\left(D_k A \left(A^T D_k^2 A\right)^{-1} \tilde{c}\right) = -D_k A \left(A^T D_k^2 A\right)^{-1} \mathscr{R}(\tilde{c}). \tag{16}$$

From (13) and (16), we get

$$d_x = \left(A^T D_k^2 A\right)^{-1} \mathscr{R}(\tilde{c}). \tag{17}$$

From $w = D_k v$,

$$d_v = D_k^{-1} d_w = -A\left(A^T D_k^2 A\right)^{-1} \mathscr{R}(\tilde{c}) = -A d_x. \tag{18}$$

After getting the search direction d_x, we need to determine the step size. Let

$$x^{k+1} = x^k + \lambda d_x, \tag{19}$$

where the step size λ should guarantee that point x^{k+1} is in the feasible region; it should satisfy the following inequalities:

$$\begin{gathered} A\left(x^k + \lambda d_x\right) < b, \\ \lambda A d_x < b - A x^k, \\ -\lambda d_v < v^k. \end{gathered} \tag{20}$$

Let

$$\alpha = \min\left\{\frac{V_i^k}{-(d_v)_i} \mid (d_v)_i < 0, i \in (1, 2, \ldots, m)\right\}. \tag{21}$$

Take

$$\lambda = \gamma\alpha, \tag{22}$$

where $\gamma \in (0, 1)$. Then, we can get x^{k+1} from x^k along the direction d_x, where $\tilde{c}^T x^{k+1} >_{\mathscr{R}} \tilde{c}^T x^k$.

4.3. Steps of the Revised Interior Point Algorithm. From the idea of revised interior point method and the derivation of calculation formula, steps of the revised interior point algorithm to solve model (9) are shown as follows.

Step 1. Give an initial interior point x^0, a safety factor parameter $\gamma \in (0, 1)$, accuracy parameter $\epsilon > 0$, and iteration $k = 0$.

Step 2. Compute

$$V^k = b - A x^k. \tag{23}$$

Step 3. Set the diagonal matrix

$$D_k = \operatorname{diag}\left(\frac{1}{V_1^k}, \frac{1}{V_2^k}, \ldots, \frac{1}{V_m^k}\right). \tag{24}$$

Step 4. Using the vector multiplication of fuzzy vectors (7), compute

$$\tilde{d}_x = \left(A^T D_k^2 A\right)^{-1} \cdot \tilde{c}; \tag{25}$$

then combined with (6), (5) and (2), the feasible direction is

$$\mathscr{R}\left(\tilde{d}_x\right) = \left(A^T D_k^2 A\right)^{-1} \mathscr{R}(\tilde{c}). \tag{26}$$

Step 5. Compute the vector

$$d_V = -A \cdot \mathscr{R}\left(\tilde{d}_x\right). \tag{27}$$

Step 6. Let

$$\lambda = \gamma \cdot \min\left\{\frac{V_i^k}{-(d_V)_i} \mid (d_V)_i < 0, i \in (1, 2, \ldots, m)\right\}. \tag{28}$$

Step 7. Compute the next point:

$$x^{k+1} = x^k + \lambda \cdot \mathscr{R}\left(\tilde{d}_x\right). \tag{29}$$

Step 8. Using the vector multiplication of fuzzy vectors (7) and ranking function (2), compare $\mathscr{R}(\tilde{c}^T x^{k+1} - \tilde{c}^T x^k)/\mathscr{R}(\tilde{c}^T x^k)$ with ϵ. If $\mathscr{R}(\tilde{c}^T x^{k+1} - \tilde{c}^T x^k)/\mathscr{R}(\tilde{c}^T x^k) < \epsilon$, then algorithm terminates and x^{k+1} is the optimal solution; else $k := k + 1$, and go to Step 2.

4.4. Choice of the Initial Interior Point. Generally, set $x^0 = (\|b\|/\|R(A\tilde{c})\|) \cdot R(\tilde{c})$ to be the initial interior point. And if $V^0 = b - Ax^0 > 0$, then go to Step 2 in Section 4.3; otherwise formulate a new fuzzy number linear programming as follows:

$$\begin{aligned} \max \quad & \tilde{z} =_{\mathscr{R}} \tilde{c}^T x + \widetilde{M} x_a \\ \text{s.t.} \quad & Ax - x_a e \le b, \end{aligned} \tag{30}$$

where $\widetilde{M} >_{\mathscr{R}} \tilde{0}$ is a big fuzzy number, $e = (1, 1, \ldots, 1)^T$ and x_a is artificial variable.

And if ${x_a}^0 > |\min\{({v_i}^0) \mid i = 1, \ldots, m\}|$, then (x^0, x_a^0) must be the interior point of (30). Now, the problem (30) can be solved by the revised interior point method.

If $x_a^k < 0$ in the kth iteration, stop solving the problem (30) and set x^k to be the initial interior point of the problem (9).

If there is the optimal solution of problem (30), and $x_a^k > 0$, then the problem (9) is not feasible.

5. Algorithm Analysis and Example Study

In this section, first we analyze the algorithm. Then, an example in the practical production is given. At last, we analyze some factors influencing the results of this method through the given example.

5.1. Algorithm Analysis. The time complexity of simplex methods [10, 11] or revised simplex algorithm [14] is exponential. Generally speaking, the simplex method has the following shortcomings.

(i) Iterations are rising rapidly as the number of planning variables and constraints increasing.

(ii) The simplex method is terminated in optimal basis of original and dual programs. Although it has reached optimal solution in the degenerate case, it often needs to iterate the basis many times in order to prove that it is optimal.

As we know, interior point methods (IPMs) are the most effective methods for solving a large-scale linear optimization problem. Since the creative work of Karmarkar [35], many researchers have proposed and analyzed various IPMs for LP and a large amount of results have been reported. And Karmarkar's IPM has a polynomial time complexity and it approaches directly the optimal solution from the feasible region through the internal. Because the iteration of

Table 1: The relationship among product demand, production capacity, and pure profit.

Product	Daily demand	Manual system			Machine system		
		Production capacity	Pure profit		Production capacity	Pure profit	
			Shift 1	Shift 2		Shift 1	Shift 2
1	5	3	(8, 10, 2, 6)	(10, 12, 1, 17)	4	(6, 8, 1, 5)	(9, 11, 1, 5)
2	10	5	(3, 5, 1, 5)	(4, 6, 2, 6)	7.5	(2, 4, 2, 6)	(4, 7, 1, 3)

Table 2: The interior point value and the corresponding objective function value of each iteration.

	Interior point value	Objective function value
1	$[1.7867\ \ 2.9185\ \ 1.8563\ \ 0.0866\ \ 0.2020\ \ 0.0878\ \ 0.7999\ \ 7.24]^T$	(81.9560 119.1517 17.6510 98.1049)
2	$[1.8604\ \ 2.9674\ \ 1.8674\ \ 0.0432\ \ 0.1220\ \ 0.0463\ \ 0.6698\ \ 7.4085]^T$	(82.4542 119.8362 17.5584 98.2900)
3	$[1.9628\ \ 2.9964\ \ 1.7165\ \ 0.0023\ \ 0.0369\ \ 0.0023\ \ 0.7878\ \ 7.4918]^T$	(82.6160 120.0960 17.7497 98.7180)
4	$[1.9975\ \ 2.9984\ \ 1.6701\ \ 0.0019\ \ 0.0022\ \ 0.0019\ \ 0.8322\ \ 7.4949]^T$	(82.6559 120.1490 17.8308 98.8179)
5	$[1.9980\ \ 2.9995\ \ 1.6694\ \ 0.0007\ \ 0.0017\ \ 0.0008\ \ 0.8314\ \ 7.4984]^T$	(82.6633 120.1615 17.8300 98.8263)

Karmarkar's interior point algorithm is less changing as the number of planning variables and constraints increases, it is more outstanding to solve the large-scale FNLP problem by using the revised interior point method proposed in this paper.

5.2. Example Study

Question. Suppose a factory produces two products representing with 1 and 2; they are made by manual system and machine system in two shift works a day. The detailed relationship between production capacity and pure profit is shown in Table 1. The daily demand of users for the products 1 and 2 is 5 and 10, respectively. So, how to arrange production to get the maximum pure profit and meet users' requirements?

Remark 23. In Table 1, the measure unit of daily demand is ton, the measure unit of production capacity is tons per shift, and the measure unit of pure profit is thousand dollars per ton.

Remark 24. In Table 1, the pure profit of each product in each shift is fuzzy. If its pure profit is about 10 after investigation, then it may be presented as a trapezoidal fuzzy number, that is (8, 10, 2, 6).

Solution. (i) Let

x_1 : the output of product 1 in shift 1 produced by manual system;

x_2 : the output of product 1 in shift 2 produced by manual system;

x_3 : the output of product 2 in shift 1 produced by manual system;

x_4 : the output of product 2 in shift 2 produced by manual system;

x_5 : the output of product 1 in shift 1 produced by machine system;

x_6 : the output of product 1 in shift 2 produced by machine system;

x_7 : the output of product 2 in shift 1 produced by machine system;

x_8 : the output of product 2 in shift 2 produced by machine system.

(ii) Now an FNLP model is established as follows:

$$\begin{aligned}
\max\quad & \tilde{z} =_{\mathscr{R}} (8,10,2,6)\,x_1 + (10,12,1,17)\,x_2 + (3,5,1,5)\,x_3 \\
& \quad + (4,6,2,6)\,x_4 + (6,8,1,5)\,x_5 + (9,11,1,5)\,x_6 \\
& \quad + (2,4,2,6)\,x_7 + (4,7,1,3)\,x_8 \\
\text{s.t.}\quad & x_1 + x_2 + x_5 + x_6 \le 5 \\
& x_3 + x_4 + x_7 + x_8 \le 10 \\
& x_1 \le 3 \\
& x_2 \le 3 \\
& x_3 \le 5 \\
& x_4 \le 5 \\
& x_5 \le 4 \\
& x_6 \le 4 \\
& x_7 \le 7.5 \\
& x_8 \le 7.5 \\
& 5x_1 + 3x_3 \le 15 \\
& 5x_2 + 3x_4 \le 15 \\
& 15x_5 + 8x_7 \le 60 \\
& 15x_6 + 8x_8 \le 60 \\
& x_i \ge 0, \quad i = 1, 2, \ldots, 8.
\end{aligned} \tag{31}$$

Table 3: The influence of the safety factor parameter γ on iterations ($\epsilon = 0.1$, $x^0 = [1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 5]^T$).

Parameter γ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
Iterations K	75	40	27	20	16	13	11	10	9	9

Table 4: The influence of the accuracy parameter ϵ on iterations ($\gamma = 0.5$, $x^0 = [1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 5]^T$).

Parameter ϵ	0.9	0.5	0.1	0.01	0.005	0.001
Iterations K	9	10	13	16	17	19

Remark 25. These inequalities $5x_1 + 3x_3 \le 15$, $5x_2 + 3x_4 \le 15$, $15x_5 + 8x_7 \le 60$, and $15x_6 + 8x_8 \le 60$ are simplified from $x_1/3 + x_3/5 \le 1$, $x_2/3 + x_4/5 \le 1$, $x_5/4 + x_7/7.5 \le 1$, and $x_6/4 + x_8/7.5 \le 1$, respectively, which is convenient for the following computation.

(iii) Then, solve the FNLP problem (31). If using simple method, it is complicated. So, we adopt the revised interior point method proposed in Section 4.3.

Step 1. Given $\epsilon = 0.1$, $x^0 = [1.7\ \ 2.9\ \ 1.9\ \ 0.1\ \ 0.2\ \ 0.1\ \ 0.7\ \ 7.1]^T$ and $\gamma = 0.95$, $k = 0$.

Step 2. Compute $V^k = b - Ax^k$, $k = 0$, then

$$V^0 = [0.1\ \ 0.2\ \ 1.3\ \ 0.1\ \ 3.1\ \ 4.9\ \ 3.8\ \ 3.9\ \ 74.3\ \ 67.9\ \ 0.8\ \ 0.2\ \ 51.4\ \ 1.7\ \ 1.7\ \ 2.9\ \ 1.9\ \ 0.1\ \ 0.2\ \ 0.1\ \ 0.7\ \ 7.1]^T. \tag{32}$$

Step 3. Set the diagonal matrix $D_k = \mathrm{diag}(1/V_1^k, 1/V_2^k, \ldots, 1/V_m^k)$, $k = 0$, then

$$D_0 = \mathrm{diag}\,[10\ \ 5\ \ 0.7692\ \ 10\ \ 0.3226\ \ 0.2041\ \ 0.2632\ \ 0.2564\ \ 0.0135\ \ 0.0147\ \ 1.25\ \ 5\ \ 0.0195\ \ 0.5882\ \ 0.5882\ \ 0.3448\ \ 0.5263\ \ 10\ \ 5\ \ 10\ \ 1.4286\ \ 0.1408]. \tag{33}$$

Step 4. Compute $\tilde{d}_x = (A^T D_k^2 A)^{-1} \cdot \tilde{c}$, $k = 0$, then

$$\tilde{d}_x = \begin{bmatrix} (0.9980 & 1.3532 & 0.2363 & 1.1384) \\ (0.0016 & 0.0021 & 0.0004 & 0.0018) \\ (-2.4963 & -1.8227 & 2.1001 & 0.4359) \\ (-0.0158 & -0.0116 & 0.0133 & 0.0028) \\ (-0.1704 & -0.1244 & 0.1433 & 0.0297) \\ (-1.1411 & -0.8332 & 0.9600 & 0.1992) \\ (-2.0983 & -1.5321 & 1.7652 & 0.3664) \\ (3.4931 & 4.7841 & 0.8353 & 4.0247) \end{bmatrix}. \tag{34}$$

Then combined with (5) and (2), the feasible direction is

$$\mathscr{R}(\tilde{d}_x) = [1.40113\ \ 0.0022\ \ -2.57555\ \ -0.01633\ \ -0.1758\ \ -1.17735\ \ -2.1649\ \ 4.93595]^T. \tag{35}$$

Step 5. Compute the vector $d_V = -A \cdot \mathscr{R}(\tilde{d}_x)$, $k = 0$; then

$$d_V = [-0.0726\ \ -0.1531\ \ -0.0726\ \ -0.0155\ \ 0.0366\ \ 0.0112\ \ -0.0017\ \ 0.0102\ \ -0.0837\ \ -0.1172\ \ -0.2532\ \ -0.0439\ \ -0.6945\ \ -0.7841\ \ 0.0726\ \ 0.0155\ \ -0.0366\ \ -0.0112\ \ 0.0017\ \ -0.0102\ \ 0.0837\ \ 0.1172]^T. \tag{36}$$

Step 6. Let $\lambda = \gamma \cdot \min\{V_i^k/[-(d_V)_i] \mid (d_V)_i < 0, i \in (1, 2, \ldots, m)\}$, $k = 0$; then step size $\lambda = 1.1944$.

Step 7. Compute $x^{k+1} = x^k + \lambda \cdot \mathscr{R}(\tilde{d}_x)$, $k = 0$; then

$$x^1 = [1.7867\ \ 2.9185\ \ 1.8563\ \ 0.0866\ \ 0.2020\ \ 0.0878\ \ 0.7999\ \ 7.24]^T. \tag{37}$$

Step 8. Compute $\mathscr{R}(\tilde{c}^T x^{k+1} - \tilde{c}^T x^k)/\mathscr{R}(\tilde{c}^T x^k) = 4.2345 > \epsilon = 0.1$; then $k = 0 + 1 = 1$ and go to Step 2. Repeat the similar calculation until $\mathscr{R}(\tilde{c}^T x^{k+1} - \tilde{c}^T x^k)/\mathscr{R}(\tilde{c}^T x^k) < \epsilon = 0.1$, and get the results.

Above all, the number of iteration is 5 and the results are listed in Table 2.

The optimal solution is $[1.9980\ \ 2.9995\ \ 1.6694\ \ 0.0007\ \ 0.0017\ \ 0.0008\ \ 0.8314\ \ 7.4984]^T$ and the optimal fuzzy value of the objective function is $(82.6633\ \ 120.1615\ \ 17.8300\ \ 98.8263)$. Then, the maximum pure profit is about $\mathscr{R}(82.6633\ \ 120.1615\ \ 17.8300\ \ 98.8263) = 121.661475$ thousand dollars.

5.3. Analysis of Factors Influencing This Method Results. Factors influencing the results of this method are mainly safety factor parameter γ, accuracy parameter ϵ, and initial interior point x^0. Take model (31) as an example.

(i) Table 3 focuses on the safety factor parameter γ, where the values of accuracy parameter ϵ and initial

Table 5: The influence of the initial interior point x^0 on iterations ($\gamma = 0.5, \epsilon = 0.01$).

Initial interior point	Iterations K	Initial interior point	Iterations K
$[0.1\ \ 0.1\ \ 0.1\ \ 0.1\ \ 0.1\ \ 0.1\ \ 0.1\ \ 0.1]^T$	33	$[1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 4]^T$	16
$[1\ \ 0.5\ \ 0.1\ \ 0.1\ \ 0.1\ \ 0.1\ \ 0.1\ \ 0.1]^T$	31	$[1\ \ 1.9\ \ 2\ \ 1\ \ 0.7\ \ 1.1\ \ 1.9\ \ 4]^T$	15
$[1\ \ 1\ \ 0.1\ \ 0.1\ \ 0.1\ \ 0.1\ \ 0.1\ \ 1]^T$	26	$[1.1\ \ 2.5\ \ 2.2\ \ 0.4\ \ 0.5\ \ 0.6\ \ 1.1\ \ 5]^T$	14
$[1\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5]^T$	21	$[1.4\ \ 2.7\ \ 2.2\ \ 0.2\ \ 0.3\ \ 0.3\ \ 0.8\ \ 6.6]^T$	13
$[1\ \ 1\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5]^T$	20	$[1.5\ \ 2.7\ \ 2\ \ 0.1\ \ 0.3\ \ 0.2\ \ 0.8\ \ 6.7]^T$	13
$[1\ \ 1\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5\ \ 1\ \ 1]^T$	19	$[1.6\ \ 2.8\ \ 2.1\ \ 0.1\ \ 0.2\ \ 0.2\ \ 0.6\ \ 7]^T$	12
$[1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1]^T$	18	$[1.6\ \ 2.8\ \ 2\ \ 0.1\ \ 0.2\ \ 0.2\ \ 0.7\ \ 6.9]^T$	12
$[1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 2]^T$	17	$[1.7\ \ 2.9\ \ 1.9\ \ 0.1\ \ 0.2\ \ 0.1\ \ 0.7\ \ 7.1]^T$	11
$[1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 3]^T$	16		

Table 6: The influence of the accuracy parameter ϵ and safety factor parameter γ on iterations ($x^0 = [1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 5]^T$).

Parameter γ		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
Iterations K	$\epsilon = 0.1$	52	29	20	16	13	11	9	8	8	7
	$\epsilon = 0.01$	75	40	27	20	16	13	11	10	9	9
	$\epsilon = 0.001$	97	50	33	25	19	16	13	12	11	11

interior point x^0 are fixed. All test problems show that the selection of a safety factor parameter plays a significant role in the fast convergence. We can see that the algorithm converges to the near-optimal solutions quickly as the safety factor parameter is increasing.

(ii) Table 4 focuses on the accuracy parameter ϵ, where the values of safety factor parameter γ and initial interior point x^0 are fixed. All test problems show that the selection of accuracy parameter plays a critical role in the fast convergence. We can see that the algorithm converges slowly to the near-optimal solutions as the safety factor parameter is decreasing. Even so, the final result is more accurate. The value of ϵ generally depends on the actual need. Therefore, this method can adjust precision to meet the requirement according to the actual need.

(iii) Table 5 focuses on the initial interior point x^0, where the values of safety factor parameter γ and accuracy parameter ϵ are fixed. All test problems show that the selection of an initial interior solution plays a significant role in the fast convergence. We can see that the algorithm converges to the near-optimal solutions quickly as the initial interior point is more and more close to the optimal solution. That is to say, the iteration is more and more small and tends to be a constant.

(iv) Table 6 focuses on the safety factor parameter γ and accuracy parameter ϵ, where the values of initial interior point x^0 is fixed. We can see that the iterations are smaller as the values of the accuracy parameter and safety factor parameter are increasing; the influence of safety factor parameter is more obvious than accuracy parameter to the iterations.

6. Conclusions

A new interior point method is presented to solve FNLP problems using linear ranking function in this paper. Compared with simplex method or revised simplex algorithm, this method is more outstanding in solving the large scale of the FNLP problem, for it has a polynomial time complexity. And some factors influencing the results of this method are analyzed. The result shows that proper safety factor parameter, accuracy parameter, and initial interior point of this method may reduce iterations and they can be selected easily according to the actual needs. Although a general method to select the initial point has been given in this paper, it is not feasible in some cases. For example, under the condition $V^0 = b - Ax^0 > 0$, the matrix $A^T D_k^2 A$ may be singular and not reversible, then the search direction cannot be obtained, thus the algorithm cannot be performed. Therefore, future work may put forward an applicable broader method for the revised initial interior point.

Acknowledgments

The authors thank the anonymous referees for their suggestions and comments to improve an earlier version of this paper. The authors are also grateful for the financial support by Scientific Research Fund of Sichuan Provincial Education Department (no. 11ZA024) and Science Foundation of Southwest Petroleum University of China (no. 2012XJZ031).

References

[1] S. Lertworasirikul, S.-C. Fang, J. A. Joines, and H. L. W. Nuttle, "Fuzzy data envelopment analysis (DEA): a possibility approach," *Fuzzy Sets and Systems*, vol. 139, no. 2, pp. 379–394, 2003.

[2] M. Wen and H. Li, "Fuzzy data envelopment analysis (DEA): model and ranking method," *Journal of Computational and Applied Mathematics*, vol. 223, no. 2, pp. 872–878, 2009.

[3] C. Kahraman, T. Ertay, and G. Büyüközkan, "A fuzzy optimization model for QFD planning process using analytic network approach," *European Journal of Operational Research*, vol. 171, no. 2, pp. 390–411, 2006.

[4] H. Tanaka, T. Okuda, and K. Asai, "On fuzzy-mathematical programming," *Journal of Cybernetics*, vol. 3, no. 4, pp. 37–46, 1973.

[5] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," *Management Science*, vol. 17, pp. B141–B164, 1970.

[6] H. J. Zimmermann, "Fuzzy programming and linear programming with several objective functions," *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 45–55, 1978.

[7] H. Rommelfanger, "Fuzzy linear programming and applications," *European Journal of Operational Research*, vol. 92, no. 3, pp. 512–527, 1996.

[8] H. R. Maleki, M. Tata, and M. Mashinchi, "Linear programming with fuzzy variables," *Fuzzy Sets and Systems*, vol. 109, no. 1, pp. 21–33, 2000.

[9] K. Ganesan and P. Veeramani, "Fuzzy linear programs with trapezoidal fuzzy numbers," *Annals of Operations Research*, vol. 143, no. 1, pp. 305–315, 2006.

[10] S. H. Nasseri, E. Ardil, A. Yazdani, and R. Zaefarian, "Simplex method for solving linear programming problems with fuzzy numbers," *Transactions on Engineering, Computing and Technology*, vol. 10, pp. 284–288, 2005.

[11] N. Mahdavi-Amiri, S. H. Nasseri, and A. Yazdani, "Fuzzy primal simplex algorithms for solving fuzzy linear programming problems," *Iranian Journal of Operations Research*, vol. 1, pp. 68–84, 2009.

[12] H. Nasseri and A. Ebrahimnejad, "A fuzzy primal simplex algorithm and its application for solving flexible linear programming problems," *European Journal of Industrial Engineering*, vol. 4, no. 3, pp. 372–389, 2010.

[13] S. H. Nasseri and B. Khabiri, "Revised fuzzy simplex algorithm for linear programming problems with fuzzy variables using linear ranking functions," *International Journal of Mathematics and Computation*, vol. 6, no. 10, pp. 44–54, 2010.

[14] S. H. Nasseri and B. Khabiri, "A revised simplex algorithm for fuzzy nonlinear linear programming problems using linear ranking functions," *International Journal of Mathematics and Computation*, vol. 8, no. 10, pp. 114–126, 2010.

[15] S. H. Nasseri, H. Attari, and A. Ebrahimnejad, "Revised simplex method and its application for solving fuzzy linear programming problems," *European Journal of Industrial Engineering*, vol. 6, no. 3, pp. 259–280, 2012.

[16] A. Ebrahimnejad, "Some new results in linear programs with trapezoidal fuzzy numbers: finite convergence of the Ganesan and Veeramani's method and a fuzzy revised simplex method," *Applied Mathematical Modelling*, vol. 35, no. 9, pp. 4526–4540, 2011.

[17] S. H. Nasseri and Z. Alizadeh, "Solving linear programming problem with fuzzy right hand sides: a penalty method," *The Journal of Mathematics and Computer Science*, vol. 3, no. 3, pp. 318–328, 2011.

[18] A. Ebrahimnejad, S. H. Nasseri, and S. M. Mansourzadeh, "Bounded primal simplex algorithm for bounded linear programming with fuzzy cost coefficients," *International Journal of Operations Research and Information Systems*, vol. 2, pp. 96–120, 2011.

[19] N. Mahdavi-Amiri and S. H. Nasseri, "Duality in fuzzy number linear programming by use of a certain linear ranking function," *Applied Mathematics and Computation*, vol. 180, no. 1, pp. 206–216, 2006.

[20] S. H. Nasseri and A. Ebrahimnejad, "A fuzzy dual simplex method for a fuzzy number linear programming problem," *Advances in Fuzzy Sets and Systems*, vol. 5, no. 2, pp. 81–95, 2010.

[21] N. Mahdavi-Amiri and S. H. Nasseri, "Duality results and a dual simplex method for linear programming problems with trapezoidal fuzzy variables," *Fuzzy Sets and Systems*, vol. 158, no. 17, pp. 1961–1978, 2007.

[22] A. Ebrahimnejad, S. H. Nasseri, F. H. Lotfi, and M. Soltanifar, "A problems with fuzzy variables," *European Journal of Industrial Engineering*, vol. 4, no. 2, pp. 189–209, 2010.

[23] A. Ebrahimnejad and S. H. Nasseri, "A dual simplex method for bounded linear programmes with fuzzy numbers," *International Journal of Mathematics in Operational Research*, vol. 2, no. 6, pp. 762–779, 2010.

[24] A. Ebrahimnejad and S. H. Nasseri, "A new approach to duality in fuzzy linear programming," *Fuzzy Engineering and Operations Research*, vol. 147, pp. 17–29, 2012.

[25] S. H. Nasseri, A. Ebrahimnejad, and S. Mizuno, "Duality in fuzzy linear programming with symmetric trapezoidal numbers," *Applications and Applied Mathematics*, vol. 5, no. 10, pp. 1467–1482, 2010.

[26] A. Ebrahimnejad and S. H. Nasseri, "Using complementary slackness property to solve linear programming with fuzzy parameters," *Fuzzy Information and Engineering*, vol. 3, pp. 233–245, 2009.

[27] L. A. Sigarpich, T. Allahviranloo, F. Hosseinzadeh, and N. A. Kiani, "Degeneracy in fuzzy linear programming and its application," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 19, no. 6, pp. 999–1012, 2011.

[28] S. Chanas, "The use of parametric programming in fuzzy linear programming," *Fuzzy Sets and Systems*, vol. 11, no. 1–3, pp. 229–241, 1983.

[29] B. Kheirfam and F. Hasani, "Sensitivity analysis for fuzzy linear programming problems with fuzzy variables," *Advanced Modeling and Optimization*, vol. 12, no. 2, pp. 257–272, 2010.

[30] S. H. Nasseri and A. Ebrahimnejad, "Sensitivity analysis on linear programming problems with trapezoidal fuzzy variables," *International Journal of Operations Research and, Information Systems*, vol. 2, pp. 22–39, 2011.

[31] A. Kumar and N. Bhatia, "Sensitivity analysis for fuzzy linear programming problems," in *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, vol. 6743, pp. 103–110, Springer, Berlin, Germany, 2011.

[32] A. Ebrahimnejad, "Sensitivity analysis in fuzzy number linear programming problems," *Mathematical and Computer Modelling*, vol. 53, no. 9-10, pp. 1878–1888, 2011.

[33] R. R. Yager, "A procedure for ordering fuzzy subsets of the unit interval," *Information Sciences*, vol. 24, no. 2, pp. 143–161, 1981.

[34] X. Wang and E. E. Kerre, "Reasonable properties for the ordering of fuzzy quantities. I," *Fuzzy Sets and Systems*, vol. 118, no. 3, pp. 375–385, 2001.

[35] N. K. Karmarkar, "A new polynomial-time algorithm for linear programming," in *Proceedings of the 16th Annual ACM Symposium*, vol. 4, pp. 373–395, 1984.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 974340, 8 pages
http://dx.doi.org/10.1155/2013/974340

Review Article

Basic Developments of Quality Characteristics Monitoring

Shahryar Sorooshian

Business School, Taylors University, 47500 Subang Jaya, Selangor, Malaysia

Correspondence should be addressed to Shahryar Sorooshian; sorooshian@gmail.com

Received 3 February 2013; Revised 27 May 2013; Accepted 27 May 2013

Academic Editor: Mohammad Khodabakhshi

Process control tools are a widely used approach in many operations and production processes. Process control chart ranks as one of the most important theories used in these disciplines. This paper reviewed the bias of quality characteristics monitoring. Specifically, this study tries to provide a comprehensive understanding of theories of process control. The text starts with a theoretical review of statistical process control theories and follows by a technical introduction to developed tools for process control.

1. Introduction

Statistical process control (SPC) is a collection of seven tools which is useful in improving the quality level by decreasing the variability and increasing the stability of the process. The most well-known tool of SPC is the control charts. Control chart is a graphical tool based on the measurement data obtained in the course of time from the process. Based on the nature of the data obtained from the process, two broad categories of control charts existed; namely, variable and attribute control charts. If the quality characteristics of the product items could be measured as a numerical scale such as weight and height, variable control chart is appropriate. On the other hand, if the quality characteristics could not be measured in numerical scale such as color and softness, attribute control chart could be utilized. By comparing these two types of control charts, we can conclude that, firstly, variable control charts need a smaller sample size than attribute control charts to construct. Secondly, in variable control charts, assignable cause could be detected sooner than attribute control charts. Thirdly, the cost and time for constructing an attribute control chart are less than a variable control chart, and finally, in attribute control charts, we could monitor more than one quality characteristic at the same time in one control chart. In the following, we technically review the attribute control charts.

2. Attribute Control Chart

Attribute control charts consist of four different control charts. If the production items are categorized into two groups based on the specification limits, the beyond statistical distribution is binomial, and each item is known as confirm or nonconfirm with the specification limits. In this case, proportion of nonconforming items (p chart) and number of nonconforming items (np chart) are appropriate. If the number of defect in a period of production time or in one production item is considered, the beyond statistical distribution is poison, and the suitable control charts are known as c chart and u chart. In the current research, we concentrated on the p chart.

2.1. The Attribute Control Charts Literature. Selecting the proper sample size for constructing the attribute control charts is so important. According to Ryan and Schwertman [1] the adequate sample size should be selected to ensure that the normality assumption is not violated. This difficulty gets more important when the proportion of nonconforming is small, because in this case the sample size should be large enough to have at least one item in the categories of nonconforming items. However, a large sample size is too hard to collect in some situations where the output rate of the

process is small, and also it is time consuming and costly. To overcome this difficulty, Schwertman and Ryan [2] proposed dual a *np* chart which consists of two charts. The first chart has a tighter control limit which requires a smaller sample size, and the second one is a CUSUM chart.

For overcoming the large sample size, Chen [3] also proposed an alternative approach. He suggested two charts which are based on discrete probability integral and arcsine transformations.

Nelson [4] also proposed an alternative approach. He suggested counting the number of conforming items between two consecutive nonconforming items. He assumed that this observation has an exponential distribution; so, by using a transformation to a normal distribution, we could monitor the process.

Several researchers discussed another topic which is the speed of detecting an abnormal shift in proportion of nonconforming items. To detect an abnormal shift like variable control charts, CUSUM chart is a good alternative approach. Reynolds and Stoumbos [5] proposed two different CUSUM charts. One is based on binomial distribution, and the second one is based on a Bernolli variable.

3. Control Charts for Categorical Data

One of the major areas in SPC is monitoring the proportion of the nonconforming units in the production processes. One of the usual control charts for such cases is the *p* chart. Instead of classifying the production units into two groups (conforming and nonconforming), suppose that they have been classified into more than two groups. As an example, they are classified into three groups: minor defect, major defect, and absent of defect. If the produced unit has a minor defect, it can be repaired by low cost and attempts. But if it has a major defect, it can be repaired by lots of cost, or it must be discarded.

If the produced units classify into more than two groups, categorical control charts could be used. In the following, categorical control charts are explained in detail.

3.1. Generalized p Chart. Suppose that Π_1, Π_2, Π_3 are the proportion of the process. This case is comparable with the *p* chart situation. Case I is when the proportions are known before. Case II is when the proportions are unknown before and at first; in phase I when the process is supposed to be in control, they must be estimated.

For monitoring a multinomial distribution, independent samples should be collected during the process. Suppose that X_{i1}, X_{i2}, X_{i3} show the number of observations in category 1, 2, and 3, respectively, in period *i*. Base period is shown with $i = 0$. n_i is the sample size for monitoring period *i*.

First, consider case I where the proportions are known before. A statistical standard approach for solving such a problem is using Pearson's goodness of fit statistic as follow [6]:

$$Y_i^2 = \sum_{j=1}^{3} \frac{\left(X_{ij} - n_i \Pi_j\right)^2}{n_i \Pi_j}, \tag{1}$$

where the process is in the state of in control and Y_i^2 has chi-square distribution with two degree of freedom.

The control chart based on (1) has an upper control limit which is determined with a percentile of chi-square distribution.

It should be noted that in processes with *c* categories, the upper control limit of summation in (1) should be *c*, and the statistic Y_i^2 has a chi-square distribution with *c-1* degree of freedom.

Now, consider the second problem (case II). The goodness of fit test is not appropriate here. An appropriate statistical approach is a consistency test between base period and other periods of the process [7]. This statistic for period *i* is as (2).

Consider the following:

$$\begin{aligned} Z_i^2 &= \sum_{k=i,0} \sum_{j=1}^{3} \frac{n_k\left(X_{kj}/n_k - \left(X_{ij} + X_{0j}\right)/\left(n_i + n_0\right)\right)^2}{\left(X_{ij} + X_{0j}\right)/\left(n_i + n_0\right)} \\ &= n_i n_0 \sum_{j=1}^{3} \frac{\left(P_{ij} + P_{0j}\right)^2}{X_{ij} + X_{0j}}, \end{aligned} \tag{2}$$

where $P_{kj} = X_{kj}/n_k$ is the ratio of each sample. If $n_i \to \infty$ where n_0/n_i is limited and greater than zero, so that Z_i^2 has a chi-square distribution with two degrees of freedom. Therefore, in case I, the control chart for this case also has an upper control limit equal to an appropriate percentile of chi-square distribution.

There is no theoretical rule for sufficient sample size for using chi-square distribution in such a case. Some rules of thumb exist to determine enough sample size. The most famous rule was proposed by Cochran [8]. He declared that the twenty percent of the frequency of each category should be greater than 5, and the expected frequency of each categories should be greater than one.

3.2. Grouped Observations. Even when the quality characteristic is variable, it is more economical to classify it into *k* categories than to measure it exactly. As measuring a variable characteristic need, cost and time, using gauge for quality inspection is suggested. As Steiner et al. [9] mentioned: "usually quality data are gathering in grouped manner."

3.3. Fuzzy Control Charts. Based on the nature of the quality characteristics, two broad categories of control charts are developed, namely, variable and attribute control charts. Variable control charts are used to monitor continuous characteristics of the products such as length, weight, and voltage which are measurable on numerical scales. However, it is not always possible to express the quality characteristics on a numerical scale. For these characteristics such as appearance, softness, and color, control charts for attribute are used. Control chart for proportion nonconforming is one of the attribute control charts. In this chart, each product unit is classified as "conforming" or "nonconforming," depending upon whether or not they meet specifications. Then, by using the principles of Shewhart control charts, this chart called *p*-chart is formed. But as Raz and Wang [10, 11] also

mentioned, the binary classification into "conforming" and "nonconforming" used in p-chart might not be appropriate in many situations where there might be a number of intermediate levels. In this case, for measuring the quality-related characteristics, it is necessary to use several intermediate levels besides conforming and nonconforming. For example, the quality of product can be classified by one of the following terms: "perfect," "good," "medium," "poor," and "fair," depending on deviation from specifications. Data obtained in this way are called categorical data, and we can use multinomial distribution instead of binary distribution. Several statistical researches have been done in this area. The early research goes back to Duncan [6, 7], who introduced a chi-square control chart for monitoring a multinomial process with categorical data. Later, this type of control chart is discussed further by Marcucci [12] and Nelson [13]. Marcucci introduced a statistical approach for a case, where the proportion of each category is not known before.

But the problem still exists. As we know, the quality level of each product is determined by the quality inspectors, and they do this task mentally. For example, one product might be classified into *perfect* category by an inspector but classified into *good* category by another inspector. It means that determining the quality level of the product mentally by the inspectors is in an uncertainty situation. As Yager and Zadeh [14] also indicated that in fact the main problem is vagueness that corresponds to the mental affect. Fuzzy set theory could be used because of the uncertainty situation and vague environment. In case of monitoring attribute data by using fuzzy set theory, several researches exist. Raz and Wang [10, 11] proposed an approach based on fuzzy set theory for monitoring attribute processes when quality characteristics are classified into mutually exclusive categories. Kanagawa et al. [15] present a control chart based on the probability density function existing behind the linguistic data, continuing the Raz and Wang approach. These approaches are discussed by Laviolette et al. [16], Almond [17], and Kandel et al. [18] and reviewed by Woodall et al. [19] and Taleb and Limam [20]. Later, Gülbay et al. [21–23] proposed an α-level fuzzy control chart for attributes in order to reflect the vagueness of data and tightness of inspection. In the following, the most famous research in the area of fuzzy attribute control charts will be illustrated in detail.

3.3.1. The Raz and Wang Approach. Constructing a control chart involves determining the center line (CL), upper control limit (UCL), and lower control limit (LCL). This is calculated based on the random sample from the process. When linguistic data are used, it is necessary to state the related fuzzy set by a representative value. In the following, several approaches to determine a representative value for a fuzzy set are explained, and after that probabilistic and fuzzy membership approach will be presented.

Representative Value. To keep the standard format of the Shewhart control chart, it is necessary to transfer the associated fuzzy set to a crisp value which we call representative value. This transformation could be done in different ways. In the following, four methods which are similar in the principle to central tendency in statistics are represented. It must be mentioned that there is no theoretical baseline to select between these four methods, and the selection is completely arbitrary. In the following definitions, F is the fuzzy subset, x is the base variable, and $\mu_F(x)$ is the membership function.

(1) The *fuzzy mode*, f_{mode}, is the value of the base variable where the membership is equal to 1:

$$\mu_F\left(f_{\text{mode}}\right) = 1\left(f_{\text{mode}} = \left\{x \mid \mu_f(x) = 1\right\},\ \forall x \in F\right). \quad (3)$$

The fuzzy mode is unique if $\mu_F(x)$ is unimodal.

In the special case where $\widetilde{A} = (a, b, c)$ is a triangular fuzzy number, the fuzzy mode is equal to b; so, we could have

$$f_{\text{mode}} = b. \quad (4)$$

(2) The *α-level fuzzy midrange*, $f_{\text{mr}}(\alpha)$, is the average of the endpoint of an α-level cut. An α-level cut of F, denoted by F_a, is a nonfuzzy subset of the base variable x containing all the values with a membership function value greater than or equal to α. Thus,

$$F_\alpha = \left\{x \mid \mu_F(x) \geq \alpha\right\}. \quad (5)$$

Note that the fuzzy mode is a special case of the α-level fuzzy midrange with $\alpha = 1$.

Suppose that $\widetilde{A}$ is a triangular fuzzy number. Applying α-cut of fuzzy set, the values of a^α and c^α are determined as follows:

$$\begin{aligned} a^\alpha &= a + \alpha(b - a), \\ c^\alpha &= c - \alpha(c - b). \end{aligned} \quad (6)$$

So, α-level fuzzy midrange for a triangular fuzzy number could be calculated as follows:

$$\begin{aligned} f_{\text{mr}}(\alpha) &= \frac{a^\alpha + c^\alpha}{2} \\ \Longrightarrow f_{\text{mr}}(\alpha) &= \frac{(a + c) + \alpha[(b - a) - (c - b)]}{2}. \end{aligned} \quad (7)$$

(3) The *fuzzy median*, f_{med}, is the point which divides the area under the membership function into two equal regions, satisfying the following equation:

$$\begin{aligned} \int_{-\infty}^{f_{\text{med}}} \mu_F(x)\,dx &= \int_{f_{\text{med}}}^{+\infty} \mu_F(x)\,dx \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} \mu_F(x)\,dx. \end{aligned} \quad (8)$$

(4) The *fuzzy average*, f_{avg}, is defined by Zadeh [24] as follows:

$$f_{\text{avg}} = A\nu(x : F) = \frac{\int_0^1 x\mu_F(x)\,dx}{\int_0^1 \mu_F(x)\,dx}. \quad (9)$$

Generally, two first approaches are simpler in calculation, especially when the membership function was nonlinear.

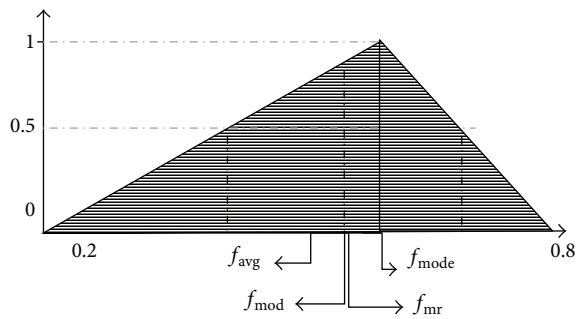

Figure 1: Representative value.

However, when the membership function is too nonsymmetrical, the result of fuzzy mode might be bias. Fuzzy midrange is more flexible, because a different level for α could be selected. When in addition to the place of membership function, the shape of the membership function is important; then, the best choice would be fuzzy average, because it has been calculated from a wide principle.

For comparison, consider a fuzzy set like $\widetilde{A}$ as follows:

$$\mu_A(x) = \begin{cases} 0 & x \le 0.2 \\ 2.5x - 0.5 & 0.2 \le x \le 0.6 \\ -5x + 4 & 0.6 \le x \le 0.8 \\ 0 & x \ge 0.8. \end{cases} \tag{10}$$

Representative value for $\widetilde{A}$ would be $f_{\text{mode}} = 0.6$, $f_{\text{med}} = 0.546$, $f_{\text{mr}} = 0.55$, and $f_{\text{avg}} = 0.533$; Figure 1 shows these results as well.

Representing a Sample. A sample could involve several observations which are selected for the inspection. Each observation is classified with a linguistic term and related to a known membership function. These separate linguistic terms need to combine to become a representative value for the sample. This combination of the observation could be done both before and after transferring the linguistic terms to representative values.

In the first case, related fuzzy sets to linguistic terms in a sample should be added together and then divided into the number of sample observations. This operation is done based on the fuzzy mathematics. The result would be a fuzzy set which might not be similar to any of the preliminary terms but is the representative of the quality of that sample. Then, a numerical value as a representative could be calculated by one of the four transformation techniques which were explained in the previous section. Suppose that t linguistic terms existed which were shown by L_i $(i = 1, 2, \ldots, t)$. For each linguistic value, a related fuzzy set such as F_i with a membership function like $\mu_i(x_i)$ is defined. Consider a sample like S with n observation $S = \{(F_1, k_1), (F_2, k_2), \ldots, (F_t, k_t)\}$, where k_i is the number of items that classify to linguistic value L_I by quality inspectors and $k_1 + k_2 + \cdots + k_t = n$. The fuzzy set which is the mean of a sample fuzzy set is shown by MFs. The membership function of MFs is $\mu_s(x_s)$ as follows [25]:

$$\mu_S(x_S) = \underset{x_S = (k_1x_1 + k_2x_2 + \cdots + k_tx_t)/n}{\text{Max}} \times \{\text{Min}[\mu_1(x_1), \mu_2(x_2), \ldots, \mu_t(x_t)]\}. \tag{11}$$

The representative value of the sample could be calculated by one of the transformation approach on the $\mu_s(x_s)$.

If the mean of the sample constructs after transferring the linguistic value to representative value, the calculation would be easier. The representative value of the F_i is shown by r_i. The sample mean, M, as the mean of the r_i could be calculated as follows:

$$M = \frac{(r_1k_1 + r_2k_2 + \cdots + r_tk_t)}{n}. \tag{12}$$

The first approach keeps fuzziness more than the second approach with the need of more calculation especially when we have a nonlinear membership function. In the following, an example is provided to show both approaches.

Consider a linguistic variable for the evaluation of the quality characteristic of a product with a set of terms such as *perfect, good, medium, poor,* and *bad*. Base variable is a level of quality which standardized in the interval $[0, 1]$. Zero shows the best quality and 1 shows the lower quality. Membership functions associated with each linguistic term are as follows.

$$\begin{aligned}
\mu_{\text{perfect}}(x) &= \begin{cases} 1 - 4x & 0 \le x \le 0.25 \\ 0 & x \ge 0.25, \end{cases} \\
\mu_{\text{good}}(x) &= \begin{cases} 4x & 0 \le x \le 0.25 \\ 2 - 4x & 0.25 \le x \le 0.5 \\ 0 & x \ge 0.5, \end{cases} \\
\mu_{\text{medium}}(x) &= \begin{cases} o & x \le 0.25;\ x \ge 0.75 \\ 4x - 1 & 0.25 \le x \le 0.5 \\ 3 - 4x & 0.5 \le x \le 0.75, \end{cases} \\
\mu_{\text{poor}}(x) &= \begin{cases} 0 & x \le 0.5 \\ 4x - 2 & 0.5 \le x \le 0.75 \\ 4 - 4x & 0.75 \le x \le 1, \end{cases} \\
\mu_{\text{bad}}(x) &= \begin{cases} 0 & x \le 0.75 \\ 4x - 3 & 0.75 \le x \le 1. \end{cases}
\end{aligned} \tag{13}$$

These membership functions are depicted in Figure 2.

Consider a sample with 10 observations as

$$\begin{aligned} S = \{&(F_{\text{perfect}}, 3), (F_{\text{good}}, 2), \\ &(F_{\text{medium}}, 2), (F_{\text{poor}}, 2), (F_{\text{bad}}, 1)\}. \end{aligned} \tag{14}$$

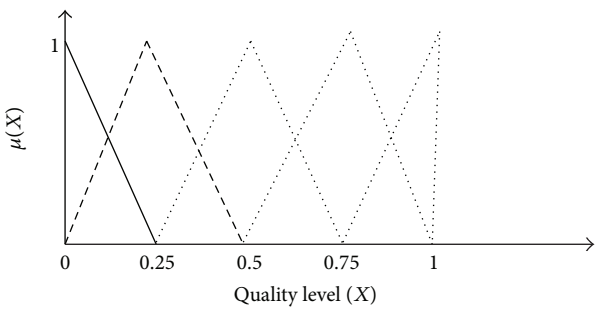

Figure 2: Membership functions for 5 linguistic terms.

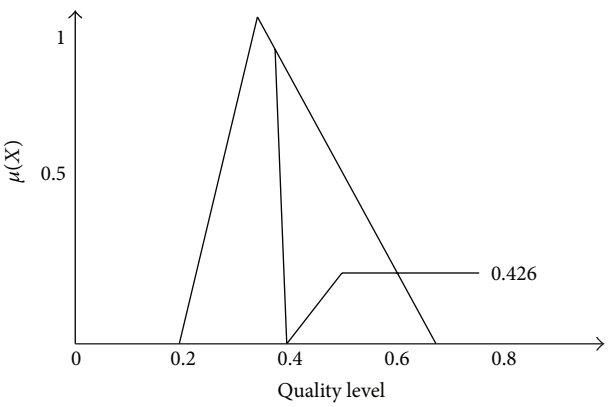

Figure 3: Combined membership function of the sample.

By combining these 10 observations based on the first approach, fuzzy set associated with the sample mean could be defined by the following membership function:

$$\mu_S(x) = \begin{cases} 0 & 0 \le x \le 0.2 \\ 5x - 1 & 0.2 \le x \le 0.4 \\ 2.333 - 3.333x & 0.4 \le x \le 0.7 \\ 0 & x \ge 0.7. \end{cases} \tag{15}$$

Figure 3 also shows this membership function. The representative value by using fuzzy median for this set would be 0.426.

By using the second approach, first, the representative value for each linguistic term must be calculated. By using the fuzzy median, we have

$$\begin{gathered} r_{\text{perfect}} = 0.146, \quad r_{\text{good}} = 0.25, \\ r_{\text{medium}} = 0.5, \quad r_{\text{poor}} = 0.75, \quad r_{\text{bad}} = 0.854. \end{gathered} \tag{16}$$

Finally, sample mean could be calculated as follows:

$$\begin{gathered} (0.146 \times 3 + 0.25 \times 2 + 0.5 \times 2 \\ +0.75 \times 2 + 0.854 \times 1)\, 10^{-1} = 0.429. \end{gathered} \tag{17}$$

Calculation of the Center Line. Normally, center line could be calculated as the average of the sample mean. Here, also both approaches could be used. Suppose m sample with n observations, then CL would be as follows:

$$\text{CL} = \frac{\sum_{j=1}^{m} M_j}{m}, \tag{18}$$

where M_j is the sample mean of the jth sample.

In the following, for determining the control limits, two approaches are explained, namely, probabilistic approach and membership approach.

Probabilistic Control Limits. In the traditional control charts, control limits were determined with a coefficient of the standard deviation of the process. So, here also we need an estimation of the standard deviation. For m sample with n observations, standard deviation is shown by SD_j for jth sample and calculated as follows:

$$\text{SD}_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{t} k_{ij}\left(r_j - M_j\right)^2}, \tag{19}$$

where t is the number of linguistic terms, r_i is the representative value associated with linguistic term L_i, and M_j is the jth sample mean. The mean of m standard deviation was shown by MSD and calculated as follows:

$$\text{MSD} = \frac{1}{m}\sum_{j=1}^{m} \text{SD}_j. \tag{20}$$

Suppose that sample distribution is approximately normal, or sample size is large enough ($n > 25$). Then, for calculating the control limits, we could use the standard method. We have

$$\begin{gathered} \text{Probabilistic LCL} = \text{Max}\left\{0, \left(\text{CL} - A_3 \cdot \text{MSD}\right)\right\}, \\ \text{Probabilistic UCL} = \text{Min}\left\{1, \left(\text{CL} + A_3 \cdot \text{MSD}\right)\right\}, \\ A_3 = \frac{3}{C_4\sqrt{n}} C_4 = \sqrt{\frac{2\left((n-2)/2\right)!}{n-1\left((n-3)/2\right)!}}. \end{gathered} \tag{21}$$

The coefficient A_3 and C_4 and table of other coefficient values could be found in Montgomery [26] and any other standard references.

Membership Control Limits. In contrast to the traditional control charts which are constructed based on the probability distribution of the sample mean, membership control limits are based on the membership function. In the following, constructing the membership control limits would be explained.

Consider a convex fuzzy set, and suppose that x_m is the fuzzy mode of the membership function. We could define an inverse membership function which consists of two parts. One part which is in the left side of x_m and is shown by $x_l(\alpha)$, and another part which is in the right side of x_m and shown by $x_r(\alpha)$. The inverse membership function is defined as follows: $x_l(\alpha)$ is the minimum value of the base variable x in which the membership value of them is equal to α, and $x_r(\alpha)$ is the maximum value of the base variable x in which

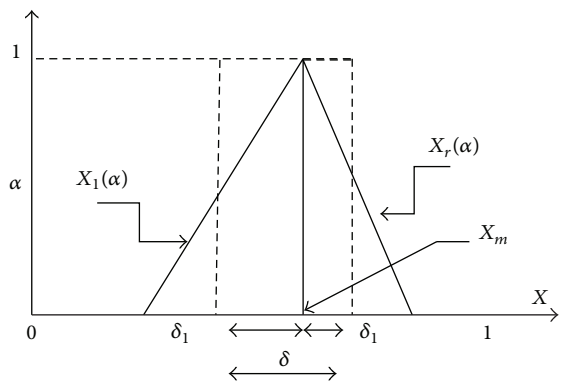

Figure 4: Deviation of mean for a fuzzy set.

the membership value of them is equal to a. In other words, $x_l(\alpha)$ and $x_r(\alpha)$ are the endpoints of α-cut. Now, a value for the deviation of fuzzy set which is called mean deviation and shown as δ could be calculated as follows by using the summation, the deviation of left mean, δ_l, and deviation of right mean, δ_r [25]:

$$\begin{aligned} \delta_l &= \int_{\alpha=0}^{1} \left[x_m - x_l(\alpha)\right] \cdot d\alpha, \\ \delta_r &= \int_{\alpha=0}^{1} \left[x_r(\alpha) - x_m\right] \cdot d\alpha, \end{aligned} \tag{22}$$

where δ_l and δ_r are left and right deviations of mean, respectively. Their values are equal to the area under the membership function at the left and right side of the mode point of fuzzy set. For fuzzy set $\widetilde{A}$, mean deviation $\delta(A)$ could be calculated as follows:

$$\begin{aligned} \delta(A) = \delta_l(A) + \delta_r(A) &= \int_{\alpha=0}^{1} \left[x_m - x_l(\alpha)\right] \cdot d\alpha \\ &\quad + \int_{\alpha=0}^{1} \left[x_r(\alpha) - x_m\right] \\ &= \int_{\alpha=0}^{1} \left[x_r(\alpha) - x_l(\alpha)\right] \cdot d\alpha. \end{aligned} \tag{23}$$

In this equation, α is the level of membership. In fact, deviation of a fuzzy set, is a numerical value which stated by the dimension of the base variable. Figure 4 presents the deviation of mean for a fuzzy set.

Suppose that we have m sample with n observations. At first, the fuzzy mean of each sample must be calculated by using fuzzy mathematics, and then, the grand mean of m sample must calculated. For determining the control limits by using the previous equation, at first, the deviation of grand mean should be calculated. The control limits a known distance from the center line. This distance is equal to a coefficient of deviation of the grand mean. We could have

$$\begin{aligned} &\text{Membership LCL} = \text{Max}\{0, (\text{CL} - K\delta)\}, \\ &\text{Membership UCL} = \text{Min}\{1, (\text{CL} + K\delta)\}, \end{aligned} \tag{24}$$

where k is a coefficient which shows the distance from center line. The value of k could be determined by using the Monte-Carlo simulation when we suppose that type I error is fixed.

3.3.2. The Kanagawa Approach. Raz and Wang [10, 11] proposed a general approach for designing control chart for monitoring the mean of the process. This approach is based on the normal assumption and just monitors the mean of the process. Kanagawa et al. [15] proposed an approach for estimating the probability density function beyond the linguistic data, and by using it, they design control charts for monitoring both the mean and the variation of the process.

Probability Density Function for Linguistic Data. The objective is to design a control chart for monitoring the variation as well as the mean of a process by using the probability density function (p.d.f). The mentioned probability density function which is beyond the linguistic variables generates the linguistic data randomly and based on the mental judgment of the inspectors.

Suppose that for standard data in the interval [0, 1], p.d.f could be determined based on the Gram-Charlier series:

$$f(x) = \phi(x)\left[1 + \alpha_1 H_1(x) + \alpha_2 H_2(x) + \cdots\right], \tag{25}$$

where $\phi(x)$ is a standard normal probability density function and H_r is the Hermite polynomial with the degree of r:

$$\begin{aligned} H_1(x) &= x \\ H_2(x) &= x^2 - 1 \\ H_3(x) &= x^3 - 3x \\ &\vdots \end{aligned} \tag{26}$$

The relationship between α_r and β_r is

$$\begin{aligned} \alpha_1 &= \beta_1 \\ \alpha_2 &= \frac{\beta_2 - 1}{2} \\ \alpha_3 &= \frac{\beta_3 - 3\beta_1}{6} \\ &\vdots \end{aligned} \tag{27}$$

Also, the relationship between β_r and K_r would be

$$\begin{aligned} K_1 &= \beta_1 \\ K_2 &= \beta_2 - \beta_1^2 \\ K_3 &= \beta_3 - 3\beta_2\beta_1 + 2\beta_1^3 \\ &\vdots \end{aligned} \tag{28}$$

Linguistic data could be considered as fuzzy data. So, based on the Gülbay and Kahraman [23] definition, the probability of a linguistic data like L_i happening is

$$\Pr\{L_i\} = \int_R \mu_i(x) f(x)\, dx. \tag{29}$$

When the p.d.f and membership function of the linguistic variable are known, previous equations helps to calculate the probability of happening the event of base variable X in the interval $[x, x+dx]$ with condition of happening the evidence of L_i, as follows:

$$\Pr(X \mid L_i)\, dx = \frac{\mu_i(x) f(x)\, dx}{\Pr(L_i)}. \tag{30}$$

In addition, if k_i is known, β_r could be calculated as follows:

$$\beta_r = \frac{1}{n}\sum_{i=1}^{t}\int_{-\infty}^{+\infty} k_i x^r \Pr(X \mid L_i)\, dx. \tag{31}$$

Based on Kanagawa et al. [15], by using the Gram-Charlier series with degree r ($r = 1, 2, \ldots, t$) and by using the following algorithm, we could estimate the p.d.f.

Step 1. By using fuzzy mode, the value of scalar number of the membership function associated with each linguistic value $X_1, X_2, \ldots, X_t$ must be determined. Then, $k_r^{(0)}$ is calculated as follows:

$$\beta_r^{(0)} = \frac{1}{n}\sum_{i=1}^{t} k_i x_i^r. \tag{32}$$

Continuously, by using (32) other torques must be determined.

Step 2. The torque which is determined in Step 1 inserts into the p.d.f.

Step 3. The values of $f(x)$ insert into (31) and update the torque.

Step 4. Repeat Steps 2 and 3 until giving the following condition:

$$\beta_r^{(0)} = \frac{1}{n}\sum_{i=1}^{t} k_i x_i^r. \tag{33}$$

Now, by using this assumption where is $X_1, X_2, \ldots, X_n$ are independent random variables from $f(x)$, α upper percent of normal distribution by using the Cornish-Fisher development method would be

$$\begin{aligned} Z_\alpha = u_\alpha &+ \frac{K_3/K_2^{3/2}}{6\sqrt{n}}\left(u_\alpha^2 - 1\right) \\ &+ \frac{K_4/K_2^2}{24n}\left(u_\alpha^3 - 3u_\alpha\right) \\ &- \frac{K_3^2/K_2^3}{36n}\left(2u_\alpha^3 - 5u_\alpha\right) + \cdots, \end{aligned} \tag{34}$$

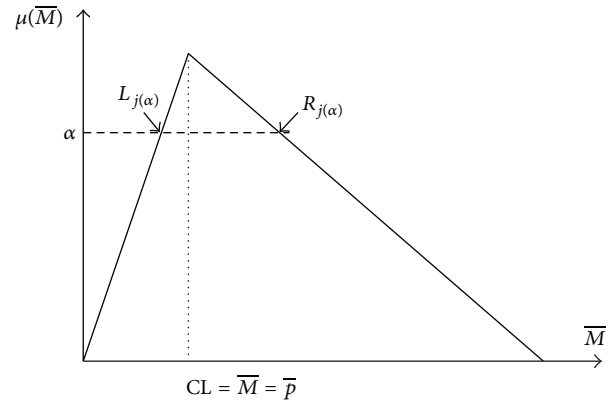

Figure 5: TFN for M and M_j for sample j.

where

$$Z = \frac{\overline{X} - K_1}{\left(K_2/n\right)^{1/2}}, \tag{35}$$

and u_α shows α upper percent of normal distribution with mean equal to zero and variance equal to 1.

3.3.3. α-Level Fuzzy Control Chart. As mentioned before in crisp state, control limits for the proportion of nonconforming could be calculated as (1). In fuzzy state, sample mean M_j and center line CL could be calculated as follows:

$$\begin{aligned} \mathrm{CL} = \overline{M}_j &= \sum_{j=1}^{m} M_j, \\ M_j &= \frac{\sum_{j=1}^{m} k_{ij} r_i}{n_j}, \quad i = 1, 2, \ldots, t. \end{aligned} \tag{36}$$

As CL is a fuzzy set, it could be stated by a triangular fuzzy number (TFN), where its fuzzy mode is equal to CL. Figure 5 depicted CL as a TFN.

4. Conclusion

To conclude, this study has technically reviewed control charts. The author in this paper covered the first phase of developments in the context of control charts. In the second phase, most of the works are based on hybrid charts as well as works which are focusing on the use of more productive charting methods [27]. The second part starts by 2000s. Clearly, developments of phase two charts are all based on pure charts which are in phase one and have been reviewed in this paper. The contribution of this study was to review pure control charts to show start points to direct further studies. Further researches could continue reviewing the developments of control charts in second phase, as well as using the pure charts of the first phase to modify the chart's productivity. The author is continuing this study to modify the current available control charts, using fuzzy theory approach.

Conflict of Interests

The author declares no possible conflict of interests.

References

[1] T. P. Ryan and N. C. Schwertman, "Optimal limits for attributes control charts," *Journal of Quality Technology*, vol. 29, no. 1, pp. 86–98, 1997.

[2] N. C. Schwertman and T. P. Ryan, "Using dual np-charts to detect changes," *Quality and Reliability Engineering International*, vol. 15, no. 4, pp. 317–320, 1999.

[3] G. Chen, "An improved p chart through simple adjustments," *Journal of Quality Technology*, vol. 30, no. 2, pp. 142–151, 1998.

[4] L. S. Nelson, "Control chart for parts-per-million nonconforming items," *Journal of Quality Technology*, vol. 26, no. 3, pp. 239–240, 1994.

[5] M. R. Reynolds Jr. and Z. G. Stoumbos, "A general approach to modeling CUSUM charts for a proportion," *IIE Transactions*, vol. 32, no. 6, pp. 515–535, 2000.

[6] A. Duncan, "A chi-square chart for controlling a set of percentages," *Industrial Quality Control*, vol. 7, pp. 11–15, 1950.

[7] A. J. Duncan, *Quality Control and Industrial Statistics*, Richard D. Irwin, Homewood, Ill, USA, 5th edition, 1986.

[8] W. G. Cochran, "Some methods for strengthening the common χ^2 tests," *Biometrics*, vol. 10, no. 4, pp. 417–451, 1954.

[9] N. R. Steiner, P. L. Geyer, and G. O. Wesolowsky, "Control charts based on grouped observations," *The International Journal of Production Research*, vol. 32, no. 1, pp. 75–91, 1994.

[10] J.-H. Wang and T. Raz, "On the construction of control charts using linguistic variables," *The International Journal of Production Research*, vol. 28, no. 3, pp. 477–487, 1990.

[11] T. Raz and J. H. Wang, "Probabilistic and membership approaches in the construction of control charts for linguistic data," *Production Planning & Control*, vol. 1, no. 3, pp. 147–157, 1990.

[12] M. Marcucci, "Monitoring multinomial processes," *Journal of Quality Technology*, vol. 17, no. 2, pp. 86–91, 1985.

[13] L. S. Nelson, "A chi-square control chart for several proportions," *Journal of Quality Technology*, vol. 19, no. 4, pp. 229–231, 1987.

[14] R. R. Yager and L. A. Zadeh, *An Introduction to Fuzzy Logic Applications in Intelligent Systems*, Kluwer Academic, 1992.

[15] A. Kanagawa, F. Tamaki, and H. Ohta, "Control charts for process average and variability based on linguistic data," *The International Journal of Production Research*, vol. 31, no. 4, pp. 913–922, 1993.

[16] M. Laviolette, J. W. Seaman, J. D. Barrett, and W. H. Woodall, "A probabilistic and statistical view of fuzzy methods," *Technometrics*, pp. 249–261, 1995.

[17] R. G. Almond, "Discussion: fuzzy logic: better science? Or better engineering?" *Technometrics*, vol. 37, no. 3, pp. 267–270, 1995.

[18] A. Kandel, A. Martins, and R. Pacheco, "Discussion: on the very real distinction between fuzzy and statistical methods," *Technometrics*, pp. 276–281, 1995.

[19] W. Woodall, K. Tsui, and G. Tucker, "A review of statistical and fuzzy control charts based on categorical data," *Frontiers in Statistical Quality Control*, vol. 5, pp. 83–89, 1997.

[20] H. Taleb and M. Limam, "On fuzzy and probabilistic control charts," *The International Journal of Production Research*, vol. 40, no. 12, pp. 2849–2863, 2002.

[21] M. Gülbay, C. Kahraman, and D. Ruan, "α-cut fuzzy control charts for linguistic data," *International Journal of Intelligent Systems*, vol. 19, no. 12, pp. 1173–1195, 2004.

[22] M. Gülbay and C. Kahraman, "Development of fuzzy process control charts and fuzzy unnatural pattern analyses," *Computational Statistics & Data Analysis*, vol. 51, no. 1, pp. 434–451, 2006.

[23] M. Gülbay and C. Kahraman, "An alternative approach to fuzzy control charts: direct fuzzy approach," *Information Sciences*, vol. 177, no. 6, pp. 1463–1480, 2007.

[24] L. A. Zadeh, "Fuzzy sets," *Information and Computation*, vol. 8, no. 3, pp. 338–353, 1965.

[25] A. Kaufmann and M. M. Gupta, *Introduction to Fuzzy Arithmetic*, Van Nostrand Reinhold Company, 1985.

[26] D. C. Montgomery, *Introduction To Statistical Quality Control*, John Wiley & Sons, New York, NY, USA, 6th edition, 2009.

[27] W. H. Woodall, "Current research on profile monitoring," *Producao*, vol. 17, no. 3, pp. 420–425, 2007.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 276245, 7 pages
http://dx.doi.org/10.1155/2013/276245

Research Article

A Global Optimization Algorithm for Sum of Linear Ratios Problem

Yuelin Gao and Siqiao Jin

Institute of Information & System Science, Beifang University of Nationalities, Yinchuan 750021, China

Correspondence should be addressed to Yuelin Gao; gaoyuelin@263.net

Received 31 January 2013; Accepted 8 May 2013

Academic Editor: Farhad Hosseinzadeh Lotfi

We equivalently transform the sum of linear ratios programming problem into bilinear programming problem, then by using the linear characteristics of convex envelope and concave envelope of double variables product function, linear relaxation programming of the bilinear programming problem is given, which can determine the lower bound of the optimal value of original problem. Therefore, a branch and bound algorithm for solving sum of linear ratios programming problem is put forward, and the convergence of the algorithm is proved. Numerical experiments are reported to show the effectiveness of the proposed algorithm.

1. Introduction

We consider the sum of linear ratios programming problem as the following form:

$$(\text{GFP}) : \begin{cases} \min f(x) = \sum_{i=1}^{p} f_i(x) = \sum_{i=1}^{p} \dfrac{n_i(x)}{d_i(x)}, \\ \text{s.t.} \quad Ax \le b, \quad x \ge 0, \end{cases} \tag{1}$$

where the feasible domain $D \triangleq \{x \in R^n \mid Ax \le b, x \ge 0\}$ is n-dimensional, nonempty, and bound, $A \in R^{m\times n}, b \in R^m$. Assume that $n_i(x) = c_i^T x + d_i \ge 0$ and $d_i(x) = e_i^T x + r_i > 0$ in some rectangle X which contains D, where $c_i, e_i \in R^n, d_i, r_i \in R$, $i = 1, 2, \ldots, p$, and $2 \le p \ll n$.

Fractional programming is an important branch of nonlinear optimization and it has attracted many researchers' concern for several decades. Sum of linear ratios problem is a special class of fractional programming problem; it has wide applications, such as investment, transportation scheme, and economic benefits [1–3]. From a research point view, sum of ratios problems challenge theoretical analysis and computation because these problems possess multiple local optima that are not globally optimal solutions; it is difficult to solve the global solution.

At present there exist a number of algorithms for globally solving sum of linear ratios problems. When $p = 2$, Konno et al. [4] constructed a similar parametric simplex algorithm which can solve large-scale optimization problems; when $p = 3$, Konno and Abe [5] developed parametric simplex algorithm and constructed an effected heuristic algorithm; when $p > 3$, the literature [6] is a sum of linear ratios problem with coefficients; by using an equivalent transformation and linearization technique, the original nonconvex programming problem reduces to a series of linear programming problems to achieve the purpose of solving it. To minimize the problem, Yanjun et al. [7] use the linearization technique twice by the nature of exponential and logarithmic functions to achieve a linear relaxation programming of the original problem. Benson [8] put forward a new branch and bound algorithm to solve the equivalent concave minimum problem of the original problem. Jiao and Feng [9] present a new pruning technique. In the literature [10], the numerator and denominator of the ratios are not necessarily positive. In this paper, we present a new branch and bound algorithm for solving the sum of linear ratios problems, and the convergence of the algorithm is proved. At last, the numerical experiments are carried out.

This paper is organized as follows. In Section 2, we show how to convert the problem (GFP) into an equivalent problem (EP) by a transformed technique. In Section 3, the linear relaxation programming problem of (EP) is constructed. The branching process of the rectangle is given in Section 4. In Section 5, the branch and bound algorithm for globally solving (EP) is presented and the convergence of

the algorithm is proved. In Section 6, some numerical results are given to show the effectiveness of the present algorithm. Finally, the conclusion is given.

2. Equivalent Transformation

Because the set D is nonempty and bound, we can construct the rectangle $X = [l, u]$, which contains the feasible region of the problem (GFP), $l = (l_1, l_2, \ldots, l_n)^T, u = (u_1, u_2, \ldots, u_n)^T$, l_j and u_j is the optimal value of the linear programming problem (2) and (3), respectively.

$$\begin{aligned} &\min \quad l(x_j) = x_j, \\ &\text{s.t.} \quad Ax \le b, \quad x \ge 0, \end{aligned} \tag{2}$$

$$\begin{aligned} &\max \quad u(x_j) = x_j, \\ &\text{s.t.} \quad Ax \le b, \quad x \ge 0. \end{aligned} \tag{3}$$

Firstly, we solve the following $2p$ linear programming problems:

$$\begin{aligned} &\min \quad d_i(x), \\ &\text{s.t.} \quad x \in D, \\ &\max \quad d_i(x), \\ &\text{s.t.} \quad x \in D, \\ &i = 1, 2, \ldots, p. \end{aligned} \tag{4}$$

The optimal solutions of (4) are x_i^1 and x_i^2 $(i = 1, 2, \ldots, p)$, and the optimal value is denoted by $\overline{l_i}$ and $\overline{u_i}$ $(i = 1, 2, \ldots, p)$ respectively. Obviously, x_i^1 and x_i^2 are feasible to (GFP). Set $W = W \cup \{x_i^1, x_i^2 : i = 1, 2, \ldots, p\}$, where W represent the set of the current feasible solution of the problem (GFP). Set

$$\begin{aligned} H^0 = \left\{ y \in R^p \mid l_i^0 \le y_i \le u_i^0, i = 1, 2, \ldots, p \right\}, \\ y = \left(y_1, y_2, \ldots, y_p \right)^T, \end{aligned} \tag{5}$$

where $l_i^0 = 1/\overline{u_i}$, $u_i^0 = 1/\overline{l_i}$. Then the problem (GFP) is converted into an equivalent nonconvex programming problem:

$$\mathrm{EP}\left(H^0\right): \begin{cases} \min \quad \varphi_0(x, y) = \sum_{i=1}^{p} y_i n_i(x) \\ \qquad\qquad = \sum_{i=1}^{p} y_i \left(\sum_{j=1}^{n} c_{ij} x_j + d_i \right), \\ \text{s.t.} \quad \varphi_i(x, y) = y_i d_i(x) \\ \qquad\qquad = y_i \left(\sum_{j=1}^{n} e_{ij} x_j + r_i \right) \\ \qquad\qquad \ge 1, \quad i = 1, 2, \ldots, p, \\ \qquad\qquad x \in D \cap X, y \in H^0. \end{cases} \tag{6}$$

Theorem 1 (see [10]). *If (x^*, y^*) is a global optimal solution of the problem EP(H^0), then x^* is a global optimal solution of the problem (GFP), and for every $i = 1, 2, \ldots, p$, when $n_i(x^*) \ge 0$, $y_i^* = 1/d_i(x^*)$; conversely, if x^* is a global optimal solution of the problem (GFP), then (x^*, y^*) is a global optimal solution of the problem EP(H^0), where $y_i^* = 1/d_i(x^*)$, $i = 1, 2, \ldots, p$.*

From Theorem 1, the problems (GFP) and EP(H^0) are equivalent; their global optimal values are equal. Therefore, in order to solve (GFP), we only need to solve EP(H^0) instead.

3. Linear Relaxation Technique

From Section 2, $X = [l, u]$ and $H^0 = [l^0, u^0]$ are rectangles; set

$$\begin{aligned} \Omega_i &= \left\{ (x, y_i) \mid l \le x \le u, l_i^0 \le y_i \le u_i^0 \right\} \\ &= \Omega_{1i} \times \Omega_{2i} \times \cdots \Omega_{ni}, \end{aligned} \tag{7}$$

where

$$\begin{aligned} \Omega_{ji} = \left\{ \left(x_j, y_i \right) \mid l_j \le x_j \le u_j, l_i^0 \le y_i \le u_i^0 \right\}, \\ j = 1, 2, \ldots, n. \end{aligned} \tag{8}$$

Because in Ω_{ji} we have $x_j - l_j \ge 0$, $y_i - l_i^0 \ge 0$, so

$$\left(x_j - l_j \right) \left(y_i - l_i^0 \right) \ge 0, \quad j = 1, 2, \ldots, n, \tag{9}$$

expanding it, then we have $x_j y_i \ge l_i^0 x_j + l_j y_i - l_j l_i^0$, $j = 1, 2, \ldots, n$.

Similarly, we can obtain that $x_j - u_j \le 0$, $y_i - u_i^0 \le 0$, so

$$\left(x_j - u_j \right) \left(y_i - u_i^0 \right) \ge 0, \quad j = 1, 2, \ldots, n, \tag{10}$$

expanding it, then we have $x_j y_i \ge u_i^0 x_j + u_j y_i - u_j u_i^0$, $j = 1, 2, \ldots, n$. Let

$$\begin{aligned} \theta_{ji}^{11}\left(x_j, y_i \right) &= l_i^0 x_j + l_j y_i - l_j l_i^0, \quad j = 1, 2, \ldots, n, \\ \theta_{ji}^{12}\left(x_j, y_i \right) &= u_i^0 x_j + u_j y_i - u_j u_i^0, \quad j = 1, 2, \ldots, n. \end{aligned} \tag{11}$$

Because $x_j y_i \ge \theta_{ji}^{11}(x_j, y_i)$, $x_j y_i \ge \theta_{ji}^{12}(x_j, y_i)$, $j = 1, 2, \ldots, n$, we have the following result:

$$x_j y_i \ge \max \left\{ \theta_{ji}^{11}\left(x_j, y_i \right), \theta_{ji}^{12}\left(x_j, y_i \right) \right\}, \quad j = 1, 2, \ldots, n. \tag{12}$$

Similarly, we have $(x_j - l_j)(y_i - u_i^0) \le 0$, $(x_j - u_j)(y_i - l_i^0) \le 0$, $j = 1, 2, \ldots, n$, expanding them, then we have $x_j y_i \le u_i^0 x_j + l_j y_i - l_j u_i^0$, $x_j y_i \le l_i^0 x_j + u_j y_i - u_j l_i^0$; let

$$\begin{aligned} \theta_{ji}^{21}\left(x_j, y_i \right) &= u_i^0 x_j + l_j y_i - l_j u_i^0, \quad j = 1, 2, \ldots, n, \\ \theta_{ji}^{22}\left(x_j, y_i \right) &= l_i^0 x_j + u_j y_i - u_j l_i^0, \quad j = 1, 2, \ldots, n. \end{aligned} \tag{13}$$

Consequently,

$$x_j y_i \le \min \left\{ \theta_{ji}^{21}\left(x_j, y_i \right), \theta_{ji}^{22}\left(x_j, y_i \right) \right\}, \quad j = 1, 2, \ldots, n. \tag{14}$$

From formulae (12) and (14), the following formula is obtained:

$$\begin{aligned} &\max\left\{\theta_{ji}^{11}\left(x_j, y_i\right), \theta_{ji}^{12}\left(x_j, y_i\right)\right\} \\ &\quad \le x_j y_i \\ &\quad \le \min\left\{\theta_{ji}^{21}\left(x_j, y_i\right), \theta_{ji}^{22}\left(x_j, y_i\right)\right\}. \end{aligned} \tag{15}$$

In the problem EP(H^0), let LB(x) and UB(x), respectively, represent the lower bound and upper bound of x; then

$$\begin{aligned} &\mathrm{LB}\left(c_{ij}x_j y_i\right) \\ &\quad = \begin{cases} c_{ij}\cdot\max\left\{\theta_{ji}^{11}\left(x_j, y_i\right), \theta_{ji}^{12}\left(x_j, y_i\right)\right\}, & c_{ij}\ge 0, \\ c_{ij}\cdot\min\left\{\theta_{ji}^{21}\left(x_j, y_i\right), \theta^{22}\left(x_j, y_i\right)\right\}, & c_{ij} < 0, \end{cases} \\ &\mathrm{UB}\left(e_{ij}x_j y_i\right) \\ &\quad = \begin{cases} e_{ij}\cdot\min\left\{\theta_{ji}^{21}\left(x_j, y_i\right), \theta_{ji}^{22}\left(x_j, y_i\right)\right\}, & e_{ij}\ge 0, \\ e_{ij}\cdot\max\left\{\theta_{ji}^{11}\left(x_j, y_i\right), \theta_{ji}^{12}\left(x_j, y_i\right)\right\}, & e_{ij} < 0. \end{cases} \end{aligned} \tag{16}$$

From formula (16), we can obtain the linear relaxation programming problem REP(H^0) of the problem EP(H^0):

$$\mathrm{REP}\left(H^0\right): \begin{cases} \min \quad \varphi_0^l(x, y) = \sum_{i=1}^{p}\left(\sum_{j=1}^{n}\mathrm{LB}\left(c_{ij}x_j y_i\right) + d_i y_i\right), \\ \text{s.t.} \quad \varphi_i^u(x, y) = \sum_{j=1}^{n}\mathrm{UB}\left(e_{ij}x_j y_i\right) + r_i y_i \ge 1, \\ \qquad\qquad i = 1, 2, \ldots, p, \\ \qquad\qquad x \in D\cap X, y \in H^0. \end{cases} \tag{17}$$

The optimal value of the problem REP(H^0) is a lower bound of the optimal value of the problem EP(H^0) in the feasible region D.

Obviously, the problem REP(H^0) can equivalently be converted into the following linear programming problem LRP(H^0):

$$\mathrm{LRP}\left(H^0\right): \begin{cases} \min \quad f(x, y, t, s) = \sum_{i=1}^{p}\left(\sum_{j=1}^{n} t_{ji} + d_i y_i\right), \\ \text{s.t.} \quad \sum_{j=1}^{n} s_{ji} + r_i y_i \ge 1, \quad i = 1, 2, \ldots p, \\ \qquad t_{ji} \ge c_{ij}\cdot\theta_{ji}^{11}\left(x_j, y_i\right), \quad c_{ij}\ge 0, \ j = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p, \\ \qquad t_{ji} \ge c_{ij}\cdot\theta_{ji}^{12}\left(x_j, y_i\right), \quad c_{ij}\ge 0, \ j = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p, \\ \qquad t_{ji} \ge c_{ij}\cdot\theta_{ji}^{21}\left(x_j, y_i\right), \quad c_{ij} < 0, \ j = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p, \\ \qquad t_{ji} \ge c_{ij}\cdot\theta_{ji}^{22}\left(x_j, y_i\right), \quad c_{ij} < 0, \ j = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p, \\ \qquad s_{ji} \le e_{ij}\cdot\theta_{ji}^{21}\left(x_j, y_i\right), \quad e_{ij}\ge 0, \ j = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p, \\ \qquad s_{ji} \le e_{ij}\cdot\theta_{ji}^{22}\left(x_j, y_i\right), \quad e_{ij}\ge 0, \ j = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p, \\ \qquad s_{ji} \le e_{ij}\cdot\theta_{ji}^{11}\left(x_j, y_i\right), \quad e_{ij} < 0, \ j = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p, \\ \qquad s_{ji} \le e_{ij}\cdot\theta_{ji}^{12}\left(x_j, y_i\right), \quad e_{ij} < 0, \ j = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p, \\ \qquad\qquad x \in D\cap X, y \in H^0. \end{cases} \tag{18}$$

The optimal value of the problem LRP(H^0) can be obtained by solving the linear programming problem LRP(H^0), which is a lower bound of the problem EP(H^0) in feasible region D.

The Determination of Upper Bound. From the process of the determination of lower bound, by solving LRP(H^0), we can obtain a global optimal solution $\overline{x}^*$; let

$$\overline{y_i}^* = \left(\sum_{j=1}^{n} e_{ij}\overline{x_j}^* + r_i\right)^{-1}. \tag{19}$$

It is obvious that $(\overline{x}^*, \overline{y}^*)$ is a feasible solution of EP(H^0). Therefore, $\varphi_0(\overline{x}^*, \overline{y}^*)$ provide an upper bound for the global optimal value $\nu(H^0)$ of the problem EP(H^0).

4. Branching

In this algorithm, the branching process is executed in the space of R^p other than in R^n. In general, when $p \ll n$, the amount of computation will decrease so that the efficiency of computation will improve. Therefore, we choose the rectangle

H^0 which contains y to branch, and the subrectangle after branching is also p-dimensional. Set

$$H = \{y \in R^p \mid L_i \le y_i \le U_i, i = 1, 2, \ldots, p\}. \tag{20}$$

Denote the initial rectangle H^0 or subrectangle of it. The branching rule is as follows:

(i) choose the longest side of H, that is, $U_s - L_s = \max\{U_i - L_i : i = 1, 2, \ldots, p\}$;

(ii) let $V_s = (U_s + L_s)/2$ and

$$\begin{aligned} H^1 &= \prod_{i=1}^{s-1} [L_i, U_i] \times [L_s, V_s] \times \prod_{i=s+1}^{p} [L_i, U_i], \\ H^2 &= \prod_{i=1}^{s-1} [L_i, U_i] \times [V_s, U_s] \times \prod_{i=s+1}^{p} [L_i, U_i]. \end{aligned} \tag{21}$$

5. Algorithm and Its Convergence

The branch and bound algorithm of the problem (GFP) is stated as follows:

Step 1. Choose $\varepsilon \ge 0$, the initial rectangle $H^0 = \{y \in R^p \mid l_i^0 \le y_i \le u_i^0, i = 1, 2, \ldots, p\}$; we can find an optimal solution x^0 and the optimal value $\mathrm{LB}(H^0)$ by solving the problem $\mathrm{LRP}(H^0)$. Set $\mathrm{LB}_0 = \mathrm{LB}(H^0)$, $x^c = x^0$. Set $y_i^c = (\sum_{j=1}^n e_{ij} x_j^c + r_i)^{-1}$, $i \in \{1, 2, \ldots, p\}$, $\mathrm{UB}_0 = \varphi_0(x^c, y^c)$.

If $\mathrm{UB}_0 - \mathrm{LB}_0 \le \varepsilon$, stop. (x^c, y^c) and x^c are global ε-optimal solutions of problems $\mathrm{EP}(H^0)$ and (GFP), respectively. Otherwise, set $P_0 = \{H^0\}$, $F = \varnothing$, $k = 1$, and go to Step 2.

Step 2. Set $\mathrm{UB}_k = \mathrm{UB}_{k-1}$. Subdivide H^{k-1} into two p-dimensional rectangles $H^{k,1}, H^{k,2} \subseteq R^p$ via the branching rule. Set $F = F \cup \{H^{k-1}\}$.

Step 3. For $j = 1, 2$, compute $\mathrm{LB}(H^{j,k})$. If $\mathrm{LB}(H^{j,k}) \ne +\infty$, find an optimal solution $x^{k,j}$ of problem $\mathrm{LRP}(\overline{H})$ with $\overline{H} = H^{j,k}$; set $t = 0$.

Step 4. Set $t = t+1$. If $t > 2$, go to Step 6. Otherwise, continue.

Step 5. If $\mathrm{UB}_k \le \mathrm{LB}(H^{k,t})$, set $F = F \cup \{H^{k,t}\}$; go to Step 4. Otherwise, set

$$y_i^{k,t} = \left(\sum_{j=1}^n e_{ij} x_j^{k,t} + r_i \right)^{-1}, \quad i \in \{1, 2, \ldots, p\}. \tag{22}$$

Let $\mathrm{UB}_k = \min\{\mathrm{UB}_k, \varphi_0(x^{k,t}, y^{k,t})\}$. If $\mathrm{UB}_k < \varphi_0(x^{k,t}, y^{k,t})$, go to Step 4. If $\mathrm{UB}_k = \varphi_0(x^{k,t}, y^{k,t})$, set $x^c = x^{k,t}$, $(x^c, y^c) = (x^{k,t}, y^{k,t})$. Let

$$F = F \cup \{H \in P_{k-1} \mid \mathrm{UB}_k \le \mathrm{LB}(H)\}. \tag{23}$$

Step 6. Set $P_k = \{H \mid H \in (P_{k-1} \cup \{H^{k,1}, H^{k,2}\}), H \notin F\}$.

Step 7. Set $\mathrm{LB}_k = \min\{\mathrm{LB}(H) \mid H \in P_k\}$. Let $H^k \in P_k$ satisfy $\mathrm{LB}_k = \mathrm{LB}(H^k)$.

If $\mathrm{UB}_0 - \mathrm{LB}_0 \le \varepsilon$, stop. (x^c, y^c) and x^c are global ε-optimal solutions of the problems $\mathrm{EP}(H^0)$ and (GFP), respectively. Otherwise, set $k = k + 1$ and go to Step 2.

Next, the convergence of the algorithm is stated in the following theorem.

Theorem 2. *(a) If the algorithm is finite, (x^c, y^c) and x^c are global ε-optimal solutions of the problems $EP(H^0)$ and (GFP), respectively.*

(b) For $k \ge 0$, let x^k denote the incumbent solution x^c at the end of step k. If the algorithm is infinite, then $\{x^k\}$ is a feasible solution sequence, whose every accumulation point is a global optimal solution of the problem (GFP), and

$$\lim_{k \to \infty} \mathrm{UB}_k = \lim_{k \to \infty} \mathrm{LB}_k = \nu. \tag{24}$$

Proof. (a) If the algorithm is finite, without loss of generality, it terminates in step k ($k \ge 0$), since (x^c, y^c) is obtained by solving problem $\mathrm{LRP}(H)$, for some $H \subseteq H^0$ and optimal solution x^c, set

$$y_i^c = \frac{1}{\sum_{j=1}^n e_{ij} x_j^c + r_i}, \quad i \in \{1, 2, \ldots, p\}, \tag{25}$$

where x^c is a feasible solution of the problem (GFP) and (x^c, y^c) is a feasible solution of problem $\mathrm{EP}(H^0)$. When $\mathrm{UB}_k - \mathrm{LB}_k \le \varepsilon$, the algorithm terminates. From Steps 1, 2, and 5, it is implied that $\varphi_0(x^c, y^c) - \mathrm{LB}_k \le \varepsilon$; by the algorithm, it shows that $\mathrm{LB}_k \le \nu$. Since (x^c, y^c) is a feasible solution of the problem $\mathrm{EP}(H^0)$, therefore, $\varphi_0(x^c, y^c) \ge \nu$.

Taken together, it is implied that

$$\nu \le \varphi_0(x^c, y^c) \le \mathrm{LB}_k + \varepsilon \le \nu + \varepsilon. \tag{26}$$

Therefore,

$$\nu \le \varphi_0(x^c, y^c) \le \nu + \varepsilon. \tag{27}$$

From the formula $y_i^c = 1/(\sum_{j=1}^n e_{ij} x_j^c + r_i)$, $i = 1, 2, \ldots, p$, we have

$$f(x^c) = \varphi_0(x^c, y^c). \tag{28}$$

From (27), this implies that

$$\nu \le f(x^c) \le \nu + \varepsilon. \tag{29}$$

The proof of (a) is complete.

(b) If the algorithm is infinite, then it generates a sequence of incumbent solutions of the problem $\mathrm{EP}(H^0)$, denoted by $\{(x^k, y^k)\}$, for each $k \ge 1$, (x^k, y^k) is obtained by solving the problem $\mathrm{LRP}(H)$. For some $H^k \subseteq H^0$ and optimal solution $x^k \in D$, set

$$y_i^k = \frac{1}{\sum_{j=1}^n e_{ij} x_j^k + r_i}, \quad i \in \{1, 2, \ldots, p\}. \tag{30}$$

Then the sequence $\{x^k\}$ consists of feasible solutions of the problem (GFP).

Suppose that $\overline{x}$ is an accumulation point of $\{x^k\}$. Assume without loss of generality that $\lim_{k\to\infty} x^k = \overline{x}$. Since D is a compact set, $\overline{x} \in D$. Furthermore, because $\{x^k\}$ is infinite, we assume without loss of generality that, for each k, $H^{k+1} \subseteq H^k$, for some point $\overline{y} \in R^p$,

$$\lim_{k\to\infty} H^k = \bigcap_k H^k = \{\overline{y}\}. \tag{31}$$

Set $\overline{H} = \{\overline{y}\}$, for each k; let $H^k = \{y \in R^p \mid L_i^k \le y_i \le U_i^k, i = 1, 2, \ldots, p\}$. Since $H^{k+1} \subseteq H^k \subseteq H^0$, from Step 5, we know that $\{\mathrm{LB}(H^k)\}$ is a nonincreasing sequence, and $\lim_{k\to\infty}\mathrm{LB}(H^k)$ is a finite number and satisfies

$$\lim_{k\to\infty} \mathrm{LB}\left(H^k\right) \le \nu. \tag{32}$$

For each k, from Step 3, we know that $\mathrm{LB}(H^k)$ is equal to the optimal value of the problem $\mathrm{LRP}(H^k)$ and that x^k is an optimal solution of this problem. From (31), we have

$$\lim_{k\to\infty} L^k = \lim_{k\to\infty} U^k = \{\overline{y}\} = \overline{H}. \tag{33}$$

Since $\lim_{k\to\infty} x^k = \overline{x}$, $L_i^k \le 1/(\sum_{j=1}^n e_{ij}x_j^k + r_i) \le U_i^k$, and the continuity of $\sum_{i=1}^p e_{ij}x_j^k + r_i$,

$$\frac{1}{\sum_{j=1}^n e_{ij}\overline{x_j} + r_i} = \overline{y_i}, \quad i = 1, 2, \ldots, p. \tag{34}$$

This implies that $(\overline{x}, \overline{y})$ is a feasible solution of the problem $\mathrm{EP}(H^0)$. Therefore,

$$\varphi_0\left(\overline{x}, \overline{y}\right) \ge \nu. \tag{35}$$

Together with (32), we have

$$\varphi_0\left(\overline{x}, \overline{y}\right) \ge \nu \ge \lim_{k\to\infty} \mathrm{LB}\left(H^k\right). \tag{36}$$

Since the branching process is bisection and the branching process of rectangle is exhaustive, we have

$$\lim_{k\to\infty} \mathrm{LB}\left(H^k\right) = \nu = \varphi_0\left(\overline{x}, \overline{y}\right). \tag{37}$$

Therefore, $(\overline{x}, \overline{y})$ is a global optimal solution of the problem $\mathrm{EP}(H^0)$. By Theorem 1, this implies that $\overline{x}$ is a global optimal solution of the problem (GFP). For each k, since x^k is the incumbent solution of the problem (GFP) at the end of step k, $\mathrm{UB}_k = f(x^k)$; by the continuity of f, we obtain that

$$\lim_{k\to\infty} f\left(x^k\right) = f\left(\overline{x}\right). \tag{38}$$

Since $\overline{x}$ is a global optimal solution of the problem (GFP),

$$f\left(\overline{x}\right) = \nu. \tag{39}$$

Therefore, $\lim_{k\to\infty}\mathrm{UB}_k = \nu$. The proof is complete. □

6. Numerical Experiment

The proposed algorithm is programmed in MATLAB 7.8 and is run in Pentium(R) 4 CPU 3.20 GHz. In order to compare with the algorithm of the literature [10], we perform three experiments to the literature [10].

Example 1 (see [10]). We choose $p = n = 2$; for each $(x_1, x_2) \in R^2$, the numerator and denominator are

$$\begin{aligned} n_1\left(x_1, x_2\right) &= 37x_1 + 73x_2 + 13, \\ n_2\left(x_1, x_2\right) &= 63x_1 - 18x_2 + 39, \\ d_1\left(x_1, x_2\right) &= 13x_1 + 13x_2 + 13, \\ d_2\left(x_1, x_2\right) &= 13x_1 + 26x_2 + 13, \end{aligned} \tag{40}$$

and all $(x_1, x_2) \in D$ satisfy

$$5x_1 - 3x_2 = 3, \quad 1.5 \le x_1 \le 3. \tag{41}$$

From our algorithm, we firstly should solve the following linear programming problems:

$$\begin{aligned} &\min \quad d_i(x), \\ &\text{s.t.} \quad Ax \le b, \\ &\max \quad d_i(x), \\ &\text{s.t.} \quad Ax \le b, \\ &i = 1, 2, \ldots, p, \end{aligned} \tag{42}$$

of which the optimal solutions denote by x_i^1, x_i^2 $(i = 1, 2)$; then

$$W = W \cup \left\{x_i^1, x_i^2 : i = 1, 2, \ldots, p\right\}, \tag{43}$$

where W represent the set of the current feasible solution of the problem $\mathrm{EP}(H^0)$, and the optimal value is denoted by $\overline{l_i}$ and $\overline{u_i}$ $(i = 1, 2)$; then the initial rectangle is

$$H^0 = \begin{bmatrix} 0.0096 & 0.0192 \\ 0.0064 & 0.0140 \end{bmatrix}. \tag{44}$$

By solving the linear relaxation programming problem $\mathrm{LRP}(H^0)$, we obtain the optimal solution $x^0 = [2.0016; 2.3360]$ and the optimal value $\mathrm{LB}(H^0) = 3.9743$; then a lower bound of the original problem is $\mathrm{LB}(H^0) = 3.9743$. Set

$$y_i^0 = \left(\sum_{j=1}^n e_{ij}x_j^0 + r_i\right)^{-1}. \tag{45}$$

Then (x^0, y^0) is a feasible solution of $\mathrm{EP}(H^0)$, $\min\{\varphi_0(x^0, y^0), f(x) : x \in W\} = 4.9126$, then it provides an upper bound for the global optimal value of the problem $\mathrm{EP}(H^0)$. Next, we choose the rectangle H^0 corresponding with the lower bound to branch; we obtain the following rectangles via our algorithm:

$$H^{0,1} = \begin{bmatrix} 0.0096 & 0.0144 \\ 0.0064 & 0.0140 \end{bmatrix}, \qquad H^{0,2} = \begin{bmatrix} 0.0144 & 0.0192 \\ 0.0064 & 0.0140 \end{bmatrix}. \tag{46}$$

Table 1

	ε	Approximate optimal value	Optimal value
	0.01	4.9027	4.9126
Example 1	$1.0e-3$	4.9116	4.9126
	$1.0e-4$	4.9125	4.9126

We solve the linear relaxation programming problem LRP in rectangles $H^{0,1}$ and $H^{0,2}$, respectively. In LRP($H^{0,1}$), the optimal solution and the optimal value are $[2.2524; 2.7540]$ and $v = 4.2345$; then in rectangle $H^{0,1}$, the lower bound of the original problem is $\mathrm{LB}(H^{0,1}) = 4.2345$, and the upper bound corresponding with the optimal solution is 4.9617 (>4.9126), so the upper bound is unchanged. In LRP($H^{0,2}$), the optimal solution and the optimal value are $[1.8019; 2.0032]$; and $v = 4.5548$; then in rectangle $H^{0,2}$, the lower bound of the original problem is $\mathrm{LB}(H^{0,2}) = 4.5548$, and the upper bound corresponding with the optimal solution is 4.9323 (>4.9126), so the upper bound is also unchanged. Then we choose the rectangle corresponding with the lower bound to branch until the 55th iteration, and we can obtain that

$$H^{55,1} = \begin{bmatrix} 0.0186 & 0.0189 \\ 0.0135 & 0.0137 \end{bmatrix}, \tag{47}$$

we solve the linear programming problem LRP in $H^{55,1}$; the lower bound is 4.9125; it satisfies the terminated rule. Therefore, the optimal value and the optimal solution of the original problem are 4.9126 and $x = [1.5000; 1.5000]$; the lower bound of the optimal value is 4.9125, which is approximate optimal value. The accuracy is $\varepsilon = 0.0001$.

The above example satisfies $(n, p) = (2, 2)$, where n denote the number of variables; our algorithm can have a good approach within accuracy. In Example 2, $(n, p) = (3, 3)$; in Example 3, $(n, p) = (3, 4)$ we still get good results. Along with the increase of n and p, the computation complexity is increasing. For example, in Example 3, $(n, p) = (3, 4)$, we can quickly obtain the approximate optimal value and the optimal value by using this paper's algorithm, but its effect is poorer than the former example. The result of Example 1 is shown in Table 1.

Example 2 (see [10]).

$$\begin{aligned} \min \quad & \frac{3x_1 + 5x_2 + 3x_3 + 50}{3x_1 + 4x_2 + 5x_3 + 50} + \frac{3x_1 + 4x_2 + 50}{4x_1 + 3x_2 + 2x_3 + 50} \\ & + \frac{4x_1 + 2x_2 + 4x_3 + 50}{5x_1 + 4x_2 + 3x_3 + 50}, \\ \text{s.t.} \quad & 2x_1 + x_2 + 5x_3 \le 10, \\ & x_1 + 6x_2 + 2x_3 \le 10, \\ & 9x_1 + 7x_2 + 3x_3 \ge 10, \\ & x_1, x_2, x_3 \ge 0. \end{aligned} \tag{48}$$

The optimal value is 2.8619.

Example 3 (see [10]).

$$\begin{aligned} \min \quad & \frac{4x_1 + 3x_2 + 3x_3 + 50}{3x_2 + 3x_3 + 50} + \frac{3x_1 + 4x_3 + 50}{4x_1 + 4x_2 + 5x_3 + 50} \\ & + \frac{x_1 + 2x_2 + 4x_3 + 50}{x_1 + 5x_2 + 5x_3 + 50} + \frac{x_1 + 2x_2 + 4x_3 + 50}{5x_2 + 4x_3 + 50}, \\ \text{s.t.} \quad & 2x_1 + x_2 + 5x_3 \le 10, \\ & x_1 + 6x_2 + 3x_3 \le 10, \\ & 9x_1 + 7x_2 + 3x_3 \ge 10, \\ & x_1, x_2, x_3 \ge 0. \end{aligned} \tag{49}$$

The optimal value is 3.7109.

We choose $\varepsilon = 1.0e-4$; then the approximate optimal solution satisfying accuracy and the iteration times and CPU running time are obtained. The results of our algorithm are shown in Table 2. But the results of the literature [10] are shown in Table 3.

According to Tables 2 and 3, in Example 1, although the optimal solution $(3, 4)^T$ of the literature [10] is feasible, its optimal value 5 is bigger than 4.9126 of our algorithm; in Example 2, the optimal solution $(0, 3.3333, 0)^T$ of the literature [10] turns out to be infeasible; in Example 2, the optimal value 4.0000 which corresponds to the optimal solution $(0, 0.625, 1.875)^T$ of the literature [10] is actually 3.8384, but it is still bigger than 3.7109 of our algorithm.

From the above comparison we know that the optimal values of our algorithm are much lesser than in the literature [10], and except for Example 1, the iterations of Examples 2 and 3 are much lesser than in the literature [10]. Although our running time is longer than the literature [10], if we can solve the more accurate optimal solution, the price we pay is acceptable.

In conclusion, our algorithm is feasible and effective, and to some degree, it is better than in the literature [10].

7. Conclusion

In this paper, the solving of the sum of linear ratios programming problem is discussed. The problem is equivalently transformed into bilinear programming problem, then by using the linear characteristics of convex envelope and concave envelope of double variables product, the linear relaxation programming of the bilinear programming problem is given, which can determine the lower bound of the optimal value of original problem. Therefore, a branch and bound algorithm for solving sum of linear ratios programming problem is proposed and the convergence of the algorithm is proved. Numerical results show the effectiveness of the algorithm, and our algorithm is better than the calculation results of the literature [10].

Table 2

Example	The optimal solution within accuracy or one solution among solutions		
	x_1	x_2	x_3
1	1.5000	1.5000	
2	5.0000	0.0000	0.0000
3	0.0000	1.6667	0.0000
Example	Approximate optimal value	The number of iterations	CPU (s)
1	4.9125	113	201.626020
2	2.8619	12	28.294344
3	3.7087	5	4.190375

Table 3

Example	The optimal solution within accuracy or one solution among solutions		
	x_1	x_2	x_3
1	3	4	
2	0	3.3333	0
3	0	0.625	1.875
Example	Approximate optimal value	The number of iterations	CPU (s)
1	5	32	1.089285
2	3.0029	80	8.566259
3	4.0000	58	2.968694

Acknowledgments

The work is supported by the Foundation of National Natural Science, China (11161001), and by the research project of Beifang University of Nationalities (2013XYZ025).

References

[1] H. Konno and H. Watanabe, "Bond portfolio optimization problems and their applications to index tracking: a partial optimization approach," *Journal of the Operations Research Society of Japan*, vol. 39, no. 3, pp. 295–306, 1996.

[2] J. E. Falk and S. W. Palocsay, "Optimizing the sum of linear fractional functions," in *Recent Advances in Global Optimization (Princeton, NJ, 1991)*, Princeton Series in Computer Science, pp. 221–258, Princeton University Press, Princeton, NJ, USA, 1992.

[3] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*, vol. 48 of *Nonconvex Optimization and Its Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2nd edition, 2000.

[4] H. Konno, Y. Yajima, and T. Matsui, "Parametric simplex algorithms for solving a special class of nonconvex minimization problems," *Journal of Global Optimization*, vol. 1, no. 1, pp. 65–81, 1991.

[5] H. Konno and N. Abe, "Minimization of the sum of three linear fractional functions," *Journal of Global Optimization*, vol. 15, no. 4, pp. 419–432, 1999.

[6] P.-P. Shen and C.-F. Wang, "Global optimization for sum of linear ratios problem with coefficients," *Applied Mathematics and Computation*, vol. 176, no. 1, pp. 219–229, 2006.

[7] W. Yanjun, S. Peiping, and L. Zhian, "A branch-and-bound algorithm to globally solve the sum of several linear ratios," *Applied Mathematics and Computation*, vol. 168, no. 1, pp. 89–101, 2005.

[8] H. P. Benson, "Solving sum of ratios fractional programs via concave minimization," *Journal of Optimization Theory and Applications*, vol. 135, no. 1, pp. 1–17, 2007.

[9] H. W. Jiao and Q. G. Feng, "Global optimization for sum of linear ratios problem using new pruning technique," *Mathematical Problem in Engineering*, vol. 2008, Article ID 646205, 12 pages, 2008.

[10] C.-F. Wang and P.-P. Shen, "A global optimization algorithm for linear fractional programming," *Applied Mathematics and Computation*, vol. 204, no. 1, pp. 281–287, 2008.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 619898, 8 pages
http://dx.doi.org/10.1155/2013/619898

Research Article

Tradeoff Analysis for Optimal Multiobjective Inventory Model

Longsheng Cheng,[1] Ching-Shih Tsou,[2] Ming-Chang Lee,[3] Li-Hua Huang,[4] Dingwei Song,[5] and Wei-Shan Teng[2]

[1] *School of Economics and Management, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China*
[2] *Institute of Information and Decision Sciences, National Taipei College of Business, Taipei 10051, Taiwan*
[3] *Department of Information Management, Yu Da University, Miaoli 36143, Taiwan*
[4] *Department of Accounting Information, National Taipei College of Business, Taipei 10051, Taiwan*
[5] *School of Management, Jiangsu University, Zhengjiang, Jiangsu 212013, China*

Correspondence should be addressed to Ching-Shih Tsou; cstsou@mail.ntcb.edu.tw

Received 1 February 2013; Accepted 14 April 2013

Academic Editor: Farhad Hosseinzadeh Lotfi

Deterministic inventory model, the economic order quantity (EOQ), reveals that carrying inventory or ordering frequency follows a relation of tradeoff. For probabilistic demand, the tradeoff surface among annual order, expected inventory and shortage are useful because they quantify what the firm must pay in terms of ordering workload and inventory investment to meet the customer service desired. Based on a triobjective inventory model, this paper employs the successive approximation to obtain efficient control policies outlining tradeoffs among conflicting objectives. The nondominated solutions obtained by successive approximation are further used to plot a 3D scatterplot for exploring the relationships between objectives. Visualization of the tradeoffs displayed by the scatterplots justifies the computation effort done in the experiment, although several iterations needed to reach a nondominated solution make the solution procedure lengthy and tedious. Information elicited from the inverse relationships may help managers make deliberate inventory decisions. For the future work, developing an efficient and effective solution procedure for tradeoff analysis in multiobjective inventory management seems imperative.

1. Introduction

Inventory control is an important activity that appears in any kind of organization. For this reason, it has been studied extensively in the past several decades. Most inventory models aggregate different cost concepts, such as ordering cost, carrying cost, and shortage cost, into a single-objective formulation. Optimal decisions about when to order and how much to order are then solved by single-objective optimization techniques. However the insight gained from the oldest inventory model, economic order quantity (EOQ), reveals that inventory management should be considered as a biobjective optimization problem to strike a balance between inventory carrying and annual orders. Practically speaking, inventory decisions involve tradeoffs related to operational efficiency and customer service.

Brown [1, 2] first examined the tradeoff between investment in working stock and annual ordering cost. He introduced the concept of exchange curve shown in Figure 1. The curve demonstrates how capital invested in working stocks can be traded for operating expenses of ordering. Points below the curve are infeasible, and decisions located above the curve are suboptimal. Suboptimal policies can be improved by moving back to the curve (i.e., seeking possible improvement from point A to B or C). Starr and Miller [3] determined tradeoffs between two performance measures: (i) number of orders per year (workload) and (ii) average investment in inventory in the case of multiple items. Gardner and Dannenbring [4] introduced customer service as another measure, along with workload and inventory investment, and generalized above exchange curve analysis to the optimal policy surface in case of probabilistic demand.

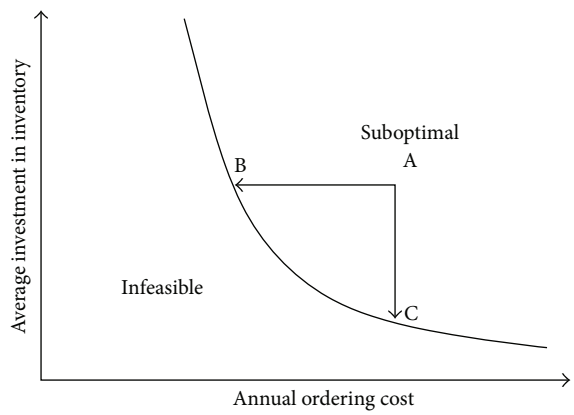

Figure 1: The exchange curve for deterministic inventory model.

The model and solution technique they used is still based on single-objective optimization. Bookbinder and Chen [5] proposed a multiobjective formulation for analyzing multi-echelon inventory and distribution systems. They argued that points on exchange curve or policy surface are equivalent to the nondominated solutions concept of multiobjective optimization. Although it could be the first multiple criteria generalizations of earlier studies, the model was solved by classical optimization techniques.

To a certain extent, the aforementioned tradeoff analysis of inventory management is developed by single-objective optimization. The motivation of this study aims to develop an intrinsically multiobjective approach for building the tradeoff surface of probabilistic inventory systems. Differences between traditional and multiobjective approach are not only in their problem formulations, but also the latter simultaneously treats several objectives analytically or heuristically under certain notion of multiobjective optimality [6, 7].

Gutiérrez et al. [8] considered a dynamic single facility single-item lot size problem. Although the total demand is assumed to be a fixed value, the distribution of this demand among different periods is unknown. They determined all the Pareto-optimal or nondominated production plans that are robust to any possible occurrence of all scenarios. Gutiérrez et al. [9] presented the characterization of the nondominated optimal solution set and use it to correct the solution method previously proposed by Bookbinder and Chen [5].

For the multiobjective exchange curve, Tsou [7] presented a two-stage framework consisting of multiobjective particle swarm optimization (MOPSO) and technique for order preference by similarity to ideal solution (TOPSIS). At the first stage, MOPSO is used to generate the tradeoff (or nondominated) front of the triobjective inventory model in Agrell [10]. Then, a preferred solution is selected by TOPSIS according to subjective preferences of decision makers. Tsou and Kao [11] also developed a metaheuristic based on electromagnetism-like mechanism (EM) to approximate the Pareto-optimal front without using any prior or interactive preference. They showed that the metaheuristic can find similar Pareto-optimal solutions as the popular interactive procedure Step method (STEM) did [12]. Tsou [13] further showed that evolutionary Pareto optimizers could generate tradeoff solutions potentially ignored by the well-known simultaneous method.

Nevertheless, we recently notice that the tradeoff solutions of the above studies actually laid on an exchange curve, instead of forming a tradeoff surface in the 3D objective space. It apparently indicates that some of the objectives, including minimization of expected annual cost, expected annual number of stockout occasions, and expected annual number of items stocked out, are not conflicting with each other. Among which, the last two objectives are redundant because they relate to the same concept of shortage but different measures. Consequently, such a kind of triobjective models was not properly justified in the above studies.

This paper first presents a triobjective model without redundancy in the next section. Nonredundancy is assured by dropping all the marginal cost parameters out of the classical fixed order model. After that, a successive approximation procedure based on the Lagrange method is utilized to iteratively search for nondominated solutions and efficient control policies. Tradeoffs among workload, inventory, and shortage are visualized by three-dimensional scatterplots. Although it is a time-consuming job to use the successive approximation to find the tradeoff surfaces of multiobjective model, all solutions found are ensured to be Pareto-optimal in comparison with other search methods, such as genetic algorithms. Finally, conclusions and directions for future research are drawn out accordingly.

2. Model Building and Solution Procedures

2.1. A Triobjective Model. The reorder point lot size system, (r, Q), is a popular control method under probabilistic demand. An order of size Q will be triggered immediately whenever the inventory position drops to the reorder point r or lower. Classical (r, Q) model minimizes a lump-sum cost including ordering cost, carrying cost, and stockout cost [14]. The triobjective model below intrinsically restores the nature of conflicts among objectives that are to minimize the workload, inventory, and shortage. Also, multiobjective (r, Q) model does not run into the incommensurate issue while aggregating objectives of different measures into a single one. The notations used here are described as follows.

D is the average annual demand,

L is the lead time,

D_L is the lead time demand. It is normally distributed with mean μ_L and standard deviation σ_L,

SS is the safety stock, which is proportional to the standard deviation of lead time demand. That is, $SS = k\sigma_L$, where k represents the safety factor,

r is the aforementioned reorder point, which equals to the average lead time demand plus the safety stock. That is, $r = \mu_L + k\sigma_L$, and $\varphi(z)$ is the probability density function of standard normal distribution.

A multiobjective (r, Q) model is formulated as follows:

$$\underset{k,Q}{\text{Min}} \quad W = \frac{D}{Q}, \tag{1}$$

$$\underset{k,Q}{\text{Min}} \quad I = \frac{Q}{2} + k\sigma_L, \tag{2}$$

$$\underset{k,Q}{\text{Min}} \quad S = \frac{D\sigma_L}{Q} \int_k^{\infty} (z-k)\,\varphi(z)\,dz \tag{3}$$

subject to the following:

$$0 \le Q \le D, \tag{4}$$

$$0 \le k \le \frac{D}{\sigma}. \tag{5}$$

Equation (1) represents the number of annual order (in cycles per year, also called workload). Equation (2) is the sum of cycle and safety stocks (in units carrying per year). Equation (3) denotes the average number of demand not covered from stock annually (in units short per year). Inequality (4) ensures that the order size (units per order) should be nonnegative and not more than the average annual demand. Inequality (5) guarantees that the safety stock (in units) will not be greater than the average annual demand and should be nonnegative.

The notion of optimality in single-objective optimization is straightforward, because the optimal solution is the one that realizes the maximum (or the minimum) of the objective function. However, the optimality for a multiobjective optimization problem is not so easy to understand because not all feasible solutions can be compared completely. Generally speaking, multiobjective optimization problems rarely have solutions that simultaneously optimize all of the objectives; as a result we are trying to optimize each objective to the greatest extent possible. There exists a set of solutions, referred as nondominated solutions, which are better than others in the search space when considering all the objectives. For the minimization problem (in Section 2.1), a control parameter $x^1 = (k_1, Q_1)$ is said to strongly dominate $x^2 = (k_2, Q_2)$ (denoted by $x^1 \prec x^2$) if and only if $W(x^1) < W(x^2)$, $I(x^1) < I(x^2)$, and $S(x^1) < S(x^2)$. That is, solution x^1 is strictly better than solution x^2 in all the cost and service objectives ([15, pp. 32]). Less stringently, a decision vector x^1 dominates x^2 (denoted by $x^1 \preceq x^2$) if and only if $W(x^1) \le W(x^2)$, $I(x^1) \le I(x^2)$, $S(x^1) \le S(x^2)$ and at least one of above inequality is strictly held ([15, pp. 28]). For other multiobjective (or multicriteria) concepts, please refer to Ehrgott [16].

Undoubtedly, we are not interested in solutions dominated by other solutions. Solutions that are not dominated by any other solutions are called nondominated in objective space or efficient in decision space. It means that the improvement of some objective could only be achieved at the expense of other objectives. This coincides with the exchange curve concept mentioned earlier. In a multiobjective optimization problem, there are normally a large number of nondominated solutions due to the conflicts among objectives. Hence, it is difficult to find the whole set of nondominated solutions. And because the nondominated set is usually unknown, most optimizers try to find a finite number of nondominated solutions to approximate the set. The successive approximation approach stated below can be used to search for the nondominated solutions of the triobjective inventory model.

2.2. Successive Approximation Based on Lagrange Method. A single objective transformation is first developed as follows. Equation (3) is kept as the objective function and treats (1) and (2) as constraints. That is,

$$\begin{aligned} &\underset{k,Q}{\text{Min}} \quad S \\ &\text{subject to} \quad W = W' \\ &\qquad\qquad\quad I = I', \end{aligned} \tag{6}$$

where W' and I' are budgets on workload and inventory.

To solve this equality constrained optimization problem, the Lagrange method is employed here. After introducing the Lagrangian multipliers λ_W and λ_I, the Lagrangian function is as follows:

$$\begin{aligned} L(k, Q, \lambda_W, \lambda_I) = {} & \frac{D\sigma_L}{Q} \int_k^{\infty} (z-k)\,\varphi(z)\,dz \\ & + \lambda_I \left(\frac{Q}{2} + k\sigma_L - I' \right) + \lambda_W \left(\frac{D}{Q} - W' \right). \end{aligned} \tag{7}$$

Some simplifying notations are introduced before presenting the successive approximation algorithm. Let

$$P = \int_k^{\infty} \varphi(z)\,dz, \tag{8}$$

$$E = \sigma_L \int_k^{\infty} (z-k)\,\varphi(z)\,dz, \tag{9}$$

P is the probability of a stockout per order cycle and E is the expected number of shortage per order cycle. Simple algebra provides the following equations used in the successive approximation:

$$\lambda_I = \frac{DP}{2(I' - k\sigma)}, \tag{10}$$

$$P = \frac{\lambda_I Q}{D}, \tag{11}$$

$$\lambda_W = \frac{1}{W'} \left(\frac{\lambda_I Q}{2} - \frac{DE}{Q} \right), \tag{12}$$

$$Q = \sqrt{\frac{2D(E + \lambda_W)}{\lambda_I}}. \tag{13}$$

To search for the efficient (k_i, Q_i) policy, the search steps are proposed as follows.

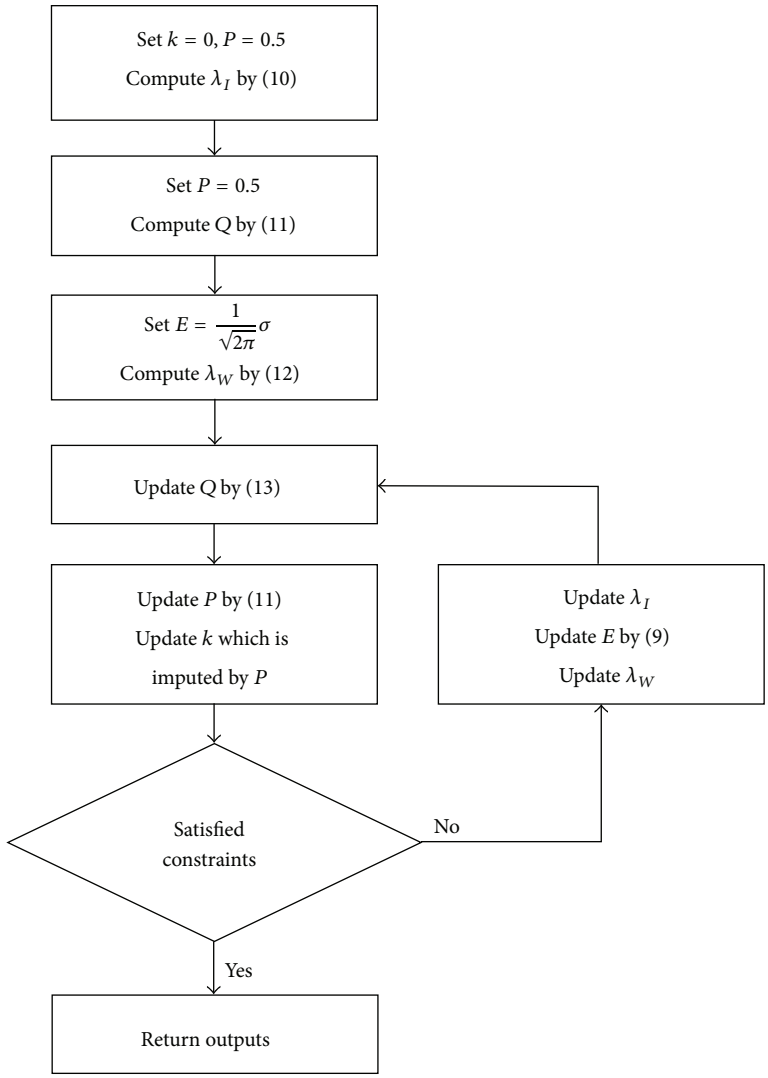

Figure 2: Flow chart of the search algorithm.

Step 1. Initialize λ_I: Compute λ_I using (10) with $k = 0$ and $P = 0.5$.

Step 2. Initialize Q: Compute Q using (11) with $P = 0.5$.

Step 3. Initialize λ_W: Compute λ_W using (12) with $E = (1/\sqrt{2\pi})\sigma$ corresponds to zero safety stock.

Step 4. Update Q: Compute Q using (13).

Step 5. Update P and k: Compute P with (11) and look up k imputed by P.

Step 6. Check constraints: If both workload and investment constraints are satisfied, then output the results. Otherwise, update λ_I, E (using (9)), and λ_W, then go back to Step 4.

The search begins with an initial guess of zero safety stock. This allows us to use (10) to derive an initial λ_I, which in turn, determines the initial Q using (11). However, (12)-(13) for λ_W and Q are interdependent, preventing their use in the initialization phase. Rearranging (11), however, we can derive an equation for Q which does not require an estimate of λ_W. The Q based on (11) is then used to provide an initial estimate of λ_W from (13). Thereafter the search progresses by iteratively updating values for Q, P (and correspondingly k), λ_I, and λ_W, using (13), (11), (10), and (12), respectively, until both the workload and investment constraints are satisfied. The flow chart of the above search algorithm is shown in Figure 2.

3. Numerical Results

Pharmaceutical inventory data with four items (Table 1) were fed into the triobjective model. The successive approximation was coded in R [17], and all computation was executed on a laptop computer. Ten representative solutions for each

Table 1: The pharmaceutical data.

Item	D	σ_L
1	3412	53.354
2	490	5.027
3	4736	57.911
4	200	2.969

Table 2: Tradeoff solutions of Item 1 generated by successive approximation.

Sol. no.	Efficient solution		Nondominated solution			Iter
	Q	k	W	I	S	
1	94.804	0.049	35.990	50.000	720.207	81*
2	162.532	0.351	20.993	99.992	277.522	20†
3	98.804	0.986	35.990	99.999	164.354	34
4	83.238	1.094	40.991	99.999	151.799	41
5	162.51	1.287	20.996	149.931	52.407	20†
6	94.798	1.923	35.992	149.993	19.999	27
7	131.234	2.517	25.999	199.92	2.635	26
8	213.250	2.686	16.000	249.915	0.948	41
9	110.065	3.653	31.000	249.922	0.052	32
10	110.065	4.590	31.000	299.911	0.001	40
Min.	83.238	0.049	16.000	50.000	0.001	20
Max.	213.250	4.590	40.991	299.911	720.207	81

† and * represent the minimum and the maximum numbers of iterations, respectively.

item generated by successive approximation are shown in Tables 2, 3, 4, and 5. The columns of nondominated solutions demonstrate that the improvement of some objective(s) could only be achieved at the expense of other objectives. For example, solution 3, compared to solution 4, in Table 2 gets better in workload at the expense of shortage.

Three-dimensional scatterplots for each item are illustrated in Figures 3, 4, 5, and 6. Any one can visually check the tradeoffs displayed in scatterplots by adding planes parallel to I-S or W-S plane. With a fixed workload, expected shortage decreases as expected inventory increases. At a fixed inventory level, increases in workload lead to a reduction of expected shortage. All these findings are intuitive and straightforward.

For the computation effort, we notice that only one solution in Table 2 can be obtained after at least twenty iterations. And the largest iterations to reach a nondominated solution is eighty-one. Ranges of other items are between eight to sixty-six iterations. Hence, creating the scatterplot of workload, inventory, and shortage by successive approximation is lengthy and tedious.

The quality of a set of tradeoff solutions is evaluated quantitatively in terms of accuracy and diversity. A metric called hypervolume (H) is used to demonstrate the accuracy of the nondominated solutions. It calculates the size of the area that is dominated by a nondominated set and is defined as follows. The idea is that the larger the area the solutions can

Table 3: Tradeoff solutions of Item 2 generated by successive approximation.

Sol. no.	Efficient solution		Nondominated solution			Iter
	Q	k	W	I	S	
1	18.8519	0.114	25.9921	9.9989	45.0181	15†
2	23.3406	0.6575	20.9935	14.9757	16.193	15†
3	30.6264	0.9159	15.9992	19.9176	7.8443	17
4	13.6139	1.6284	35.9927	14.9929	3.9316	21
5	11.9541	1.794	40.9901	14.9954	2.9865	22
6	18.8469	2.0863	25.999	19.9113	0.8779	21
7	15.8077	2.3897	30.9976	19.917	0.4372	21
8	11.9532	2.781	40.9934	19.9568	0.1671	22
9	18.8462	3.0794	26	24.9033	0.0376	30
10	30.625	3.8973	16	34.9042	0.0009	60*
Min.	11.9532	0.114	15.9992	9.9989	0.0009	15
Max.	30.6264	3.8973	40.9934	34.9042	45.0181	60

† and * represent the minimum and the maximum numbers of iterations, respectively.

Table 4: Tradeoff solutions of Item 3 generated by successive approximation.

Sol. no.	Efficient solution		Nondominated solution			Iter
	Q	k	W	I	S	
1	296.134	0.032	15.993	149.924	354.824	8†
2	430.546	0.598	11	249.913	107.77	25
3	789.333	0.954	6	449.91	31.58	59*
4	152.82	1.27	30.991	149.964	87.038	21
5	225.525	1.505	21	199.918	35.237	25
6	152.781	2.133	30.999	199.911	10.597	24
7	115.534	2.455	40.992	199.945	5.463	24
8	296	2.623	16	299.901	1.26	50
9	131.556	3.18	36	249.913	0.416	30
10	152.774	3.86	31	299.91	0.024	41
Min.	115.512	0.032	6	149.924	0.024	8
Max.	789.333	3.86	40.992	499.91	354.824	59

† and * represent the minimum and the maximum numbers of iterations, respectively.

dominate in the objective space, the better it is [18]:

$$\prod_{i=1}^{M}\left(f_i^{\max}-f_i^{\min}\right), \tag{14}$$

where M is the number of objectives. Figure 7 shows the pictorial explanation of H in which O' represents the reference point and S is the nondominated set.

Keeping the nondominated set as diverse as possible is very important. Here spacing (S) and maximum spread (D) are used to evaluate the distribution and spread of

Table 5: Tradeoff solutions of Item 4 generated by successive approximation.

Sol. no.	Efficient solution		Nondominated solution			Iter
	Q	k	W	I	S	
1	18.1965	0.2896	10.9911	9.9582	8.8351	10†
2	5.5568	0.7482	35.9916	5	14.059	36
3	33.3333	1.0901	6	19.9032	1.2465	28
4	12.5064	1.2415	15.9918	9.9393	2.4446	15
5	9.527	1.7531	20.993	9.9686	1.0003	18
6	7.6947	2.0674	25.9918	9.9854	0.546	20
7	5.5571	2.4315	35.9902	9.9978	0.2642	24
8	33.3333	2.7747	6	24.9047	0.0148	66⋆
9	12.5	2.9139	16	14.9015	0.0245	28
10	7.6925	3.7241	25.9993	14.9032	0.0018	25
Min.	5.5568	0.2896	6	5	0.0018	10
Max.	33.3333	3.7241	35.9916	24.9047	14.059	66

† and ⋆ represent the minimum and the maximum numbers of iterations, respectively.

the nondominated fronts generated by successive approximation:

$$S = \sqrt{\frac{1}{|\widetilde{A}|}\sum_{i=1}^{|\widetilde{A}|}\left(d_i - \overline{d}\right)^2},$$
$$D = \sqrt{\sum_{k=1}^{3}\left(\max_{i=1}^{|\widetilde{A}|} f_k^i - \min_{i=1}^{|\widetilde{A}|} f_k^i\right)^2}, \tag{15}$$

where $d_i = \min_{j\in\widetilde{A}\wedge j\neq i}\sum_{k=1}^{3}|f_k^i - f_k^j|$, f_k^i represents the kth criterion function value of the nondominated solution i, and $\overline{d}$ is the mean value of the absolute distance measure where $\overline{d} = \sum_{i=1}^{|\widetilde{A}|}(d_i/|\widetilde{A}|)$.

Larger above measures are better except for the spacing. Table 6 shows the results of the successive approximation method. If there is a reference solutions set known to decision makers or generated by other solution procedures, one can use the figures in Table 6 to compare successive approximation with the benchmark that he/she is interested in. After verifying the validity of nondominated (W, I, S) solutions generated by successive approximation, they can be used to construct a tradeoff surface for inventory control. It helps managers choose an appropriate control policy for probabilistic demand, such as the fund level tied in inventory versus the service level under fixed workload.

4. Conclusions and Suggestions

Tradeoff analysis in inventory management is useful in quantifying what the firm must pay in terms of workload and investment to meet the desired customer service. A triobjective model is presented first to generate the efficient (k, Q) policies in decision space that correspond to the nondominated (W, I, S) solutions in objective space. Nonredundancy is assured by dropping all the marginal cost parameters out of the classical fixed order model. Such that

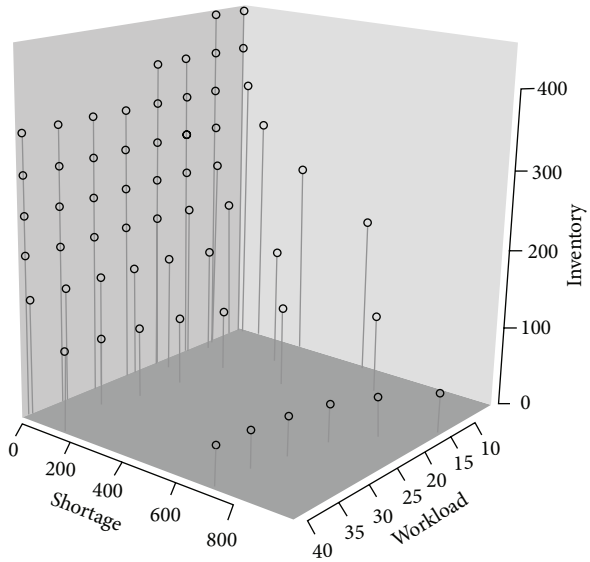

Figure 3: Scatterplot of Item 1 nondominated solutions from successive approximation.

Table 6: Performance measures of the successive approximation approach.

ID	Accuracy	Distribution	
	H	Spacing	Spread
1	42,694,515	0.091	1.532
2	7,215,908	0.097	1.232
3	68,798,878	0.121	1.637
4	476,041	0.105	1.362

leads to a triobjective model which intends to minimize the workload, inventory, and shortage that are all conflicting with each other.

To solve the triobjective model, successive approximation approach is employed in this paper. The successive approximation is in an attempt to build the optimal tradeoff surface under probabilistic demand. Successive approximation can be used to obtain the nondominated solutions of workload, inventory, and shortage. Specifically, lots of trial and error involve in deriving the edge of each objective with a Lagrangian model. Several iterations needed to reach a nondominated solution for all items make the creating of scatterplot of workload, inventory, and shortage by successive approximation lengthy and tedious. However, visualization of the tradeoffs displayed by the scatterplots of Figures 3, 4, 5, and 6 justifies the computation effort done in the experiment. The inverse relationship among workload, inventory, and shortage conforms to our intuition.

The quality of a set of tradeoff solutions has to be evaluated quantitatively when comparing to other benchmarks, although developing an efficient and effective solution procedure for tradeoff analysis of multiobjective inventory management will be our future work. After verifying the validity of nondominated (W, I, S) solutions, they can be used to construct a tradeoff surface that helps managers choose

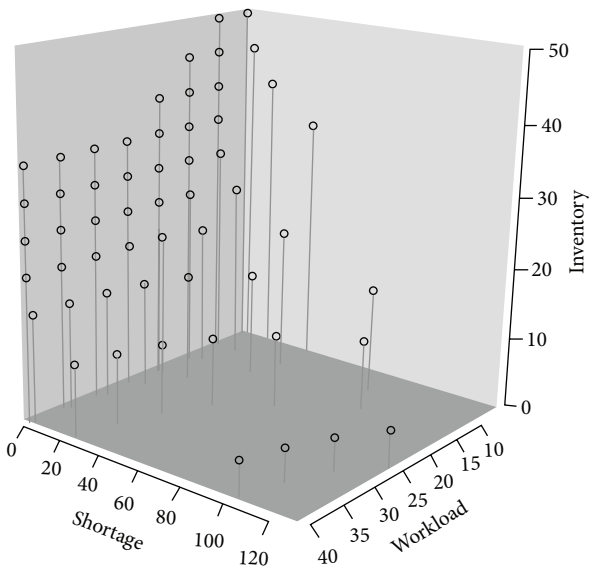

Figure 4: Scatterplot of Item 2 nondominated solutions from successive approximation.

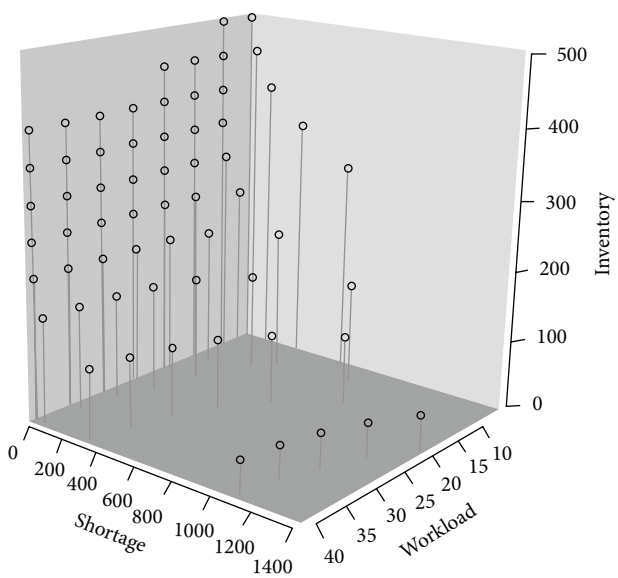

Figure 5: Scatterplot of Item 3 nondominated solutions from successive approximation.

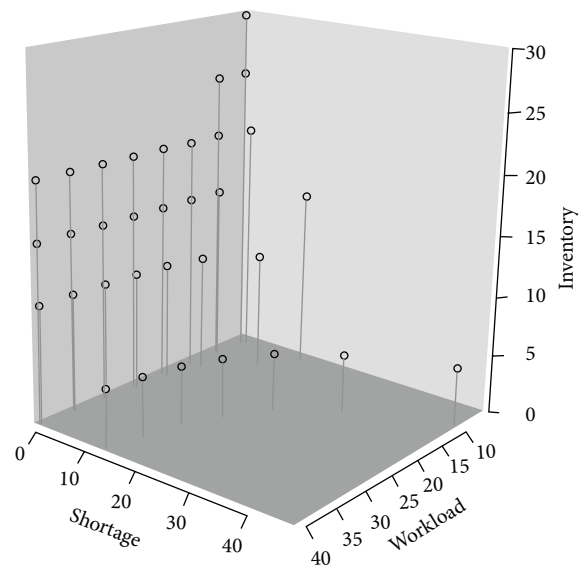

Figure 6: Scatterplot of Item 4 nondominated solutions from successive approximation.

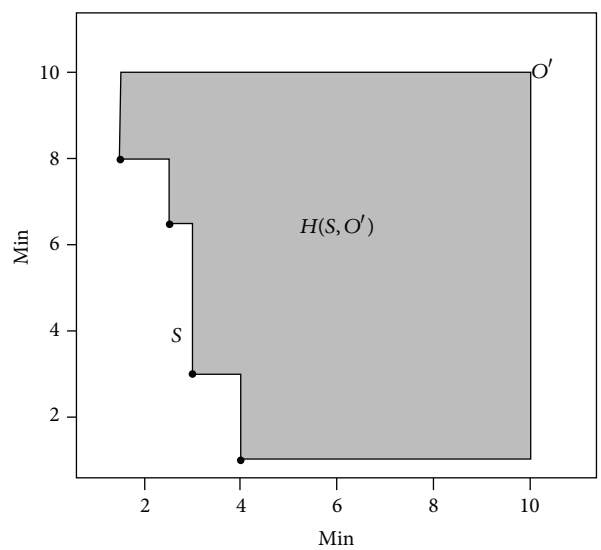

Figure 7: The hypervolume index in the minimization problem.

an appropriate control policy for probabilistic demand. For example, a common debate on the fund level tied in inventory versus the service level under fixed workload usually arises between warehousing and sales departments. By utilizing the information coming from the tradeoff surface, deliberate decisions among conflicting objectives can be easily made.

Moreover, the tradeoff surface under multi-item context deserves closer attention to bridge the gap between inventory theory and managerial practice, because too much attention was focused on the single-item problem in view of the past literature. Finally, multiechelon inventory and/or distribution systems are very common in business logistics. It is worthwhile to study the multiobjective inventory policies of different parties in a supply chain.

Acknowledgments

The authors would like to express special thanks to the National Science Council of Taiwan for financial support in NSC100-2628-H-141-001-MY2. This research was also partially supported by the National Natural Science Foundation of China (Grant 71271114).

References

[1] R. G. Brown, "Use of carrying charge to control cycle stocks," *APICS Quarterly Bulletin*, vol. 2, pp. 29–46, 1961.

[2] R. G. Brown, *Decision Rules for Inventory Management*, Dryden Press, Illinois, Ill, USA, 1967.

[3] M. K. Starr and D. W. Miller, *Inventory Control: Theory and Practice*, Prentice Hall, New Jersey, NJ, USA, 1962.

[4] E. S. Gardner and D. G. Dannenbring, "Using optimal policy surfaces to analyze aggregate inventory tradeoffs," *Management Science*, vol. 25, no. 8, pp. 709–720, 1979.

[5] J. Bookbinder and V. Y. X. Chen, "Multicriteria trade-offs in a warehouse/retailer system," *Journal of the Operational Research Society*, vol. 43, no. 7, pp. 707–720, 1992.

[6] J. Puerto and F. R. Fernández, "Pareto-optimality in classical inventory problems," *Naval Research Logistics*, vol. 45, no. 1, pp. 83–98, 1998.

[7] C. S. Tsou, "Multi-objective inventory planning using MOPSO and TOPSIS," *Expert Systems with Applications*, vol. 35, no. 1-2, pp. 136–142, 2008.

[8] J. Gutiérrez, J. Puerto, and J. Sicilia, "The multiscenario lot size problem with concave costs," *European Journal of Operational Research*, vol. 156, no. 1, pp. 162–182, 2004.

[9] J. Gutiérrez, J. Puerto, and J. Sicilia, "Determining the Pareto set in a bicriteria two-echelon inventory/distribution system," *Optimization*, vol. 59, no. 2, pp. 253–271, 2010.

[10] P. J. Agrell, "A multicriteria framework for inventory control," *International Journal of Production Economics*, vol. 41, no. 1–3, pp. 59–70, 1995.

[11] C. S. Tsou and C. H. Kao, "Multi-objective inventory control using electromagnetism-like meta-heuristic," *International Journal of Production Research*, vol. 46, no. 14, pp. 3859–3874, 2008.

[12] R. Benayoun, J. de Montgolfier, J. Tergny, and O. Laritchev, "Linear programming with multiple objective functions: step method (stem)," *Mathematical Programming*, vol. 1, no. 1, pp. 366–375, 1971.

[13] C. S. Tsou, "Evolutionary Pareto optimizers for continuous review stochastic inventory systems," *European Journal of Operational Research*, vol. 195, no. 2, pp. 364–371, 2009.

[14] E. A. Silver, D. F. Pyke, and R. Peterson, *Inventory Management and Production Planning and Scheduling*, John Wiley & Sons, New York, NY, USA, 1998.

[15] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, Chichester, UK, 2001.

[16] M. Ehrgott, *Multicriteria Optimization*, Springer, Berlin, Germany, 2nd edition, 2005.

[17] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009, http://www.r-project.org/.

[18] T. Okabe, Y. Jin, and B. Sendhoff, "A critical survey of performance indices for multi-objective optimization," in *Proceedings of the IEEE Congress on Evolutionary Computation*, vol. 2, pp. 878–885, IEEE Press, New Jersey, NJ, USA, 2003.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 827268, 8 pages
http://dx.doi.org/10.1155/2013/827268

Research Article

A Fuzzy Multicriteria Group Decision-Making Method with New Entropy of Interval-Valued Intuitionistic Fuzzy Sets

Xiaohong Chen, Li Yang, Pei Wang, and Wei Yue

Business School, Central South University, Changsha 410083, China

Correspondence should be addressed to Li Yang; yangli@csu.edu.cn

Received 14 September 2012; Accepted 30 March 2013

Academic Editor: Farhad Hosseinzadeh Lotfi

A new entropy measure of interval-valued intuitionistic fuzzy set (IVIFS) is proposed by using cotangent function, which overcomes several limitations in the existing methods for calculating entropy of IVIFS. The efficiency of the new entropy is demonstrated by comparing it with several classical entropies. Moreover, an entropy weight model is established to determine the entropy weights for fuzzy multicriteria group decision-making (FMCGDMs) problems, which depends on incomplete weight information of criteria in IVIFSs setting. Finally, an illustrative supplier selection problem is used to demonstrate the practicality and effectiveness of the proposed method. It is capable of the handling the FMCGDM problems with incomplete known weights for criteria.

1. Introduction

The theory of fuzzy sets (FSs) proposed by Zadeh [1] has achieved a great success in various fields. A lot of generalized forms of FSs have been proposed. The classical sets include interval-valued fuzzy sets (IVFSs) [2], intuitionistic fuzzy sets (IFSs) [3, 4], interval-valued intuitionistic fuzzy sets (IVIFSs) [5], R-fuzzy sets [6], and intuitionistic linguistic fuzzy sets (ILFSs) [7]. From [4, 8–10], it turns out that IVFS theory is equivalent to IFS theory, and IVIFS theory extends IFS theory.

As an important topic in the theory of fuzzy sets, entropy measures of IFSs have been investigated widely by many researchers from different views. Burillo and Bustince [11] introduced the notion that entropy of IVFSs and IFSs can be used to evaluate the degree of intuitionism of an IVFS or IFS. Szmidt and Kacprzyk [12] proposed a nonprobabilistic-type entropy measure with a geometric interpretation of IFSs. Hung and Yang [13] gave their axiomatic definitions of entropy of IFSs and IVFSs by using the concept of probability. Wei et al. [14] gave a new entropy measure for IVIFSs to overcome the disadvantages of those three entropy measures defined independently by Szmidt and Kacprzyk [12], Wang and Lei [15], and Huang and Liu [16]. Different entropy formulas for IFS [15, 17], IVFS [18, 19], and vague set [16, 20, 21] were also proposed by other researchers.

The entropy of IFSs has been applied widely in decision making [22, 23]. On the one hand, due to the increasing complexity of the social-economic environment and a lack of information about the problem domains, the decision information may be provided with IVIFSs, whose membership degree and nonmembership degree are intervals, instead of real numbers. Entropy is concerned as a measure of fuzziness. Therefore, it is highly necessary and significant to study the entropy of IVIFSs. And on the other hand, a proper assessment of attribute weights plays an essential role in the MADM process [24]. In terms of determining weights, the entropy method is one of the most representative approaches, which expresses the relative intensities of attribute importance to signify the average intrinsic information transmitted to the DM [22, 25, 26]. The following are some of the research findings.

Ye [27] proposed two entropy measures for IVIFSs and established an entropy weight model, which could be used to determine the criteria weights on alternatives. Zhang et al. [28] proposed a new information entropy measure of IVIFS by using membership interval and nonmembership interval of IVIFSs, which complied with the extended form of De Luca and Termini [29] axioms for fuzzy entropy. Wei et al. [14] also proposed an entropy measure for IVIFSs, and they applied the new entropy measure to solve problem on multicriteria fuzzy decision making.

However, the entropy is seldom applied in multiexpert and multicriteria decision making, which is one of the most important branches of decision-making method. Due to limited investigation on multiexpert and multicriteria decision-making issues, this study proposed a novel formula to calculate the entropy of an IVIFS on the basis of the argument on the relationship among the entropies of IFSs given in [27, 30]. For interval-valued intuitionistic fuzzy multicriteria group decision-making problem, in which the information on the weights of criteria is incomplete, a linear fuzzy programming model based on intuitionistic fuzzy entropy is constructed to obtain the criteria weights.

The rest of this paper is organized as follows. In Section 2, we introduce some basic notions of IFS and IVIFSs. In Section 3, we propose a new entropy measure of interval-valued intuitionistic fuzzy set by using cotangent function. In Section 4, the method and procedure for solving FMCGDM problems with the new entropy measure of interval-valued intuitionistic fuzzy set are developed in detail. An illustrative supplier selection problem was employed to demonstrate how to apply the proposed approach in Section 5. Short conclusion is given in Section 6.

2. Interval-Valued Intuitionistic Fuzzy Sets

Definition 1 (see [3]). Let X be an ordinary finite nonempty set. An intuitionistic fuzzy set (IFS) in X is an object of the form:

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle \mid x \in X\}, \tag{1}$$

where

$$\mu_A : X \longrightarrow [0,1], \qquad \nu_A : X \longrightarrow [0,1] \tag{2}$$

with the condition $0 \le \mu_A(x) + \nu_A(x) \le 1$, for all $x \in X$.

The numbers $\mu_A(x)$, $\nu_A(x)$ denote the degree of membership and nonmembership of the element x in the set A, respectively.

For each IFS, we call $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$ the intuitionistic index of the element x in the set A. It also denotes the hesitancy degree of x to A.

Definition 2 (see [5]). Let $D[0,1]$ be the set of all closed subintervals of the interval $[0,1]$, and let X be an ordinary finite nonempty set. An interval-valued intuitionistic fuzzy set (IVIFS) in X is an object of the form:

$$A = \{\langle x, \tilde{\mu}_A(x), \tilde{\nu}_A(x)\rangle \mid x \in X\}, \tag{3}$$

x in the set A, where $\tilde{\mu}_A : X \to D[0,1]$, $\tilde{\nu}_A : X \to D[0,1]$, with the condition $0 \le \sup(\tilde{\mu}_A(x)) + \sup(\tilde{\nu}_A(x)) \le 1$, for all $x \in X$.

The intervals $\tilde{\mu}_A(x)$, $\tilde{\nu}_A(x)$ denote the degree of membership and nonmembership of the element x in the set A, respectively.

For convenience, let $\tilde{\mu}_A(x) = [\mu_{AL}(x), \mu_{AU}(x)]$, $\tilde{\nu}_A(x) = [\nu_{AL}(x), \nu_{AU}(x)]$, then

$$A = \{\langle x, [\mu_{AL}(x), \mu_{AU}(x)], [\nu_{AL}(x), \nu_{AU}(x)]\rangle \mid x \in X\}, \tag{4}$$

where $0 \le \mu_{AU}(x) + \nu_{AU}(x) \le 1$, $\mu_{AL}(x) \ge 0$, and $\nu_{AL}(x) \ge 0$.

For each element x, we can compute the intuitionistic index of an intuitionistic fuzzy interval of $x \in X$ in A defined as follows:

$$\begin{aligned}\tilde{\pi}_A(x) &= 1 - \tilde{\mu}_A(x) - \tilde{\nu}_A(x)\\ &= [1 - \mu_{AU}(x) - \nu_{AU}(x), 1 - \mu_{AL}(x) - \nu_{AL}(x)].\end{aligned} \tag{5}$$

For convenience, an IVIFS value is denoted by $A = ([a,b], [c,d])$.

Definition 3 (see [5]). Assume $A, B \in \mathrm{IVIFS}(X)$, then some operations can be defined as follows:

$$\begin{aligned}A \cup B = {} & \{\langle x_i, [\mu_{AL}(x_i) \vee \mu_{BL}(x_i), \mu_{AU}(x_i) \vee \mu_{BU}(x_i)],\\ & [\nu_{AL}(x_i) \wedge \nu_{BL}(x_i), \nu_{AU}(x_i) \wedge \nu_{BU}(x_i)]\rangle\},\\ A \cap B = {} & \{\langle x_i, [\mu_{AL}(x_i) \wedge \mu_{BL}(x_i), \mu_{AU}(x_i) \wedge \mu_{BU}(x_i)],\\ & [\nu_{AL}(x_i) \vee \nu_{BL}(x_i), \nu_{AU}(x_i) \vee \nu_{BU}(x_i)]\rangle\}.\end{aligned} \tag{6}$$

The following expressions are defined in [5] for all $A, B \in \mathrm{IVFSs}(X)$:

$A \subseteq B$ if and only if $\mu_{AL} \le \mu_{BL}$, $\mu_{AU} \le \mu_{BU}$, $\nu_{AL} \ge \nu_{BL}$, and $\nu_{AU} \ge \nu_{BU}$ for all $x \in X$;

$A = B$ if and only if $A \subseteq B$ and $B \subseteq A$ for all $x \in X$;

$A^C = \{\langle x, [\nu_{AL}(x), \nu_{AU}(x)], [\mu_{AL}(x), \mu_{AU}(x)]\rangle \mid x \in X\}$.

In the following, we introduce two weighted aggregation operators related to IVIFSs.

Definition 4 (see [31]). Let A_j $(j = 1, 2, \ldots, n) \in \mathrm{IVIFS}(X)$. The weighted geometric average operator (IVIF-WGA operator) is defined by

$$\begin{aligned}&F_w(A_1, A_2, \ldots, A_n)\\ &= \left(\left[\prod_{j=1}^{n} \mu_{A_jL}(x)^{w_j}, \prod_{j=1}^{n} \mu_{A_jU}(x)^{w_j}\right],\right.\\ &\qquad \left.\left[1 - \prod_{j=1}^{n}\left(1 - \nu_{A_jL}(x)\right)^{w_j}, 1 - \prod_{j=1}^{n}\left(1 - \nu_{A_jU}(x)\right)^{w_j}\right]\right),\end{aligned} \tag{7}$$

where w_j is the weight of A_j $(j = 1, 2, \ldots, n)$, $w_j \in [0,1]$ and $\sum_{j=1}^{n} w_j = 1$. Particularly, if $w_j = 1/n$ $(j = 1, 2, \ldots, n)$, then F_w is called a geometric average operator for IVIFSs.

Definition 5 (see [31]). Let A_j $(j = 1, 2, \ldots, n) \in \mathrm{IVIFS}(X)$. The hybrid averaging operator (IVIF-HA operator) is defined by

$$H_w(A_1, A_2, \ldots, A_n) = \left(\left[\prod_{j=1}^{n} \mu_{B_jL}(x)^{w_j}, \prod_{j=1}^{n} \mu_{B_jU}(x)^{w_j} \right], \left[1 - \prod_{j=1}^{n} \left(1 - \nu_{B_jL}(x)\right)^{w_j}, 1 - \prod_{j=1}^{n} \left(1 - \nu_{B_jU}(x)\right)^{w_j} \right] \right), \tag{8}$$

where w_j is the weight of B_j $(j = 1, 2, \ldots, n)$, $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$. B_j is the jth largest one of all values $(nw_1A_1, nw_2A_2, \ldots, nw_nA_n)$, and ω_k is the weight of A_k $(k = 1, 2, \ldots, n)$, satisfying $\omega_k \in [0, 1]$ and $\sum_{k=1}^{n} \omega_k = 1$. n is a balance factor. Here, $W = (w_1, w_2, \ldots, w_n)^T$ can be obtained by the following formula:

$$w_j = Q\left(\frac{j}{l}\right) - Q\left(\frac{(j-1)}{l}\right), \quad j = 1, 2, \ldots, n, \tag{9}$$

where

$$Q = \begin{cases} 0, & r < a, \\ \dfrac{r-a}{a-b}, & a \le r \le b, \\ 1, & r > b. \end{cases} \tag{10}$$

We utilize the principle of antonym pairs *most*, *at least half*, *as many as possible*, where the parameters (a, b) are equal to (0.3,0.8), (0,05), and (0.5,1), respectively.

Definition 6 (see [32]). Let $A = ([a, b], [c, d])$ be an interval-valued intuitionistic fuzzy number. A score function S of an interval-valued intuitionistic fuzzy value can be represented as follows:

$$S(A) = \frac{a - c + b - d}{2}. \tag{11}$$

Definition 7 (see [32]). Let $A_1 = ([a_1, b_1], [c_1, d_1])$ and $A_2 = ([a_2, b_2], [c_2, d_2])$ be two interval-valued intuitionistic fuzzy values, and let $S(A_1) = (a_1 - c_1 + b_1 - d_1)/2$ and $S(A_2) = (a_2 - c_2 + b_2 - d_2)/2$ be the scores of A_1 and A_2, respectively, then if $S(A_1) < S(A_2)$, A_1 is smaller than A_2, denoted by $A_1 < A_2$.

3. Interval-Valued Intuitionistic Fuzzy Entropy

Definition 8 (see [23]). A real-valued function $E : \mathrm{IVIFS}(X) \to [0, 1]$ is called an entropy measure on $\mathrm{IVIFS}(X)$ if it satisfies the following axiomatic requirements:

(P1) $E(A) = 0$, if and only if A is a crisp set;

(P2) $E(A) = 1$, if and only if $\tilde{\mu}_A(x_i) = \tilde{\nu}_A(x_i)$ for all $x_i \in X$;

(P3) $E(A) = E(A^C)$ for all $x_i \in \mathrm{IVIFS}(X)$;

(P4) $E(A) \le E(B)$ if A is less fuzzy than B; that is, $\mu_{AL}(x_i) \le \nu_{AL}(x_i)$, $\mu_{AU}(x_i) \le \nu_{AU}(x_i)$, and $A \subseteq B$ for all $x_i \in X$ or $\mu_{AL}(x_i) \ge \nu_{AL}(x_i)$, $\mu_{AU}(x_i) \ge \nu_{AU}(x_i)$, and $B \subseteq A$ for all $x_i \in X$.

3.1. The Limitations of the Existing Interval-Valued Intuitionistic Fuzzy Entropy. Let us suppose that $E(A_i)$ is the entropy of IVIFSs.

Vlachos' entropy measure [30] is as follows

$$E^1(A) = 1 - \sqrt{\frac{1}{2n} \sum_{i=1}^{n} \left(\left|\mu_{AL}(x_i) - \nu_{AL}(x_i)\right|\right)^2 + \left(\left|\mu_{AU}(x_i) - \nu_{AU}(x_i)\right|\right)^2}. \tag{12}$$

Example 9. Let $A = ([0.4, 0.5], [0.3, 0.4])$ and $B = ([0.1, 0.2], [0, 0.1])$ be two IVIFSs in X.

Intuitively, we can see that B is more fuzzy than A. If we calculate the $E^1(A)$ and $E^1(B)$ by (12), then we can obtain

$$\begin{aligned} E^1(A) &= 1 - \sqrt{\frac{1}{2}\left(|0.4 - 0.3|^2 + |0.5 - 0.4|^2\right)} = 0.9, \\ E^1(B) &= 1 - \sqrt{\frac{1}{2}\left(|0.1 - 0|^2 + |0.2 - 0.1|^2\right)} = 0.9, \end{aligned} \tag{13}$$

which indicate that $E^1(A) = E^1(B)$ and is not consistent with our intuition.

Ye's entropy measures [27] are as follows

$$E^2(A) = \left\{ \sin \frac{\pi \times \left[1 + \mu_{AL}(x_i) + pW_{\mu A}(x_i) - \nu_{AL}(x_i) - qW_{\nu A}(x_i)\right]}{4} + \sin \frac{\pi \times \left[1 - \mu_{AL}(x_i) - pW_{\mu A}(x_i) + \nu_{AL}(x_i) + qW_{\nu A}(x_i)\right]}{4} - 1 \right\} \times \frac{1}{\sqrt{2} - 1}, \tag{14}$$

where $p, q \in [0, 1]$ are two fixed numbers, $W_{\mu A}(x) = \mu_{AU}(x) - \mu_{AL}(x)$, and $W_{\nu A}(x) = \nu_{AU}(x) - \nu_{AL}(x)$.

Example 10. Let $A = ([0.5, 0.5], [0.1, 0.1])$ and $B = ([0.6, 0.6], [0.2, 0.2])$ be two IVIFSs in X.

Intuitively, A is more fuzzy than B. Now the $E^2(A)$ and $E^2(B)$ can be gained by (14) and take $p = q = 0.5$, and the results are

$$\begin{aligned} E^2(A) &= \left\{\sin\frac{\pi\times[1+0.5-0.1]}{4} + \sin\frac{\pi\times[1-0.5+0.1]}{4} - 1\right\} \times \frac{1}{\sqrt{2}-1} = 0.833, \\ E^2(B) &= \left\{\sin\frac{\pi\times[1+0.6-0.2]}{4} + \sin\frac{\pi\times[1-0.6+0.2]}{4} - 1\right\} \times \frac{1}{\sqrt{2}-1} = 0.833, \end{aligned} \tag{15}$$

which indicate that $E^2(A) = E^2(B)$ and are not consistent with our intuition.

3.2. New Interval-Valued Intuitionistic Fuzzy Entropy Based on Cotangent Function. A new interval-valued intuitionistic fuzzy entropy measure is introduced as follows.

Definition 11. Assuming that $A \in \mathrm{IVIFS}(X)$, then an interval-valued intuitionistic fuzzy entropy measure can be defined as

$$\begin{aligned} E(A) &= \frac{1}{n}\sum_{i=1}^{n}\cot\left(\frac{1}{4}\pi + \left(\frac{1}{2}\left|\mu_{AL}(x_i) - \nu_{AL}(x_i) + \mu_{AU}(x_i) - \nu_{AU}(x_i)\right|\right) \times \left(4\left(1+\pi_A(x_i)\right)\right)^{-1}\pi\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\cot\left(\frac{1}{4}\pi + \left(\left|\mu_{AL}(x_i) - \nu_{AL}(x_i) + \mu_{AU}(x_i) - \nu_{AU}(x_i)\right|\right) \times \left(4\left(4-\mu_{AL}(x_i) - \nu_{AL}(x_i) - \mu_{AU}(x_i) - \nu_{AU}(x_i)\right)^{-1}\right)\pi\right). \end{aligned} \tag{16}$$

Theorem 12. *The mapping $E(A)$, defined by* (16), *is an entropy measure for IVIFSs.*

Proof. In order for (16) to be qualified as a sensible measure of interval-valued intuitionistic fuzzy entropy, it must satisfy the conditions (P1)–(P4) in Definition 8.

Let $0 \le \tilde{\mu}_A(x_i)$, $\tilde{\nu}_A(x_i)$, and $\tilde{\pi}_A(x_i) \le 1$; we have $0 \le |\mu_{AL}(x_i) - \nu_{AL}(x_i) + \mu_{AU}(x_i) - \nu_{AU}(x_i)| \le 2$. It follows that $(0 \le |\mu_{AL}(x_i) - \nu_{AL}(x_i) + \mu_{AU}(x_i) - \nu_{AU}(x_i)|/8(1+\pi_A(x_i)))\pi \le (1/4)\pi$. Thus $0 \le E(A) \le 1$.

(P1) Let A be a crisp set. Then we have $\tilde{\pi}_A(x_i) = 0$, $\tilde{\mu}_A(x_i) = 1$, and $\tilde{\nu}_A(x_i) = 0$ or $\tilde{\pi}_A(x_i) = 0$, $\tilde{\mu}_A(x_i) = 0$, and $\tilde{\nu}_A(x_i) = 1$. So $E(A) = 0$.

(P2) Let $\tilde{\mu}_A(x_i) = \tilde{\nu}_B(x_i)$, then $E(A) = 1$.

(P3) It is clear that $A^C = \{\langle x_i, [\nu_{AL}(x_i), \nu_{AU}(x_i)], [\mu_{AL}(x_i), \mu_{AU}(x_i)]\rangle \mid x_i \in X\}$. By applying (11), we have $E(A) = E(A^C)$.

(P4) In order to show that (11) fulfill the requirement of (P4), it is suffice to to prove the following function:

$$\frac{\left|\mu_{AL}(x_i) - \nu_{AL}(x_i) + \mu_{AU}(x_i) - \nu_{AU}(x_i)\right|}{4\left(4-\mu_{AL}(x_i) - \nu_{AL}(x_i) - \mu_{AU}(x_i) - \nu_{AU}(x_i)\right)} \ge \frac{\left|\mu_{BL}(x_i) - \nu_{BL}(x_i) + \mu_{BU}(x_i) - \nu_{BU}(x_i)\right|}{4\left(4-\mu_{BL}(x_i) - \nu_{BL}(x_i) - \mu_{BU}(x_i) - \nu_{BU}(x_i)\right)}, \tag{17}$$

where function E is monotonic decreasing.

Suppose that $\mu_{BL}(x_i) \le \nu_{BL}(x_i)$, $\mu_{BU}(x_i) \le \nu_{BU}(x_i)$, and $A \subseteq B$, in order to prove (17); namely, we prove that

$$\left(2-\mu_{AL}(x_i) - \mu_{AU}(x_i)\right)\left(2-\nu_{BL}(x_i) - \nu_{BU}(x_i)\right) \ge \left(2-\mu_{BL}(x_i) - \mu_{BU}(x_i)\right)\left(2-\nu_{AL}(x_i) - \nu_{AU}(x_i)\right). \tag{18}$$

If $\mu_{AL}(x_i) \le \mu_{BL}(x_i)$ and $\mu_{AU}(x_i) \le \mu_{BU}(x_i)$, then $2-\mu_{AL}(x_i)-\mu_{AU}(x_i) \ge 2-\mu_{BL}(x_i)-\mu_{BU}(x_i) \ge 0$. If $\nu_{AL}(x_i) \ge \nu_{BL}(x_i)$ and $\mu_{AU}(x_i) \ge \mu_{BU}(x_i)$, then we have $2-\nu_{BL}(x_i)-\nu_{BU}(x_i) \ge 2-\nu_{AL}(x_i)-\nu_{AU}(x_i) \ge 0$. So we can get that (18) holds. Therefore, $E(A) \le E(B)$.

Similarly, when $\mu_{BL}(x_i) \ge \nu_{BL}(x_i)$, $\mu_{BU}(x_i) \ge \nu_{BU}(x_i)$, and $B \subseteq A$, we can also prove that $E(A) \le E(B)$. □

Example 13. Let $A = ([0.4, 0.5], [0.3, 0.4])$ and $B = ([0.1, 0.2], [0, 0.1])$ be two IVIFSs in X.

Intuitively, we can see that B is more fuzzy than A. Now we calculate the $E(A)$ and $E(B)$ by (16), and we can obtain that

$$\begin{aligned} E(A) &= \cot\left(\frac{1}{4}\pi + \frac{|0.4-0.3+0.5-0.4|}{4(4-0.4-0.3-0.5-0.4)}\pi\right) = 0.877, \\ E(B) &= \cot\left(\frac{1}{4}\pi + \frac{|0.1-0+0.2-0.1|}{4(4-0.1-0-0.2-0.1)}\pi\right) = 0.916, \end{aligned} \tag{19}$$

which indicate that $E(A) < E(B)$ and are consistent with our intuition.

Example 14. Let $A = ([0.5, 0.5], [0.1, 0.1])$ and $B = ([0.6, 0.6], [0.2, 0.2])$ be two IVIFSs in X.

Intuitively, we can see that A is more fuzzy than B. Now we calculate the $E(A)$ and $E(B)$ by (16), and we can obtain that

$$\begin{aligned} E(A) &= \cot\left(\frac{1}{4}\pi + \frac{|0.5-0.1+0.5-0.1|}{4(4-0.5-0.1-0.5-0.1)}\pi\right) = 0.628, \\ E(B) &= \cot\left(\frac{1}{4}\pi + \frac{|0.6-0.2+0.6-0.2|}{4(4-0.6-0.2-0.6-0.2)}\pi\right) = 0.577, \end{aligned} \tag{20}$$

which indicate that $E(A) > E(B)$ and are consistent with our intuition.

4. Fuzzy Multicriteria Group Decision-Making Method Based on the New Interval-Valued Intuitionistic Fuzzy Entropy

In this section, we propose a method for fuzzy group decision-making problems based on the interval-valued intuitionistic fuzzy entropy.

Consider an interval-valued intuitionistic fuzzy multicriteria group decision-making problem. Assume that there are n alternatives $X = \{x_1, x_2, \ldots, x_n\}$ and m decision criteria $C = \{c_1, c_2, \ldots, c_m\}$ with weight vector $W = (w_1, w_2, \ldots, w_m)^T$ associated with C, where $w_j \in [0,1]$ and $\sum_{j=1}^{m} w_j = 1$. Assume that there are t decision makers $D = \{d_1, d_2, \ldots, d_t\}$ whose corresponding weight vector is $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_t)^T$, where $\lambda_k \in [0,1]$ and $\sum_{k=1}^{t} \lambda_k = 1$. In this case, the characteristic of the alternative A_i of the kth decision maker d_k is represented by the following IVIFS:

$$A_i^k = \left\{\left\langle C_j^k, \left[\mu_{A_i^k L}\left(C_j^k\right), \mu_{A_i^k U}\left(C_j^k\right)\right], \left[\nu_{A_i^k L}\left(C_j^k\right), \nu_{A_i^k U}\left(C_j^k\right)\right]\right\rangle \mid C_j^k \in C\right\}, \tag{21}$$

where $0 \le \mu_{A_i^k U}(C_j^k) + \nu_{A_i^k U}(C_j^k) \le 1$, $\mu_{A_i^k L}(C_j^k) \ge 0$, $\mu_{A_i^k L}(C_j^k) \ge 0$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, and $k = 1, 2, \ldots, t$. The IVIFS value that is the pair of intervals $\mu_{A_i^k}(C_j^k) = [a_{ij}^k, b_{ij}^k]$, $\nu_{A_i^k}(C_j^k) = [c_{ij}^k, d_{ij}^k]$ for $C_j^k \in C$ is denoted by $r_{ij}^k = ([a_{ij}^k, b_{ij}^k], [c_{ij}^k, d_{ij}^k])$. Here, we can elicit the interval-valued intuitionistic fuzzy decision matrix $R^k = (r_{ij}^k)_{n\times m}$.

If the information about weight w_j of the criterion C_j $(j = 1, 2, \ldots, m)$ is incomplete, for determining the criterion weight from the decision matrix we can establish a model of interval-valued intuitionistic fuzzy entropy weights.

For the criteria C_j, the entropy of the alternative A_i of the kth decision maker can be given as

$$\begin{aligned} E\left(r_{ij}^k\right) &= \cot\Big(\frac{1}{4}\pi + \left(\left|\mu_{AL}\left(r_{ij}^k\right) - \nu_{AL}\left(r_{ij}^k\right) + \mu_{AU}\left(r_{ij}^k\right) - \nu_{AU}\left(r_{ij}^k\right)\right|\right) \\ &\quad \times \Big(4\Big(4 - \mu_{AL}\left(r_{ij}^k\right) - \nu_{AL}\left(r_{ij}^k\right) \\ &\quad - \mu_{AU}\left(r_{ij}^k\right) - \nu_{AU}\left(r_{ij}^k\right)\Big)\Big)^{-1}\pi\Big). \end{aligned} \tag{22}$$

And the entropy for the alternative A_i^k of the kth decision maker is given as

$$\begin{aligned} E\left(A_i^k\right) &= \sum_{j=1}^{m} \cot\Big(\frac{1}{4}\pi + \left(\left|\mu_{AL}\left(r_{ij}^k\right) - \nu_{AL}\left(r_{ij}^k\right) + \mu_{AU}\left(r_{ij}^k\right) - \nu_{AU}\left(r_{ij}^k\right)\right|\right) \\ &\quad \times \Big(4\Big(4 - \mu_{AL}\left(r_{ij}^k\right) - \nu_{AL}\left(r_{ij}^k\right) \\ &\quad - \mu_{AU}\left(r_{ij}^k\right) - \nu_{AU}\left(r_{ij}^k\right)\Big)\Big)^{-1}\pi\Big). \end{aligned} \tag{23}$$

As each alternative is made in a fair competitive environment, and the fuzzy entropy of each alternative is from a same criteria weight coefficient, the alternatives should be combined. The overall entropy for the alternative A_i is given as

$$\begin{aligned} E\left(A_i\right) &= \sum_{k=1}^{t}\sum_{j=1}^{m} \cot\Big(\frac{1}{4}\pi + \left(\left|\mu_{AL}\left(r_{ij}^k\right) - \nu_{AL}\left(r_{ij}^k\right) + \mu_{AU}\left(r_{ij}^k\right) - \nu_{AU}\left(r_{ij}^k\right)\right|\right) \\ &\quad \times \Big(4\Big(4 - \mu_{AL}\left(r_{ij}^k\right) - \nu_{AL}\left(r_{ij}^k\right) \\ &\quad - \mu_{AU}\left(r_{ij}^k\right) - \nu_{AU}\left(r_{ij}^k\right)\Big)\Big)^{-1}\pi\Big). \end{aligned} \tag{24}$$

According to the entropy theory, if the entropy value for an alternative is smaller across alternatives, it can provide decision makers with the useful information. Therefore, the criteria should be assigned to a bigger weight value. Then the smaller the value of (24) is, the better the weight we should assign to the criteria.

Let H be the set of incomplete information about criteria weights; to get the optimal weight vector, the following model can be constructed:

$$\begin{aligned} \min \quad & E(A) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{t}\sum_{j=1}^{m} w_j \\ &\quad \times \cot\Big(\frac{1}{4}\pi + \left(\left|\mu_{AL}\left(r_{ij}^k\right) - \nu_{AL}\left(r_{ij}^k\right) + \mu_{AU}\left(r_{ij}^k\right) - \nu_{AU}\left(r_{ij}^k\right)\right|\right) \\ &\quad \times \Big(4\Big(4 - \mu_{AL}\left(r_{ij}^k\right) - \nu_{AL}\left(r_{ij}^k\right) \\ &\quad - \mu_{AU}\left(r_{ij}^k\right) - \nu_{AU}\left(r_{ij}^k\right)\Big)\Big)^{-1}\pi\Big), \\ \text{s.t.} \quad & w \in H, \quad \sum_{j=1}^{m} w_j = 1, \quad w_j \ge 0, \ j = 1, 2, \ldots, m. \end{aligned} \tag{25}$$

By solving model (25) with Lingo software, we get the optimal solution $(w_1, w_2, \ldots, w_m)^T$.

In summary, the main procedure of the decision method proposed is listed in the following.

Step 1. Calculate the weight vector $W = (w_1, w_2, \ldots, w_m)^T$ by solving model (25).

Step 2. Aggregate the values of each decision maker, and utilize the IVIF-WGA operator to derive the values z_i^k ($i = 1, 2, \ldots, n$) of each decision maker as follows:

$$z_i^k = F_w\left(r_{i1}^k, r_{i2}^k, \ldots, r_{im}^k\right), \quad i = 1, 2, \ldots, n, \ k = 1, 2, \ldots, t. \tag{26}$$

Step 3. Aggregate the values of all decision makers, and utilize the IVIF-HA operator to derive the collective values z_i of the alternative x_i as follows

$$z_i = H_\lambda\left(z_i^1, z_i^2, \ldots, z_i^t\right), \quad i = 1, 2, \ldots, n, \tag{27}$$

where the associated weighting vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_t)$ of the IVIF-HA operator.

Step 4. Calculate the scores $S(z_i)$ ($i = 1, 2, \ldots, n$) of the collective overall values to rank all the alternatives A_i ($i = 1, 2, \ldots, n$) and then to select the best one(s).

5. An Illustration for Solving the Supplier Selection Problem

This section adopts a supplier selection problem in [33, 34] to demonstrate how to apply the proposed approach. With continual business development, globalized markets become increasingly competitive. Establishing effective supply chain management (SCM) becomes a critical activity because a sound SCM system can reduce supply chain risk, maximize revenue, optimize business processes, and allow a company to maintain a dominant position in the market [34, 35]. On the other hand, it is a hard problem since supplier selection is typically a multicriteria group decision-making problem involving several conflicting criteria on which decision maker's knowledge is usually vague and imprecise [34]. Previous research concerning supplier selection often used exact numbers to measure criterion weights. In this study, considering that the decision maker may have difficulty in eliciting precise criterion weights, the proposed approach is proposed to select appropriate supplier in group decision-making environment. It should be noted that, as suggested and illustrated by Merigo and Gil-Lafuente [36], the proposed approach can be easily applied to a host of practical decision problems that involve choosing an optimal alternative from a list of alternatives when multiple attributes must be considered.

Suppose that a high-tech company which manufactures electronic products intends to evaluate and select a supplier of USB connectors. There are four suppliers x_1, x_2, x_3, and x_4 which are chosen as candidates. A committee of three decision makers d_1, d_2, and d_3 is established, which are an engineering expert, financial expert, and quality control expert, respectively. Four evaluated criteria are considered, including finance (c_1), performance (c_2), technique (c_3), and organizational culture (c_4). The expert weight vector is given by $\lambda = (0.35, 0.35, 0.3)^T$. The interval-valued intuitionistic fuzzy decision matrices of criterion values are constructed as follows:

$$R^1 = \begin{bmatrix} ([0.3, 0.4], [0.4, 0.6]) & ([0.6, 0.7], [0.1, 0.2]) & ([0.5, 0.7], [0.2, 0.3]) & ([0.7, 0.8], [0.0, 0.1]) \\ ([0.7, 0.8], [0.1, 0.2]) & ([0.6, 0.7], [0.2, 0.3]) & ([0.2, 0.3], [0.4, 0.6]) & ([0.5, 0.6], [0.1, 0.3]) \\ ([0.5, 0.8], [0.1, 0.2]) & ([0.7, 0.8], [0.0, 0.1]) & ([0.5, 0.5], [0.4, 0.5]) & ([0.2, 0.3], [0.2, 0.4]) \\ ([0.2, 0.3], [0.4, 0.5]) & ([0.5, 0.7], [0.1, 0.3]) & ([0.6, 0.7], [0.1, 0.2]) & ([0.4, 0.5], [0.1, 0.3]) \end{bmatrix},$$

$$R^2 = \begin{bmatrix} ([0.4, 0.5], [0.3, 0.4]) & ([0.5, 0.6], [0.2, 0.2]) & ([0.6, 0.7], [0.2, 0.3]) & ([0.7, 0.8], [0.1, 0.2]) \\ ([0.6, 0.8], [0.1, 0.2]) & ([0.5, 0.6], [0.3, 0.4]) & ([0.4, 0.5], [0.3, 0.4]) & ([0.4, 0.6], [0.3, 0.4]) \\ ([0.5, 0.6], [0.3, 0.4]) & ([0.5, 0.7], [0.1, 0.2]) & ([0.5, 0.6], [0.3, 0.4]) & ([0.3, 0.4], [0.2, 0.5]) \\ ([0.5, 0.6], [0.3, 0.4]) & ([0.7, 0.8], [0.0, 0.1]) & ([0.4, 0.5], [0.2, 0.4]) & ([0.5, 0.7], [0.1, 0.2]) \end{bmatrix},$$

$$R^3 = \begin{bmatrix} ([0.4, 0.6], [0.3, 0.4]) & ([0.5, 0.7], [0.1, 0.2]) & ([0.5, 0.6], [0.2, 0.4]) & ([0.6, 0.8], [0.1, 0.2]) \\ ([0.5, 0.8], [0.1, 0.2]) & ([0.3, 0.5], [0.2, 0.3]) & ([0.3, 0.6], [0.2, 0.4]) & ([0.4, 0.5], [0.2, 0.4]) \\ ([0.5, 0.6], [0.0, 0.1]) & ([0.5, 0.8], [0.1, 0.2]) & ([0.4, 0.7], [0.2, 0.3]) & ([0.2, 0.4], [0.2, 0.3]) \\ ([0.5, 0.7], [0.1, 0.3]) & ([0.4, 0.6], [0.0, 0.1]) & ([0.3, 0.5], [0.2, 0.4]) & ([0.7, 0.9], [0.0, 0.1]) \end{bmatrix}. \tag{28}$$

The incomplete information about the criterion weights are as follows (in this problem, the criterion weights are incomplete information. The specific weight calculation method can be found in [34]):

$$\begin{aligned} H = \{&0.228 \le w_1 \le 0.8758, 0.2285 \le w_2 \le 0.8789, \\ &0.1642 \le w_3 \le 0.7979, 0.1419 \le w_4 \le 0.7824\}. \end{aligned} \tag{29}$$

Step 1. Calculate the weight vector $W = (w_1, w_2, w_3, w_4)^T$ by solving model (25) as follows:

$$W = (0.228, 0.4659, 0.1642, 0.1419)^T. \tag{30}$$

Step 2. Aggregate the values of each decision maker, andutilize the IVIF-WGA operator to derive the values z_i^k ($i = 1, 2, \ldots, n$) of each decision maker.

The integrated values for alternatives x_1, x_2, x_3 and x_4 of decision maker d_1 are, respectively,

$$\begin{aligned} z_1^1 &= ([0.508, 0.628], [0.183, 0.320]), \\ z_2^1 &= ([0.506, 0.614], [0.203, 0.342]), \\ z_3^1 &= ([0.514, 0.644], [0.130, 0.249]), \\ z_4^1 &= ([0.405, 0.550], [0.179, 0.337]). \end{aligned} \tag{31}$$

The integrated values for alternatives x_1, x_2, x_3 and x_4 of decision maker d_2 are, respectively,

$$\begin{aligned} z_1^2 &= ([0.514, 0.615], [0.211, 0.267]), \\ z_2^2 &= ([0.487, 0.622], [0.259, 0.359]), \\ z_3^2 &= ([0.439, 0.584], [0.198, 0.331]), \\ z_4^2 &= ([0.546, 0.649], [0.124, 0.259]). \end{aligned} \tag{32}$$

The integrated values for alternatives x_1, x_2, x_3 and x_4 of decision maker d_3 are, respectively,

$$\begin{aligned} z_1^3 &= ([0.488, 0.672], [0.166, 0.285]), \\ z_2^3 &= ([0.351, 0.573], [0.178, 0.312]), \\ z_3^3 &= ([0.423, 0.664], [0.111, 0.211]), \\ z_4^3 &= ([0.435, 0.639], [0.06, 0.205]). \end{aligned} \tag{33}$$

Step 3. Aggregate the values of all decision makers, and utilize the IVIF-HA operator to derive the collective values z_i of the alternative x_i.

The weighting vector of the IVIL-HA operator is obtained by the principle of antonym pairs many and (a, b) equal to (0.3, 0.8). So we can obtain $\omega = (0.243, 0.514, 0.243)$ and the integrated values as follows:

$$\begin{aligned} z_1 &= ([0.489, 0.617], [0.202, 0.297]), \\ z_2 &= ([0.436, 0.593], [0.236, 0.357]), \\ z_3 &= ([0.430, 0.625], [0.144, 0.262]), \\ z_4 &= ([0.463, 0.606], [0.129, 0.278]). \end{aligned} \tag{34}$$

Step 4. Calculate the scores $S(z_i)$ $(i = 1, 2, 3, 4)$ of the collective overall values z_i $(i = 1, 2, 3, 4)$ as follows:

$$\begin{aligned} S(z_1) &= 0.607, \qquad S(z_2) = 0.436, \\ S(z_3) &= 0.649, \qquad S(z_4) = 0.663. \end{aligned} \tag{35}$$

Step 5. Rank all the alternatives A_i $(i = 1, 2, 3, 4)$ in accordance with the score $S(z_i)$ of the collective overall interval-valued intuitionistic fuzzy values z_i $(i = 1, 2, 3, 4)$: $A_4 \succ A_3 \succ A_1 \succ A_2$, and thus the best alternative is A_4.

6. Conclusion

In this paper, a new entropy measure of IVIFS is proposed by using cotangent function, which can overcome limitations of some existing methods. And we provide several numerical examples to illustrate its validity. For interval-valued intuitionistic fuzzy multicriteria group decision-making problem with incomplete information on the weights of criteria, an entropy weight model is established to determine the entropy weights. In addition, the method and procedure are developed to solve FMCGDM problems. Finally, the supplier selection problem is used as an example to demonstrate how to apply the proposed multicriteria group decision-making approach.

Acknowledgments

The authors would like to thank Yufeng Xu, Fuqiang Wang, Mengjun Liao, Siying Wang, and the anonymous reviewers for their valuable comments on this paper. This paper was supported by the National Natural Science Foundation of China (no. 70921001).

References

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–356, 1965.

[2] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[3] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[4] K. Atanassov, *Intuitionistic Fuzzy Sets*, Springer, Physica-Verlag, Heidelberg, Germany, 1999.

[5] K. Atanassov and G. Gargov, "Interval valued intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 31, no. 3, pp. 343–349, 1989.

[6] Y. Yang and C. Hinde, "A new extension of fuzzy sets using rough sets: R-fuzzy sets," *Information Sciences*, vol. 180, no. 3, pp. 354–365, 2010.

[7] J. Q. Wang and H. B. Li, "Multi-criteria decision-making method based on aggregation operators for intuitionistic linguistic fuzzy numbers," *Control and Decision*, vol. 25, no. 10, pp. 1571–1584, 2010.

[8] H. Bustince and P. Burillo, "Vague sets are intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 79, no. 3, pp. 403–405, 1996.

[9] C. Cornelis, K. T. Atanassov, and E. E. Kerre, "Intuitionistic fuzzy sets and interval-valued fuzzy sets: a critical comparison," in *Proceedings of the 3rd European Conference on Fuzzy Logic and Technology (EUSFLAT '03)*, pp. 159–163, Zittau, Germany, 2003.

[10] G. Deschrijver and E. E. Kerre, "On the relationship between some extensions of fuzzy set theory," *Fuzzy Sets and Systems*, vol. 133, no. 2, pp. 227–235, 2003.

[11] P. Burillo and H. Bustince, "Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 78, no. 3, pp. 305–316, 1996.

[12] E. Szmidt and J. Kacprzyk, "Entropy for intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 118, no. 3, pp. 467–477, 2001.

[13] W. L. Hung and M. S. Yang, "Fuzzy entropy on intuitionistic fuzzy sets," *International Journal of Intelligent Systems*, vol. 21, no. 4, pp. 443–451, 2006.

[14] C.-P. Wei, P. Wang, and Y. Z. Zhang, "Entropy, similarity measure of interval-valued intuitionistic fuzzy sets and their applications," *Information Sciences*, vol. 181, no. 19, pp. 4273–4286, 2011.

[15] Y. Wang and Y. J. Lei, "A technique for constructing intuitionistic fuzzy entropy," *Control and Decision*, vol. 22, no. 12, pp. 1390–1394, 2007.

[16] G. S. Huang and Y. S. Liu, "The fuzzy entropy of vague sets based on non-fuzzy sets," *Computer Applications and Software*, vol. 22, no. 6, pp. 16–17, 2005.

[17] I. K. Vlochos and G. D. Sergiadis, "Intuitionistic fuzzy information-applications to pattern recognition," *Pattern Recognition Letter*, vol. 28, no. 2, pp. 197–206, 2007.

[18] W. Zeng and H. Li, "Relationship between similarity measure and entropy of interval valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 157, no. 11, pp. 1477–1484, 2006.

[19] H. Zhang, W. Zhang, and C. Mei, "Entropy of interval-valued fuzzy sets based on distance and its relationship with similarity measure," *Knowledge-Based Systems*, vol. 22, no. 6, pp. 449–454, 2009.

[20] G. S. Huang and Y. S. Liu, "On the fuzzy entropy of vague sets," *Computer Engineering and Applications*, vol. 41, no. 33, pp. 48–50, 2005.

[21] Q. S. Zhang and S. Y. Jiang, "A note on information entropy measures for vague sets and its applications," *Information Sciences*, vol. 178, no. 21, pp. 4184–4191, 2008.

[22] T. Y. Chen and C. H. Li, "Determining objective weights with intuitionistic fuzzy entropy measures: a comparative analysis," *Information Sciences*, vol. 180, no. 21, pp. 4207–4222, 2010.

[23] E. Szmidt and J. Kacprzyk, "A new concept of a similarity measure for intuitionistic fuzzy sets and its use in group decision making," in *Modelling Decision For Artificial Intelligence*, V. Torra, Y. Narukawa, and S. Miyamoto, Eds., vol. 3558 of *Lecture Notes in Artificial Intelligence*, pp. 272–282, Springer, 2005.

[24] C. L. Hwang and K. Yoon, *Multiple Attribute Decision Making-Methods and Applications, A State-of-the-Art Survey*, Springer, New York, NY, USA, 1981.

[25] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, NY, USA, 1980.

[26] M. Zeleny, *Multiple Criteria Decision Making*, Springer, New York, NY, USA, 1982.

[27] J. Ye, "Multicriteria fuzzy decision-making method using entropy weights-based correlation coefficients of interval-valued intuitionistic fuzzy sets," *Applied Mathematical Modelling*, vol. 34, no. 12, pp. 3864–3870, 2010.

[28] Q.-S. Zhang, S. Jiang, B. Jia, and S. Luo, "Some information measures for interval-valued intuitionistic fuzzy sets," *Information Sciences*, vol. 180, no. 24, pp. 5130–5145, 2010.

[29] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," *Information and Control*, vol. 20, no. 4, pp. 301–312, 1972.

[30] I. K. Vlachos and G. D. Sergiadis, "Subsethood, entropy, and cardinality for interval-valued fuzzy sets-An algebraic derivation," *Fuzzy Sets and Systems*, vol. 158, no. 12, pp. 1384–1396, 2007.

[31] Z. S. Xu, "Methods for aggregating interval-valued intuitionistic fuzzy information and their application to decision making," *Control and Decision*, vol. 22, no. 2, pp. 215–219, 2007.

[32] Z. S. Xu and J. Chen, "An approach to group decision making based on interval-valued intuitionistic judgment matrices," *System Engineering Theory and Practice*, vol. 27, no. 4, pp. 126–133, 2007.

[33] F. E. Boran, S. Genç, M. Kurt, and D. Akay, "A multi-criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method," *Expert Systems with Applications*, vol. 36, no. 8, pp. 11363–11368, 2009.

[34] T. Y. Chen, H. P. Wang, and Y. Y. Lu, "A multicriteria group decision-making approach based on interval-valued intuitionistic fuzzy sets: a comparative perspective," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7647–7658, 2011.

[35] S. H. Ha and R. Krishnan, "A hybrid approach to supplier selection for the maintenance of a competitive supply chain," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1303–1311, 2008.

[36] J. M. Merigo and A. M. Gil-Lafuente, "The induced generalized OWA operator," *Information Sciences*, vol. 179, no. 6, pp. 729–741, 2009.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 497586, 7 pages
http://dx.doi.org/10.1155/2013/497586

Research Article

Properties of Expected Residual Minimization Model for a Class of Stochastic Complementarity Problems

Mei-Ju Luo[1] and Yuan Lu[2]

[1] *School of Mathematics, Liaoning University, Liaoning 110036, China*
[2] *School of Sciences, Shenyang University, Liaoning 110044, China*

Correspondence should be addressed to Mei-Ju Luo; meiju_luo@yahoo.cn

Received 31 January 2013; Revised 10 May 2013; Accepted 14 May 2013

Academic Editor: Farhad Hosseinzadeh Lotfi

Expected residual minimization (ERM) model which minimizes an expected residual function defined by an NCP function has been studied in the literature for solving stochastic complementarity problems. In this paper, we first give the definitions of stochastic P-function, stochastic P_0-function, and stochastic uniformly P-function. Furthermore, the conditions such that the function is a stochastic $P(P_0)$-function are considered. We then study the boundedness of solution set and global error bounds of the expected residual functions defined by the "Fischer-Burmeister" (FB) function and "min" function. The conclusion indicates that solutions of the ERM model are robust in the sense that they may have a minimum sensitivity with respect to random parameter variations in stochastic complementarity problems. On the other hand, we employ quasi-Monte Carlo methods and derivative-free methods to solve ERM model.

1. Introduction

Given a vector-valued function $F : \mathbf{R}^n \times \Omega \to \mathbf{R}^n$, the stochastic complementarity problems, denoted by SCP$(F(x,\omega))$, are to find a vector x^* such that

$$\begin{gathered} x^* \geq 0, \quad F(x^*,\omega) \geq 0, \\ (x^*)^T F(x^*,\omega) = 0, \quad \omega \in \Omega \text{ a.s.}, \end{gathered} \tag{1}$$

where $\omega \in \Omega \subseteq \mathbf{R}^m$ is a random vector with given probability distribution $\mathscr{P}$ and "a.s." means "almost surely" under the given probability measure. Particularly, when F is an affine function of x for any ω, that is,

$$F(x,\omega) = M(\omega)x + q(\omega), \quad \omega \in \Omega, \tag{2}$$

where $M(\omega) \in \mathbf{R}^{n\times n}$ and $q(\omega) \in \mathbf{R}^n$, the SCP$(F(x,\omega))$ is called stochastic linear complementarity problems, denoted by SLCP$(M(\omega), q(\omega))$. Correspondingly, problem (1) is called stochastic nonlinear complementarity problem, denoted by SNCP$(F(x,\omega))$ if F can not be denoted by an affine function of x for any ω. The deterministic problems, which are called complementarity problems (denoted by CP$(F(x))$), have been intensively studied. More information about theoretical analysis, numerical algorithms and applications especially in economics and engineering can be found in comprehensive books [1, 2].

In practical applications, some elements may involve stochastic factors. In fact, due to stochastic factors, the function value of F depends not only on x, but also on random vectors. Hence, problem (1) does not have solution in general for almost all $\omega \in \Omega$. To solve these problems, researchers focus on giving reasonable deterministic reformulations for SCP$(F(x,\omega))$. Certainly, different deterministic formulations may yield different solutions that are optimal in different senses. In the study of SCP$(F(x,\omega))$, three types of formulations have been proposed; the expected value (EV) formulation, the expected residual minimization (ERM) formulation, and the SMPEC formulation [3].

The EV formulation is studied by Gürkan et al. [4]. The problem considered in [4] is actually a stochastic variational inequality problem. When applied to the SCP$(F(x,\omega))$, the EV model can be stated as follows:

$$x^* \geq 0, \quad \mathbf{E}[F(x^*,\omega)] \geq 0, \quad (x^*)^T \mathbf{E}[F(x^*,\omega)] = 0, \tag{3}$$

where $\mathbf{E}$ means expectation with respect to ω.

The ERM model is first proposed by Chen and Fukushima [5] for solving the SLCP$(M(\omega), q(\omega))$. By employing an NCP function ϕ, the SCP$(F(x,\omega))$ (1) is transformed equivalently to the stochastic equations

$$\Phi(x,\omega) = 0, \quad \omega \in \Omega \text{ a.s.}, \tag{4}$$

where $\Phi : \mathbf{R}^n \times \Omega \to \mathbf{R}^n$ is defined by

$$\Phi(x,\omega) := \begin{pmatrix} \phi(x_1, F_1(x,\omega)) \\ \vdots \\ \phi(x_n, F_n(x,\omega)) \end{pmatrix}, \tag{5}$$

and x_i denotes the ith component of the vector x. Here $\phi : \mathbf{R}^n \to \mathbf{R}$ is an NCP function which has the property

$$\phi(a,b) = 0 \Longleftrightarrow a \geq 0, \quad b \geq 0, \ ab = 0. \tag{6}$$

Then the ERM formulation for (1) is given by

$$\min_{x \in \mathbf{R}^n_+} \theta(x) := \mathbf{E}\left[\|\Phi(x,\omega)\|^2\right]. \tag{7}$$

The NCP functions employed in [5] include the Fischer-Burmeister function, which is defined by

$$\phi_{\mathrm{FB}}(a,b) := \sqrt{a^2+b^2} - (a+b) \tag{8}$$

and the min function

$$\phi_{\min}(a,b) := \min\{a,b\}. \tag{9}$$

In particular, it is known [6, 7] that there exist the following relations between these two functions:

$$\frac{2}{\sqrt{2}+2}\left|\phi_{\min}\right| \leq \left|\phi_{\mathrm{FB}}\right| \leq \left(\sqrt{2}+2\right)\left|\phi_{\min}\right|. \tag{10}$$

As observed in [5], the ERM formulations with different NCP functions may have different properties. Subsequently, the ERM formulation for SCP$(F(x,\omega))$ has been studied in [6, 8–13]. Note that Fang et al. [8] propose a new concept of stochastic matrice: $M(\cdot)$ is called a stochastic R_0 matrix if

$$\mathscr{P}\left\{\omega : x \geq 0, \ M(\omega)x \geq 0, \ x^T M(\omega) x = 0\right\} = 1 \Longrightarrow x = 0. \tag{11}$$

Moreover, Zhang and Chen [11] introduce a new concept of stochastic R_0 function, which can be regarded as a generalization of the stochastic R_0 matrix given in [8].

Throughout this paper, we suppose that the sample space Ω is nonempty and compact set and that the function $F(x,\omega)$ is continuous with respect to x and ω. On the other hand, we will use the following notations: $I(x) = \{i : x_i = 0\}$ and $J(x) = \{i : x_i \neq 0\}$ for a given vector $x \in \mathbf{R}^n$. $\langle l, n\rangle$ represents the set $\{l, l+1, \ldots, l+n\}$ for natural numbers l and u with $l < u$. $x_+ = \max\{x, 0\}$ for any given vector x. $\|\cdot\|$ refers to the Euclidean norm.

The remainder of the paper is organized as follows: in Section 2, we introduce the concepts of a stochastic P-function, a stochastic P_0-function, and a stochastic uniformly P-function, which can be regarded as a generalization of the deterministic P, P_0-function, and uniformly P-function or an extension of stochastic P matrix and stochastic P_0 matrix [14]. In addition, some properties of a stochastic $P(P_0)$-function are given. In Section 3, we show the sufficient conditions for the solution set of ERM problem to be nonempty and bounded. In Section 4, we discuss error bounds of SCP$(F(x,\omega))$. In Section 5, an algorithm will be given to solve ERM model. We then give conclusions in Section 6.

2. Stochastic $P(P_0)$-Function

It is well known that the P-function, P_0-function, and uniformly P-function play an important role in the nonlinear complementarity problems theory [1]. We will introduce a new concept of stochastic P-function, P_0-function, and uniformly P-function, which can be regarded as a generalization of their deterministic form or stochastic P matrix and stochastic P_0 matrix.

Definition 1 (see [14]). $M(\cdot)$ is called a stochastic $P(P_0)$-matrix if there exists $i \in J(x)$ such that, for every $x \neq 0$ in $\mathbf{R}^n$,

$$\mathscr{P}\left\{\omega : x_i (M(\omega)x)_i > 0 \ (\geq 0)\right\} > 0. \tag{12}$$

Definition 2. A function $F : \mathbf{R}^n \times \Omega \to \mathbf{R}^n$ is a stochastic $P(P_0)$-function if there exist $i \in J(x,y)$, $i \in \langle 1, n\rangle$ such that, for every $x \neq y$ in $\mathbf{R}^n$,

$$\begin{gathered} x_i \neq y_i, \\ \mathscr{P}\left\{\omega : (x_i - y_i)\left(F_i(x,\omega) - F_i(y,\omega)\right) > 0 \ (\geq 0)\right\} > 0. \end{gathered} \tag{13}$$

Definition 3. A function $F : \mathbf{R}^n \times \Omega \to \mathbf{R}^n$ is a stochastic uniformly P-function if there exists a positive constant α and $i \in J(x,y)$, $i \in \langle 1, n\rangle$ such that, for every $x \neq y$ in $\mathbf{R}^n$,

$$\mathscr{P}\left\{\omega : (x_i - y_i)\left(F_i(x,\omega) - F_i(y,\omega)\right) \geq \alpha\|x-y\|^2\right\} > 0. \tag{14}$$

Clearly, every stochastic uniformly P-function must be a stochastic P-function, which in turn must be a stochastic P_0-function. We further cite the definition of "equicoercive" in [11]. More information about this definition can be found in [11].

Definition 4 (see [11]). We say that $F : \mathbf{R}^n \times \Omega \to \mathbf{R}^n$ is equicoercive on $\mathscr{D} \subseteq \mathbf{R}^n$, if, for any $\{x^k\} \subseteq \mathscr{D}$ satisfying $\|x^k\| \to \infty$, the existence of $\{\omega^k\} \subseteq \operatorname{supp}\Omega$ with $\lim_{k\to\infty} F_i(x^k,\omega^k) = \infty (\lim_{k\to\infty}(-F_i(x^k,\omega^k))_+ = \infty)$ for some $i \in \langle 1, n\rangle$ implies that

$$\begin{aligned} &\mathscr{P}\left\{\omega : \lim_{k\to\infty} F_i\left(x^k,\omega\right) = \infty\right\} \\ &\quad > 0\left(\mathscr{P}\left\{\omega : \lim_{k\to\infty}\left(-F_i\left(x^k,\omega\right)\right)_+ = \infty\right\} > 0\right), \end{aligned} \tag{15}$$

where

$$\operatorname{supp}\Omega := \left\{\overline{\omega} \in \Omega : \int_{B_\omega(\overline{\omega},\nu)\cap\Omega} dF(\omega) > 0 \text{ for any } \nu > 0\right\} \tag{16}$$

and $B_\omega(\overline{\omega},\nu) := \{\omega : \|\omega - \overline{\omega}\| < \nu\}$ and $F(\omega)$ is the distribution function of ω.

More details about $\operatorname{supp}\Omega$ were included in [8].

Proposition 5. *If F is a stochastic P_0-function, then $F + \varepsilon x$ is a stochastic P-function for every $\varepsilon > 0$.*

Proof. From the definition of stochastic P_0-function, there exist $i \in J(x,y)$, $i \in \langle 1,n\rangle$ such that, for every $x \neq y$,

$$x_i \neq y_i, \quad \mathscr{P}\{\omega : (x_i - y_i)(F_i(x,\omega) - F_i(y,\omega)) \geq 0\} > 0. \tag{17}$$

Hence, we have

$$\begin{aligned}
&\mathscr{P}\{\omega : (x_i - y_i)(F_i(x,\omega) + \varepsilon x_i - (F_i(y,\omega) + \varepsilon y_i)) > 0\} \\
&\quad = \mathscr{P}\left\{\omega : (x_i - y_i)(F_i(x,\omega) - F_i(y,\omega)) + \varepsilon(x_i - y_i)^2 > 0\right\} \\
&\quad \geq \mathscr{P}\{\omega : (x_i - y_i)(F_i(x,\omega) - F_i(y,\omega)) \geq 0\} > 0.
\end{aligned} \tag{18}$$

□

This proposition gives the relationship between stochastic P_0-function and stochastic P-function.

Proposition 6. *Let F be an affine function of x for any $\omega \in \Omega$ defined by (2). Then F is a stochastic $P(P_0)$-function if and only if $M(\cdot)$ is a stochastic $P(P_0)$ matrix.*

Proof. By the definition of stochastic $P(P_0)$-function, we have that there exist $i \in J(x,y)$, $i \in \langle 1,n\rangle$ such that, for every $x \neq y$,

$$\begin{gathered} x_i \neq y_i, \\ \mathscr{P}\{\omega : (x_i - y_i)(F_i(x,\omega) - F_i(y,\omega)) > 0\,(\geq 0)\} > 0, \end{gathered} \tag{19}$$

which is equivalent to

$$\begin{gathered} x_i \neq y_i, \\ \mathscr{P}\{\omega : (x_i - y_i)(M(\omega)(x-y))_i > 0\,(\geq 0)\} > 0, \end{gathered} \tag{20}$$

when F is defined by (2). Set $z = x - y$; then $z \neq 0$, and we have that formulation (20) holds if and only if there exists $i \in J(z)$ such that, for every $z \neq 0$,

$$\mathscr{P}\{\omega : z_i(M(\omega)z)_i > 0\,(\geq 0)\} > 0. \tag{21}$$

Hence, $M(\cdot)$ is a stochastic $P(P_0)$ matrix. □

Proposition 7. *F is a stochastic $P(P_0)$-function if and only if there exists a $\overline{\omega} \in \operatorname{supp}\Omega$ such that $F(\cdot,\overline{\omega})$ is a $P(P_0)$-function.*

Proof. For the "if" part, suppose on the contrary that F is not a stochastic $P(P_0)$-function, and then there exist $\tilde{x}, \tilde{y}, \tilde{x} \neq \tilde{y}$ in $\mathbf{R}^n$ for any $i \in \langle 1,n\rangle$ satisfying

$$\begin{gathered} \tilde{x}_i \neq \tilde{y}_i, \\ \mathscr{P}\{\omega : (\tilde{x}_i - \tilde{y}_i)(F_i(\tilde{x},\omega) - F_i(\tilde{y},\omega)) > 0\,(\geq 0)\} = 0. \end{gathered} \tag{22}$$

On the other hand, since $F(\cdot,\overline{\omega})$ is a $P(P_0)$-function, then for $\tilde{x}, \tilde{y}$ there exist $i \in J(\tilde{x},\tilde{y})$, $i \in \langle 1,n\rangle$ such that

$$(\tilde{x}_i - \tilde{y}_i)(F_i(\tilde{x},\overline{\omega}) - F_i(\tilde{y},\overline{\omega})) > 0\,(\geq 0). \tag{23}$$

Notice that $\overline{\omega} \in \operatorname{supp}\Omega$, by the definition of $\operatorname{supp}\Omega$ in (16), we have

$$\mathscr{P}\{\overline{\omega} : (\tilde{x}_i - \tilde{y}_i)(F_i(\tilde{x},\overline{\omega}) - F_i(\tilde{y},\overline{\omega})) > 0\,(\geq 0)\} > 0. \tag{24}$$

This contradicts formulation (22). Therefore, F is a stochastic $P(P_0)$-function.

Now for the "only if" part, suppose on the contrary that there does not exist a $\overline{\omega} \in \operatorname{supp}\Omega$ such that $F(\cdot,\overline{\omega})$ is a $P(P_0)$-function. Then for any $i \in \langle 1,n\rangle$, $\omega \in \operatorname{supp}\Omega$, there exists $\overline{x}, \overline{y}, \overline{x} \neq \overline{y}$ in $\mathbf{R}^n$, such that

$$\begin{gathered} \overline{x}_i \neq \overline{y}_i, \\ (\overline{x}_i - \overline{y}_i)(F_i(\overline{x},\omega) - F_i(\overline{y},\omega)) \leq 0\,(< 0), \end{gathered} \tag{25}$$

which means that

$$\begin{gathered} \overline{x}_i \neq \overline{y}_i, \\ \mathscr{P}\{\omega \in \operatorname{supp}\Omega : (\overline{x}_i - \overline{y}_i)(F_i(\overline{x},\omega) - F_i(\overline{y},\omega)) \\ > 0\,(\geq 0)\} = 0. \end{gathered} \tag{26}$$

By the definition of $\operatorname{supp}\Omega$ in (16), we have $\mathscr{P}\{\omega \in \Omega \setminus \operatorname{supp}\Omega\} = 0$. Hence, formulation (26) is equivalent to

$$\begin{gathered} \overline{x}_i \neq \overline{y}_i, \\ \mathscr{P}\{\omega \in \Omega : (\overline{x}_i - \overline{y}_i)(F_i(\overline{x},\omega) - F_i(\overline{y},\omega)) > 0\,(\geq 0)\} = 0, \end{gathered} \tag{27}$$

which contradicts definition (20). Therefore, there exists a $\overline{\omega} \in \operatorname{supp}\Omega$ such that $F(\cdot,\overline{\omega})$ is a $P(P_0)$-function. □

Theorem 8. *Suppose that $f(x) := \mathbf{E}[F(x,\omega)]$ is a $P(P_0)$-function. Then F is a stochastic $P(P_0)$-function.*

Proof. Suppose on the contrary that F is not a stochastic $P(P_0)$-function, then there exist $\tilde{x}, \tilde{y}, \tilde{x} \neq \tilde{y}$ in $\mathbf{R}^n$ for any $i \in \langle 1,n\rangle$ satisfying

$$\begin{gathered} \tilde{x}_i \neq \tilde{y}_i, \\ \mathscr{P}\{\omega : (\tilde{x}_i - \tilde{y}_i)(F_i(\tilde{x},\omega) - F_i(\tilde{y},\omega)) > 0\,(\geq 0)\} = 0. \end{gathered} \tag{28}$$

This means that

$$(\tilde{x}_i - \tilde{y}_i)(F_i(\tilde{x},\omega) - F_i(\tilde{y},\omega)) \leq 0\,(< 0) \tag{29}$$

always holds for any $i \in \langle 1, n\rangle$ and $\omega \in \Omega$. Furthermore, following from (29), we have

$$\mathbf{E}\left[\left(\tilde{x}_i - \tilde{y}_i\right)\left(F_i\left(\tilde{x},\omega\right) - F_i\left(\tilde{y},\omega\right)\right)\right] \le 0\,(<0), \tag{30}$$

that is

$$\left(\tilde{x}_i - \tilde{y}_i\right)\left(f_i\left(\tilde{x}\right) - f_i\left(\tilde{y}\right)\right) \le 0\,(<0), \tag{31}$$

which contradicts the definition of $P(P_0)$-function. Therefore, F is a stochastic $P(P_0)$-function. □

Note that there is at most one solution (may not be a solution) for the EV model stochastic complementarity problems if $f(x) := \mathbf{E}[F(x,\omega)]$ is a $P(P_0)$-function.

3. Boundedness of Solution Set

Theorem 9. *Suppose that F is a stochastic uniformly P-function and F is equicoercive on $\mathbf{R}^n$. Then the solution set of ERM model (7) defined by $\phi_{\min}$ and ϕ_{FB} is nonempty and bounded.*

Proof. Suppose on the contrary that the ERM model defined by $\phi_{\min}$ is not bounded. Thus there exist a sequence $\{x^k\} \subset \mathbf{R}^n_+$ with $\|x^k\| \to \infty$ $(k \to \infty)$ and a constant $c \in \mathbf{R}_+$, such that

$$\theta\left(x^k\right) \le c, \quad \text{for } \forall k. \tag{32}$$

Define the index set $I \subseteq \{1,\ldots,n\}$ by

$$I := \left\{i \mid \left\{x_i^k\right\} \text{ is unbounded}\right\}. \tag{33}$$

By assumption, we have $I \neq \emptyset$. We now define a sequence $\{y^k\} \subseteq \mathbf{R}^n$ as follows:

$$y_i^k := \begin{cases} 0 & \text{if } i \in I, \\ x_i^k & \text{if } i \notin I. \end{cases} \tag{34}$$

From the definition of y^k and the fact that F is a stochastic uniformly P-function, we obtain that for any x^k, y^k, there exists i such that

$$\begin{aligned} &\mathscr{P}\left\{\omega : \left(x_i^k - y_i^k\right)\left(F_i\left(x^k,\omega\right) - F_i\left(y^k,\omega\right)\right)\right. \\ &\qquad \left. \ge \alpha\left\|x^k - y^k\right\|^2\right\} > 0, \end{aligned} \tag{35}$$

and hence there are $\overline{\omega}^k \in \operatorname{supp}\Omega$ satisfying

$$\left(x_i^k - y_i^k\right)\left(F_i\left(x^k,\overline{\omega}^k\right) - F_i\left(y^k,\overline{\omega}^k\right)\right) \ge \alpha\left\|x^k - y^k\right\|^2. \tag{36}$$

Take subsequence x^{k_i}, y^{k_i} such that the corresponding subscript of (36) is j. Noting that $j \in I$ and taking (36) into account, we have

$$\begin{aligned} &\alpha\sum_{j\in I}\left(x_j^{k_i}\right)^2 \\ &\quad \le x_j^{k_i}\left(F_j\left(x^{k_i},\overline{\omega}^{k_i}\right) - F_j\left(y^{k_i},\overline{\omega}^{k_i}\right)\right) \\ &\quad \le \sqrt{\sum_{j\in I}\left(x_j^{k_i}\right)^2}\cdot\sqrt{\sum_{j\in I}\left|F_j\left(x^{k_i},\overline{\omega}^{k_i}\right) - F_j\left(y^{k_i},\overline{\omega}^{k_i}\right)\right|}, \end{aligned} \tag{37}$$

from which we get

$$\alpha\sqrt{\sum_{j\in I}\left(x_j^{k_i}\right)^2} \le \sqrt{\sum_{j\in I}\left|F_j\left(x^{k_i},\overline{\omega}^{k_i}\right) - F_j\left(y^{k_i},\overline{\omega}^{k_i}\right)\right|}. \tag{38}$$

By definition, the sequence $\{y^{k_i}\}$ remains bounded. From the continuity of F, it follows that the sequence $\{F_j(y^{k_i},\omega^{k_i})\}$ is also bounded for every $j \in I$. Hence, taking a limit in (38), we obtain that there is at least one index $\overline{j} \in I$ such that

$$x_{\overline{j}}^{k_i} \longrightarrow \infty, \qquad F_{\overline{j}}\left(x^{k_i},\overline{\omega}^{k_i}\right) \longrightarrow \infty. \tag{39}$$

Since F is equicoercive on $\mathbf{R}^n$, we have

$$\mathscr{P}\left\{\omega : \lim_{k_i\to\infty} F_{\overline{j}}\left(x^{k_i},\omega\right) = \infty\right\} > 0. \tag{40}$$

Let

$$\Omega_1 = \left\{\omega : \lim_{k_i\to\infty} \min\left(x_{\overline{j}}^{k_i}, F_{\overline{j}}\left(x^{k_i},\omega\right)\right) = \infty\right\}. \tag{41}$$

Then $\mathscr{P}\{\Omega_1\} > 0$. By Fatou's Lemma [15], we have

$$\begin{aligned} &\mathbf{E}_{\Omega_1}\left[\liminf_{k_i\to\infty}\left(\min\left(x_{\overline{j}}^{k_i}, F_{\overline{j}}\left(x^{k_i},\omega\right)\right)\right)^2\right] \\ &\quad \le \liminf_{k_i\to\infty}\mathbf{E}_{\Omega_1}\left[\left(\min\left(x_{\overline{j}}^{k_i}, F_{\overline{j}}\left(x^{k_i},\omega\right)\right)\right)^2\right]. \end{aligned} \tag{42}$$

Since

$$\liminf_{k_i\to\infty}\left(\min\left(x_{\overline{j}}^{k_i}, F_{\overline{j}}\left(x^{k_i},\omega\right)\right)\right)^2 = \infty \tag{43}$$

on Ω_1 and $\mathscr{P}\{\Omega_1\} > 0$, then the left-hand side of formulation (42) is infinite. Therefore,

$$\liminf_{k_i\to\infty}\mathbf{E}_{\Omega_1}\left[\left(\min\left(x_{\overline{j}}^{k_i}, F_{\overline{j}}\left(x^{k_i},\omega\right)\right)\right)^2\right] = \infty. \tag{44}$$

Moreover, it is easy to find

$$\begin{aligned} \theta\left(x^{k_i}\right) &= \mathbf{E}\left[\left\|\Phi(x^{k_i}, F(x^{k_i},\omega))\right\|^2\right] \\ &\ge \mathbf{E}_{\Omega_1}\left[\left(\min\left(x_{\overline{j}}^{k_i}, F_{\overline{j}}\left(x^{k_i},\omega\right)\right)\right)^2\right] \\ &\longrightarrow \infty \end{aligned} \tag{45}$$

as $k_i \to \infty$. This contradicts formulation (32). Hence, the solution set of ERM model (7) defined by $\phi_{\min}$ is nonempty and bounded. Similar results about ϕ_{FB} can be obtained by relation formulation (10). □

4. Robust Solution

As we show, both EV model and ERM model give decisions by a deterministic formulation. However, the decisions may not be the best or may be even infeasible for each individual event. In fact, we should take risk into account to make

a priori decision in many cases. Naturally, it is necessary to know how good or how bad the decision which we have given can be. In this section, we study the robustness of solutions of the ERM model. Let $\mathrm{SOL}(F(x,\omega))$ denote the solution set of $\mathrm{SCP}(F(x,\omega))$, and define the distance from a point x to the set $\mathrm{SOL}(F(x,\omega))$ by

$$\operatorname{dist}(x, \mathrm{SOL}(F(x,\omega))) := \inf_{x' \in \mathrm{SOL}(F(x,\omega))} \left\| x - x' \right\|. \tag{46}$$

Theorem 10. *Assume that $\Omega = \{\omega^1, \omega^2, \ldots, \omega^N\} \subset \mathbf{R}^m$, and ω takes values $\omega^1, \ldots, \omega^N$ with respective probabilities $p_1, \ldots, p_N$. Furthermore, suppose that for every $\omega \in \Omega$, $F(x,\omega)$ is uniformly P-function and Lipschitz continuous with respect to x. Then there is a constant $C > 0$ such that*

$$\mathbf{E}\left[\operatorname{dist}(x, SOL(F(x,\omega)))\right] \le C \cdot \sqrt{\theta(x)}, \tag{47}$$

where $\theta(x)$ is defined by $\phi_{\min}$ or ϕ_{FB}.

Proof. For any fixed ω^i, since $F(x,\omega^i)$ is uniformly P-function and Lipschitz continuous, from Corollary 3.19 of [16], we have unique solution $\hat{x}(\omega^i)$ of $\mathrm{CP}(F(x,\omega^i))$, and there exists a constant C_i such that

$$\left\| x - \hat{x}\left(\omega^i\right) \right\| \le C_i \left\| \min\left\{x, F\left(x,\omega^i\right)\right\} \right\|. \tag{48}$$

Letting $C := ((\sqrt{2}+2)/2)\max\{C_1, \ldots, C_N\}$, we have

$$\begin{aligned}
&\mathbf{E}^2\left[\operatorname{dist}(x, \mathrm{SOL}(F(x,\omega)))\right] \\
&\quad = \mathbf{E}^2\left[\left\| x - \hat{x}\left(\omega^i\right)\right\|\right] \\
&\quad \le \mathbf{E}\left[\left\| x - \hat{x}\left(\omega^i\right)\right\|^2\right] \\
&\quad \le \sum_{i=1}^{N} p_i \cdot C_i^2 \left\| \min\left\{x, F\left(x,\omega^i\right)\right\}\right\|^2 \\
&\quad \le \sum_{i=1}^{N} p_i \cdot C_i^2 \cdot \left(\frac{\sqrt{2}+2}{2}\right)^2 \\
&\qquad \times \sum_{j=1}^{n}\left(\sqrt{F_j^2\left(x,\omega^i\right) + x_j^2} - \left(F_j\left(x,\omega^i\right) + x_j\right)\right)^2 \\
&\quad \le C^2 \sum_{i=1}^{N} p_i \sum_{j=1}^{n}\left(\sqrt{F_j^2\left(x,\omega^i\right) + x_j^2} - \left(F_j\left(x,\omega^i\right) + x_j\right)\right)^2 \\
&\quad = C^2 \theta(x),
\end{aligned} \tag{49}$$

where the first inequality follows from Cauchy-Schwarz inequality, the second inequality follows from formulation (48), and the third inequality follows from formulation (10). This completes the proof of the theorem. □

Theorem 10 particularly shows that for the solution x^* of (7),

$$\mathbf{E}\left[\operatorname{dist}\left(x^*, \mathrm{SOL}(F(x,\omega))\right)\right] \le C \cdot \sqrt{\theta(x^*)}. \tag{50}$$

This inequality indicates that the expected distance to the solution set $\mathrm{SOL}(F(x,\omega))$ for $\omega \in \Omega$ is also likely to be small at the solution x^* of (7). In other words, we may expect that a solution of the ERM formulation (7) has a minimum sensitivity with respect to random parameter variations in $\mathrm{SCP}(F(x,\omega))$. In this sense, solutions of (7) can be regarded as robust solutions for $\mathrm{SCP}(F(x,\omega))$.

5. Quasi-Monte Carlo and Derivative-Free Methods for Solving ERM Model

Note that the ERM model (7) included an expectation function, which is generally difficult to be evaluated exactly. Hence in this section, we first employ a quasi-Monte Carlo method to obtain approximation problems of (7) for numerical integration. Then, we consider derivative-free methods to solve these approximation problems.

By the quasi-Monte Carlo method, we obtain the following approximation problem of (7):

$$\min_{x \in \mathbf{R}_+^n} \theta^N(x) := \frac{1}{N} \sum_{\omega^i \in \Omega_N} \left\| \Phi\left(x,\omega^i\right)\right\|^2 \rho\left(\omega^i\right), \tag{51}$$

where $\Omega_N = \{\omega^i \mid i = 1, 2, \ldots, N\}$ is a set of observations generated by a quasi-Monte Carlo method such that $\Omega_N \subseteq \Omega$ and $\rho(\omega)$ stands for the probability density function. In the rest of this paper, we assume that the probability density function ρ is continuous on Ω. For each N, $\theta^N(x)$ is continuously differentiable function. We denote by x^N the optimal solutions of approximation problems (51). We are interested in the situation where the first-order derivatives of $\theta^N(x)$ cannot be explicitly calculated or approximated.

Condition 1. Given a point $x_0 \ge 0$, the level set

$$L := \{x \ge 0 \mid f(x) \le f(x_0)\} \tag{52}$$

is compact.

Condition 2. If $\{x_k^N\}$ and $\{y_k^N\}$ are sequences of points such that $x_k^N \ge 0$, $y_k^N \ge 0$ converging to some $\overline{x}^N$ and $I_k^N \subseteq I(\overline{x}^N) := \{i \mid \overline{x}_i^N = 0\}$ for all k, then

$$\left\{\operatorname{dist}\left(T_{I_k^N}\left(x_k^N\right), T_{I_k^N}\left(y_k^N\right)\right)\right\} \longrightarrow 0, \tag{53}$$

where $\operatorname{dist}(T_1, T_2) = \max_{d_1 \in T_1 : \|d_1\| = 1}\{\min_{d_2 \in T_2} \|d_1 - d_2\|\}$ and $T_{I_k^N}(x) := \{d_k^N \in \mathbf{R}^n \mid d_{k,i}^N \ge 0, \forall i \in I_k^N\}$.

Condition 3. For every $\overline{x}^N \ge 0$ there exist scalars $\delta > 0$, and $\eta > 0$ such that

$$\begin{gathered}
\min_{z \ge 0} \|z - x\| \le \eta \sum_{i=1}^{n} \max(-x_i, 0), \\
\forall x \in \left\{x \in \mathbf{R}^n \mid \|x - \overline{x}\| \le \delta\right\}.
\end{gathered} \tag{54}$$

Condition 4. Given x_k^N and $\epsilon_k^N > 0$, the set of search directions

$$D_k^N = \left\{d_k^{N,j}, j = 1, \ldots, r_k^N\right\}, \quad \text{with } \left\|d_k^{N,j}\right\| = 1, \tag{55}$$

satisfing r_k^N is uniformly bounded and cone$\{D_k^N\} = T(x_k^N; \epsilon_k^N)$. Here,

$$\begin{aligned} &\operatorname{cone}\left\{D_k^N\right\} \\ &\quad = \left\{d_k^{N,1}\beta^1 + \cdots + d_k^{N,r_k^N}\beta^{r_k^N} : \beta^1 \geq 0, \ldots, \beta^{r_k^N} \geq 0\right\}, \\ &\quad T\left(x_k^N; \epsilon_k^N\right) = \left\{d_k^N \in \mathbf{R}^n \mid d_{k,i}^N \geq 0, x_{k,i}^N \leq \epsilon_k^N\right\}. \end{aligned} \tag{56}$$

Under Conditions 1, 2, and 3 and by choosing D_k^N satisfying Condition 4 with $\epsilon_k^N \to 0$, then the following generated iterates have at least one cluster point that is a stationary point of (51) for each N.

Algorithm 11. Parameters: $x_0^N \geq 0$, $\tilde{\alpha}_0^N > 0$, $\gamma^N > 0$, $\theta_1^N \in (0,1)$, $\theta_2^N \in (0,1)$, $\epsilon_0^N > 0$.

Step 1. Set $k^N = 0$.

Step 2. Choose a set of directions $D_k^N = \{d_k^{N,j}, j = 1, \ldots, r_k^N\}$ satisfying Condition 4.

Step 3.

(a) Set $j = 1$, $y_k^{N,j} = x_k^N$, $\tilde{\alpha}_k^{N,j} = \tilde{\alpha}_k^N$.

(b) Compute the maximum stepsize $\overline{\alpha}_k^{N,j}$ such that $y_{i,k}^{N,j} + \overline{\alpha}_k^{N,j}\hat{d}_{i,k}^{N,j} \geq 0$ for all i. Set $\hat{\alpha}_k^{N,j} = \min\{\overline{\alpha}_k^{N,j}, \tilde{\alpha}_k^{N,j}\}$.

(c) If $\hat{\alpha}_k^{N,j} > 0$ and $\theta^N(\hat{y}_k^{N,j}) \leq \theta^N(y_k^{N,j}) - \gamma(\hat{\alpha}_k^{N,j})^2$, set $\tilde{\alpha}_k^{N,j+1} = \alpha_k^{N,j}$; otherwise set $\alpha_k^{N,j} = 0$, $y_k^{N,j+1} = y_k^{N,j}$, $\tilde{\alpha}_k^{N,j+1} = \theta_1^N\tilde{\alpha}_k^{N,j}$.

(d) If $\alpha_k^{N,j} = \overline{\alpha}_k^{N,j}$, set $\epsilon_{k+1}^N = \epsilon_k^N$, and go to Step 4.

(e) If $j < r_k^N$, set $j = j+1$, and go to Step 3(b). Otherwise set $\epsilon_{k+1}^N = \theta_2^N\epsilon_k^N$ and go to Step 4.

Step 4. Find $x_{k+1}^N \geq 0$ such that $\theta^N(x_{k+1}^N) \leq \theta^N(y_k^{N,j+1})$. Set $\tilde{\alpha}_{k+1}^N = \tilde{\alpha}_k^{N,j+1}$, $\tilde{r}_k = j$, $k = k+1$, and go to Step 2.

For this algorithm, it is easy to proof that if x_k^N is the sequence produced by algorithm under Conditions 1–4, then x_k^N is bounded and there exists at least one cluster point which is a stationary point of problem (51) for each N.

6. Conclusions

The SCP$(F(x,\omega))$ has a wide range of applications in engineering and economics. Therefore, it is meaningful and interesting to study this problem. In this paper, we give the definitions of stochastic P-function, stochastic P_0-function and stochastic uniformly P-function, which can be regarded as a generalization of the deterministic formulation or an extension of a stochastic R_0 function given in [11]. Moreover, we consider the conditions when the function is a stochastic $P(P_0)$-function. Furthermore, we show that the involved function being a stochastic uniformly P-function and equicoercive [11] are sufficient conditions for the solution set of the expected residual minimization problem to be nonempty and bounded. Finally, we illustrate that the ERM formulation produces robust solutions with minimum sensitivity in violation of feasibility with respect to random parameter variations in SCP$(F(x,\omega))$. On the other hand, we employ a quasi-Monte Carlo method to obtain approximation problems of (7) for dealing numerical integration and further consider derivative-free methods to solve these approximation problems.

Acknowledgments

This work was supported by NSFC Grants no. 11226238 and no. 11226230, and predeclaration fund of state project of Liaoning university 2012, 2012LDGY01, and University Scientific Research Projects of School of Education Department of Liaoning Province 2012, 2012427.

References

[1] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, NY, USA, 2003.

[2] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem*, Computer Science and Scientific Computing, Academic Press, Boston, Mass, USA, 1992.

[3] G.-H. Lin and M. Fukushima, "New reformulations for stochastic nonlinear complementarity problems," *Optimization Methods & Software*, vol. 21, no. 4, pp. 551–564, 2006.

[4] G. Gürkan, A. Y. Özge, and S. M. Robinson, "Sample-path solution of stochastic variational inequalities," *Mathematical Programming*, vol. 84, no. 2, pp. 313–333, 1999.

[5] X. Chen and M. Fukushima, "Expected residual minimization method for stochastic linear complementarity problems," *Mathematics of Operations Research*, vol. 30, no. 4, pp. 1022–1038, 2005.

[6] X. Chen, C. Zhang, and M. Fukushima, "Robust solution of monotone stochastic linear complementarity problems," *Mathematical Programming*, vol. 117, no. 1-2, pp. 51–80, 2009.

[7] P. Tseng, "Growth behavior of a class of merit functions for the nonlinear complementarity problem," *Journal of Optimization Theory and Applications*, vol. 89, no. 1, pp. 17–37, 1996.

[8] H. Fang, X. Chen, and M. Fukushima, "Stochastic R_0 matrix linear complementarity problems," *SIAM Journal on Optimization*, vol. 18, no. 2, pp. 482–506, 2007.

[9] G.-H. Lin, X. Chen, and M. Fukushima, "New restricted NCP functions and their applications to stochastic NCP and stochastic MPEC," *Optimization*, vol. 56, no. 5-6, pp. 641–953, 2007.

[10] C. Ling, L. Qi, G. Zhou, and L. Caccetta, "The SC^1 property of an expected residual function arising from stochastic complementarity problems," *Operations Research Letters*, vol. 36, no. 4, pp. 456–460, 2008.

[11] C. Zhang and X. Chen, "Stochastic nonlinear complementarity problem and applications to traffic equilibrium under uncertainty," *Journal of Optimization Theory and Applications*, vol. 137, no. 2, pp. 277–295, 2008.

[12] C. Zhang and X. Chen, "Smoothing projected gradient method and its application to stochastic linear complementarity problems," *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 627–649, 2009.

[13] G. L. Zhou and L. Caccetta, "Feasible semismooth Newton method for a class of stochastic linear complementarity problems," *Journal of Optimization Theory and Applications*, vol. 139, no. 2, pp. 379–392, 2008.

[14] X. L. Li, H. W. Liu, and Y. K. Huang, "Stochastic P matrix and P_0 matrix linear complementarity problem," *Journal of Systems Science and Mathematical Sciences*, vol. 31, no. 1, pp. 123–128, 2011.

[15] K. L. Chung, *A Course in Probability Theory*, Academic Press, New York, NY, USA, 2nd edition, 1974.

[16] B. Chen and P. T. Harker, "Smooth approximations to nonlinear complementarity problems," *SIAM Journal on Optimization*, vol. 7, no. 2, pp. 403–420, 1997.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 437637, 6 pages
http://dx.doi.org/10.1155/2013/437637

Research Article

Car Delay Model near Bus Stops with Mixed Traffic Flow

Yang Xiaobao, Huan Mei, and Gao Ziyou

MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Yang Xiaobao; yangxb@bjtu.edu.cn

Received 27 January 2013; Accepted 23 April 2013

Academic Editor: Xiaoning Zhang

This paper proposes a model for estimating car delays at bus stops under mixed traffic using probability theory and queuing theory. The roadway is divided to serve motorized and nonmotorized traffic streams. Bus stops are located on the nonmotorized lanes. When buses dwell at the stop, they block the bicycles. Thus, two conflict points between car stream and other traffic stream are identified. The first conflict point occurs as bicycles merge to the motorized lane to avoid waiting behind the stopping buses. The second occurs as buses merge back to the motorized lane. The average car delay is estimated as the sum of the average delay at these two conflict points and the delay resulting from following the slower bicycles that merged into the motorized lane. Data are collected to calibrate and validate the developed model from one site in Beijing. The sensitivity of car delay to various operation conditions is examined. The results show that both bus stream and bicycle stream have significant effects on car delay. At bus volumes above 200 vehicles per hour, the curbside stop design is not appropriate because of the long car delays. It can be replaced by the bus bay design.

1. Introduction

As the first point of contact between the passenger and the public transit service, the bus stop is a critical element in a transit system's overall goal of providing timely, safe, and convenient transportation. In the past several decades, traffic planners, designers, and scholars have paid much attention to the location, design, and operations of bus stops [1–5]. A prominent achievement of this research is a set of guidelines for use in designing and locating bus stops, sponsored by TCRP in the United States [1]. Other researchers mainly focused on the effects of bus stops on traffic flow. For example, Wong et al. analyzed the delay at a signal-controlled intersection with a bus stop upstream [6]. Fernández applied the microscopic traffic simulation model to study operational impacts on bus stops such as capacity, delays, queues, and waiting times [7]. Tang et al. used the macrodynamic model to analyze the effects of bus stop on traffic flow [8]. Most existing research of bus stops analyzes only the mixed traffic flow between buses and cars without including nonmotorized vehicles. This might be the reason that the car-bus conflict at bus stops is usually regarded as more important than the motorized vehicle-bicycle conflict. Another reason might be the low proportion of cycling among all travel modes in developed countries.

As a developing country, China has its own traffic characteristics. A mix of nonmotorized and motorized vehicles is an important traffic type in China. Some surveys show that the nonmotorized vehicle, especially the bicycle, is one of the most widely used traffic tools in Chinese daily travel activity. Typically, there are three types of bus stops in urban areas: curbside stops, bus bays, and bus boarders [9]. The curbside stop is the most common type on minor roadways in many Chinese cities. Figure 1 shows the mixed traffic streams at a typical curbside stop. There are two lanes on the urban roadway: the bicycle lane and the motorized lane, and there are three types of traffic streams: bicycle, bus, and car. Bus stops are usually located on the bicycle lane. When a bus dwells at the curbside stop, it blocks the bicycles. Bicycles merge to the motorized lane to avoid waiting behind the stopping bus. Thus, the presence of a stopped bus creates a temporary conflict between bicycles and cars. In addition, the car-bus conflict takes place when a bus departs from the stop to the motorized lane. Similar phenomena may be found in other Asian developing countries, for example, India, Malaysia, Vietnam, and Cambodia.

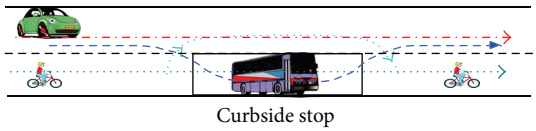

Figure 1: Curbside stop with mixed traffic streams of buses, cars, and bicycles.

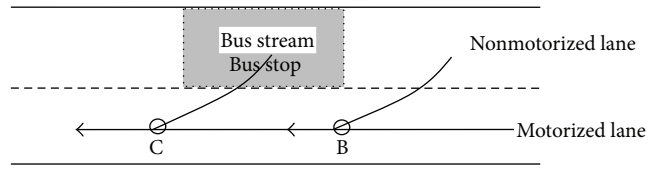

Figure 2: Conflict among cars, buses and bicycles at a curbside stop.

Due to the special features of mixed traffic, the application of existing traffic models for bus stops, developed by developed countries, has not produced a clear effect on Chinese traffic management and control. Therefore, it is necessary to deeply study the mixed traffic flow between nonmotorized vehicles and motorized vehicles.

On the mixed traffic, till now, much research has been conducted on basic segments and intersections [10, 11], but the correlative research on bus stops is much less in the literature. In recent years, some researchers have realized this, and correlative work is being done, but it still has a long way to go. Koshy and Arasan used simulation technique to study the impact of bus stop type on the speeds of other vehicles under heterogeneous conditions [9]. Yang et al. established car capacity models near a curbside stop with bicycles based on gap acceptance theory and conflict technique [12, 13]. However, little information was found in the literature on delay time near bus stops with mixed traffic flow. This paper investigates car delay time near a curbside stop under mixed traffic conditions. Firstly, mixed traffic flow characteristics near bus stops are analyzed. Then, the delay model based on probability and queuing theory is proposed. Next, the delay model is validated by field data in Beijing. In addition, the sensitivity of car delay to various operation conditions is examined. Finally, conclusions and future researches are given.

2. Mixed Traffic Flow Characteristics near Curbside Stops

2.1. Bus Stream: M/M/k Queuing Model. Consider a road link near the bus stop as shown in Figure 1. A sophisticated queuing theory model can be developed on the assumption that the simple bus stream system can be represented by an M/M/k queue. The service counter is the bus stop. The input into the system in equilibrium, as well as the output, is formed by the buses approaching from upstream, which are assumed to arrive at random; that is, there are negative exponentially distributed arrival headways with mean $1/\lambda_b$ seconds. The dwelling time at the stop is the service time, which is also assumed to be independent and negative exponential distributed random variables with mean t_b seconds. Finally, the "k" in M/M/k stands for k identical servers, that is, the number of existing berths at the bus stop.

For the M/M/k system as a general property, the probability of the busy system is given by the following equation. That is, the probability of one or more buses at the stop is

$$p_s = 1 - \left(\sum_{j=0}^{k-1} \frac{\rho_b^j}{j!} + \frac{\rho_b^k}{k!} \cdot \frac{k}{k - \rho_b} \right)^{-1}. \quad (1)$$

Note that, here, $\rho_b = \lambda_b t_b$, and for the existence of a steady-state solution, $\lambda_b < k/t_b$. The subscript "b" stands for bus stream passing the stop and the subscript "s" stands for bus stream at the stop.

And the expected number in the system at steady state; that is, the expected number of buses both in service and in queue at the stop is

$$L = E(N) = \frac{\rho_b^k}{k!} \cdot \frac{k\rho_b}{(k - \rho_b)^2} (1 - p_s) + \rho_b. \quad (2)$$

2.2. Conflict between Different Streams. At the curbside stop with mixed traffic flow, there are three streams among buses, bicycles, and cars. As shown in Figure 2, car stream is directly affected by two conflicts. One is the interaction between car stream and bicycle stream at point B when one or more buses dwell at the stop. The other is the conflict between car stream and bus stream at point C as buses merge back to the motorized lane.

When a bus dwells at the curbside stop in the nonmotorized lane, the nonmotorized lane is blocked by a stopped bus. A lane change for bicycles from the nonmotorized lane to the motorized lane is "essential" when bicyclists approach to the last stopped bus. As a result, forced lane changing maneuvers take place. Bicycles in the nonmotorized lane would force the subsequent car in the motorized lane to slow down for the lane-changing execution. Field observations indicate that this cooperative lane changing and priority-sharing behavior is prevalent between bicycles and cars near bus stops [13]. As the acceptable gap of bicycles is approximate to the follow-up time of successive cars, the bicycle-car conflict near a bus top is similar to the conflict at merges under low speed or high flow conditions. In the saturated traffic flow, gap acceptance theory completely loses its applicability; waiting vehicles generally perform forced lane-changing maneuvers and pass the conflict point alternately [14, 15]. That is to say, they comply with the FIFO (first-in-first-out) discipline under low speed condition. Near a curbside stop, the vehicles usually pass the conflicting areas with a low speed because there is a serious conflict among different streams. The conflicting areas near a curbside stop can also be considered in such a way that the FIFO discipline is applied.

3. Delay Model for Car Stream near the Stop under Mixed Traffic Flow

The average delay to car stream near the stop under mixed traffic conditions is estimated as the sum of the average delay at these two conflict points and the delay resulting

from following the slower bicycle traffic that merged into the motorized lanes.

3.1. Delay Resulting from following the Slower Bicycles from Point B to Point C. Both car stream and bicycle stream are assumed to arrive at random, that is, negative exponentially distributed arrival headways. Let t_n denote the time that a bicycle (nonmotorized vehicle) arrives at point B, and t_c denote the time that a car arrives at point B behind the bicycle. Let Z be a nonnegative random variable representing the difference between t_c and t_n:

$$Z = t_c - t_n. \tag{3}$$

Then, the probability that Z is not more than a given time z can be expressed as

$$\begin{aligned} P(Z \le z) &= \iint_{t_c - t_n \le z} f(t_n) f(t_c)\, dt_c\, dt_n \\ &= \int_0^{+\infty} \int_0^{z+t_n} f(t_n) f(t_c)\, dt_c\, dt_n \\ &= \int_0^{+\infty} \int_0^{z+t_n} \lambda_n e^{-\lambda_n t_n} \cdot \lambda_c e^{-\lambda_c t_c} dt_c\, dt_n \\ &= 1 - \frac{\lambda_n}{\lambda_n + \lambda_c} e^{-\lambda_c z}. \end{aligned} \tag{4}$$

Then, the probability density function of Z is

$$f(z) = \frac{dF(z)}{dz} = \frac{dP(Z \le z)}{dz} = \frac{\lambda_n \lambda_c}{\lambda_n + \lambda_c} e^{-\lambda_c z}. \tag{5}$$

As shown in Figure 2, it may occur that a car follows the slower bicycles when a bicyclist rides at section BC. The distance from B to C can be obtained by the formula:

$$l_{\mathrm{BC}} = l_b \cdot L, \tag{6}$$

where l_b is the minimum headway distance of successively stopped buses at the stop and L is the expected number of buses at the stop which can be given as (2).

The phenomenon that a car follows the slower bicycle takes place only when one or more buses berth the stop and Z falls within limits, $0 \le Z \le z_{\max}$. The minimum value, 0, represents the condition that a car begins to follow a bicycle from point B. The maximum value, $z_{\max}$, represents the condition that a car begins to follow a bicycle from point C. If Z is less than $z_{\max}$, a car must decelerate to follow the preceding bicycle before point C. $z_{\max}$ can be calculated as

$$z_{\max} = l_{\mathrm{BC}} \left(\frac{1}{v_{1n}} - \frac{1}{v_{1c}} \right), \tag{7}$$

where v_{1c} and v_{1n} are the free-flow velocity for car stream and bicycle stream near a curbside stop, respectively. It is noted that $(1/v_{1n} - 1/v_{1c})$ is the delay for a car driving one meter when it decelerates to follow the slower bicycle.

Let a car with the free-flow velocity catch up with its preceding bicycle at the time t. In this case, the distance that the car drives from time t_c to t is equal to the distance that the bicyclist rides from time t_n to t, which can be expressed as

$$l_1 = v_{1n}(t - t_n) = v_{1c}(t - t_c). \tag{8}$$

Combined with (3), (7), and (8), l_1 can be given as

$$l_1 = \frac{Z}{(1/v_{1n} - 1/v_{1c})} = Z \cdot \frac{l_{\mathrm{BC}}}{z_{\max}}. \tag{9}$$

Thus, the remaining distance that a car must decelerate to follow its preceding bicycle in section BC is

$$l_2 = l_{\mathrm{BC}} - l_1 = l_{\mathrm{BC}} - Z \cdot \frac{l_{\mathrm{BC}}}{z_{\max}}. \tag{10}$$

The expected value that a car decelerates to follow the preceding bicycle can be calculated as

$$\begin{aligned} E(l_2) &= l_{\mathrm{BC}} - \frac{l_{\mathrm{BC}}}{z_{\max}} \int_0^{z_{\max}} z f(z)\, dz \\ &= l_{\mathrm{BC}} - \frac{\lambda_n l_{\mathrm{BC}}}{(\lambda_n + \lambda_c)\lambda_c z_{\max}} \\ &\quad \times \left(1 - e^{-\lambda_c z_{\max}} - \lambda_c z_{\max} e^{-\lambda_c z_{\max}}\right). \end{aligned} \tag{11}$$

It occurs that a car follows the slower bicycles only when one or more buses berth the stop. Thus, combined with (1), (7), and (11), the delay resulting from following the slower bicycles can be calculated as

$$\begin{aligned} d_{\mathrm{BC}} &= p_s \cdot E(l_2) \cdot \left(\frac{1}{v_{1n}} - \frac{1}{v_{1c}} \right) \\ &= p_s \left[z_{\max} - \frac{\lambda_n}{(\lambda_n + \lambda_c)\lambda_c} \left(1 - e^{-\lambda_c z_{\max}} - \lambda_c z_{\max} e^{-\lambda_c z_{\max}}\right) \right]. \end{aligned} \tag{12}$$

3.2. Delay at Two Conflict Points. As shown in Figure 2, for the conflicting point B, if one or more buses berth the stop, the car-bicycle conflict takes place. The conflicting areas near bus stops can be considered in such a way that the FIFO discipline is applied. Because different stream has different service time passing the conflicting area, the car-bicycle conflict at point B can be represented by the advanced Markovian model with no priorities but unequal arrival rates and unequal service rates for customers of two major types. Similarly, the car-bus conflict at point C can be described by the advanced Markovian model.

It is assumed that each stream arrives as a Poisson process to a single exponential channel, and there are two types of customers with no priorities but unequal arrival rates (λ_1, λ_2) and unequal service rates (μ_1, μ_2, and $\mu_2 > \mu_1$). Here, service rate (μ) is the reciprocal value of the mean service time (s), that is, $\mu_1 = 1/s_1$, $\mu_2 = 1/s_2$. Then, the expected waiting time for each type of customers in queue at steady state [16] is

$$W_{q_1} = \frac{(\lambda_1 + \lambda_2) s_1^2 \left[1 - \lambda_2 s_2 (1 - s_2/s_1)\right]}{(1 - \lambda_1 s_1 - \lambda_2 s_2)}, \tag{13}$$

$$W_{q_2} = \frac{(\lambda_1 + \lambda_2) s_1^2 \left[s_2^2/s_1^2 + (1 - s_2/s_1)(\lambda_1 s_2)\right]}{(1 - \lambda_1 s_1 - \lambda_2 s_2)}. \tag{14}$$

3.2.1. Delay Resulting from the Car-Bicycle Conflict at Point B. Traffic conditions near a curbside stop are classified into two types: inexistence and existence of stopped bus at the stop. The probabilities of these two conditions can be obtained by using (1). Under the former condition, the bicycle stream and the car stream at point B have no conflict and car travel time is not affected by the bicycle stream. Under the latter condition, the car-bicycle conflict at point B leads to an effect on car travel time by the bicycle stream. In this case, the car-bicycle conflict at point B can be represented by the advanced Markovian model. The car delay caused by the car-bicycle conflict is equal to the expected waiting time in the queue system for mixed streams between cars and bicycles, $W_{c,\mathrm{B}}$. As the mean service time for car stream at point B is larger than that for bicycle stream, $W_{c,\mathrm{B}}$ can be obtained by (13). Thus, the car delay at the point B can be given as

$$\begin{aligned} d_{\mathrm{B}} &= p_s W_{c,\mathrm{B}} + (1 - p_s) \cdot 0 \\ &= p_s s_c^2 (\lambda_c + \lambda_n) \frac{1 - (1 - s_n/s_c) \lambda_n s_n}{1 - \lambda_c s_c - \lambda_n s_n}, \end{aligned} \tag{15}$$

where λ_n and λ_c are the arrival rate of bicycle stream and car stream approaching the conflicting point B, respectively. s_n and s_c are the mean service time of bicycle stream and car stream passing the point B, respectively.

3.2.2. Delay Resulting from the Car-Bus Conflict at Point C. As buses merge back to the motorized lane, the car-bus conflict takes place. The car-bus conflict at point C can be represented by the advanced Markovian model. The car delay caused by the car-bus conflict is equal to the expected waiting time in the queue system for mixed streams between cars and buses, $W_{c,\mathrm{C}}$. As the mean service time for car stream at point C is less than that for bus stream, $W_{c,\mathrm{C}}$ can be obtained by (14). Thus, the car delay at the point C can be given as

$$d_{\mathrm{C}} = W_{c,\mathrm{C}} = s_b'^2 (\lambda_c + \lambda_b) \frac{s_c^2/s_b'^2 + \left(1 - s_c/s_b'\right) \lambda_b s_c}{1 - \lambda_b s_b' - \lambda_c s_c}, \tag{16}$$

where λ_b and λ_c are the arrival rate of bus stream and car stream, respectively. s_b' and s_c are the mean service time of bus stream and car stream passing the point C, respectively.

4. Model Validation and Comparison

In order to calibrate the proposed model of car delay at the curbside stop under mixed traffic conditions, field data collected at a bus stop in Beijing were employed. Video cameras were used to record traffic operations at the bus stop. Vehicle type, flows, and travel times were recorded for each vehicle passing through the stop. In addition, the dwell time of bus stream and the headways in the conflicting area were also recorded. Data were collected in the spring of 2008 in one direction over 3 minutes categorized into a group.

The basic parameter used to compute car delay at a stop under mixed traffic is the service time for each stream. The service time was directly measured for each vehicle at the conflicting point using video during this study. The service time in this paper is the follow-up headway for vehicles in this approach if no vehicle is waiting on the conflicting approach and is equal to the minimum saturation headway. On the basis of field survey and video process, the minimum saturation headways for bicycle stream, car stream, and bus stream are 0.90 s, 2.04 s, and 4.27 s, respectively. Here, the relatively low value for bicycle stream is the result of cycling parallel behavior and group behavior. In addition, $l_b = 12$ m, $v_{1n} = 4.5$ m/s, and $v_{1c} = 10$ m/s on the surveyed curbside stops in Beijing.

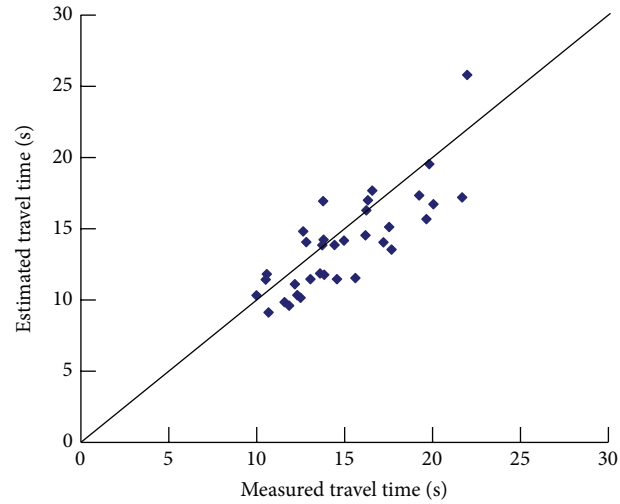

Figure 3: Comparison of measured average travel time and estimated average travel time.

The data collected in field study are used to validate the model, as shown in Figure 3. To facilitate comparison, the line where the measured average travel time equals to the estimated average travel time is superimposed on each figure. And it is found that scatter dots fluctuates narrowly around the line. In addition, the mean percent error between the estimated travel times and the measured times is −6.6%, and the mean absolute percentage error is 12.7%. Thus, the proposed delay model at the stop with mixed traffic flow is desirable. Before the applications, however, it is noted that the model should be estimated using the specified field data.

5. Effects of Individual Traffic Stream on Car Delay

Differences in the arrival rate of bicycle stream affect car delay time near the curbside stop are shown in Figure 4. Here, the curbside stop has two berths; that is to say, bus stream system can be considered as an M/M/2 queue. In addition, $\lambda_b = 0.03$ veh/s, $s_b = 25$ s, $l_b = 12$ m, $v_{1n} = 4.5$ m/s, and $v_{1c} = 10$ m/s, and the minimum saturation headways for bicycle stream, car stream, and bus stream are 0.90 s, 2.04 s, and 4.27 s, respectively. Bicycle stream headways follow the negative exponential distribution. At the same car flow rate, the probability of the car-bicycle conflict increases with the increasing bicycle stream, which finally lead to the increase of car delays.

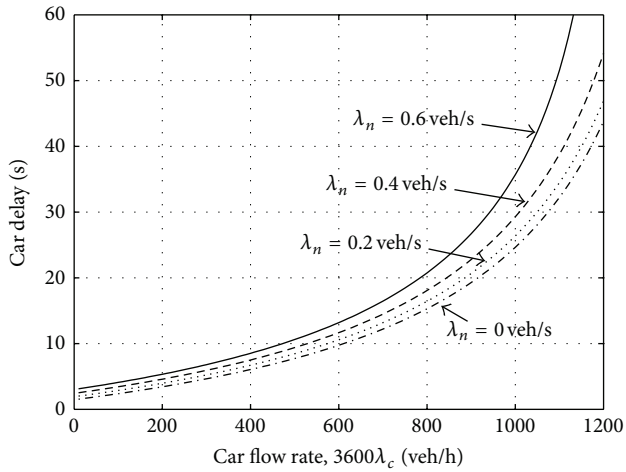

Figure 4: Car delay with different arrival rates for bicycle stream, $\lambda_b = 0.03$ veh/s and $s_b = 25$ s.

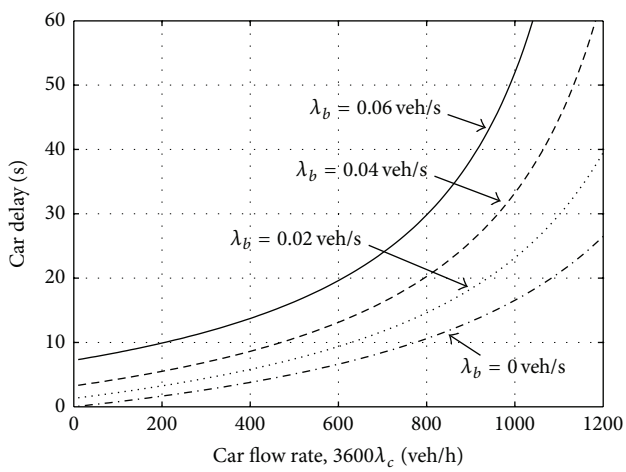

Figure 5: Car delay with different arrival rates for bus stream, $\lambda_n = 0.3$ veh/s and $s_b = 25$ s.

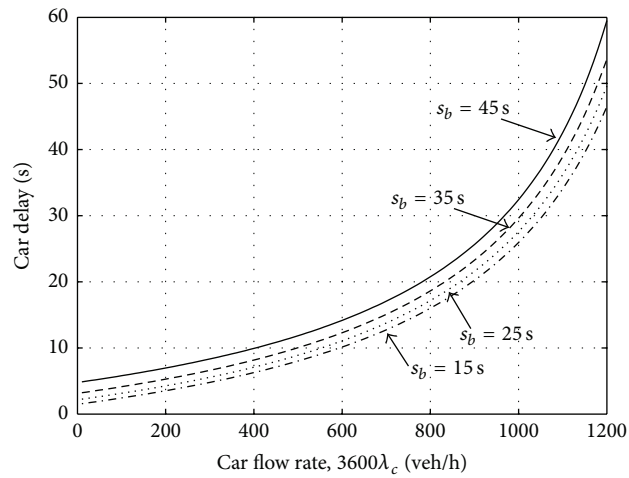

Figure 6: Car delay with different dwelling times for bus stream, $\lambda_n = 0.3$ veh/s and $\lambda_b = 0.03$ veh/s.

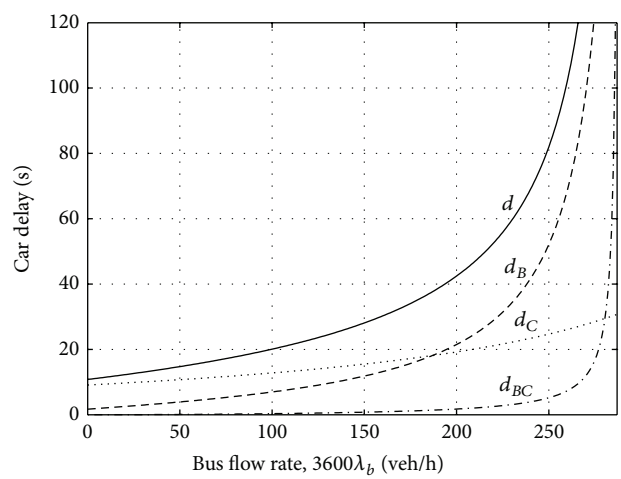

Figure 7: Effects of bus stream on car delay, $\lambda_n = 0.6$ veh/s, $\lambda_c = 2/9$ veh/s, and $s_b = 25$ s.

Figures 5 and 6 give the effects of bus stream on car delay time. Firstly, as shown in Figure 5, the probability of one or more stopped buses at the stop increases with the increasing bus flow rate, and so does also the probability of the car-bus conflict. This extends car travel time. Similarly, Figure 6 displays car delay with different dwelling times of bus stream. As the dwelling time of bus stream increases, the probability of one or more stopped buses at the stop increases. This increases the delay time for car stream on the basis of (12) and (15).

In addition, Figures 4, 5, and 6 all show that car delay increases with the creasing car flow rate. Meanwhile, a comparison of these three figures indicates that bus flow rate has the most significant effect on car delays. This is because all parts of car delay (including d_{B}, d_{BC}, and d_{C}) increase as bus flow rate increases. As shown in Figure 7, especially, at high bus flow rates, as the queuing systems are near the threshold values of equilibrium conditions; that is, $\lambda_c s_c + \lambda_n s_n$ or $\lambda_b s_b / k$ is near to one, d_{B} and d_{BC} will become infinite. The results are consistent with observed phenomena that car delays are very long under high flow and slow speed conditions. In this case, most drivers would give way to bicyclists merging into the motorized lane as bicycles have advantages of small size, light weight, flexible action, and so on [12].

6. Conclusion

Delay time to cars at a bus stop with mixed traffic flow is investigated on the basis of queuing theory and probability theory. Bus stream system can be represented by an M/M/k queue. Meanwhile, the conflict between different streams at the stop can be described by the advanced Markovian model with no priorities but unequal service rates for customers. The delay resulting from following the slower bicycles can be obtained by the joint distribution of bivariate continuous random variable. The analysis shows that both bus stream and bicycle stream have significant effects on car delay. Car delay near the stop under mixed traffic condition is the function of

three types of traffic streams with buses, cars, and bicycles. At bus volumes above approximately 200 vehicles per hour, the curbside stop design is unreasonable because of the long car delays. Therefore, in this case, it can be replaced by the bus bay design. The proposed model may be applicable to design and operational analysis of bus stops in other Asian developing countries.

Although this study has given valuable insights into car delay at the bus stop with mixed traffic flow, possible further research work is suggested. Firstly, we assume that bicycles have no priority over cars when bus berths at the stop in this paper. However, field observations show that sometimes a few bicyclists or drivers politely allow others to proceed. Bunker and Troutbeck [17] studied minor stream delays at a limited priority freeway merge. We need to present a new delay model which assumes limited priority for different streams near bus stops. In addition, the plan and design problems of bus stops with mixed traffic flow should be further researched.

Acknowledgments

This work was supported by the National Basic Research Program of China (Grant no. 2012CB725401), National Natural Science Foundation of China (Grants nos. 70901005, 71131001), and Fundamental Research Funds for the Central Universities (Grant no. 2013JBM043).

References

[1] Federal Transit Administration, *Transit Cooperative Research Program Report 19: Guidelines for the Location and Design of Bus Stops*, Transportation Research Board, Washington, DC, USA, 1996.

[2] W. Y. Szeto and Y. Z. Wu, "A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong," *European Journal of Operational Research*, vol. 209, no. 2, pp. 141–155, 2011.

[3] Á. Ibeas, L. Dell'Olio, B. Alonso, and O. Sainz, "Optimizing bus stop spacing in urban areas," *Transportation Research E*, vol. 46, no. 3, pp. 446–458, 2010.

[4] W. H. Gu, Y. W. Li, M. J. Cassidy, and J. B. Griswold, "On the capacity of isolated, curbside bus stops," *Transportation Research B*, vol. 45, no. 4, pp. 714–723, 2011.

[5] X. B. Yang, B. F. Si, and M. Huan, "Mixed traffic flow modeling near Chinese bus stops and its applications," *Journal of Central South University*, vol. 19, no. 9, pp. 2697–2704, 2012.

[6] S. C. Wong, H. Yang, W. S. A. Yeung, S. L. Cheuk, and M. K. Lo, "Delay at signal-controlled intersection with bus stop upstream," *Journal of Transportation Engineering*, vol. 124, no. 3, pp. 229–234, 1998.

[7] R. Fernández, "Modelling public transport stops by microscopic simulation," *Transportation Research C*, vol. 18, no. 6, pp. 856–868, 2010.

[8] T. Q. Tang, Y. Li, and H. J. Huang, "The effects of bus stop on traffic flow," *International Journal of Modern Physics C*, vol. 20, no. 6, pp. 941–952, 2009.

[9] R. Z. Koshy and V. T. Arasan, "Influence of bus stops on flow characteristics of mixed traffic," *Journal of Transportation Engineering*, vol. 131, no. 8, pp. 640–643, 2005.

[10] D. H. Wang, T. J. Feng, and C. Y. Liang, "Research on bicycle conversion factors," *Transportation Research A*, vol. 42, no. 8, pp. 1129–1139, 2008.

[11] H. W. Guo, W. H. Wang, W. W. Guo, X. B. Jiang, and H. Bubb, "Reliability analysis of pedestrian safety crossing in urban traffic environment," *Safety Science*, vol. 50, no. 4, pp. 968–973, 2012.

[12] X. B. Yang, Z. Y. Gao, X. M. Zhao, and B. F. Si, "Road capacity at bus stops with mixed traffic flow in China," *Transportation Research Record*, vol. 2111, pp. 18–23, 2009.

[13] X. B. Yang, Z. Y. Gao, B. F. Si, and L. Gao, "Car capacity near bus stops with mixed traffic derived by additive-conflict-flows procedure," *Science China Technological Sciences*, vol. 54, no. 3, pp. 733–740, 2011.

[14] S. Teply, M. I. Abou-Henaidy, and J. D. Hunt, "Gap acceptance behavior: aggregate and logit perspective," *Traffic Engineering and Control*, vol. 38, no. 9, pp. 474–482, 1997.

[15] A. Polus and S. Shmueli, "Analysis and evaluation of the capacity of roundabouts," *Transportation Research Record*, vol. 1572, pp. 99–104, 1997.

[16] J. Medhi, *Stochastic Models in Queueing Theory*, Academic Press, San Diego, Calif, USA, 2nd edition, 2002.

[17] J. Bunker and R. Troutbeck, "Prediction of minor stream delays at a limited priority freeway merge," *Transportation Research B*, vol. 37, no. 8, pp. 719–735, 2003.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 696481, 11 pages
http://dx.doi.org/10.1155/2013/696481

Research Article

Stochastic Congestion Pricing among Multiple Regions: Competition and Cooperation

Hua Wang,[1] Wei Mao,[2] and Hu Shao[3]

[1] *Centre for Maritime Studies, National University of Singapore, Singapore 118411*
[2] *School of Business, Central South University, Changsha, Hunan 410000, China*
[3] *Department of Mathematics, School of Sciences, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China*

Correspondence should be addressed to Hu Shao; hu_shao2005@yahoo.com

Received 17 January 2013; Accepted 3 March 2013

Academic Editor: Xiaoning Zhang

Previous studies of road congestion pricing problem assume that transportation networks are managed by a central administrative authority with an objective of improving the performance of the whole network. In practice, a transportation network may be comprised of multiple independent local regions with relative independent objectives. In this paper, we investigate the cooperative and competitive behaviors among multiple regions in congestion pricing considering stochastic conditions; especially demand uncertainty is taken into account in transportation modelling. The corresponding congestion pricing models are formulated as a bilevel programming problem. In the upper level, congestion pricing model either aims to maximize the regional social welfare in competitive schemes or attempts to maximize the total social welfare of multiple regions in cooperative schemes. In the lower level, travellers are assumed to follow a reliability-based stochastic user equilibrium principle considering risks of late arrival under uncertain conditions. Numerical examples are carried out to compare the effects of different pricing schemes and to analyze the impact of travel time reliability. It is found that cooperative pricing strategy performs better than competitive strategy in improving network performance, and the pricing effects of both schemes are quite sensitive to travel time reliability.

1. Introduction

Congestion pricing is widely regarded as an effective strategy to alleviate traffic congestion in transportation networks. It is also viewed as one of the most efficient means by transportation economists as it employs the price mechanism. Congestion pricing, thus, has been paid extensive attention in the literature [1]. In the congestion pricing scheme, the decision maker (e.g. road authority) aims to optimize system performance, where the travelers' path choice decisions are considered.

Previous studies of congestion pricing mainly focused on enhancing the system performance of the entire transportation network [2–8] and few of them considered interactions among different stakeholders [9]. Researchers also studied parking pricing and parking permission management, based on the assumption that all parking facilities are controlled by a central authority [6, 7, 10, 11]. Actually, a transportation network may be comprised of several regions, where the authority in each region manages its own subnetwork independently. Thus conventional pricing models may be inapplicable for the pricing problem in a real network spanning multiple regions, because the regional authorities optimize the toll levels in order to achieve their selfish objectives rather than improve the performance of the whole network [9]. There are a number of significant efforts on the investigation of such competition of road pricing problem (e.g., [9, 12–14]). However, these analyses are carried out under deterministic condition; that is, the demand and supply sides of a transportation network are assumed to be deterministic. As such, traffic demand is either treated as a deterministic value in the fixed demand case or assumed as a deterministic function of the average travel time/cost in the elastic demand case [15]. On the supply side, the link capacity is also treated as a fixed value. The congestion pricing problem under deterministic condition could be perfectly solved by the theory of marginal cost pricing (MPC) [16, 17]. Specifically, if a marginal cost toll is allowed to be charged on each link

of the network, the corresponding traffic flow pattern will be driven toward a social optimum (SO) under deterministic user equilibrium (UE) path choice principle [18] or stochastic user equilibrium (SUE) path choice principle [19]. Regardless of the perfect features of MPC, congestion pricing under deterministic condition cannot capture the uncertain factors of transportation networks.

It is well known that transportation networks are inherently stochastic in reality. In recent studies, more and more researchers found that network uncertainty might exert an important impact on the transportation management and overlooking this factor leads to a suboptimal optimization scheme [20, 21]. Therefore, in this paper, we aim to study the pricing problem of multiple regions on a stochastic network with demand uncertainty. The literature review provided here by no means presents a comprehensive survey to general pricing problems; instead, it focuses on the pricing model and the traffic equilibrium with network uncertainty, particularly demand uncertainty.

In the past decades, network uncertainties have been widely recognized and extensively studied in the literature of transportation field. In a realistic transportation network, there are a variety of uncertainty sources in both demand and supply sides. In demand side, traffic demand of a study period (e.g. morning peak) fluctuates from day to day due to travelers' variant activities. In supply side, the link capacities degrade due to various incidents, such as traffic accidents, road works, earthquakes, and signal failures. Under uncertain conditions, the travel time during a particular period varies from day to day. Confronted with the travel time variations, travelers have to consider the risk of being late to their destination while making their path choice decisions. To resist the disturbance of network uncertainties, travelers take into account not only the travel time/cost but also the travel time reliability for path choices. That is to say, travel time reliability also exerts an important impact on travelers' path choice decisions. The corresponding reliability-based traffic assignment problems have been attracting an increasing attention in the literature [22–28]. From the viewpoint of a decision maker, the enhancement of network reliability is also an important target while making road pricing schemes. Thus, under network uncertainties, how to formulate the congestion pricing model is an interesting and practical topic.

A few studies have been carried out to account for the congestion pricing problem under uncertainty. For example, Li et al. [29] proposed a reliability-based optimal toll design model with respect to stochastic link capacities and OD demand with varied toll levels. Boyles et al. [30] obtained first-best tolls in static transportation networks with day-to-day variation in network capacity. Sumalee and Xu [15] proposed a closed-form formulation to calculate the first-best marginal cost toll for the stochastic network under demand uncertainty. Gardner et al. [31] presented a road pricing framework for representing uncertainty in long-term travel demand and in day-to-day network capacity. All these studies indicated that the congestion pricing scheme under network uncertainties is different from that under a deterministic condition. Therefore, the investigation of congestion pricing problem under network uncertainties could help the decision maker make appropriate toll design.

The task of this study is to model the competitive and cooperative behaviors of pricing problem among multiple regions under demand uncertainty. The network uncertainty may bring two challenges for the road pricing problem among multiple regions. On the one hand, road pricing can alleviate the traffic congestion in terms of minimizing the expected total travel time related to the stochastic flows. On the other hand, travelers' reliability-based path choice behaviors may play an important role in the pricing scheme by influencing the flow pattern and the optimization objective simultaneously. To account for the two potential impacts, we formulated the stochastic pricing models among multiple regions based on the conceptual framework proposed by Zhang et al. [9].

The rest of the paper is organized as follows. In the next section, the road congestion pricing optimization models are formulated to characterize the competitive and cooperative behaviors among multiple regions. Then, a heuristic solution algorithm is proposed in Section 3. Numerical examples and results are discussed in Section 4. Finally, conclusions and further studies are given in Section 5.

2. Model Formulation

This section builds the road congestion pricing models among multiple regions on the stochastic network with day-to-day demand fluctuations. It first investigates the impact of demand uncertainty on the flow patterns in terms of the mean and covariance of the flow distributions. A reliability-based traffic assignment model based on the stochastic flow patterns is proposed to characterize the travelers' path choice behaviors considering their own risk preference. Subsequently, we propose the competitive and cooperative pricing schemes for stochastic road pricing problem among multiple regions. The optimization objective for the pricing model is to maximize the social welfare (equal to the total user benefit minus the mean total travel time).

2.1. Demand Uncertainty and Flow Distribution. It is assumed that the daily traffic demands during the same study period (e.g. morning peak, 8:00 am-9:00 am) between all OD pairs are multivariate random variables. For each OD pair $rs \in \mathbf{R}$, the random traffic demand is expressed as

$$\begin{aligned} Q_{rs} &= q_{rs} + \varepsilon_{rs}, \quad \forall rs \in \mathbf{R}, \\ cv_{rs} &= \frac{\sigma^{q}_{rs}}{q_{rs}}, \quad \forall rs \in \mathbf{R}, \end{aligned} \tag{1}$$

where q_{rs} is the mean (or expected) OD demand between OD pair rs, $E[Q_{rs}] = q_{rs}$; ε_{rs} is the random term with $E[\varepsilon_{rs}] = 0$; $\mathbf{R}$ is the set of OD pairs. For convenience, denote $\mathbf{Q}$ and $\mathbf{q}$ as the $|\mathbf{R}|$-vectors of $(\ldots, Q_{rs}, \ldots)^T$ and $(\ldots, q_{rs}, \ldots)^T$ for all $rs \in \mathbf{R}$, respectively. The covariance between OD demand Q_{rs} and $Q_{rs'}$ is denoted as

$$\sigma^{q}_{rs,rs'} = \operatorname{cov}\left[Q_{rs}, Q_{rs'}\right], \quad \forall rs,\ rs' \in \mathbf{R}. \tag{2}$$

The corresponding covariance matrix of traffic demands of all OD pairs can be expressed as

$$\Sigma^{\mathbf{q}} = \left\{\sigma^{q}_{rs,rs'}\right\}_{|\mathbf{R}|\times|\mathbf{R}|}. \tag{3}$$

The value of $\sigma^{q}_{rs,rs'}$ may be positive, zero, or negative. It is assumed that the OD demands of all OD pairs follow multivariate normal distribution, that is $\mathbf{Q} \sim \mathrm{MVN}(\mathbf{q}, \Sigma^{\mathbf{q}},)$, where $\mathbf{q}$ and $\Sigma^{\mathbf{q}}$ are fixed and known. Let F^{k}_{rs} be the random traffic flow on path $k \in \mathbf{K}_{rs}$ with its mean $f^{k}_{rs} = E[F^{k}_{rs}]$, where $\mathbf{K}_{rs}$ is the path set between OD pair rs and $\mathbf{K} = \bigcup_{rs\in\mathbf{R}} \mathbf{K}_{rs}$. For convenience, $\mathbf{F}$ and $\mathbf{f}$ are denoted as the $|\mathbf{K}|$-vectors of $(\ldots, F^{k}_{rs}, \ldots)^{T}$ and $(\ldots, f^{k}_{rs}, \ldots)^{T}$ for all $k \in \mathbf{K}_{rs}$ and $rs \in \mathbf{R}$, respectively. The path flows and OD demands satisfy the following flow conservation condition:

$$\mathbf{Q} = \Lambda\mathbf{F}, \tag{4}$$

where Λ is the OD-path incidence matrix. Then, the following conservation conditions hold:

$$\mathbf{q} = E[\mathbf{Q}] = E[\Lambda\mathbf{F}] = \Lambda E[\mathbf{F}] = \Lambda\mathbf{f}. \tag{5}$$

Equation (5) can be rewritten as

$$q_{rs} = \sum_{k\in\mathbf{K}_{rs}} f^{k}_{rs}, \quad \forall rs \in \mathbf{R}. \tag{6}$$

It is assumed that the path flow is a product of the corresponding path choice proportion and the OD demand as follows [27]:

$$F^{k}_{rs} = p^{k}_{rs} Q_{rs}, \quad \forall k \in \mathbf{K}_{rs},\ rs \in \mathbf{R}, \tag{7}$$

where p^{k}_{rs} is the path choice proportion of the traffic flow on path $k \in \mathbf{K}_{rs}$ of vehicle, which is assumed to be constant of the probability [24]. Then, it follows from (7) that

$$f^{k}_{rs} = p^{k}_{rs} q_{rs}, \quad \forall k \in \mathbf{K}_{rs},\ rs \in \mathbf{R}. \tag{8}$$

The covariance between F^{k}_{rs} and $F^{k'}_{rs'}$ can be deduced as

$$\begin{aligned}\sigma^{f,k,k'}_{rs,rs'} &= \mathrm{cov}\left[F^{k}_{rs}, F^{k'}_{rs'}\right] \\ &= p^{k}_{rs} p^{k'}_{rs'} \mathrm{cov}\left[Q_{rs}, Q_{rs'}\right] = p^{k}_{rs} p^{k'}_{rs'} \sigma^{q}_{rs,rs'}, \\ &\qquad \forall k \in \mathbf{K}_{rs},\ k' \in \mathbf{K}_{rs'},\ rs, rs' \in \mathbf{R}.\end{aligned} \tag{9}$$

The corresponding covariance matrix of path flows can be expressed as

$$\Sigma^{\mathbf{f}} = \left\{\sigma^{f,k,k'}_{rs,rs'}\right\}_{|\mathbf{K}|\times|\mathbf{K}|}. \tag{10}$$

According to (4), the covariance conservation condition between path flows and OD demands is expressed as

$$\Sigma^{\mathbf{q}} = \Lambda\Sigma^{\mathbf{f}}\Lambda^{T}. \tag{11}$$

Denote $\delta^{k,a}_{rs}$ as the element of the link-path incidence matrix. $\delta^{k,a}_{rs} = 1$, if path k uses link a; otherwise, $\delta^{k,a}_{rs} = 0$. Then, the conservation condition of the estimated link and path flows is expressed as

$$V_a = \sum_{rs\in\mathbf{R}} \sum_{k\in\mathbf{K}_{rs}} \delta^{k,a}_{rs} F^{k}_{rs}, \quad \forall a \in \mathbf{A}, \tag{12}$$

where V_a is the random traffic flow on link a. The mean link flow is denoted as $v_a = E[V_a]$. It follows from (12) that

$$\begin{aligned} v_a &= \sum_{rs\in\mathbf{R}} \sum_{k\in\mathbf{K}_{rs}} \delta^{k,a}_{rs} f^{k}_{rs} \\ &= \sum_{rs\in\mathbf{R}} \sum_{k\in\mathbf{K}_{rs}} \delta^{k,a}_{rs} p^{k}_{rs} q_{rs}, \quad \forall a \in \mathbf{A}. \end{aligned} \tag{13}$$

It can be seen from (13) that the mean link flow is a linear function with respect to the path choice proportions, which can be expressed as follows:

$$v_a = v_a(\mathbf{p}), \quad \forall a \in \mathbf{A}, \tag{14}$$

where $\mathbf{p}$ is denoted as the $|\mathbf{K}|$-vector of path choice proportions $(\ldots, p^{k}_{rs}, \ldots)^{T}$, all $k \in \mathbf{K}_{rs}$ and $rs \in \mathbf{R}$. Also, the conservation condition of the link and path flow covariance can be obtained as

$$\begin{aligned} \sigma^{v}_{a,a'} &= \mathrm{cov}\left[V_a, V_{a'}\right] = \mathrm{cov}\left[\sum_{rs\in\mathbf{R}}\sum_{k\in\mathbf{K}} \delta^{k,a}_{rs} F^{k}_{rs}, \sum_{rs'\in\mathbf{R}}\sum_{k'\in\mathbf{K}} \delta^{k',a'}_{rs'} F^{k'}_{rs'}\right] \\ &= \sum_{rs\in\mathbf{R}}\sum_{k\in\mathbf{K}}\sum_{rs'\in\mathbf{R}}\sum_{k'\in\mathbf{K}} \delta^{k,a}_{rs}\delta^{k',a'}_{rs'} \mathrm{cov}\left[F^{k}_{rs}, F^{k'}_{rs'}\right] \\ &= \sum_{rs\in\mathbf{R}}\sum_{k\in\mathbf{K}}\sum_{rs'\in\mathbf{R}}\sum_{k'\in\mathbf{K}} \delta^{k,a}_{rs}\delta^{k',a'}_{rs'} \sigma^{f,k,k'}_{rs,rs'} \\ &= \sum_{rs\in\mathbf{R}}\sum_{k\in\mathbf{K}}\sum_{rs'\in\mathbf{R}}\sum_{k'\in\mathbf{K}} \delta^{k,a}_{rs}\delta^{k',a'}_{rs'} p^{k}_{rs} p^{k'}_{rs'} \sigma^{q}_{rs,rs'}, \quad \forall a, a' \in \mathbf{A}, \end{aligned} \tag{15}$$

where $\sigma^{v}_{a,a'}$ is the covariance between link flows V_a and $V_{a'}$, $a, a' \in \mathbf{A}$. It can be seen from (15) that the covariance of link flows is a function with respect to the path choice proportions, which is expressed as:

$$\sigma^{v}_{a,a'} = \sigma^{v}_{a,a'}(\mathbf{p}), \quad \forall a, a' \in \mathbf{A}. \tag{16}$$

2.2. Reliability-Based Traffic Assignment Problem. Under demand uncertainty, the link and path travel times stochastically fluctuate from day to day, indicated as $T_a(V_a(\mathbf{p}), u_a)$ and T^{k}_{rs}, respectively. Let $\overline{\mathbf{A}} \subseteq \mathbf{A}$ be a subset of links which are implemented the congestion pricing scheme and $\mathbf{u} = (\ldots, u_a, \ldots)^{T} a \in \overline{\mathbf{A}}$ denote the vector of link tolls, where u_a is the toll on link a $\in \overline{\mathbf{A}}$. Then, the link travel time function for link $a \in \overline{\mathbf{A}}$ can be defined as the following modified Bureau of Public Roads (BPR) function:

$$T_a(V_a(\mathbf{p}), u_a) = t^{0}_{a}\left(1 + \alpha\left(\frac{V_a(\mathbf{p})}{c_a}\right)^{n}\right) + \frac{1}{\beta}u_a, \quad \forall a \in \overline{\mathbf{A}}, \tag{17}$$

where α is a parameter of link performance function and $\beta > 0$ is a constant which represents the value of time (VOT). To facilitate the presentation of the essential idea, it is assumed that the VOTs of all travelers are same. For other links, the original BPR function is used as the link time function:

$$T_a\left(V_a(\mathbf{p}), u_a\right) = t_a^0\left(1 + \alpha\left(\frac{V_a(\mathbf{p})}{c_a}\right)^n\right), \quad \forall a \in \overline{\mathbf{A}} \setminus \mathbf{A}, \tag{18}$$

where $\overline{\mathbf{A}} \setminus \mathbf{A}$ represents the link set, for which toll is not charged on the link. According to the method in Clark and Watling [24], (16) and (17), the mean and covariance of link travel times can be deduced, which are the functions with respect to $\mathbf{p}$ and $\mathbf{u}$. The mean path travel time can be expressed as a function with respect to $\mathbf{p}$ and $\mathbf{u}$ as follows:

$$t_{rs}^k(\mathbf{p}, \mathbf{u}) = \sum_{a \in \mathbf{A}} \delta_{rs}^{k,a} t_a(\mathbf{p}, \mathbf{u}), \quad \forall k \in \mathbf{K}_{rs},\ rs \in \mathbf{R}, \tag{19}$$

where $t_a(\mathbf{p}, \mathbf{u})$ is the travel cost on $a \in \mathbf{A}$; $t_{rs}^k(\mathbf{p}, \mathbf{u})$ is the travel cost on path $k \in \mathbf{K}_{rs}$. Similarly, the variance of path travel times can be expressed as

$$\begin{aligned}
\sigma_{rs}^{t,k}(\mathbf{p}, \mathbf{u}) &= \operatorname{cov}\left[T_{rs}^k, T_{rs}^k\right] \\
&= \operatorname{cov}\left[\sum_{a \in \mathbf{A}} \delta_{rs}^{k,a} T_a(\mathbf{p}, \mathbf{u}), \sum_{a' \in \mathbf{A}} \delta_{rs}^{k,a'} T_{a'}(\mathbf{p}, \mathbf{u})\right] \\
&= \sum_{a \in \mathbf{A}} \sum_{a' \in \mathbf{A}} \delta_{rs}^{k,a} \delta_{rs}^{k,a'} \operatorname{cov}\left[T_a(\mathbf{p}, \mathbf{u}), T_{a'}(\mathbf{p}, \mathbf{u})\right] \\
&= \sum_{a \in \mathbf{A}} \sum_{a' \in \mathbf{A}} \delta_{rs}^{k,a} \delta_{rs}^{k,a'} \sigma_{a,a'}^t(\mathbf{p}, \mathbf{u}), \quad \forall k \in \mathbf{K},\ rs \in \mathbf{R},
\end{aligned} \tag{20}$$

where T_{rs}^k is the path travel time on path $k \in \mathbf{K}_{rs}$, $E[T_{rs}^k] = t_{rs}^k$; T_a and $T_{a'}$ are link travel times on links $a \in \mathbf{A}$ and $a' \in \mathbf{A}$, respectively; $\sigma_{a,a'}^t$ is the covariance of link travel times on $a \in \mathbf{A}$ and $a' \in \mathbf{A}$.

Since travelers' path choice decisions will be influenced by the uncertain OD demand variations, the decision of road pricing is also dependent on such stochasticity. To consider this effect, a reliability-based stochastic user equilibrium (RSUE) model [27] is adopted to account for the travelers' reliability-based path choice behaviors in the road pricing problem. In this RSUE model, the effective travel time, $\hat{t}_{rs}^k$, is used as the path choice criterion, which is defined as the summation of the mean travel time, t_{rs}^k, and the safety margin s_{rs}^k [32]:

$$\hat{t}_{rs}^k(\mathbf{p}, \mathbf{u}) = t_{rs}^k(\mathbf{p}, \mathbf{u}) + s_{rs}^k(\mathbf{p}, \mathbf{u}), \quad \forall k \in \mathbf{K},\ rs \in \mathbf{R}. \tag{21}$$

The value of s_{rs}^k can be obtained by solving the following chance-constrained minimization problem:

$$\begin{aligned}
&\min_{s_{rs}^k} \quad \hat{t}_{rs}^k = t_{rs}^k + s_{rs}^k \\
&\text{s.t.} \quad \Pr\left[T_{rs}^k \le \hat{t}_{rs}^k\right] \ge \rho \quad \forall k \in \mathbf{K},\ rs \in \mathbf{R},
\end{aligned} \tag{22}$$

where ρ is the confidence level of travel time reliability. A high value of ρ means that the travelers would prefer setting a large safety margin of path travel time in order to guarantee a high on-time arrival probability. On the other hand, a low value of ρ means the travelers would prefer tolerating a high risk of on-time arrival. The effective path travel time can be calculated that

$$\hat{t}_{rs}^k(\mathbf{p}, \mathbf{u}) = t_{rs}^k(\mathbf{p}, \mathbf{u}) + \Phi^{-1}(\rho)\, \sigma_{rs}^{t,k}(\mathbf{p}, \mathbf{u}), \quad \forall k \in \mathbf{K},\ rs \in \mathbf{R}, \tag{23}$$

where $\Phi^{-1}(\cdot)$ is the inverse of cumulative function of standard normal distribution.

The reliability-based stochastic user equilibrium could be reached, in which for each OD pair no traveler can decrease his/her perceived disutility by unilaterally changing their paths. In the Logit-based RSUE model, the path choice proportion p_{rs}^k, which is defined in (7), can thus be specified by the following formula:

$$\begin{aligned}
p_{rs}^k &= \frac{\exp\left(-\theta \hat{t}_{rs}^k(\mathbf{p}, \mathbf{u})\right)}{\sum_{j \in \mathbf{K}_{rs}} \exp\left(-\theta \hat{t}_{rs}^j(\mathbf{p}, \mathbf{u})\right)} \\
&= w_{rs}^k(\mathbf{p}, \mathbf{u}), \quad \forall k \in \mathbf{K},\ rs \in \mathbf{R},
\end{aligned} \tag{24}$$

where θ is the dispersion parameter on travelers' perception errors of effective travel time. For the sake of convenience, we denote the path choice function p_{rs}^k as $w_{rs}^k(\mathbf{p}, \mathbf{u})$ and the set notation can be denoted as:

$$W(\mathbf{p}, \mathbf{u}) = \left(\ldots, w_{rs}^k(\mathbf{p}, \mathbf{u}), \ldots\right)^T. \tag{25}$$

Meanwhile, it is well known that level of service on the network would exert an impact on the OD demand [33]. Therefore, the mean OD demand is viewed as a function with respect to the expected disutility of travel for each OD pair:

$$q_{rs}(\pi) = \overline{q_{rs}} \exp(-\eta\pi), \tag{26}$$

where $\overline{q_{rs}}$ indicates the potential demand for OD pair rs, and η denotes the elastic coefficient. The expected disutility can be attained by

$$\pi_{rs} = -\frac{1}{\theta} \operatorname{In} \sum_i \exp\left(-\theta \cdot \hat{t}_{rs}^k(\mathbf{p}, \mathbf{u})\right). \tag{27}$$

Then, the vector form of (24) considering elastic demand yields the following fixed point problem:

$$\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} W(\mathbf{p}, \mathbf{u}) \\ \mathbf{q}\left(\hat{\mathbf{t}}_{rs}^k(\mathbf{p}, \mathbf{u})\right) \end{pmatrix}. \tag{28}$$

2.3. Optimization of Congestion Charges in a Single Region. Recently, an increasing number of transportation researchers have recognized that demand uncertainty plays an important role in the decision making of transportation management. Under demand uncertainty, the expected value of total system travel time, as an important measure, is always adopted to evaluate the system performance. Meanwhile, the variation

of total network travel time could bring troubles for transportation planning as well as the mean total travel time. Theoretically, the uncertainty of demand will directly lead to variation of the travel time, which influences decision making for administrators and travelers. For decision maker, it is difficult to evaluate the level of service of the road as travel time stochastically fluctuates from day to day. On the other hand, for the travelers, the variation of the travel time may result in late arrival to the destination, which influences their path and departure time choices. In this regard, the variation of total travel time should be considered in the optimization objective for the pricing problem. To facilitate the model formulation, we only take into account the minimization of mean total travel time as the objective of the pricing problem. This is not to deny the importance of the higher moment of the total travel time.

The congestion models in previous studies are assumed to alleviate the traffic congestion over the whole network, which is managed by a central authority. Therefore, the central authority will be concerned with the mean total travel time cost on the network. The corresponding road pricing is objectived comprises of the total user benefit and the mean total travel time cost:

$$\max_{\mathbf{u}} \mathrm{SW} = \sum_{rs\in\mathbf{R}} \int_0^{q_{rs}(\mathbf{u})} q_{rs}^{-1}(\omega)\, d\omega - E\left[\mathrm{TTT}(\mathbf{p},\mathbf{u})\right], \quad (29)$$

where at RSUE the mean of total travel time ($E[\mathrm{TTT}(\mathbf{p},\mathbf{u})]$) can be calculated as

$$E\left[\mathrm{TTT}(\mathbf{p},\mathbf{u})\right] = E\left[\sum_{a\in\mathbf{A}} \left(T_a(\mathbf{p}) - \frac{1}{\alpha}u_a\right)\cdot V_a(\mathbf{p}(\mathbf{u}))\right]. \quad (30)$$

Equation (30) can be calculated using the method proposed by Clark and Watling [24]. It is noticeable that the following mathematical inequality generally holds:

$$\begin{aligned} &E\left[\sum_{a\in\mathbf{A}} \left(T_a(\mathbf{p}) - \frac{1}{\alpha}u_a\right)\cdot V_a(\mathbf{p}(\mathbf{u}))\right] \\ &\quad \neq \sum_{a\in\mathbf{A}} E\left[T_a(\mathbf{p}) - \frac{1}{\alpha}u_a\right]\cdot E\left[V_a(\mathbf{p}(\mathbf{u}))\right]. \end{aligned} \quad (31)$$

In practice, the regional road systems may be freely managed by the separate transportation authorities on the optimizations of toll level and toll location. Here, we introduce the stochastic pricing problem where only one local region performs pricing scheme. As stated by Zhang et al. [9], the optimization objective function for single region pricing scheme is different from that of the centric pricing scheme with a central authority. The former aims to maximize the social welfare just for its own residents, who live in its local administrative region whereas the later concerns the social benefit of all users on the whole network. In this study, we extend the objective function to the stochastic traffic network and consider the stochastic flows resulted from the demand uncertainty. The regional authority only takes into account the trips with origins that locate at its own regime. The pricing problem considering reliability-based user equilibrium can be formulated as follows (Model A):

$$\begin{aligned} \max_{\mathbf{u}} \mathrm{SW} = &\sum_{rs\in\mathbf{R}^i} \int_0^{q_{rs}(\mathbf{u})} q_{rs}^{-1}(\omega)\, d\omega \\ &+ E\left[\sum_{a\in\overline{\mathbf{A}}^i} \mathbf{u}_a \cdot V_a\left(\mathbf{p}\left(\mathbf{u}^i\right)\right)\right. \\ &\left. - \sum_{rs\in\mathbf{R}^i}\sum_{k\in\mathbf{K}} F_{rs}^k(\mathbf{p}(\mathbf{u}))\cdot T_{rs}^k(\mathbf{p}(\mathbf{u}))\right] \end{aligned} \quad (32)$$

subject to

$$\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix} W(\mathbf{p},\mathbf{u}) \\ \mathbf{q}\left(\hat{\mathbf{t}}_{rs}^k(\mathbf{p},\mathbf{u})\right)\end{pmatrix}, \quad (33)$$

$$\underline{u}_a \le u_a \le \overline{u}_a, \quad \forall a \in \overline{\mathbf{A}}^i, \quad (34)$$

where $\mathbf{R}^i$ denotes the OD pairs with the origins in region; $\overline{\mathbf{A}}^i$ denotes the links in region i. In the optimization problem, $\mathbf{u}$ is the decision variable, and the path choice proportion ($\mathbf{p}$) is the constant probability, which can be determined by the fixed point problem in (33). The constraint (34) sets the upper and lower bounds for the toll charges.

It should be stressed that the social benefit expressed in (32) concerns only the local region i that implements congestion pricing, while the stochastic equilibrium flow pattern is characterized by the choice decisions of all users in all regions in the network.

2.4. Competitive Behavior of Pricing Problem among Multiple Regions. In this section, we study the competitive behavior of the scenario that several local authorities implement congestion pricing independently. Let $I = \{1, 2, \ldots, |I|\}$ indicate the set of regions in which congestion pricing schemes are implemented, and $i \in I$. Let $\mathbf{R}^i$ denote the set of O-D pairs for all residents living in this region, which is a subset of $\mathbf{R}$, and $\mathbf{R} = \cup_{i\in I}\mathbf{R}^i$. The set of candidate toll links $\overline{\mathbf{A}}^i$ is a subset of $\overline{\mathbf{A}}$. Let $\mathbf{u}^i$ be the set of tolls u_a on the links $a \in \overline{\mathbf{A}}^i$, and $\mathbf{u} = \cup_{i\in I}\mathbf{u}^i$. Based on the current decisions of other regions, each region designs its own pricing scheme with the objective of maximizing its own social benefit. Once an authority makes change of it's toll levels, other authorities will make their best responses of adjusting their toll levels. The competition with theses mutual responses can be characterized as a Nash game.

The optimal toll levels for a specific regional authority i can be obtained by solving a similar optimization model as proposed in Section 2.3. For region i, its authority aims to maximize its own social welfare by setting toll charges in its regime, in which the network users follow the reliability-based stochastic user equilibrium principle. Taking the viewpoint of region i, the Nash equilibrium model is formulated as follows (Model B):

$$\max_{\mathbf{u}^i} \mathrm{SW}\left(\mathbf{u}^i \mid \mathbf{u}^i \setminus \mathbf{u}\right)$$

$$= \sum_{rs\in\mathbf{R}^i} \int_0^{q_{rs}(\mathbf{u}^i)} q_{rs}^{-1}(\omega)\, d\omega + E\left[\sum_{a\in\bar{\mathbf{A}}^i} u_a \cdot V_a\left(\mathbf{p}\left(\mathbf{u}^i\right)\right) - \sum_{rs\in\mathbf{R}^i}\sum_{k\in\mathbf{K}} F_{rs}^k\left(\mathbf{p}\left(\mathbf{u}^i\right)\right)\cdot T_{rs}^k\left(\mathbf{p}\left(\mathbf{u}^i\right)\right)\right] \tag{35}$$

subject to constraints (33) and (34).

In this pricing scheme, eventually, all the regions are self-best responding to each other, named a Nash equilibrium, that no player can change his pricing strategy unilaterally to obtain a better result. The outcome of the Nash game can be obtained by iteratively solving the above pricing problem for all the players.

2.5. Cooperative Congestion Pricing among Multiple Regions. In Section 2.4, we propose a Nash equilibrium model to capture the competitive behavior among the authorities in different regions. Each authority sets its own pricing objective to improve the social welfare independently. But such competition may actually be detrimental to the travelers from other regions of the increasing travel burden. In this regard, the competitive pricing scheme may do harms to traffic efficiency of the whole network. Therefore, the cooperative manner is recommended. We propose a pricing model in which the local authorities cooperate to maximize the total social welfare of travelers in the regions that implement the pricing schemes. All elements used in this subsection are the same as those in Section 2.4, except that the regional authorities behave in a cooperative manner. The congestion pricing model is formulated as follows (Model C):

$$\max_{\mathbf{u}^i} \mathrm{SW} = \sum_{i\in I}\left\{\sum_{rs\in\mathbf{R}^i}\int_0^{q_{rs}(\mathbf{u}^i)} q_{rs}^{-1}(\omega)\, d\omega\right\} + \sum_{i\in I}\left\{E\left[\sum_{a\in\bar{\mathbf{A}}^i} u_a \cdot V_a\left(\mathbf{p}\left(\mathbf{u}^i\right)\right) - \sum_{rs\in\mathbf{R}^i}\sum_{k\in\mathbf{K}} F_{rs}^k\left(\mathbf{p}\left(\mathbf{u}^i\right)\right)\cdot T_{rs}^k\left(\mathbf{p}\left(\mathbf{u}^i\right)\right)\right]\right\}, \tag{36}$$

where the constraints are also the traffic flow equilibrium constraint and the boundary constraint for the design variable, which is formulated by (33) and (34).

It should be stressed that, as pointed out by Zhang et al. [9], whether such cooperation can be achieved depends on the benefit gain of each player under the cooperative manner. If all regional authorities can benefit from the cooperation, the agreement of the cooperation would be achieved easily. In contrast, when some regions suffer a loss in the cooperative scheme, they prefer competing with other regions. However, if the regions that gain more benefit are willing to compensate the regions that suffer loss, the alliance may still hold.

In this section, we propose the competitive and cooperative congestion pricing models under demand uncertainty. To the best of our knowledge, Zhang et al. [9] have made comprehensive analyses of noncooperative behaviors of the road pricing problem among multiple regions. The proposed models differ from the most related studies, such as Zhang et al. [9], of the traffic conditions. Our models are based on the stochastic traffic flows and the corresponding RSUE principle, while the model in Zhang et al. [9] is based on the deterministic traffic flows following the user equilibrium principle. This study makes twofold contribution to the literature. First, travelers' reliability-based path choice behaviors can be reflected by the RSUE constraint (33). It makes an obvious difference from the user equilibrium principle in previous studies by considering travelers' risk-taking behavior. Generally, the risk-taking preference is important for a stochastic transport system because it exerts an important impact on both individual travel activity and system-level decision making. Furthermore, the competitive and cooperative behaviors of the pricing problem on a stochastic network have been discussed by incorporating the mean total travel time into the optimization objectives.

3. Solution Algorithm

The proposed pricing optimization models are inherently bilevel programs, in which the upper level is to optimize the pricing objective and the lower level is the RSUE traffic assignment. As the proposed bilevel program is by nature nonlinear and non-convex, the global optimal solution is difficult to be obtained by using the conventional optimization algorithms. Existing effective algorithms for solving the nonconvex bilevel programming problems are meta-heuristic, including the genetic algorithm, simulated annealing method, and sequential quadratic programming, to name but a few. These methods search the local optimal solution in an evolutionary manner based on the traffic flow patterns in the lower-level program. The lower-level traffic assignment problem formulated as the fixed-point problem can be solved by the Method of Successive Averages [25, 26].

Since the transportation assignment problem is integrated as a nonlinear constraint, we develop a heuristic solution algorithm, which combines the penalty function method, to solve the proposed models. The penalty function method is used to cope with the equilibrium constraint and the boundary constraint. The constrained optimization problem can be further transformed into an unconstrained one as follows:

$$\min_{\mathbf{u}} \mu_1^{(j)}\left\|\min\{\underline{\mathbf{u}},\mathbf{u}\}\right\|^2 + \mu_2^{(j)}\left\|\min\{\mathbf{u},\overline{\mathbf{u}}\}\right\|^2 + \mu_3^{(j)}\left\|\mathbf{p} - W(\mathbf{p},\mathbf{u})\right\|^2 - \mathrm{SW}(\mathbf{p},\mathbf{u}), \tag{37}$$

where μ_1, μ_2, and μ_3 are three positive penalty coefficients, $\|\cdot\|$ is the Euclidean norm of a vector. For convenience, denote the penalty term as follows:

$$y = \mu_1^{(j)}\left\|\min\{\underline{\mathbf{u}}, \mathbf{u}\}\right\|^2 + \mu_2^{(j)}\|\min\{\mathbf{u}, \overline{\mathbf{u}}\}\|^2 + \mu_3^{(j)}\|\mathbf{p} - W(\mathbf{p}, \mathbf{u})\|^2. \tag{38}$$

Obviously, it is difficult to obtain the gradient of the objective function (37) due to the complexity of $g_1(\mathbf{p}, \mathbf{u})$, $g_2(\mathbf{p}, \mathbf{u})$, and $W(\mathbf{p}, \mathbf{u})$. Therefore, some derivative-free optimization methods could be employed for solving the minimization problem (37), such as the simplex search method [34] and generalized pattern search methods [35, 36]. In this paper, the simplex search method [34] is used to solve the unconstrained optimization problem (37), which is available in the Matlab optimization toolbox by the subroutine "fminsearch". The flowchart of this method is shown in Figure 1.

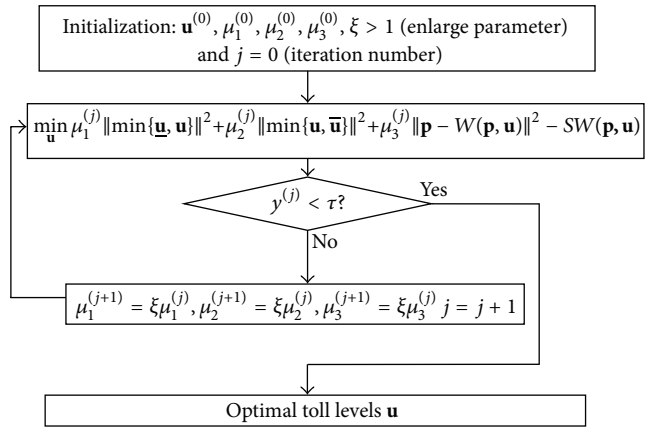

Figure 1: The flowchart of the proposed solution algorithm.

4. Numerical Examples

4.1. Preliminary. The numerical examples are used to illustrate: (a) the difference between two pricing strategies (b) effects of travelers' reliability-based path choice behavior on pricing schemes. In the numerical study, a small network shown in Figure 2 is employed to demonstrate the property of the proposed model. There are 6 nodes, 10 links, and 4 O-D pairs on the network. The network is partitioned into two regions, A and B, by dash line X-X. The candidate links to be charged are depicted with block dash-dot lines. The potential traffic demand for each OD pair and the coefficient of standard deviation (S.D.) of the actual demand are given in Table 1. Table 2 provides the link performance parameters, t_a^0 and c_a.

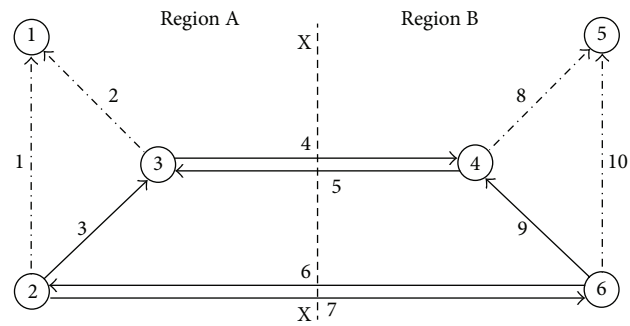

Figure 2: A small network for numerical examples.

Table 1: Potential OD demands in the network.

OD pair	$2 \to 1$	$2 \to 5$	$6 \to 5$	$6 \to 1$
$\overline{q}_{rs}$	800	500	500	600
Coefficient of S.D.	0.5	0.5	1.0	1.0

The elastic coefficient of demand function, η, is set as 0.05 and the dispersion parameter, θ, is set as 0.1. For the sake of simplicity, it is assumed that the traffic demands between two OD pairs are independent with each other (i.e. $\sigma^q_{rs,r's'} = 0$). $\alpha = 0.15$, $n = 1$, $\beta = 1$, and $\eta = 0.05$ in (17), (18), and (26). Meanwhile, the path flows, and link and path travel times are all assumed mutually independent and follow normal distributions. The convergence stopping tolerance τ is set as 10^{-3}. The solution code is run on Windows 7 system with the following attributes: Intel Core i5-2520 2.5 GHz×2 and 4 GB RAM.

4.2. Pricing Outcomes of Cooperative and Competitive Schemes. The pricing outcomes of cooperative and competitive schemes can be obtained by solving Models B and C, respectively. Let SWA, SWB, SWT denote the social welfare for Region A, B, and the whole network respectively. The network is comprised of two regions that both are considered in the cooperative pricing scheme. So the pricing objective in Model C is equivalent to that in (29), which is proposed for the conventional stochastic pricing model with a central authority. As mentioned by Shao et al. [25], the mean values of stochastic flows will be equal to those in the deterministic traffic scenario if the following conditions are satisfied: a linear link travel time function is adopted and the travel time reliability is set as risk neutral ($\rho = 50\%$). It should be stressed that, in spite of this, the optimization objective for stochastic pricing scheme is still different from deterministic pricing scheme due to the fact that the deviations of both flow and travel time are taken into account in the objective function. Two scenarios of different travelers' risk-taking behaviors, namely $\rho = 50\%$ and $\rho = 90\%$, are constructed to test the pricing effects of two pricing schemes. The pricing results for cooperative and competitive schemes are provided in Table 3. Before comparing the results of two pricing schemes, we first introduce the equilibrium of competitive pricing.

In the competitive scheme, both regions compete with each other to maximize their own social benefit. The authority in Region A sets pricing scheme on links 1 and 2, and the authority in Region B implements a pricing scheme on links 8 and 10. Each region makes a response to the actions in other regions by updating its tolls. If the toll levels in one region change, another region would respond by adjusting its own pricing scheme. This iterative process continues until reaching a Nash equilibrium. As the competition initiator,

Table 2: Parameters used in link performance functions.

Number	1	2	3	4	5	6	7	8	9	10
t_a^0	8.0	4.0	4.0	10.0	10.0	14.0	14.0	4.0	4.0	8.0
c_a	300	300	300	400	400	400	400	300	300	200

Table 3: Pricing results for cooperative and competitive schemes.

Confidence level (ρ)	Pricing scheme	Optimal solutions				Social welfare of region A	Social welfare of region B	Total Social welfare
		u_1	u_2	u_8	u_{10}	(SWA)	(SWB)	(SWT)
50%	cooperation	16.36	7.81	7.55	17.78	12997.12	9374.28	22371.40
	competition	13.83	12.32	13.37	14.36	12754.40	9259.83	22014.23
90%	cooperation	15.79	7.03	6.81	17.20	12965.86	9416.70	22382.56
	competition	13.40	12.13	13.19	13.99	12708.41	9268.52	21976.93

Region A makes the toll scheme in the first instance. At first competition round, tolls in Region B are zero.

In this example, we illustrate the iterative competition process under the scenario that the confidence level of the travel time reliability is set as 90%. The competition reaches an equilibrium state after seven iterations. The reaction process and the optimal solution of the Nash game are given in Table 4. At equilibrium, social welfares for Region A and B are 12708.41 and 9268.52, respectively, and the total social welfare for the whole network is 21976.93.

As shown in Table 3, the pricing results in terms of optimal solution and social welfares are different between cooperative and competitive schemes since different objectives are considered in the two pricing optimizations. When the confidence level is fixed at 50%, under cooperation, the social welfare for Region A is 12997.12, the social welfare for Region B is 9374.28, and the total social welfare is 22371.40. However, under competition, the social welfare for Region A is 12754.40, the social welfare for Region B is 9259.83, and the total social welfare is 22014.23. Compared with cooperation, the competition makes Region, A and B suffer benefit losses about 242.7 and 114.5, respectively, and leads a degradation of the network system performance and so do the pricing results of the risk-aversion case with high travelers' travel time reliability ($\rho = 90\%$). Therefore, it clearly reveals that the pricing effect in terms of system performance of the cooperative scheme is better than that of competitive scheme. Moreover, by comparing the results between two scenarios represented different travel time reliabilities, it can be seen that the pricing outcomes are different since the stochastic patterns depend on the travel time reliability. In the next subsection, we will discuss the impact of travel time reliability on the pricing effects of two pricing schemes.

4.3. Effects on Travelers' Reliability-Based Path Choice Behaviors. The proposed model was carried out under different values of ρ that represent different risk-taking path choice behaviors. In this impact analysis, the travel time reliability increases from 10% to 90% that each step is 10%. Three travelers' risk-taking behaviors are considered, namely risk-prone behavior ($\rho < 50\%$), risk-neutral behavior ($\rho = 50\%$), and risk-averse behavior ($\rho > 50\%$). A higher confidence level for travel time reliability means travelers will pay more attention to guaranteeing the on-time arrival by setting larger safety margins. For the case of $\rho = 50\%$, the proposed pricing model can be regarded to be carried out under conventional path choice behavior assumptions; that is, travelers take the mean path travel time as the path choice criteria. Figure 3 depicts the variation of the social welfare for each region with different travel time reliability. The variation of whole network performance with travel time reliability is shown in Figure 4. Benefit gain/loss of the cooperative scheme changing with travel time reliability is given in Figure 5.

From Figure 3, it can be found that the social welfare of Region B increases gradually with the travel time reliability no matter what the pricing scheme is the Meanwhile, the increase of social welfare in the cooperative scheme is more apparent and dramatical than that in the competitive scheme. The contrast, the social welfare of Region A decreases monotonically as the increase of the travel time reliability under both pricing schemes. Moreover, the reduction of social welfare under competition is more quick and obvious than that under cooperation. To evaluate the impact of travel time reliability on the network performance properly, we can pay attention to Figure 4, namely the variation of total welfare. For cooperation, the total social welfare increases from 22352.15 to 22382.56 as the travelers put more emphasis on the travel time reliability. However, for competition, the total social welfare decreases from 22039.82 to 21976.91 as the increase of the travel time reliability. On the whole, the total social welfare under the cooperative scheme is more than that of the competitive scheme no matter what travelers' risk-taking behavior is. Although these impact analysis tests are network-specific, they cannot deny the fact that cooperation is more beneficial to improving the network system performance for all users.

The difference of social welfare between two pricing schemes is plotted in Figure 5 to illustrate the impact of travel time reliability. The difference of the social welfare is defined as the value of social welfare in cooperation minus social welfare in competition. A positive value means a benefit gain in cooperation, and vice visa. As shown in Figure 5, both

Table 4: The process of the competitive congestion pricing (ρ = 90%).

Iteration	Reactor	Optimal solutions				Social welfare of region A (SWA)	Social welfare of region A (SWB)	Total Social welfare (SWT)
		u_1	u_2	u_8	u_{10}			
1	A	13.31	12.17	0.00	0.00	13900.31	6912.76	20813.06
2	B	13.31	12.17	13.19	13.98	12707.82	9265.94	21973.76
3	A	13.41	12.12	13.19	13.98	12708.11	9269.08	21977.19
4	B	13.41	12.12	13.18	13.97	12708.50	9269.28	21977.78
5	A	13.40	12.13	13.18	13.98	12708.80	9268.41	21977.21
6	B	13.40	12.13	13.19	13.99	12708.41	9268.52	21976.93
7	A	13.40	12.13	13.19	13.99	12708.41	9268.52	21976.93

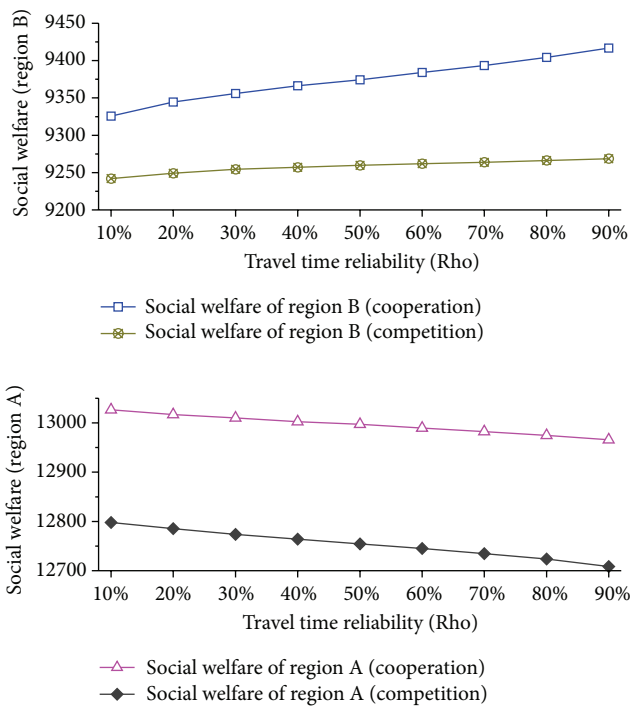

Figure 3: Social welfare for each region with different travel time reliability.

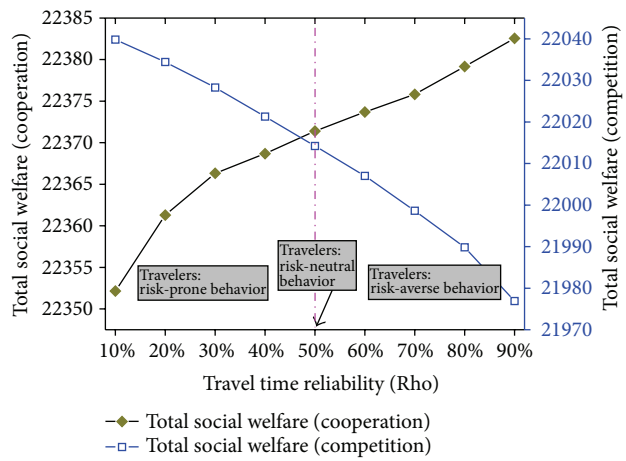

Figure 4: Total social welfare with different travel time reliability.

regions and the whole network benefit from the cooperative manner under all different travel time reliability scenarios. Obviously, the higher travel time reliability, the higher benefit gain. In this regard, both regions would prefer a cooperative manner so as to obtain more benefit for themselves. Here, it should be pointed out that the result in this example does not mean that the regions will always benefit from the cooperative schemes, which have been demonstrated by the analyses in Zhang et al. [9]. From the above discussion, it is clear that the resultant competitive and cooperative pricing schemes are different for different risk-taking path choice behaviors. This finding indicates that different risk-taking path choice behaviors lead to different optimal toll levels and clearly proves that the reliability-based path choice behaviors should be considered in congestion pricing problems, particular for transportation network with demand uncertainty.

5. Conclusions and Further Studies

This paper proposed two new optimization models for congestion pricing problem on stochastic transportation networks with demand uncertainty. We analyze the stochastic road pricing schemes on a network with multiple regions. In practice, there may be several independent regions in a transportation network; regional authorities either prefer to maximize the system performance for the whole network or to maximize their own benefits separately. The cooperative and competitive behaviors among multiple regional decision makers have been investigated. Two pricing strategies, cooperation and competition, can be formulated as bi-level programs, in which stochastic flow equilibrium is considered. Different from most conventional modelling approaches of congestion pricing, the traffic demand of the study period was assumed to fluctuate from day to day. As a result, the travel time also varies accordingly. Under such circumstance, the conventional second-best congestion pricing model was extended to capture the effects of uncertainty for both the decision makers and the travelers. On one hand, the authorities aim to maximize the social welfare through setting the toll charges on the candidate links. The social welfare here is comprised of the total user benefit and the mean total travel time cost, which are dependent on the stochastic flow

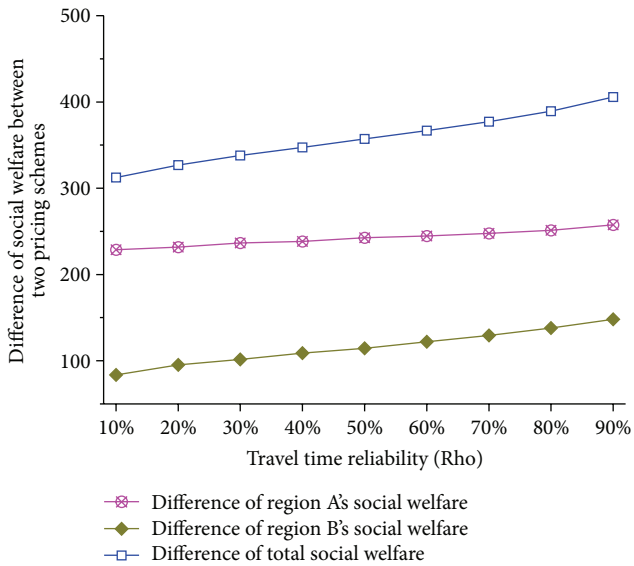

Figure 5: Benefit losses/gains in cooperation with different travel time reliability.

pattern. On the other hand, the travelers were assumed to minimize their effective travel time for path choices, which explicitly accounts for the risk-taking path choice behaviors under uncertainty condition. Such path choice behavior is formulated as an equality constraint for the congestion pricing optimization problem by a fixed point formulation.

A heuristic solution algorithm is proposed in this paper. The proposed algorithm employs the penalty function method for constrained optimization problem, namely equilibrium flow constraint. Numerical examples demonstrated that, on stochastic network, the cooperative pricing scheme is more beneficial to improve the system performance than the competitive scheme. Meanwhile, both two pricing schemes were quite sensitive to the travelers' risk-taking path choice behaviors; that is, the travel time reliability plays an important role in determining the pricing effects.

Further studies could be carried out to extend the proposed model in the following aspects. First, the proposed model is formulated under demand uncertainty. How to simultaneously consider demand and supply uncertainties in the congestion pricing problem reveals important investigations. Furthermore, in the proposed model, the toll charges schemes are determined on the fixed locations of links. How to optimize toll charge locations as well as toll levels could be an interesting extension. Finally, the proposed solution algorithm is heuristic by nature. It is necessary to propose a more efficient solution algorithm in further investigations.

Acknowledgments

This research was supported by research Grants from the National Natural Science Foundation of China (Projects nos. 71125004, 70901073, and 71271205) and the National Basic Research Program (973 Program) of China (Project no. 2012CB725404).

References

[1] H. Yang and H. J. Huang, *Mathematical and Economic Theory of Road Pricing*, Elsevier, 2005.

[2] H. Yang and X. Zhang, "Multiclass network toll design problem with social and spatial equity constraints," *Journal of Transportation Engineering*, vol. 128, no. 5, pp. 420–428, 2002.

[3] H. Yang and X. Zhang, "Optimal toll design in second-best link-based congestion pricing," *Transportation Research Record*, no. 1857, pp. 85–92, 2003.

[4] H. Yang, X. Zhang, and Q. Meng, "Modeling private highways in networks with entry-exit based toll charges," *Transportation Research Part B*, vol. 38, no. 3, pp. 191–213, 2004.

[5] X. Zhang and H. Yang, "The optimal cordon-based network congestion pricing problem," *Transportation Research Part B*, vol. 38, no. 6, pp. 517–537, 2004.

[6] X. Zhang, H. J. Huang, and H. M. Zhang, "Integrated daily commuting patterns and optimal road tolls and parking fees in a linear city," *Transportation Research Part B*, vol. 42, no. 1, pp. 38–56, 2008.

[7] X. Zhang, H. Yang, and H. J. Huang, "Multiclass multicriteria mixed equilibrium on networks and uniform link tolls for system optimum," *European Journal of Operational Research*, vol. 189, no. 1, pp. 146–158, 2008.

[8] X. Zhang and B. van Wee, "Enhancing transportation network capacity by congestion pricing with simultaneous toll location and toll level optimization," *Engineering Optimization*, vol. 44, no. 4, pp. 477–488, 2012.

[9] X. N. Zhang, H. M. Zhang, H. J. Huang, L. J. Sun, and T. Q. Tang, "Competitive, cooperative and Stackelberg congestion pricing for multiple regions in transportation networks," *Transportmetrica*, vol. 7, no. 4, pp. 297–320, 2011.

[10] X. Zhang and B. van Wee, "Efficiency comparison of various parking charge schemes considering daily travel cost in a linear city," *European Journal of Transport and Infrastructure Research*, vol. 11, no. 2, pp. 234–255, 2011.

[11] X. N. Zhang, H. Yang, and H. J. Huang, "Improving travel efficiency by parking permits distribution and trading," *Transportation Research Part B*, vol. 45, no. 7, pp. 1018–1034, 2011.

[12] A. De Palma and R. Lindsey, "Private toll roads: competition under various ownership regimes," *Annals of Regional Science*, vol. 34, no. 1, pp. 13–35, 2000.

[13] B. Ubbels and E. T. Verhoef, "Governmental competition in road charging and capacity choice," *Regional Science and Urban Economics*, vol. 38, no. 2, pp. 174–190, 2008.

[14] A. Yuen, L. J. Basso, and A. Zhang, "Effects of gateway congestion pricing on optimal road pricing and hinterland," *Journal of Transport Economics and Policy*, vol. 42, no. 3, pp. 495–526, 2008.

[15] A. Sumalee and W. Xu, "First-best marginal cost toll for a traffic network with stochastic demand," *Transportation Research Part B*, vol. 45, no. 1, pp. 41–59, 2011.

[16] A. C. Pigou, *The Economics of Welfare*, MacMillan, London, UK, 1920.

[17] F. H. Knight, "Some fallacies in the interpretation of social costs," *Quarterly Journal of Economics*, vol. 38, pp. 582–606, 1924.

[18] H. Yang and H. J. Huang, "Principle of marginal-cost pricing: how does it work in a general road network?" *Transportation Research Part A*, vol. 32, no. 1, pp. 45–54, 1998.

[19] H. Yang, "System optimum, stochastic user equilibrium, and optimal link tolls," *Transportation Science*, vol. 33, no. 4, pp. 354–360, 1999.

[20] S. T. Waller and A. K. Ziliaskopoulos, "Stochastic dynamic network design problem," *Transportation Research Record*, no. 1771, pp. 106–113, 2001.

[21] G. R. Patil and S. V. Ukkusuri, "System-optimal stochastic transportation network design," *Transportation Research Record*, no. 2029, pp. 80–86, 2007.

[22] A. Chen, Z. Ji, and W. Recker, "Travel time reliability with risk-sensitive travelers," *Transportation Research Record*, no. 1783, pp. 27–33, 2002.

[23] S. Nakayama and J. Takayama, "A traffic network equilibrium model for uncertain demands," in *Proceedings of the 82nd Transportation Research Board Annual Meeting*, 2003.

[24] S. Clark and D. Watling, "Modelling network travel time reliability under stochastic demand," *Transportation Research Part B*, vol. 39, no. 2, pp. 119–140, 2005.

[25] H. Shao, W. H. K. Lam, Q. Meng, and M. L. Tam, "Demand-driven traffic assignment problem based on travel time reliability," *Transportation Research Record*, no. 1985, pp. 220–230, 2006.

[26] H. Shao, W. H. K. Lam, and M. L. Tam, "A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand," *Networks and Spatial Economics*, vol. 6, no. 3-4, pp. 173–204, 2006.

[27] W. H. K. Lam, H. Shao, and A. Sumalee, "Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply," *Transportation Research Part B*, vol. 42, no. 10, pp. 890–910, 2008.

[28] B. Y. Chen, W. H. K. Lam, A. Sumalee, and H. Shao, "An efficient solution algorithm for solving multi-class reliability-based traffic assignment problem," *Mathematical and Computer Modelling*, vol. 54, no. 5-6, pp. 1428–1439, 2011.

[29] H. Li, M. C. J. Bliemer, and P. H. L. Bovy, "Network reliability-based optimal toll design," *Journal of Advanced Transportation*, vol. 42, no. 3, pp. 311–332, 2008.

[30] S. D. Boyles, K. M. Kockelman, and S. T. Waller, "Congestion pricing under operational, supply-side uncertainty," *Transportation Research Part C*, vol. 18, no. 4, pp. 519–535, 2010.

[31] L. M. Gardner, S. D. Boyles, and S. T. Waller, "Quantifying the benefit of responsive pricing and travel information in the stochastic congestion pricing problem," *Transportation Research Part A*, vol. 45, no. 3, pp. 202–218, 2011.

[32] R. W. Hall, "Travel outcome and performance: the effect of uncertainty on accessibility," *Transportation Research Part B*, vol. 17, no. 4, pp. 275–290, 1983.

[33] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice Hall, Englewood Cliff, NJ, USA, 1985.

[34] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1999.

[35] V. Torczon, "On the convergence of pattern search algorithms," *SIAM Journal on Optimization*, vol. 7, no. 1, pp. 1–25, 1997.

[36] C. Audet and J. E. Dennis, "Analysis of generalized pattern searches," *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 889–903, 2003.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 976509, 9 pages
http://dx.doi.org/10.1155/2013/976509

Research Article

Global Convergence of a New Nonmonotone Filter Method for Equality Constrained Optimization

Ke Su, Wei Liu, and Xiaoli Lu

College of Mathematics and Computer Science, Hebei University, Baoding 071002, China

Correspondence should be addressed to Ke Su; pigeonsk@163.com

Received 6 December 2012; Accepted 5 March 2013

Academic Editor: Hadi Nasseri

A new nonmonotone filter trust region method is introduced for solving optimization problems with equality constraints. This method directly uses the dominated area of the filter as an acceptability criterion for trial points and allows the dominated area decreasing nonmonotonically. Compared with the filter-type method, our method has more flexible criteria and can avoid Maratos effect in a certain degree. Under reasonable assumptions, we prove that the given algorithm is globally convergent to a first order stationary point for all possible choices of the starting point. Numerical tests are presented to show the effectiveness of the proposed algorithm.

1. Introduction

We analyze an algorithm for solving optimization problems where a smooth objective function is to be minimized subject to smooth nonlinear equality constraints. More formally, we consider the problem,

$$\begin{aligned} &\min \quad f(x), \\ &\text{s.t.} \quad c_i(x) = 0, \quad i \in I = \{1, 2, \ldots, m\}, \end{aligned} \tag{P}$$

where $x \in R^n$, the functions $f : R^n \to R$ and $c_i\ (i \in I) : R^n \to R$ are all twice continuously differentiable. For convenience, let $g(x) = \nabla f(x)$, $c(x) = (c_1(x), c_2(x), \ldots, c_m(x))^T$ and $A(x) = (\nabla c_1(x), \nabla c_2(x), \ldots, \nabla c_m(x))$, and f_k refers to $f(x_k)$, c_k to $c(x_k)$, g_k to $g(x_k)$ and A_k to $A(x_k)$, and so forth.

There are many trust region methods for equality constrained nonlinear programming (P), for example, Byrd et al. [1], Dennis Jr. et al. [2] and Powell and Yuan [3], but in these works, a penalty or augmented Lagrange function is always used to test the acceptability of the iterates. However, there are several difficulties associated with the use of penalty function, and in particular the choice of the penalty parameter. Hence, in 2002, Fletcher and Leyffer [4] proposed a class of filter method, which does not require any penalty parameter and has promising numerical results. Consequently, filter technique has been employed to many approaches, for instance, SLP methods [5], SQP methods [6–8], interior point approaches [9], bundle techniques [10], and so on.

Filter technique, in fact, exhibits a certain degree of nonmonotonicity. The nonmonotone technique was proposed by Grippo et al. in 1986 [11] and combined with many other methods. M. Ulbrich and S. Ulbrich [12] proposed a class of penalty-function-free nonmonotone trust region methods for nonlinear equality constrained optimization without filter technique. Su and Pu [13] introduced a nonmonotone trust region method which used the nonmonotone technique in the traditional filter criteria. Su and Yu [14] presented a nonmonotone method without penalty function or filter. Gould and Toint [15] directly used the dominated area of the filter as an acceptability criteria for trial points and obtained the global convergence properties. We refer the reader [16–18] for some works about this issue.

Motivated by the ideas and methods above, we propose a modified nonmonotone filter trust region method for solving problem (P). Similar to the Byrd-Omojokun class of algorithms, each step is decomposed into the sum of two distinct components, a quasi-normal step and a tangential step. The main contribution of our paper is to employ the nonmonotone idea to the dominated area of the filter so that the new and more flexible criteria is given, which is different from that of Gould and Toint [15] and Su and Pu [13].

Under usual assumptions, we prove that the given algorithm is globally convergent to first order stationary points.

This paper is organized as follows. In Section 2, we introduce the fraction of Cauchy decrease and the composite SQP step. The new nonmonotone filter technique is given in Section 3. In Section 4, we propose the new nonmonotone filter method and present the global convergence properties in Section 5. Some numerical results are reported in the last section.

2. The Fraction of Cauchy Decrease and the Composite SQP Step

Consider the following unconstraint minimization optimization problem:

$$\min_{x\in R^n} f(x), \tag{1}$$

where $f : R^n \to R$ is a continuously differentiable function. A trust region algorithm for solving the above problem is an iterate procedure that computes a trial step as an approximate solution to the following subproblems:

$$\begin{aligned} \min \quad & q(d) = g^T d + \frac{1}{2} d^T H d, \\ \text{s.t.} \quad & \|d\| \le \Delta, \end{aligned} \tag{2}$$

where H is the Hessian matrix $\nabla^2 f(x)$ or an approximate to it and $\Delta > 0$ is a given trust region radius.

To assure the global convergence, the step is required only to satisfy a fraction of Cauchy decrease condition. This means that d must predict via the quadratic model function $q(d)$ at least as much as a fraction of the decreased given by the Cauchy step on $q(d)$; that is, there exists a constant $\sigma > 0$ fixed across all iterations, such that

$$q(0) - q(d) \ge \sigma \left(q(0) - q\left(d^{cp}\right)\right), \tag{3}$$

where d^{cp} is the steepest descent step for $q(d)$ inside the trust region.

Lemma 1. *If the trial step d satisfies a fraction of Cauchy decrease condition, then*

$$q(0) - q(d) \ge \frac{\sigma}{2} \|\nabla f(x)\| \min\left\{\Delta, \frac{\|\nabla f(x)\|}{\|H\|}\right\}. \tag{4}$$

Proof (see Powell [19] for the proof). Now, we turn to explain the composite SQP step. Given an approximate estimate of the solution x_k at kth iteration, following Dennis Jr. et al. [2] and M. Ulbrich and S. Ulbrich [12], we obtain the trial step $d_k = d_k^n + d_k^t$ by computing a quasi-normal step d_k^n and a tangential step d_k^t. The purpose of the quasi-normal step d_k^n is to improve feasibility. To improve optimality, we seek d_k^t in the tangential space of the linearized constraints in such a way that it provides sufficient decrease for a quadratic model of the objective function $f(x)$. Let $q_k(d) = g_k^T d + (1/2) d^T H_k d$, where H_k is a symmetric approximation of $\nabla^2 f(x)$.

d_k^n is the solution to the subproblem

$$\begin{aligned} \min \quad & \frac{1}{2}\left\|c_k + A_k^T d^n\right\|^2 + \frac{\xi}{2}\left\|d^n\right\|^2, \\ \text{s.t.} \quad & \left\|d^n\right\| \le \Delta_k, \end{aligned} \tag{5}$$

where Δ_k is a trust region radius and $A_k = \nabla c(x_k) \in R^{n\times m}$, $\xi > 0$. In order to improve the value of the objective function, we solve the following subproblem to get d_k^t :

$$\begin{aligned} \min \quad & q_k\left(d_k^n + d^t\right), \\ \text{s.t.} \quad & A_k^T d^t = 0 \\ & \left\|d^t\right\| \le \Delta_k. \end{aligned} \tag{6}$$

Then we get the current trial step $d_k = d_k^n + d_k^t$. Let $d_k^t = W_k \overline{d}_k^t$, where $\overline{d}_k^t \in R^{n-m}$ and $W_k \in R^{n\times(n-m)}$ denote a matrix whose columns form a basis of the null space of A_k^T. We refer to [2] for a more detailed discussion of this issue.

In usual way that impose a trust region in step-decomposition methods, the quasi-normal step d_k^n and the tangential step d_k^t are required to satisfy

$$\left\|d^n\right\| \le \kappa\Delta_k, \qquad \left\|d_k^n + d^t\right\| \le \Delta_k, \tag{7}$$

where $0 < \kappa < 1$. Here, to simplify the proof, we only impose a trust region on $\| d^n \| \le \Delta_k$ and $\| d^t \| \le \Delta_k$, which is natural.

Note that $W_k^T \nabla q_k(d^t)$ is the reduced gradient of q_k in terms of the representation $d^t = W_k s$ of the tangential step:

$$\nabla_s\left(q_k\left(W_k s\right)\right) = W_k^T \nabla q_k\left(W_k s\right) = W_k^T \nabla q_k\left(d^t\right). \tag{8}$$

Define

$$\hat{g}(x) = W(x)^T g(x). \tag{9}$$

Then the first order necessary optimality conditions (Karush-Kuhn-Tucker or KKT conditions) at a local solution $\overline{x} \in R^n$ of problem (P) can be written as

$$c(\overline{x}) = 0, \qquad \hat{g}(\overline{x}) = 0. \tag{10}$$

□

3. A New Nonmonotone Filter Technique

In filter method, originally proposed by Fletcher and Leyffer [4], the acceptability of iterates is determined by comparing the value of constraint violation and the objective function with previous iterates collected in a filter. Define the violation function $h(x)$ by $h(x) = \| c(x)\|_2^2$, it is easy to see that $h(x) = 0$ if and only if x is a feasible point, so a trial point should reduce either the value of constraint violation or that of the objective function f.

In the process of the algorithm, we need to decide whether the trial point x_k^+ is any better than x_k as an approximate solution to the problem (P). If we decide that this is the case, we say that the iteration k is successful and choose x_k^+ as

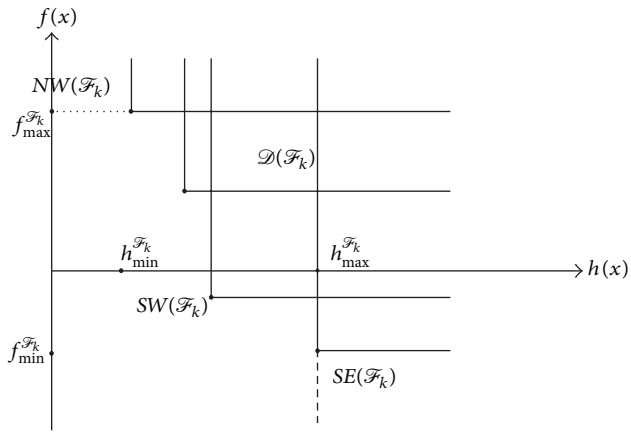

Figure 1

the next iterate. Let us denote by $\mathcal{S}$ the set of all successful iterations, that is,

$$\mathcal{S} = \{k \mid x_{k+1} = x_k^+\}. \tag{11}$$

In traditional filter method, a point x is called acceptable to the filter if and only if

$$h(x) \le \beta h_j \text{ or } f(x) \le f_j - \gamma h_j, \quad \forall (h_j, f_j) \in \mathcal{F}, \tag{12}$$

where $0 < \gamma < \beta < 1$, $\mathcal{F}$ denotes the filter set. Define

$$\mathcal{D}(\mathcal{F}) = \left\{(h, f) \,\middle|\, h > h_j \text{ and } f > f_j,\ \exists j \in \mathcal{F}\right\}. \tag{13}$$

A trial point x_k^+ is accepted if and only if $(h_k^+, f_k^+) \notin \mathcal{D}(\mathcal{F}_k)$.

Now, similar to the idea of Gould and Toint [15], we give a new modified nonmonotone filter technique. For any (h, f)-pair, define an area that represents its contribution to the area of $\mathcal{D}(\mathcal{F})$, we hope this contribution is positive; that is, the area of $\mathcal{D}(\mathcal{F})$ is increasing. For convenience, we partition the right half-plane $[0, +\infty] \times [-\infty, +\infty]$ into four different regions (see Figure 1). Define $\mathcal{D}(\mathcal{F}_k)^c$ to be the complement of $\mathcal{D}(\mathcal{F}_k)$. Let

$$\begin{aligned} h_{\min}^{\mathcal{F}_k} &\stackrel{\text{def}}{=} \min_{j \in \mathcal{F}_k} h_j, & h_{\max}^{\mathcal{F}_k} &\stackrel{\text{def}}{=} \max_{j \in \mathcal{F}_k} h_j, \\ f_{\min}^{\mathcal{F}_k} &\stackrel{\text{def}}{=} \min_{j \in \mathcal{F}_k} f_j, & f_{\max}^{\mathcal{F}_k} &\stackrel{\text{def}}{=} \max_{j \in \mathcal{F}_k} f_j. \end{aligned} \tag{14}$$

These four parts are

(1) the dominated part of the filter: $NE(\mathcal{F}_k) \stackrel{\text{def}}{=} \mathcal{D}(\mathcal{F}_k)$;

(2) the undominated part of lower left corner of the half plane:

$$SW(\mathcal{F}_k) \stackrel{\text{def}}{=} \mathcal{D}(\mathcal{F}_k)^c \cap \left[0, h_{\max}^{\mathcal{F}_k}\right] \times \left[-\infty, f_{\max}^{\mathcal{F}_k}\right], \tag{15}$$

(3) the undominated upper left corner: $NW(\mathcal{F}_k) \stackrel{\text{def}}{=} [0, h_{\min}^{\mathcal{F}_k}) \times (f_{\max}^{\mathcal{F}_k}, +\infty]$;

(4) the undominated lower right corner: $SE(\mathcal{F}_k) \stackrel{\text{def}}{=} (h_{\max}^{\mathcal{F}_k}, +\infty] \times [-\infty, f_{\min}^{\mathcal{F}_k})$.

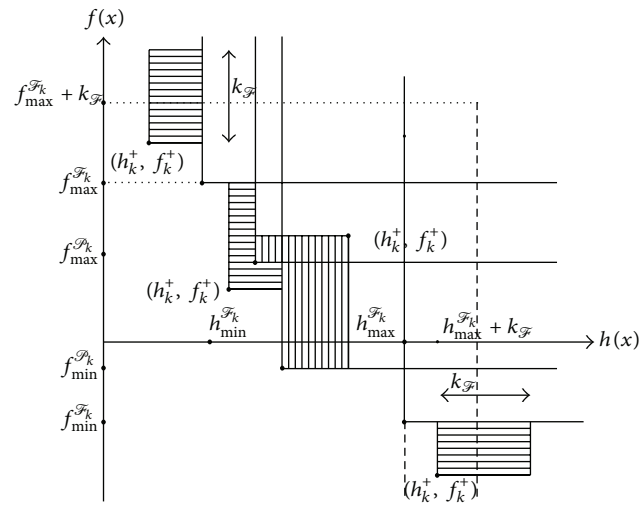

Figure 2

Consider the trial point x_k^+, if the filter is empty, then define its contribution to the area of the filter by

$$\alpha(x_k^+, \mathcal{F}_k) \stackrel{\text{def}}{=} \kappa_F{}^2, \tag{16}$$

where $\kappa_F > 0$ is a constant. If the filter is not empty, then define the contribution of x_k^+ to the area of the filter by four different formulae.

If $(h_k^+, f_k^+) \in SW(\mathcal{F}_k)$, assume

$$\begin{aligned} \alpha(x_k^+, \mathcal{F}_k) \stackrel{\text{def}}{=} \text{area}\Big(&\mathcal{D}(\mathcal{F}_k)^c \cap \left[h_k^+, h_{\max}^{\mathcal{F}_k} + \kappa_F\right] \\ &\times \left[f_k^+, f_{\max}^{\mathcal{F}_k} + \kappa_F\right]\Big). \end{aligned} \tag{17}$$

If $(h_k^+, f_k^+) \in NW(\mathcal{F}_k)$, assume

$$\alpha(x_k^+, \mathcal{F}_k) \stackrel{\text{def}}{=} \kappa_F \left(h_{\min}^{\mathcal{F}_k} - h_k^+\right). \tag{18}$$

If $(h_k^+, f_k^+) \in SE(\mathcal{F}_k)$, assume

$$\alpha(x_k^+, \mathcal{F}_k) \stackrel{\text{def}}{=} \kappa_F \left(f_{\min}^{\mathcal{F}_k} - f_k^+\right). \tag{19}$$

If $(h_k^+, f_k^+) \in NE(\mathcal{F}_k) = \mathcal{D}(\mathcal{F}_k)$, assume

$$\begin{aligned} \alpha(x_k^+, \mathcal{F}_k) \stackrel{\text{def}}{=} -\text{area}\Big(&\mathcal{D}(\mathcal{F}_k) \cap \left[h_k^+ - h_{\min}^{\mathcal{P}_k}\right] \\ &\times \left[f_k^+ - f_{\min}^{\mathcal{P}_k}\right]\Big), \end{aligned} \tag{20}$$

where $\mathcal{P}_k = \{(h, f) \in \mathcal{F}_k \mid h_j < h_k^+, f_j < f_k^+\}$, and $h_{\min}^{\mathcal{P}_k} \stackrel{\text{def}}{=} \min_{j \in \mathcal{P}_k} h_j$, $f_{\min}^{\mathcal{P}_k} \stackrel{\text{def}}{=} \min_{j \in \mathcal{P}_k} f_j$.

Figure 2 illustrate the corresponding areas in the filter. Horizontally dashed surfaces indicate a positive contribution and vertically dashed ones a negative contribution. Note that $\alpha(x, \mathcal{F})$ is a continuous function of $(h(x), f(x))$.

Next, we should consider the updating of the filter. If $(h_k, f_k) \notin \mathscr{D}(\mathscr{F}_k)$, then

$$\mathscr{F}_{k+1} \longleftarrow \mathscr{F}_k \cup (h_k, f_k). \tag{21}$$

If $(h_k, f_k) \in \mathscr{D}(\mathscr{F}_k)$, then

$$\mathscr{F}_{k+1} \longleftarrow (\mathscr{F}_k \setminus \mathscr{P}_k) \cup \left(h_{\min}^{\mathscr{P}_k}, f_k\right) \cup \left(h_k, f_{\min}^{\mathscr{P}_k}\right). \tag{22}$$

We now return to the question of deciding whether a trial point x_k^+ is acceptable for the filter or not. We will insist that this is a necessary condition for the iteration k to be successful in the sense that $x_{k+1} = x_k^+$. If we consider an iterate x_k, there must exist a predecessor iteration such that $x_{p(k)}^+ = x_{p(k)+1} = x_k$. Under the monotonic situation, a trial point x_k^+ would be accepted whenever it results in an sufficient increase in the dominated area of the filter, that means x_k^+ would be accepted whenever

$$\alpha_k \geq \gamma_{\mathscr{F}} \left(h_k^+\right)^2, \tag{23}$$

where $\alpha_k \overset{\text{def}}{=} \alpha(x_k^+, \mathscr{F}_k)$, $\gamma_{\mathscr{F}} \in (0,1)$ is a constant. Under the nonmonotonic situation, we relax condition (23) to be

$$\max\left\{\alpha_k, \sum_{\substack{j=r(k)+1 \\ j\in\mathscr{U}}}^{k} \lambda_{p(j)}\alpha_{p(j)}\right\} \geq \gamma_{\mathscr{F}} \left(h_k^+\right)^2, \tag{24}$$

where $\alpha_{p(j)} \overset{\text{def}}{=} \alpha(x_j, \mathscr{F}_{p(j)})$, $\mathscr{U} = \{k|$ filter is updated for $(h_k, f_k)\}$, $r(k) \leq k$ is some reference iteration for $\mathscr{U}$, $r(k) \in \mathscr{U}$, $\mathscr{U} \subseteq \mathscr{S}$, $\lambda_{p(j)} \in [0,1]$, $\sum_{j=r(k)+1,\, j\in\mathscr{U}}^{k} \lambda_{p(j)} = 1$.

Compared to condition (2.21) [15], our condition (24) is more flexible if α_k is negative.

According to condition (24), it is possible to accept x_k^+ even though it may be dominated. Then x_k^+ will be accepted if either (23) or (24) holds.

4. The New Nonmonotone Filter Trust Region Algorithm

Our algorithm is based on the usual trust region technique; define the predict reduction for the function $q_k(x)$ to be

$$\operatorname{pred}(d_k) = q_k(0) - q_k(d_k) \tag{25}$$

and the actual reduction

$$\operatorname{ared}(d_k) = f(x_k) - f(x_k^+). \tag{26}$$

Moreover, let $r_k = \operatorname{ared}(d_k)/\operatorname{pred}(d_k)$, if there exists a nonzero constant η_1 such that $r_k \geq \eta_1$ and condition (23) and (24) hold, the trial point x_k^+ will be called acceptable. Then the next trial point x_{k+1} is obtained, and for its feasibility, we consider the condition

$$h_{k+1} \leq \eta_3 \min\left\{\mu, \alpha_1 \Delta_k^{2+\alpha_2}\right\} \tag{27}$$

is true or not, where η_3 and μ are all positive constants, if it is not true, then turn to the feasibility restoration phase and define

$$r_k^j = \frac{\left\|c(x_k^j)\right\|^2 - \left\|c(x_k^j + d_k^j)\right\|^2}{\left\|c(x_k^j)\right\|^2 - \left\|c(x_k^j) + A(x_k^j)^T d_k^j\right\|^2}. \tag{28}$$

A formal description of the algorithm is given as follows.

Algorithm A Step 0. Choose an initial point $x_0 \in R^n$, a symmetric matrix $H_0 \in R^{n\times n}$, let $\Delta_0 > 0$, $\epsilon_t > 0$, $\eta_2, \eta_3 \in (0,1)$, $\alpha_1, \alpha_2 \in [0,1]$, $\eta_1 > 0$, $0 < \gamma_0 < \gamma_1 < 1 \leq \gamma_2 < \gamma_3 \leq 2$, $0 < \xi < 1$, $\gamma_{\mathscr{F}} \in (0,1)$, $\mu > 0$, compute $f(x_0)$, $h(x_0)$, let $k = 0$, $\mathscr{F}_0 = \emptyset$.

Step 1. Compute $c_k, A_k, W_k, f_k, h_k, g_k, \widehat{g}_k = W_k^T g_k$.

Step 2. If $\| \widehat{g}_k \| + h_k \leq \epsilon_t$, stop.

Step 3. Solve the subproblems (5) and (6) to get the quasi-normal step d_k^n and the tangential step d_k^t. Let $d_k = d_k^n + d_k^t$, $x_k^+ = x_k + d_k$.

Step 4. If $r_k < \eta_1$, let $x_{k+1} = x_k$, then go to Step 8.

Step 5. If $x_k + d_k$ is not acceptable to the filter, go to Step 8. Otherwise $x_{k+1} = x_k^+$ and update the filter according to (21) and (22), the trust region radius and H_k, then get the corresponding h_{k+1}, f_{k+1}.

Step 6. If $h_{k+1} \leq \eta_3 \min\{\mu, \alpha_1 \Delta_k^{2+\alpha_2}\}$, $k = k+1$ and go to Step 1, otherwise go to Step 7.

Step 7. By restoration Algorithm B to get d_k^r, then the trial point $x_k^r = x_k + d_k^r$.

Step 8. Update the trust region radius by Algorithm C, let $k = k+1$ and go to Step 3.

We aim to reduce the value of $h(x)$ in the restoration Algorithm B, that is to get $c(x_k^r) = 0$ by Newton-type method.

Algorithm B Step 0. Let $x_k^0 = x_k$, $\Delta_k^0 = \Delta_k$, $j = 0$, $\eta_2, \eta_3 \in (0,1)$, $\alpha_1, \alpha_2 \in [0,1]$, $\mu > 0$.

Step 1. If $h(x_k^j) \leq \eta_3 \min\{h_k^{\mathscr{F}_k}, \alpha_1 \Delta_k^{2+\alpha_2}\}$ and x_k^j is acceptable by the filter, then $x_k^r = x_k^j$, stop.

Step 2. Compute

$$\begin{aligned} \min \quad & \left\|c(x_k^j) + A(x_k^j)^T d_k^j\right\|^2, \\ \text{s.t.} \quad & \left\|d_k^j\right\| \leq \Delta_k^j \end{aligned} \tag{29}$$

to get d_k^j, then compute r_k^j.

Step 3. If $r_k^j \leq \eta_2$, $x_k^{j+1} = x_k^j$, $\Delta_k^{j+1} = \Delta_k^j/2$, $j = j+1$ and go to Step 2.

Step 4. If $r_k^j > \eta_2$, $x_k^{j+1} = x_k^j + d_k^j$, $\Delta_k^{j+1} = 2\Delta_k^j$, $j = j+1$, compute A_k^{j+1} and go to Step 1, where $h_k^{\mathcal{F}_k} = \min\{\min_{j\in\mathcal{F}_k}\{h_j | h_j > 0\}, \mu\}$.

Algorithm C (updating the trust region radius). Given $\eta_1 > 0$, $0 < \gamma_0 < \gamma_1 < 1 \le \gamma_2 < \gamma_3 \le 2$, we have the following.

(1) If $r_k < \eta_1$ or x_k^+ is not acceptable to the filter, $\Delta_{k+1} \in [\gamma_0\Delta_k, \gamma_1\Delta_k]$.

(2) If x_k^+ is acceptable to the filter but does not satisfiy condition (27), $\Delta_{k+1} \in (\Delta_k, \gamma_2\Delta_k)$.

(3) If x_k^+ is acceptable to the filter and satisfies (27), $\Delta_{k+1} \in [\gamma_2\Delta_k, \gamma_3\Delta_k]$.

From the description above and the idea of the algorithm, we can see that our algorithm is more flexible. Every successful iterate must be any better than the predecessor one in some degree according to the traditional filter method. But our algorithm relaxes this demand by using the nonmonotone technique and also avoids Maratos effect in a certain degree. Moreover, Algorithm C allows a relatively wide choice of the trust region.

5. The Convergence Properties

In this section, to present a proof of global convergence of algorithm, we always assume that the following conditions hold.

Assumption

(A1) The objective function f and the constraint functions c_i $(i \in I = \{1, 2, \ldots, m\})$ are twice continuously differentiable.

(A2) For all k, x_k, and $x_k + d_k$ all remain in a closed, bounded convex subset $\Omega \subset R^n$.

(A3) The matrix $A(x) = \nabla c(x)$ is nonsingular matrix for all $x \in \Omega$.

(A4) The matrices $(A(x)^T A(x))^{-1}$, $W(x)$, $(W^T W)^{-1}$ are uniformly bounded in Ω, where $W(x)$ denotes a matrix whose columns form a basis of the null space of $A(x)^T$.

(A5) The matrix H_k is uniformly bounded.

By the assumptions, we can suppose there exist constants v_0, v_1, v_2, v_3 such that $\| f(x) \| \le v_0$, $\| g(x) \| \le v_0$, $\| c(x) \| \le v_0$, $\| A(x) \| \le v_0$, $\| A(x)^T A(x)^{-1} \| \le v_1$, $\| W(x) \| \le v_2$, $\| H_k \| \le v_3$, $\| W_k^T H_k \| \le v_3$, $\| W_k^T H_k W_k \| \le v_3$.

By (A1) and (A2), it holds

$$f_{\min} \le f_k, \quad 0 \le h_k \le h_{\max} \quad \forall k, \tag{30}$$

where $f_{\min}, h_{\max} > 0$, hence in the (h, f)-plane, the (h, f)-pair lies in the area $[0, h_{\max}] \times [f_{\min}, +\infty]$.

From (A1), (A2), and (A3), it exists a constant $\bar{v}$ such that

$$\left| f(x_k + d_k) - q_k(d_k) \right| \le \bar{v}\Delta_k^2. \tag{31}$$

Lemma 2. *At the current iterate x_k, let the trial point component d_k^n actually be normal to the tangential space. Under the problem assumptions, there exists a constant $k_1 > 0$ independent of the iterates such that*

$$\| d_k^n \| \le \alpha_1 \| c_k \|. \tag{32}$$

Proof. It is similar to the proof of Lemma 2 in [13]. □

Lemma 3. *Under Assumptions, there exist positive constants k_2, k_3, k_4 independent of the iterates such that*

$$\begin{aligned} &\| c_k \|^2 - \| c_k + A_k^T d_k^n \|^2 \ge k_2 \| c_k \| \min\{k_3 \| c_k \|, \Delta_k\}, \\ &q_k(d_k^n) - q_k(d_k) \\ &\quad \ge \frac{\sigma}{2} \| W_k^T \nabla q_k(d_k^n) \| \min\left\{k_4 \| W_k^T \nabla q_k(d_k^n) \|, \Delta_k\right\}. \end{aligned} \tag{33}$$

Proof. The proof is an application of Lemma 1 to the two subproblems (5) and (6). □

Lemma 4. *Suppose that Assumptions hold, then restoration Algorithm B is well defined.*

Proof. The conclusion is obvious, if $h_k^j \to 0$. Otherwise it exists $\epsilon > 0$ such that for all j, it holds $h_k^j > \epsilon$. Consider the set

$$K = \left\{ j \,\middle|\, r_k^j = \frac{\| c_k^j \|^2 - \| c(x_k^j + d_k^j) \|^2}{\| c_k^j \|^2 - \| c_k^j + (A_k^j)^T d_k^j \|^2} > \eta_2 > 0 \right\}, \tag{34}$$

where $c_k^j = c(x_k^j)$, $A_k^j = A(x_k^j)$. By Lemma 3 and the definition of h_k, we have

$$\begin{aligned} +\infty > \sum_{j=1}^{\infty} \left(h_k^{j-1} - h_k^j\right) &\ge \sum_{j\in K} \left(\| c_k^j \|^2 - \left\| c_k^j + \left(A_k^j\right)^T d_k^j \right\|^2 \right) \\ &\ge \eta_2 k_2 \sum_{j\in K} \| c_k \| \min\left\{k_3 \| c_k \|, \Delta_k^j\right\}. \end{aligned} \tag{35}$$

By $h_k^j > \epsilon$, it holds $\Delta_k^j \to 0$ for $j \in K$. From Algorithm B, we can obtain that $\Delta_k^j \to 0$ for all j.

On the other side,

$$\| c_k^j \|^2 - \| c(x_k^j + d_k^j) \|^2 = \| c_k^j \|^2 - \left\| c_k^j + \left(A_k^j\right)^T d_k^j \right\|^2 + o\left(\Delta_k^j\right) \tag{36}$$

for $\Delta_k^j \to 0$. By the algorithm, the radius Δ_k^j should be satisfied $\Delta_k^{j+1} > \Delta_k^j$, that is contradicted to $\Delta_k^j \to 0$. The proof is complete. □

Now, we analyze the impact of the criteria (23) and (24). Once a trial point is accepted as a new iterate, it must be provided some improvement, and we formalize this by saying that iterate $x_k = x_{p(k)+1}$ improves on iterate $x_{i(k)}$. That is

the trial point x_k is accepted at iterate $p(k)$; it happens under two situations, one is by the criteria (23), that is,

$$i(k) = p(k) \quad \text{if } p(k) \notin \mathscr{A} \tag{37}$$

the other is by the criteria (24), that is,

$$i(k) = p(k) \quad \text{if } p(k) \notin \mathscr{A}. \tag{38}$$

Now consider any iterate x_k, it improved on $x_{i(k)}$, which was itself accepted because it improved on $x_{i(i(k))}$, and so on, until back to x_0. Hence we may construct a chain of successful iterations indexed by $\mathscr{C}_k = \{l_1, l_2, \ldots, l_q\}$ for each k, such that

$$x_{l_1} = x_0, \quad x_{l_q} = x_k, \quad x_{l_j} = x_{i(l_{j+1})}, \quad j = 1, 2, \ldots, q-1, \tag{39}$$

where l_1 is the smallest index in the chain of successful iterations.

Lemma 5. *Suppose that Assumptions hold and Algorithm A does not terminate finitely, apply Algorithm A to the problem (P), then for all k and $\mathscr{C}_k = \{l_1, l_2, \ldots, l_q\}$, it holds*

$$\text{area}\left(\mathscr{D}\left(\mathscr{F}_k\right)\right) \geq \gamma_{\mathscr{F}} \sum_{j=1}^{q} h_{l_j}^2. \tag{40}$$

Proof. For $\forall l_j \in \mathscr{C}_k$, if $p(l_j) \in \mathscr{A}$, by (24), $i(l_j) = r(p(l_j)) = l_{j-1}$, then

$$\max\left\{\alpha_{p(l_j)}, \sum_{\substack{i=l_{j-1}+1 \\ i\in\mathscr{U}}}^{l_j} \lambda_{p(i)}\alpha_{p(i)}\right\} \geq \gamma_{\mathscr{F}} h_{l_j}^2. \tag{41}$$

If $\max\{\alpha_{p(l_j)}, \sum_{i=l_{j-1}+1, i\in\mathscr{U}}^{l_j} \lambda_{p(i)}\alpha_{p(i)}\} = \alpha_{p(l_j)}$,

$$\alpha_{p(l_j)} \geq \gamma_{\mathscr{F}} h_{l_j}^2. \tag{42}$$

If $\max\{\alpha_{p(l_j)}, \sum_{i=l_{j-1}+1, i\in\mathscr{U}}^{l_j} \lambda_{p(i)}\alpha_{p(i)}\} = \sum_{i=l_{j-1}+1, i\in\mathscr{U}}^{l_j} \lambda_{p(i)} \alpha_{p(i)}$,

$$\sum_{\substack{i=l_{j-1}+1 \\ i\in\mathscr{U}}}^{l_j} \alpha_{p(i)} \geq \gamma_{\mathscr{F}} h_{l_j}^2. \tag{43}$$

If $p(l_j) \notin \mathscr{A}$, $l_{j-1} = p(l_j)$. By $\mathscr{U} \subseteq \mathscr{S}$, it holds

$$\left\{l_{j-1}+1, \ldots, l_j\right\} \cap \mathscr{U} \subseteq \left\{l_{j-1}+1, \ldots, l_j\right\} \cap \mathscr{S} = \left\{l_j\right\}. \tag{44}$$

Then from (23), $\alpha_{p(i)} \geq \gamma_{\mathscr{F}} h_{l_j}^2$. It implies (43). Moreover

$$\text{area}\left(\mathscr{D}\left(\mathscr{F}_k\right)\right) \geq \sum_{\substack{i=0 \\ i\in\mathscr{U}}}^{k} \alpha_{p(i)} = \sum_{j=0}^{q} \left(\sum_{\substack{i=l_{j-1}+1 \\ i\in\mathscr{U}}}^{l_j} \alpha_{p(i)}\right). \tag{45}$$

Together with (42) and (43), the result follows. □

Lemma 6. *Suppose that Assumptions hold. If Algorithm A does not terminate finitely and the filter contains infinite iterates, then* $\lim_{k\to\infty} h_k = 0$.

Proof. Suppose by contradiction that there exists a constant $\epsilon > 0$ and infinite sequence $\{k_i\} \subseteq \mathscr{S}$ such that $h_{k_i} \geq \epsilon$ for all i. Because there are infinite iterations in the filter, we have $|\mathscr{S}| = \infty$, then $h_{l_q} \geq \epsilon$ for $\forall q$.

$$\text{area}\left(\mathscr{D}\left(\mathscr{F}_k\right)\right) \geq \gamma_{\mathscr{F}} \cdot q \cdot \epsilon^2. \tag{46}$$

Then by (31), area$(\mathscr{D}(\mathscr{F}_k))$ is upper bounded for each k. That means it exists $\kappa_{\mathscr{F}}^{\max} \geq 0$ such that area$(\mathscr{D}(\mathscr{F}_k)) \leq \kappa_{\mathscr{F}}^{\max}$, so $i \leq (\kappa_{\mathscr{F}}^{\max}/\gamma_{\mathscr{F}}\epsilon^2)$. Hence i must be finite, it contradicts to the infinity of $\{k_i\}$. The proof is complete. □

Lemma 7. *Suppose that Assumptions hold and Algorithm A terminate finitely, then* $h_k = 0$.

Proof. From the Algorithm A and the definition of filter, the conclusion follows. □

Lemma 8. *For any trial point $x_{k+1} \neq x_k$, there must be one accepted by the filter.*

Lemma 9. *Suppose that Assumptions hold, there exists $k_5 > 0$ independent of the iterates such that*

$$q_k(0) - q_k\left(d_k^n\right) \geq -k_5 \left\|c_k\right\|. \tag{47}$$

Proof. By (32), the assumptions and $\left\|d_k^n\right\| \leq \Delta_{\max}$, it is obvious that

$$\begin{aligned} q_k(0) - q_k\left(d_k^n\right) &= -g_k^T d_k^n - \frac{1}{2}\left(d_k^n\right)^T H_k d_k^n \\ &\geq -\left\|d_k^n\right\| \left(\left\|g_k\right\| + \frac{1}{2}\left\|H_k\right\| \left\|d_k^n\right\|\right) \\ &\geq -k_1 \left\|c_k\right\| \left(\nu_0 + \nu_3 \Delta_{\max}\right) \\ &\overset{\text{def}}{=} -k_5 \left\|c_k\right\|. \end{aligned} \tag{48}$$

The proof is complete. □

Lemma 10. *Suppose that Assumptions hold and $\left\|\widehat{g}_k\right\| \geq \epsilon_t$, if*

$$\begin{aligned} \Delta_k \leq \min\Bigg\{&\left(\frac{\epsilon_t}{2k_1\nu_3\sqrt{\eta_3\alpha_1}}\right)^{2/(2+\alpha_2)}, \frac{k_4\epsilon_t}{2}, \\ &\left(\frac{\nu_0}{k_1\nu_3\sqrt{\eta_3\alpha_1}}\right)^{2/(2+\alpha_2)}, \\ &\left(\frac{\sigma\epsilon_t}{16k_1\nu_0\sqrt{\eta_3\alpha_1}}\right)^{2/\alpha_2}\Bigg\} \overset{\text{def}}{=} \delta_1 \end{aligned} \tag{49}$$

one can deduce

$$q_k(0) - q_k\left(d_k\right) \geq \frac{\sigma\epsilon_t}{8}\Delta_k \overset{\text{def}}{=} \overline{\sigma}\Delta_k. \tag{50}$$

Proof. By the assumptions and the definition of d_k^n, it holds

$$\begin{aligned}\left\|W_k^T\nabla q_k\left(d_k^n\right)\right\| &= \left\|W_k^T\left(g_k+H_kd_k^n\right)\right\| \geq \left\|\widehat{g}_k\right\| - \left\|W_k^TH_kd_k^n\right\| \\ &\geq \epsilon_t - \nu_3k_1\left\|c_k\right\| \\ &\geq \epsilon_t - \nu_3k_1\sqrt{\eta_3\alpha_1}\Delta_k^{1+(\alpha_2/2)} \geq \frac{\epsilon_t}{2}.\end{aligned} \tag{51}$$

From Lemma 3, $q_k(d_k^n) - q_k(d_k) \geq (\sigma/4)\epsilon_t \min\{\Delta_k, k_4\epsilon_t/2\} \geq (\sigma\epsilon_t/4)\Delta_k$. Together with

$$\begin{aligned}q_k(0) - q_k\left(d_k^n\right) &= -g_k^Td_k^n - \frac{1}{2}\left(d_k^n\right)^TH_kd_k^n \\ &\geq -\left\|g_k\right\|\left\|d_k^n\right\| - \frac{1}{2}\left\|H_k\right\|\left\|d_k^n\right\|^2 \\ &\geq -k_1\nu_0\sqrt{\eta_3\alpha_1}\Delta_k^{1+(\alpha_2/2)} - k_1^2\nu_3\eta_3\alpha_1\Delta_k^{2+\alpha_2} \\ &\geq -2k_1\nu_0\sqrt{\eta_3\alpha_1}\Delta_k^{1+(\alpha_2/2)} \\ &\geq -\frac{\sigma\epsilon_t}{8}\Delta_k\end{aligned} \tag{52}$$

then $q_k(0) - q_k(d_k) \geq -(\sigma\epsilon_t/8)\Delta_k + (\sigma\epsilon_t/4)\Delta_k = (\sigma\epsilon_t/8)\Delta_k = \overline{\sigma}\Delta_k$. It is the conclusion. □

Lemma 11. *Suppose the conditions of Lemma 10 hold, if*

$$\Delta_k \leq \min\left\{\delta_1, \frac{(1-\eta_1)\overline{\sigma}}{\overline{\nu}}\right\} \stackrel{\text{def}}{=} \delta_2 \tag{53}$$

then $r_k \geq \eta_1$.

Proof. From the definition of r_k and Lemma 10, together with (31), we have

$$\left|r_k - 1\right| = \frac{\left|f\left(x_k+d_k\right) - q_k\left(d_k\right)\right|}{\left|q_k(0) - q_k\left(d_k\right)\right|} \leq \frac{\overline{\nu}\Delta_k^2}{\overline{\sigma}\Delta_k} \leq 1-\eta_1. \tag{54}$$

It is obvious that $r_k \geq \eta_1$. □

Lemma 12. *Suppose the conditions of Lemmas 10 and 11 hold, if*

$$h_k \leq \left(\eta_3\alpha_1\right)^{-1/(1+\alpha_2)}\left(\frac{\eta_1\overline{\sigma}}{\sqrt{\gamma_{\mathcal{F}}}}\right)^{(2+\alpha_2)/(1+\alpha_2)} \tag{55}$$

then

$$f\left(x_k^+\right) \leq f\left(x_k\right) - \sqrt{\gamma_{\mathcal{F}}}h_k. \tag{56}$$

Proof. By Lemmas 3, 10, and 11, together with $\| \widehat{g}_k \| \geq \epsilon_t$, it holds

$$f\left(x_k\right) - f\left(x_k^+\right) \geq \eta_1\left(q_k(0) - q_k\left(d_k\right)\right) \geq \eta_1\overline{\sigma}\Delta_k. \tag{57}$$

From the Algorithm, $h_k \leq \eta_3\alpha_1\Delta_k^{2+\alpha_2}$, then

$$f\left(x_k\right) - f\left(x_k^+\right) \geq \eta_1\overline{\sigma}\left(\frac{h_k}{\eta_3\alpha_1}\right)^{1/(2+\alpha_2)}. \tag{58}$$

Hence $f(x_k) - f(x_k^+) \geq \sqrt{\gamma_{\mathcal{F}}}h_k$. □

Theorem 13. *Suppose the assumptions hold, there must exist* $\Delta_{\min} > 0$ *such that for each k, it holds*

$$\Delta_k \geq \Delta_{\min}. \tag{59}$$

Proof. Let $\overline{k}_1$ be large enough such that $h_{\overline{k}_1} \leq \epsilon_t$, it is true by Lemmas 6 and 7. Suppose by contradiction that the index j is the first one after $\overline{k}_1$, which satisfies

$$\Delta_j \leq \gamma_0\min\left\{\delta_2, \left(\frac{\left(1-\sqrt{\gamma_{\mathcal{F}}}\right)h^F}{\eta_3\alpha_1}\right)^{1/(2+\alpha_2)}, \Delta_{\overline{k}_1}\right\} \stackrel{\text{def}}{=} \gamma_0\delta_3, \tag{60}$$

where $h^F \stackrel{\text{def}}{=} \min_{i\in\mathcal{U}}h_i$ is the smallest value of violation function in filter. Then $\Delta_j \leq \gamma_0\Delta_{\overline{k}_1}$. By the above analysis, we know $j \geq \overline{k}_1 + 1$, that is $j-1 \geq \overline{k}_1$. From the Algorithm and (60), it concludes

$$\Delta_{j-1} \leq \frac{1}{\gamma_0}\Delta_j \leq \delta_3. \tag{61}$$

By (60) and (61), (53) can be obtained. In Lemma 11, let $j-1$ instead of k, it deduces

$$r_{j-1} \geq \eta_1. \tag{62}$$

Based on Lemma 12, together with (60), (61), and the algorithm, we can see

$$h_{j-1}^+ \leq \eta_3\alpha_1\Delta_{j-1}^{2+\alpha_2} \leq \left(1-\sqrt{\gamma_{\mathcal{F}}}\right)h^F. \tag{63}$$

It can be seen that (53) is true for $j-1 \geq k$, with (55), we can deduce

$$f_{j-1}^+ \leq f_{j-1} - \sqrt{\gamma_{\mathcal{F}}}h_{j-1}. \tag{64}$$

That means x_{j-1}^+ can be accepted by the filter. From above and (55), we know $\Delta_j \geq \Delta_{j-1}$. Hence the index j is not the first one after $\overline{k}_1$ which satisfied (60), that is a contradiction. So, for any $k > \overline{k}_1$, it holds $\Delta_k \geq \gamma_0\delta_3$. Define

$$\Delta_{\min} = \min\left\{\Delta_0, \ldots, \Delta_{\overline{k}_1}, \gamma_0\delta_3\right\} \tag{65}$$

we can see that

$$\Delta_k \geq \Delta_{\min} \tag{66}$$

holds for each k. The proof is complete. □

Lemma 14. *Suppose that Assumptions hold and Algorithm A does not terminate finitely, then* $\liminf_{k\to\infty}\left\|\widehat{g}_k\right\| = 0$.

Proof. Suppose by contradiction that for ϵ_t, there exists a constant $\overline{k} > 0$ such that $\| \widehat{g}_k \| \geq \epsilon_t$.

By Assumption (A3) and (A4), $\left\|W_k^T\nabla q_k(d_k^n)\right\| \geq \left\|\widehat{g}_k\right\| - \nu_3k_1 \| c_k \|$. From Lemma 6, we know $h_k \to 0$. Hence there exists $\widetilde{k} > 0$ such that

$$\left\|c_k\right\| \leq \frac{2\epsilon_t}{3\nu_3k_1} \tag{67}$$

for $k > \widetilde{k}$. Then $\| W_k^T \nabla q_k(d_k^n) \| \geq (1/3) \| \widehat{g}_k \| \geq (1/3)\epsilon_t$ for $k > \widehat{k} \stackrel{\text{def}}{=} \max\{\overline{k}, \widetilde{k}\}$.

It is obvious that

$$q_k(0) - q_k(d_k) = q_k(0) - q_k(d_k^n) + q_k(d_k^n) - q_k(d_k). \tag{68}$$

By the proof of Lemma 9, it holds $|q_k(0) - q_k(d_k^n)| \leq \nu_0 \| d_k^n \| + (1/2)\nu_3 \| d_k^n \|^2$. Together with Lemma 6 and the definition of d_k, we have

$$\lim_{k\to\infty} (q_k(0) - q_k(d_k^n)) = 0. \tag{69}$$

By the Algorithm, we can get

$$+\infty > \sum_{k=0}^{\infty} (f_k - f_{k+1}) \geq \eta_1 \sum_{k=0}^{\infty} (q_k(0) - q_k(d_k)). \tag{70}$$

Then

$$\lim_{k\to\infty} (q_k(0) - q_k(d_k)) = 0. \tag{71}$$

By (68), (69), and (71), it deduces

$$\lim_{k\to\infty} (q_k(d_k^n) - q_k(d_k)) = 0. \tag{72}$$

Based on the assumptions, Lemma 3 and Theorem 13, for $k > \widehat{k}$, it holds

$$\begin{aligned} & q_k(d_k^n) - q_k(d_k) \\ & \quad \geq \frac{\sigma}{2} \left\| W_k^T \nabla q_k(d_k^n) \right\| \min\left\{\Delta_k, k_4 \left\| W_k^T \nabla q_k(d_k^n) \right\|\right\} \\ & \quad \geq \frac{\sigma\epsilon_t}{6} \min\left\{\Delta_{\min}, \frac{k_4\epsilon_t}{6}\right\} > 0, \end{aligned} \tag{73}$$

which contradicts (72). The conclusion follows. □

Theorem 15. *Suppose the assumptions hold, and apply the algorithm to problem* (P), *then*

$$\liminf_{k\to\infty} (h_k + \|\widehat{g}_k\|) = 0, \tag{74}$$

where $\widehat{g}_k = W_k^T g_k$, $g_k = \nabla f(x_k)$, $W(x)$ *denotes a matrix whose columns form a basis of the null space of* $A(x)^T$.

Proof. If the algorithm terminates finitely, it is obvious that it holds. Otherwise, by Lemmas 6 and 14, the conclusion also can be obtained. □

Theorem 16. *Suppose the assumptions hold, and* $\{x_k\}$ *is the infinite sequence obtained by the algorithm, then there must exist a subsequence such that*

$$\lim_{j\to\infty} x_{k_j} = x^* \tag{75}$$

and x^* *satisfies the one order KKT condition of* (P).

Table 1

Problem	n	m	NF	NG	L's NF	L's NG
HS6	2	1	11	11	20	12
HS7	2	1	9	3	15	11
HS8	2	2	7	4	—	—
HS9	2	1	6	6	6	6
HS26	3	1	24	24	42	24
HS39	4	2	15	9	26	19
HS40	4	3	7	5	7	5
HS42	4	2	8	8	11	11
HS78	5	3	6	6	8	7

Proof. By Assumption (A1), there exist a subsequence $\{k_j\}$ and x^*, such that $\lim_{j\to\infty} x_{k_j} = x^*$. Together with Assumption (A3) and (A4), it hods $\lim_{j\to\infty} W_{k_j}^T g_{k_j} = 0$, which means for large enough j, g_{k_j} lies in the space spaned by the columns of $A_{k_j}^T$. That is there exists λ_{k_j} such that

$$\lim_{j\to\infty} g_{k_j} + A_{k_j}^T \lambda_{k_j} = 0. \tag{76}$$

The conclusion follows. □

6. Some Numerical Experiments

(1) Updating of H_k is done by $H_{k+1} = H_k + (y_k^T y_k / y_k^T s_k) - (H_k s_k s_k^T H_k / s_k^T H_k s_k)$, where $y_k = \theta_k \widehat{y}_k + (1-\theta_k) H_k s_k$

$$\theta_k = \begin{cases} 1 & s_k^T \widehat{y}_k \geq 0.2 s_k^T H_k s_k \\ \dfrac{0.8 s_k^T H_k s_k}{s_k^T H_k s_k - s_k^T \widehat{y}_k} & \end{cases} \tag{77}$$

and $\widehat{y}_k = g_{k+1} - g_k + (A_{k+1} - A_k)\rho_k$, $s_k = x_{k+1} - x_k \rho_k$ is the multipluser of corresponding quadratic subproblems.

(2) We assume the error toleration is 10^{-5}.

(3) The algorithm parameters were set as follows: $H_0 = I \in R^{n\times n}$, $\eta_1 = 0.25$, $\eta_2 = 0.25$, $\eta_3 = 0.1$, $\alpha_1 = \alpha_2 = 0.5$, $\gamma = 0.02$, $\rho = 0.5$, $\xi = 10^{-6}$, $\gamma_0 = 0.1$, $\gamma_1 = 0.5$, $\gamma_2 = 2$, $\gamma_F = 10^{-4}$, $\Delta_0 = 1$. The program is written in Matlab.

The numerical results for the test problems are listed in Table 1.

In Table 1, the problems are numbered in the same way as in Schittkowski [20] and Hock and Schittkowski [21]. For example, "S216" is the problem (216) in Schittkowski [20] and "HS6" is the problem (6) in Hock and Schittkowski [21]. NF, NG represent the numbers of function and gradient calculations and "L's" is the solution in [22]. The numerical results show that the our algorithm is more effective than the L's for most test examples. Moreover, the higher the level of nonmonotonic, the better the numerical results. The results show that the new algorithm is robust and effective, and more flexible for the acceptance of the the trial iterate.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (no. 11101115) and the Natural Science Foundation of Hebei Province (nos. A2010000191, Y2012021).

References

[1] R. H. Byrd, R. B. Schnabel, and G. A. Shultz, "A trust region algorithm for nonlinearly constrained optimization," *SIAM Journal on Numerical Analysis*, vol. 24, no. 5, pp. 1152–1170, 1987.

[2] J. E. Dennis Jr., M. El-Alem, and M. C. Maciel, "A global convergence theory for general trust-region-based algorithms for equality constrained optimization," *SIAM Journal on Optimization*, vol. 7, no. 1, pp. 177–207, 1997.

[3] M. J. D. Powell and Y. Yuan, "A trust region algorithm for equality constrained optimization," *Mathematical Programming A*, vol. 49, no. 2, pp. 189–211, 1991.

[4] R. Fletcher and S. Leyffer, "Nonlinear programming without a penalty function," *Mathematical Programming A*, vol. 91, no. 2, pp. 239–269, 2002.

[5] C. M. Chin and R. Fletcher, "On the global convergence of an SLP-filter algorithm that takes EQP steps," *Mathematical Programming A*, vol. 96, no. 1, pp. 161–177, 2003.

[6] R. Fletcher, N. I. M. Gould, S. Leyffer, P. L. Toint, and A. Wächter, "Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming," *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 635–659, 2002.

[7] R. Fletcher, S. Leyffer, and P. L. Toint, "On the global convergence of a filter-SQP algorithm," *SIAM Journal on Optimization*, vol. 13, no. 1, pp. 44–59, 2002.

[8] S. Ulbrich, "On the superlinear local convergence of a filter-SQP method," *Mathematical Programming B*, vol. 100, no. 1, pp. 217–245, 2004.

[9] M. Ulbrich, S. Ulbrich, and L. N. Vicente, "A globally convergent primal-dual interior-point filter method for nonlinear programming," *Mathematical Programming A*, vol. 100, no. 2, pp. 379–410, 2004.

[10] R. Fletcher and S. Leyffer, "A bundle filter method for nonsmooth nonlinear optimization," Tech. Rep. NA/195, Department of Mathematics, University of Dundee, Dundee, Scotland, December 1999.

[11] L. Grippo, F. Lampariello, and S. Lucidi, "A nonmonotone line search technique for Newton's method," *SIAM Journal on Numerical Analysis*, vol. 23, no. 4, pp. 707–716, 1986.

[12] M. Ulbrich and S. Ulbrich, "Non-monotone trust region methods for nonlinear equality constrained optimization without a penalty function," *Mathematical Programming B*, vol. 95, no. 1, pp. 103–135, 2003.

[13] K. Su and D. Pu, "A nonmonotone filter trust region method for nonlinear constrained optimization," *Journal of Computational and Applied Mathematics*, vol. 223, no. 1, pp. 230–239, 2009.

[14] K. Su and Z. Yu, "A modified SQP method with nonmonotone technique and its global convergence," *Computers & Mathematics with Applications*, vol. 57, no. 2, pp. 240–247, 2009.

[15] N. I. M. Gould and P. L. Toint, "Global convergence of a nonmonotone trust-region filter algorithm for nonlinear programming," in *Multiscale optimization methods and applications*, vol. 82, pp. 125–150, Springer, New York, NY, USA, 2006.

[16] Z. Chen and X. Zhang, "A nonmonotone trust-region algorithm with nonmonotone penalty parameters for constrained optimization," *Journal of Computational and Applied Mathematics*, vol. 172, no. 1, pp. 7–39, 2004.

[17] Z. W. Chen, "A penalty-free-type nonmonotone trust-region method for nonlinear constrained optimization," *Applied Mathematics and Computation*, vol. 173, no. 2, pp. 1014–1046, 2006.

[18] F. A. M. Gomes, M. C. Maciel, and J. M. Martínez, "Nonlinear programming algorithms using trust regions and augmented Lagrangians with nonmonotone penalty parameters," *Mathematical Programming A*, vol. 84, no. 1, pp. 161–200, 1999.

[19] M. J. D. Powell, "Convergence properties of a class of minimization algorithms," in *Nonlinear Programming*, pp. 1–27, Academic Press, New York, NY, USA, 1974.

[20] K. Schittkowski, *More Test Examples for Nonlinear Programming Codes*, vol. 282, Springer, Berlin, Germany, 1987.

[21] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, vol. 187, Springer, Berlin, Germany, 1981.

[22] M. Lalee, J. Nocedal, and T. Plantenga, "On the implementation of an algorithm for large-scale equality constrained optimization," *SIAM Journal on Optimization*, vol. 8, no. 3, pp. 682–706, 1998.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 607190, 8 pages
http://dx.doi.org/10.1155/2013/607190

Research Article

Modified Malmquist Productivity Index Based on Present Time Value of Money

Farhad Hosseinzadeh Lotfi,[1] Golamreza Jahanshahloo,[1] Mohsen Vaez-Ghasemi,[1] and Zohreh Moghaddas[2]

[1] *Department of Mathematics, Islamic Azad University, Science and Research Branch, P.O. Box 14155/4933, Tehran, Iran*
[2] *Department of Electrical, Computer and Biomedical Engineering, Islamic Azad University, Qazvin Branch, P.O. Box 34185/1416, Qazvin, Iran*

Correspondence should be addressed to Farhad Hosseinzadeh Lotfi; farhad@hosseinzadeh.ir

Received 9 September 2012; Revised 11 December 2012; Accepted 30 December 2012

Academic Editor: Mohammad Khodabakhshi

Data envelopment analysis (DEA) models can calculate the Malmquist Productivity Index (MPI). Classic Malmquist Productivity Index shows regress and progress of a DMU in different periods with efficiency and technology variations without considering the present value of money. This issue is of major importance since while a currency of in previous year is not equal to that of now this would yield bias results which can affect the correct interpretation. The index developed here is defined in terms of Modified Malmquist Productivity Index model, which can calculate progress and regress by using the factor of present time value of money. The incorporation of present time value of money is also calculated within the framework of data envelopment analysis. This factor is fundamental and should be considered in DEA Malmquist Productivity Index. Moreover, here, differences between presented models are compared to those of previous ones indeed, biased results will be shown in the case study in banks, and problem and solution have been investigated in the literature.

1. Introduction

Data envelopment analysis is mathematical programming technique for obtaining relative efficiency of a set of decision making units (DMUs). Nowadays DEA is widely used in various fields. Utilizing data envelopment analysis (DEA) methodology it is also possible to estimate the Malmquist Productivity Index. As one of the major sources of economic development is productivity growth thus having a comprehensive interpretation of those factors affects productivity is very influential and leading.

Malmquist [1], in 1953, published a quantity index for use in consumption analysis. In this index input distance functions are used to make comparison among two or more consumption bundles. Later in 1982, in production analysis Caves et al. [2], introduced Malmquist Productivity Index on basis of what malmquist has proposed. Nowadays applications which use the Malmquist Productivity Index have become widespread in the literature. In recent years, among researchers who are studying firm performance, the measurement and analysis of productivity change have enjoyed a great deal of attention.

As measuring productivity change gains an important attention in the literature Färe et al. [3] in a paper completely discussed productivity growth, technical progress, and efficiency change. They applied these factors in evaluating industrialized countries. Maniadakis and Thanassoulis [4] developed a productivity index that is an extension of the work on malmquist indexes. They evolved a productivity index which is applicable when input prices are known and producers are cost minimisers. In doing so, they developed a productivity index that accounts not only for technical efficiency and technological variations but also for allocative efficiency and for the effects of input price variations. Grifell-Tatjé and Lovell [5] provided a paper in order to adopt a different approach to the use of DEA with panel data and

create a malmquist index of productivity change and provide a new decomposition for it. Grifell-Tatjé et al. [6] provided a new Malmquist Productivity Index called a quasi-Malmquist productivity index which incorporates all slacks on the selected side and replaces conventional radial efficiency measures with the new nonradial efficiency ones. Also, Chen [7], on bases of the fact that DEA-based Malmquist Productivity Index measures the technical and productivity changes over time, has extended the Malmquist Productivity Index into a nonradial index where the decision maker's preference over performance improvement can also be incorporated. The advantage of this index is that by the nonzero slacks it eliminates possible inefficiency.

Since malmquist indexes of productivity are generally estimated using index number techniques or nonparametric frontier approaches Fuentes et al. [8] aimed to estimate malmquist indexes in a similar way using parametric-deterministic or parametric-stochastic frontier approaches. They adopted an output distance function and showed that using the estimated parameters, several radial distance functions can be calculated and moreover combined for estimating and decomposing the productivity indexes. Orea [9] in his paper provided a parametric decomposition of a generalized Malmquist Productivity Index which considers scale economies. As he said in his research the contribution of scale economies to productivity change is evaluated without recourse to scale efficiency measures, which are neither bounded for globally increasing, decreasing, or constant returns to scale technologies nor for ray-homogeneous technologies. Lin et al. [10] in their article considered 117 branches of a certain bank in Taiwan and introduced data envelopment analysis to assess the operating performances of business units of this bank. Their work, in determining operation strategies, provides the reference for a bank's managers. In their investigation Wang and Lin [11] established an analytical hierarchy framework for helping banks in order to choose merger strategies. Also, The consistent fuzzy preference relation is used for improving effectiveness and decision-making consistency. The obtained analytical results shed light on the issue that, in strategy selection, risk management and financial composition of banks are the main considerations. Wu et al. [12] for banking performance evaluation proposed a fuzzy multiple criteria decision-making (FMCDM) approach. Also, the three MCDM analytical tools of SAW, TOPSIS, and VIKOR were respectively adopted to rank the banking performance and improve the gaps with three banks as an empirical example. Ng et al. [13] indicated that in the banking industry, it is desirable to identify potential bank failure or high-risk banks. Thus, in their paper they have proposed a fuzzy CMAC (cerebellar model articulation controller) model based on compositional rule of inference, called FCMAC-CRI(S), as an innovative way for tackling the problem using localized learning.

Here the aim is to become more precise in calculating Malmquist Productivity Index since in this subject inaccurate inputs would lead to biased results of efficiency. Considering the Malmquist Productivity Index which is used to compute the progress and regress of entities in successive periods we emphasize that it is of major significance to pay concentration while the Malmquist Productivity Index is being calculated for DMUs which have similar performances in time t and time $t+1$. It would be definitely not fair enough to merely consider efficiency variations and technological variations. The fact is that a specific value of money in time t is not equal to that value in time $t+1$, that is, $(10\$)_t \neq (10\$)_{t+1}$. Thus if technological variations and efficiency variations in time t and time $t+1$ have the same performances, then, the interest rate needs to be considered in time $t+1$. The index developed here is defined in terms of Modified Malmquist Productivity Index model (MMPI), which can calculate progress and regress by using the factor of present time value of money. The incorporation of present time value of money is also calculated within the framework of data envelopment analysis.

The current paper proceeds as follows. In the next section, Malmquist Productivity Index will be briefly reviewed. Then, in Section 3, the proposed method, Modified Malmquist Productivity Index, which is based on the present time value of money, will be discussed. An illustrative example is documented in Section 4 in which main findings are highlighted, and Section 5 concludes the paper with conclusions and recommendation.

2. Malmquist Productivity Index

Utilizing DEA methodology it is possible to estimate the Malmquist Productivity Index. As is, DEA models are linear programming (LP) models with which the production frontier can be estimated. Those DMUs located onto this frontier are called efficient and others referred to as inefficient. The degree of efficiency for each DMUs can be obtained on the basis of the Euclidean distance of their input-output ratio from the estimated frontier. Since efficient DMUs construct production frontier thus it can obviously change over time. What Malmquist DEA approach does is to calculate the efficiency measure for one year relative to that of the prior year, while the frontier may change from time to time (time t and time $t+1$). Thus it can be said that the frontier function has shifted from frontier t to frontier $t+1$.

Let DMU_l denote a unit from a total n units that relative efficiency is being evaluated. Define $x_l \in R_+^m$ and $y_l \in R_+^s$ as semipositive input and output vectors of DMU_l. The most general way of characterization of production technology is production possibility set T, which is defined with a set of semipositive (x, y) as

$$T = \left\{ (x, y) \mid x \geq \sum_{j=1}^{n} \lambda_j x_j,\ \ y \leq \sum_{j=1}^{n} \lambda_j y_j,\ \ \lambda_j \geq 0,\ \ j = 1, \ldots, n \right\}. \tag{1}$$

As existed in the literature Malmquist Productivity Index can be calculated via several functions, such as distance function:

$$D(X_l, Y_l) = \mathrm{Min}\left\{\theta : (\theta X_l, Y_l) \in T\right\}. \tag{2}$$

The resultant distance function can be computed by solving linear programming problems. Consider an input-oriented CCR model as follows:

$$\begin{aligned} D^f\left(x_l^k, y_l^k\right) = \ & \min \theta \\ & \text{s.t.} \quad \sum_{j=1}^{n} \lambda_j x_{ij}^f \le \theta x_{il}^k, \quad i = 1, \ldots, m, \\ & \sum_{j=1}^{n} \lambda_j y_{rj}^f \ge y_{rl}^k, \quad r = 1, \ldots, s, \\ & \lambda_j \ge 0, \quad j = 1, \ldots, n, \end{aligned} \tag{3}$$

in which l is the unit under assessment and each of k and f varies between time t and time $t + 1$. As an instance for assessing DMU_l consider $k = t$ and $f = t+1$, $D^{t+1}(x_l^t, y_l^t)$; this means that DMU_l is considered in time t while technology is considered in time $t + 1$. Considering this notification, four LP problems can be defined.

In regards of this subject, Caves et al. [2] have introduced the Malmquist Productivity Index as follows in which the results obtained from the mentioned models are being used:

$$M\left(x_l^{t+1}, y_l^{t+1}, x_l^t, y_l^t\right) = \left(\frac{D^t\left(x_l^{t+1}, y_l^{t+1}\right) D^{t+1}\left(x_l^{t+1}, y_l^{t+1}\right)}{D^t\left(x_l^t, y_l^t\right) D^{t+1}\left(x_l^t, y_l^t\right)}\right)^{1/2}, \tag{4}$$

in which x_l^t and y_l^t are the input and output vectors for unit l, used in period t. Also, x_l^{t+1} and y_l^{t+1} are the input and output vectors for unit l, used in period $t + 1$. This index measures the productivity of unit l at the production (x_l^{t+l}, y_l^{t+l}) relative to (x_l^t, y_l^t).

The previously equation can be further decomposed into two components mentioned: one for measuring the change in technical efficiency and the other for measuring the technical change which means the technology frontier shift between the two time periods, t and $t + l$:

$$M\left(x_l^{t+1}, y_l^{t+1}, x_l^t, y_l^t\right) = \frac{D^{t+1}\left(x_l^{t+1}, y_l^{t+1}\right)}{D^t\left(x_l^t, y_l^t\right)} \left[\frac{D^t\left(x_l^{t+1}, y_l^{t+1}\right) D^t\left(x_l^t, y_l^t\right)}{D^{t+1}\left(x_l^{t+1}, y_l^{t+1}\right) D^{t+1}\left(x_l^t, y_l^t\right)}\right]^{1/2}. \tag{5}$$

The interpretation of this equation is that $M(x_l^{t+1}, y_l^{t+1}, x_l^t, y_l^t) > 1$ indicates an improvement in total productivity, $M(x_l^{t+1}, y_l^{t+1}, x_l^t, y_l^t) < 1$ indicates a decline, and $M(x_l^{t+1}, y_l^{t+1}, x_l^t, y_l^t) = 1$ shows an unchanged productivity growth, see Caves et al. [2], and Chen [7].

3. Main Subject

In performance assessment inaccurate inputs would lead to biased results of efficiency. Malmquist Productivity Index is used for computing the progress and regress of entities in successive periods. It is of great importance to pay attention when Malmquist Productivity Index is being calculated for DMUs with similar performances in time t and time $t + 1$. Thus, a question is brought forth for discussion: would it be fair enough to merely consider efficiency variations and technological variations? Of course not. The fact is that an specific value of money in time t is not equal to that value in time $t + 1$, that is, $(10\$)_t \neq (10\$)_{t+1}$. Thus if technological variations and efficiency variations in time t and time $t + 1$ have the same performances, then, the interest rate needs to be considered in time $t + 1$.

For instance consider a bank with a large financial capital in a year which has a performance lower than the interest rate in the country; it would definitely have regressed even if it have a high efficiency and positive technological variations. In this case the corresponding Malmquist Productivity Index is greater than one.

Here "single payment compound" is utilized for calculating the time value of money in two successive years. If one has $A\$$ in time f, corresponding value will be $A \times (1 + e)^n$ in time l where $n = l - f$ and e is the interest rate in time f to l. If $n > 0$ then $A_l \times (1 + e)^n = A_f$ and if $n < 0$ then $A_l = A_f \times (1 + e)^n$ which means that $A_l \times (1/(1 + e)^{-n}) = A_f$. It makes no difference to multiply $(1 + e)^n$ to A_f or divide A_l by $(1 + e)^n$. This means those DMUs have inputs and (or) outputs influenced by time value of money should be compared on equal terms with one an other. Thus it is necessary to make these changes first and then consider the observations and compare them to the efficient frontiers. As said before in order to make these values equal it is possible to make the changes in either side of the equation. Consider $f = t$ and $l = t + 1$; in this case $n = t + 1 - t$; thus the vale of money will be $A \times (1 + e)$ in time $t + 1$. As in Malmquist Productivity Index times t and $t + 1$ are compared with each other thus always $n = 1$.

For clarity consider the following example. If one has 12\$ in time t and 14\$ in time $t + 1$, while all the factors, specially time value of money, are the same in these two time periods, thus progress had happed. But, if the value of 12\$ in time t is equal to the value of 15\$ in time $t + 1$, therefore a regress had happened. Thus, it is necessary to consider time value of money for those factors which is impressible while evaluating the progress or the regress of units.

It should be noted that if productivity is calculated in successive months the interest rate has been computed on basis of months.

This procedure will be performed for those factors on which time value of money is impressive.

Therefore, consider a situation in time t in which from the x units of inputs, with the interest value of e, y units of outputs have been produced. In this situation, certainly, in time $t + 1$ with the interest value of $\acute{e}$ the inputs (x) and the outputs (y) are not the same as those of in time t. Thus, considering the time value of money for those factors on which it leaves impression, the results may be different to those acquired without regarding the time value of money. As a result, at first, the interest rate of money is expected to be accounted for them, and efficiency variations and technological variations

should be calculated. For those factors on which interest value is not impressive, such as number of personals and equipment, there is no need to be dealt with like this, and they should be treated similar to the precedent.

Consider the previously-mentioned discussion with the time value of money is being incorporated into the analysis, the following four LPs will be presented for assessing Modified Malmquist Productivity Index.

Under constant returns to scale, the LP for $D^t(x_l^t, y_l^t)$, with m inputs and s outputs, is as follows:

$$\begin{aligned} \overline{D}^t\left(x_l^t, y_l^t\right) = \min\ & \theta \\ \text{s.t.}\quad & \sum_{j=1}^{n} \lambda_j x_{ij}^t \le \theta x_{il}^t, \quad i = 1, \ldots, m, \\ & \sum_{j=1}^{n} \lambda_j y_{rj}^t \ge y_{rl}^t, \quad r = 1, \ldots, s, \\ & \lambda_j \ge 0, \quad j = 1, \ldots, n. \end{aligned} \tag{6}$$

Similarly, the other three LP problems become

$$\begin{aligned} \overline{D}^{t+1}\left(x_l^t, y_l^t\right) = \min\ & \theta \\ \text{s.t.}\quad & \sum_{j=1}^{n} \lambda_j x_{ij}^{t+1} \le \theta x_{il}^t, \quad i \in I_1, \\ & \sum_{j=1}^{n} \lambda_j y_{rj}^{t+1} \ge y_{rl}^t, \quad r \in R_1, \\ & \sum_{j=1}^{n} \lambda_j x_{ij}^{t+1} \le \theta (1+e)^1 x_{il}^t, \quad i \in I_2, \\ & \sum_{j=1}^{n} \lambda_j y_{rj}^{t+1} \ge (1+e)^1 y_{rl}^t, \quad r \in R_2, \\ & \lambda_j \ge 0, \quad j = 1, \ldots, n, \end{aligned} \tag{7}$$

where I_1 and R_1 show the subsets of inputs and outputs, respectively for which time value of money the nonimpressible and I_2 and R_2 shows the subsets of inputs and outputs, respectively, for which is the time value of money is influential. It also should be mentioned that $I = \{1, \ldots, m\}$, $R = \{1, \ldots, s\}$ and $I = I_1 \cup I_2$, $R = R_1 \cup R_2$

$$\begin{aligned} \overline{D}^{t+1}\left(x_l^{t+1}, y_l^{t+1}\right) = \min\ & \theta \\ \text{s.t.}\quad & \sum_{j=1}^{n} \lambda_j x_{ij}^{t+1} \le \theta x_{il}^{t+1}, \quad i = 1, \ldots, m, \\ & \sum_{j=1}^{n} \lambda_j y_{rj}^{t+1} \ge y_{rlo}^{t+1}, \quad r = 1, \ldots, s, \\ & \lambda_j \ge 0, \quad j = 1, \ldots, n, \end{aligned} \tag{8}$$

$$\begin{aligned} \overline{D}^{t}\left(x_l^{t+1}, y_l^{t+1}\right) = \min\ & \theta \\ \text{s.t.}\quad & \sum_{j=1}^{n} \lambda_j x_{ij}^{t} \le \theta x_{il}^{t+1}, \quad i \in I_1, \\ & \sum_{j=1}^{n} \lambda_j y_{rj}^{t} \ge y_{rl}^{t+1}, \quad r \in R_1, \\ & \sum_{j=1}^{n} \lambda_j (1+e)^1 x_{ij}^{t} \le \theta x_{il}^{t+1}, \quad i \in I_2, \\ & \sum_{j=1}^{n} \lambda_j (1+e)^1 y_{rj}^{t} \ge y_{rl}^{t+1}, \quad r \in R_2, \\ & \lambda_j \ge 0, \quad j = 1, \ldots, n. \end{aligned} \tag{9}$$

In model (7) subsets of inputs and outputs are the same as what has been discussed previously.

It is noteworthy of attention that in models (6) and (8) time value of money is not included. Time value of money does not influence the procedure since two similar periods are being compared with each other and since time value of money is fixed in a period. Moreover, according to the aforesaid formula $(1+e)^n$, when n is equal to zero, one is multiplied to the input and output parameters. But, in models (7) and (9), which are considered in various periods, the time value of money, for the indexes under the influence of it, is calculated by "single payment compound" factor. The Modified Malmquist Productivity Index is calculated like the preceding classic analysis through the following formula:

$$\overline{M}\left(x_i^{t+1}, y_i^{t+1}, x_i^t, y_i^t\right) = \frac{\overline{D}^{t+1}\left(x_i^{t+1}, y_i^{t+1}\right)}{\overline{D}^{t}\left(x_i^{t}, y_i^{t}\right)} \left[\frac{\overline{D}^{t}\left(x_i^{t+1}, y_i^{t+1}\right) \overline{D}^{t}\left(x_i^{t}, y_i^{t}\right)}{\overline{D}^{t+1}\left(x_i^{t+1}, y_i^{t+1}\right) \overline{D}^{t+1}\left(x_i^{t}, y_i^{t}\right)} \right]^{1/2}. \tag{10}$$

Considering the aforesaid discussion, in regards of (10) it can be concluded that $\overline{M}(x_i^{t+1}, y_i^{t+1}, x_i^t, y_i^t) > 1$ indicates productivity gain, $\overline{M}(x_i^{t+1}, y_i^{t+1}, x_i^t, y_i^t) < 1$ indicates productivity loss, and $\overline{M}(x_i^{t+1}, y_i^{t+1}, x_i^t, y_i^t) = 1$ means no change in productivity from time t to time $t+1$.

4. Application

In early work in this field, productivity change was explained in terms of technical change, and efficiency change but in this paper according to the mentioned discussion it has been convinced that present time value of money plays an influential role in showing the progress or regress of an entity; thus this factor should also be accounted for.

Here an application of the methodology to the Iranian banks in the period of 2006 to 2009 has been examined. Employing the Malmquist Productivity Index which is calculated based on data envelopment analysis' models,

Table 1: Description.

Index	Status
Input	
Asses quality (I_1)	Nonimpressible
Rate of deposit growth (I_2)	Nonimpressible
Total possessing (I_3)	Impressible
Personal costs (I_4)	Impressible
Interest payment (I_5)	Impressible
Output	
Profit marginal (O_1)	Nonimpressible
Rate of revenue growth (O_2)	Nonimpressible
Received commission (O_3)	Impressible
Share-holders equity (O_4)	Impressible
Acquired interest (O_5)	Impressible
Total revenue (O_6)	Impressible

productivity measure can be computed. The incorporation of present time value of money is also calculated within the framework of data envelopment analysis as showed in previous section.

Over the last years, the standard structural analysis that has taken place in the productivity measurement has been developed in terms of technical change and efficiency change, but the actuality is that present time value of money should also be incorporated into the analysis. In Table 1 we give a brief explanation about variables. The input-output indexes are listed in Tables 2–5. Also, it is specified as to whether they are under the influence of the time value of money. As you can see, for some indexes like "Asses quality" and "rate of deposit growth" time value of money is not influential and they are indicated as "nonimpressible" and for some other as "total possessing" and "personal costs", it is observable and it should be considered into the analysis. These indexes are indicated as "impressible."

According to the presented models and aforesaid discussions, the present time value of money is also incorporated into the analysis within the framework of data envelopment analysis. As shown in previous section, Modified Malmquist Productivity Index has been calculated and the results of these two analysis are gathered in Tables 6–10.

As it was shown in the following tables MMPI model yields different results in comparison to those of MPI. On regards of the interest rate in 2006-2007, 2007-2008, and 2008-2009 it can be found out that on basis of the first wrong picture which shows a progress in some of the banks, all of them in the first period of analysis have made regress. That means that those banks have shown lower performance in contrast to that of classic model. Thus, one of the influential factors which should be incorporated while progress and regress of organizations are being analyzed is to calculate the interest rate and time value of money. It is worthy of attention that in developing countries interest rate has a great amount, and its effect on economics transactions has a significant role. In this application the interest rates of 2006-2007, 2007-2008, and 2008-2009 are 16%, 18.4%, and 12.5%, respectively.

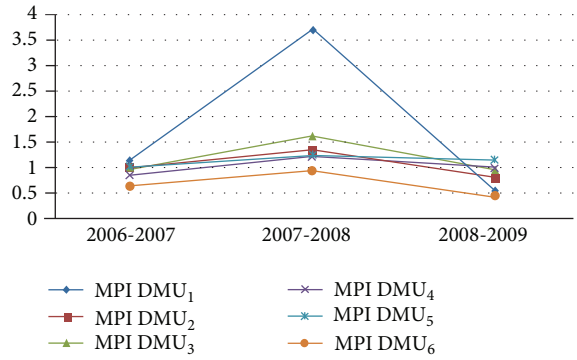

Figure 1: Malmquist Productivity Index.

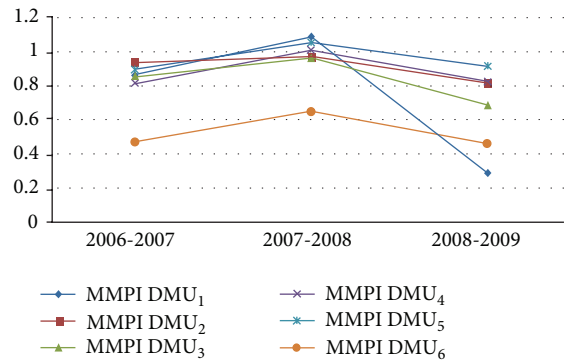

Figure 2: Modified Malmquist Productivity Index.

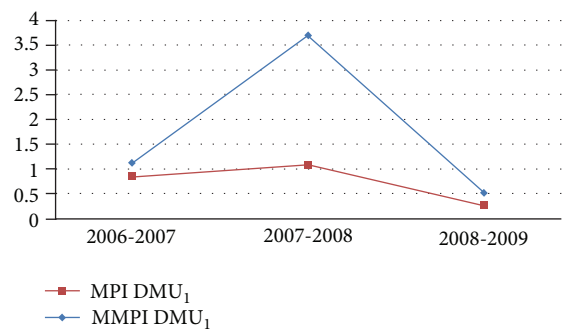

Figure 3: Malmquist changes for DMU_1.

In the second period due to the reduction of interest value and corresponding variations of time value of money, the performance of banks has been improved somehow. But, while the acquired results have being compared to those obtained from classic model, which shows five banks has made progress, in modified analysis only three banks have progressed.

Considering the acquired results from modified analysis in the third period it has been revealed that all banks have regressed. By inclusion of the interest rate in modified model for those banks which are under evaluation, a warning bell rings which shows the weak performance of Iranian banks in successive periods while this factor has been considered.

Table 2: Inputs and outputs (2006).

DMUs	I_1	I_2	I_3	I_4	I_5	O_1	O_2	O_3	O_4	O_5	O_6
1	0.824	0.350	90906777	1546117	3733535	0.021	0.359	474259	2314028	5456846	6242343
2	0.916	0.381	42765690	761666	1531782	0.039	0.381	147729	1227237	2986501	3281831
3	0.848	0.297	61415068	1012123	2713555	0.037	0.364	172220	2192410	4774258	5165554
4	0.914	0.280	31843148	562000	1322229	0.027	0.417	136994	1116026	2109188	2380064
5	0.857	0.419	39809905	612876	1580745	0.028	0.397	188265	1323499	2583767	2862649
6	0.882	0.360	9190113	209150	322760	0.056	0.854	42873	650130	772256	860719

Table 3: Inputs and outputs (2007).

DMUs	I_1	I_2	I_3	I_4	I_5	O_1	O_2	O_3	O_4	O_5	O_6
1	0.845	0.259	106959115	2045491	4746010	0.016	0.272	460303	2292258	6279449	7943232
2	0.912	0.240	52281855	988163	2202181	0.029	0.227	215136	1125702	3568242	4026108
3	0.866	0.240	66852215	1267093	3450791	0.027	0.106	178679	1962481	5209039	5711620
4	0.915	0.273	38858011	719412	1737239	0.030	0.323	181317	1234487	2815229	3148523
5	0.956	0.361	66933174	987139	2039642	0.027	0.328	222179	2262840	3456352	3802313
6	0.887	0.735	13969634	317444	329968	0.060	0.376	76364	998289	1025285	1184574

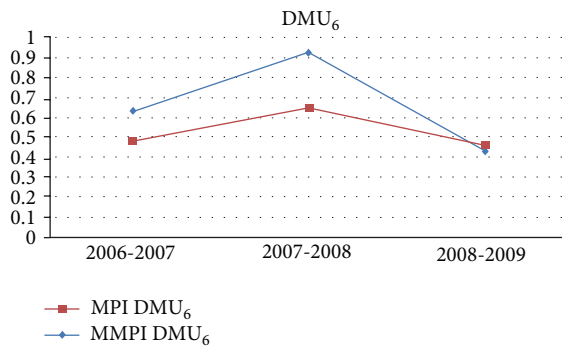

Figure 4: Malmquist changes for DMU_6.

While the increasing interest rate is being incorporated, in the event that the technology has not changed, banks encountered a regress, and this difficulty should be prevented and an immediate action must be taken.

In the following, performance of each bank is being compared to that of itself in different periods. It can also be discussed that if the performance in 2006 is being compared to that of 2009 in the corresponding model n should be replaced with 3; that means a computation of three periods for interest rate.

As it can be seen in the following figures, variations in classic and Modified Malmquist Productivity Indexes have major differences. In classic analysis, except DMU_1, other DMUs have similar variations, but in that of modified one variations have various procedures. Variations in classic Malmquist Productivity Index are described in Figure 1.

Also, variations in Modified Malmquist Productivity Index are depicted in Figure 2. In the following, variations in classic and Modified Malmquist Productivity Indexes will be specifically discussed for two DMUs (DMU_1 and DMU_6).

For the first bank (DMU_1), variations of Malmquist Productivity Index is as what has been seen in Figure 3. The progress that DMU_1, in classic models, has made is totally different from that of the modified analysis, and the variance of variations in the modified approach is more rationale. That means, all of the under-assessment banks in years of analysis do not have significant technological variations. Thus, the corresponding Malmquist Index has a more stable procedure. This fact in modified analysis is considerable. Now, consider Figure 4 which shows variations in classic and modified approaches for DMU_6. Modified Malmquist Productivity Index in the third period has revealed a lower regress in comparison to that of second period. Whereas, in classic analysis it witnessed an intense decrease while being compared to the second period. As a consequence of considering the present time value of money according to the aforesaid discussion it has been shown that regarding the modified analysis has led to different results while Malmquist Productivity Index is being calculated.

5. Conclusion

Classic Malmquist Productivity Index, in different periods, without considering the present value of money, shows regress and progress of a DMU while considering efficiency and technology variations. This shortcoming would yield biased results which can affect the correct interpretation since a currency in last year in not equal to the that of this. Noted that performance assessment with inaccurate inputs would lead to biased results of efficiency. This shortcoming would affect Malmquist Productivity Index which is used to compute the progress and regress of entities in successive periods. Thus it is obvious that it would not be fair enough to merely consider efficiency and technological variations. The index developed here has been defined in terms of Modified Malmquist Productivity Index (MMPI) model, which can calculate progress and regress by using the factor of present

Table 4: Inputs and outputs (2008).

DMUs	I_1	I_2	I_3	I_4	I_5	O_1	O_2	O_3	O_4	O_5	O_6
1	0.838	0.268	16281551	2622188	6131088	0.025	0.494	716748	8012504	9522348	11869855
2	0.922	0.498	94278569	1240252	3380231	0.030	0.617	330604	3972909	5549420	6509109
3	0.787	0.406	128550383	1903395	4582403	0.041	0.728	595662	5648607	8656018	9870337
4	0.940	0.327	62867728	990467	2241437	0.034	0.461	328427	2365168	3985100	4600389
5	0.960	0.326	88157665	1258469	2891489	0.026	0.464	384005	2267367	4916408	5567726
6	0.790	0.464	20262710	491521	732181	0.056	0.606	110347	1088278	1691760	1902707

Table 5: Inputs and outputs (2009).

DMUs	I_1	I_2	I_3	I_4	I_5	O_1	O_2	O_3	O_4	O_5	O_6
1	0.872	0.243	215200038	3624698	8301843	0.014	0.319	2189673	6770928	11133284	15660622
2	0.967	0.388	146756030	1688724	4416677	0.037	0.307	470243	4327269	7275909	8507807
3	0.823	0.188	149243454	2223659	5216403	0.035	0.166	756999	5142175	10133005	11504037
4	0.933	0.226	83332310	1124923	3350167	0.026	0.416	501502	2639362	5286830	6512891
5	0.971	0.245	114430158	1516034	3951486	0.024	0.327	701409	2940119	6342139	7387085
6	0.853	0.385	27618519	651419	1256218	0.031	0.221	154685	1136311	2005444	2323583

Table 6: Malmquist index comparison of 2007 to 2006.

DMUs	MPI	MPI status	MMPI	MMPI status	Differences
1	1.134	Progress	0.864	Regress	Changed
2	0.988	Regress	0.935	Regress	Equable
3	0.997	Regress	0.855	Regress	Equable
4	0.844	Regress	0.816	Regress	Equable
5	1.025	Progress	0.897	Regress	Changed
6	0.634	Regress	0.480	Regress	Equable

Table 7: Malmquist index comparison of 2008 to 2007.

DMUs	MPI	MPI status	MMPI	MPI status	Differences
1	3.704	Progress	1.090	Progress	Equable
2	1.338	Progress	0.967	Regress	Changed
3	1.609	Progress	0.963	Regress	Changed
4	1.199	Progress	1.010	Progress	Equable
5	1.243	Progress	1.056	Progress	Equable
6	0.927	Regress	0.649	Regress	Equable

Table 8: Malmquist index comparison in 2009 to 2008.

DMUs	MPI	MPI Status	MMPI	MMPI Status	Differences
1	0.530	Regress	0.289	Regress	Equable
2	0.804	Regress	0.816	Regress	Equable
3	0.944	Regress	0.687	Regress	Equable
4	0.996	Regress	0.828	Regress	Equable
5	1.150	Progress	0.917	Regress	Changed
6	0.430	Regress	0.461	Regress	Equable

time value of money. It should be noted that the incorporation of present time value of money is also calculated within the framework of data envelopment analysis. In the case study presented here the major concentration is showing the true progress and regress of bank branches. Moreover, those factors on which the time value of money is impressible are mainly financial ones that are under the influence of the interest rate. Thus while considering Time Value of Money, further investigations of other concepts relevant to DEA can also be considered from this point of view.

Table 9: Malmquist Productivity Index.

DMUs	MPI (2006-2007)	MPI (2007-2008)	MPI (2008-2009)
1	1.134	3.704	0.530
2	0.988	1.338	0.804
3	0.997	1.609	0.944
4	0.844	1.199	0.996
5	1.025	1.243	1.150
6	0.634	0.927	0.430

Table 10: Modified Malmquist Productivity Index.

DMUs	MMPI (2006-2007)	MMPI (2007-2008)	MMPI (2008-2009)
1	0.864	1.090	0.289
2	0.935	0.967	0.816
3	0.855	0.963	0.687
4	0.816	1.010	0.828
5	0.897	1.056	0.917
6	0.480	0.649	0.461

References

[1] S. Malmquist, "Index numbers and indifference surfaces," *Trabajos de Estadistica*, vol. 4, no. 2, pp. 209–242, 1953.

[2] D. W. Caves, L. R. Christensen, and W. E. Diewert, "The economic theory of index numbers and the measurement of input, output and productivity," *Econometrica*, vol. 50, no. 6, pp. 1393–1414, 1982.

[3] R. Färe, S. Grosskopf, M. Norris, and Z. Zhang, "Productivity growth, technical progress, and efficiency change in industrialized countries," *The American Economic Review*, vol. 84, no. 1, pp. 66–83, 1994.

[4] N. Maniadakis and E. Thanassoulis, "A cost Malmquist productivity index," *European Journal of Operational Research*, vol. 154, no. 2, pp. 396–409, 2004.

[5] E. Grifell-Tatjé and C. A. K. Lovell, "A DEA-based analysis of productivity change and intertemporal managerial performance," *Annals of Operations Research*, vol. 73, pp. 177–189, 1997.

[6] E. Grifell-Tatjé, C. A. K. Lovell, and J. T. Pastor, "A Quasi-Malmquist productivity index," *Journal of Productivity Analysis*, vol. 10, no. 1, pp. 7–20, 1998.

[7] Y. Chen, "A non-radial Malmquist productivity index with an illustrative application to Chinese major industries," *International Journal of Production Economics*, vol. 83, no. 1, pp. 27–35, 2003.

[8] H. J. Fuentes, E. Grifell-Tatjé, and S. Perelman, "A parametric distance function approach for Malmquist productivity index estimation," *Journal of Productivity Analysis*, vol. 15, no. 2, pp. 79–94, 2001.

[9] L. Orea, "Parametric decomposition of a generalized Malmquist productivity index," *Journal of Productivity Analysis*, vol. 18, no. 1, pp. 5–22, 2002.

[10] T. T. Lin, C.-C. Lee, and T.-F. Chiu, "Application of DEA in analyzing a bank's operating performance," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8883–8891, 2009.

[11] T.-C. Wang and Y.-L. Lin, "Applying the consistent fuzzy preference relations to select merger strategy for commercial banks in new financial environments," *Expert Systems with Applications*, vol. 36, no. 3, pp. 7019–7026, 2009.

[12] H.-Y. Wu, G.-H. Tzeng, and Y.-H. Chen, "A fuzzy MCDM approach for evaluating banking performance based on Balanced Scorecard," *Expert Systems with Applications*, vol. 36, no. 6, pp. 10135–10147, 2009.

[13] G. S. Ng, C. Quek, and H. Jiang, "FCMAC-EWS: a bank failure early warning system based on a novel localized pattern learning and semantically associative fuzzy neural network," *Expert Systems with Applications*, vol. 34, no. 2, pp. 989–1003, 2008.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 480962, 9 pages
http://dx.doi.org/10.1155/2013/480962

Research Article

A Game-Theoretic Analysis of Bandwidth Allocation under a User-Grouping Constraint

Shun-Pin Hsu,[1] Shun-Liang Hsu,[2] and Alan Shenghan Tsai[1]

[1] *Department of Electrical Engineering, National Chung Hsing University, 250 Kuo-Kuang Road, Taichung 402, Taiwan*
[2] *Department of Law and Graduate Institute of Technology Management, National Chung Hsing University, 250 Kuo-Kuang Road, Taichung 402, Taiwan*

Correspondence should be addressed to Shun-Pin Hsu; shsu@nchu.edu.tw

Received 2 October 2012; Revised 22 December 2012; Accepted 26 December 2012

Academic Editor: Xiaoning Zhang

A new bandwidth allocation model is studied in this paper. In this model, a system, such as a communication network, is composed of a finite number of users, and they compete for limited bandwidth resources. Each user adopts the decision that maximizes his or her own benefit characterized by the utility function. The decision space of each user is subject to constraints. In addition, some users form a group, and their joint decision space is also subject to constraints. Under the assumption that each user's utility function satisfies some continuity and concavity conditions, the existence, uniqueness, and fairness, in some appropriate sense, of the Nash equilibrium point in the allocation game are proved. An algorithm yielding a sequence converging to the equilibrium point is proposed. Finally, a numerical example with detailed analysis is provided to illustrate the effectiveness of our work.

1. Introduction

With the widespread use of internet and the increasing popularity of mobile devices, more and more people can get online at almost anytime and anywhere. An immediate challenge facing the significant increment of online users is the support of quality of service (QoS). Over the past decade considerable efforts have been made to ensure the smooth operation of the networking systems. For example, the load balancing problems were considered by Anselmi et al. [1] and Ayesta et al. [2]. The routing problems were studied by La and Anantharam [3], Richman and Shimkin [4], Boulogne et al. [5], and Korilis et al. [6–8]. Niyato and Hossain [9] studied the practical issue such as the admission control for the wireless broadband standard. Ganesh et al. [10] and Yaïche et al. [11] considered the pricing issues. In modeling the networking problems, quite often the bandwidth availability is the main concern and has its role in the associated performance measure. The network system quantifies the results caused by different operation scenarios and seeks the approach leading to the greatest benefits in some sense. Since the benefit of any network user inevitably involves that of other users, its evaluation is mostly carried out in the context of game theory. In the survey paper Altman et al. collected a long list of networking models based on game theoretic formulation. Interested readers are referred to [12] and the rich reference therein.

As mentioned above, the bandwidth availability is the major concern in many networking problems. The bandwidth allocation is thus the core issue as far as the quality of service is concerned. While the bandwidth allocation problem was considered by many authors in different contexts of networking protocols or communication standards (see e.g., [9, 11, 13, 14]), at a high level of abstraction the problem can be regarded as the classical resources distribution problem studied in many professional fields such as economics, management science, and operations research. Given a finite number of units competing for the limited resources, how does each unit decide its share based on its own utility function? Lazar et al. [15] formulated this problem for the network composed of noncooperative users. Under certain monotonicity, differentiability, and convexity assumptions on the cost function the unique existence and certain fairness property of the Nash equilibrium point (NEP) were proved. An algorithm based on

Gauss-Seidel and Jacobi schemes was proposed and proved to yield a sequence converging to the NEP. However, the framework in [15] assumes only the natural constraint for the bandwidth allocation. That is, the feasible bandwidth of any user falls within the interval lower bounded by zero and upper bounded by the bandwidth available for that user. In practice some techniques such as the bandwidth throttling and bandwidth/traffic shaping [16] are available to provide more adaptive bandwidth control. To address this issue, Rhee and Konstantopoulos [17, 18] relaxed the assumption and allowed some prespecified numbers for the upper and lower bounds of the bandwidth. This relaxation increases the flexibility of flow control and helps the QoS satisfaction by the networking system. On the other hand, in modern broadband communication systems some users might form a group and expect the group-wise QoS, in addition to the user-wise one. For example, the customers of an internet service provider (ISP) might include the individuals and a company with many employees. To maintain the QoS, parameters would be assigned to bound the bandwidth of each individual and each employee. Furthermore the ISP and the company would set the constraint for the employee-averaged, or equivalently, employee-totaled bandwidth, as shown in Figure 1. A similar concept of group constraint can be seen in the cost-effective broadband access network such as the Ethernet-based passive optical network (EPON) [19, 20]. This system is composed of an optical line terminal (OLT) and many optical network units (ONUs). The OLT is situated in the central office and ONUs are distributed over the remote areas for multimedia communication with the subscribers. In the upload process each ONU adopts the time division multiplexing access (TDMA) protocol to transmit data frames to the OLT, in the sense that each ONU only transmits the data during the time slots specifically scheduled for it [21]. The protocol avoids the frame collisions between different ONUs at the cost of imposing the upper bound for the time-average flow of each ONU and the upper bound for the total flow of all ONUs.

In light of the bandwidth sharing mechanism in EPON and other similar systems, we extend the existing results to include the user-grouping constraint for better model fitting. Suppose each user has his or her own utility function that describes the relation between the allocated bandwidth and the resultant benefit to that user. Following the standard assumptions, (1) the function depends on the bandwidth of the user, and on the bandwidth of other users only through their total bandwidth, and (2) the function satisfies certain continuity and concavity properties; we show the unique existence and the fairness with some appropriate sense, of the NEP in the allocation game. The contributions of our work are twofold. First, a novel concept called user-grouping NEP is proposed. This concept is corresponding to the new introduction of the group constraint, under which the uniqueness of NEP proved in [15, 17] no longer holds. Based on this concept we give a new definition for the equilibrium point and prove its uniqueness under our assumptions. The fairness of the allocation based on the user-grouping NEP is also proved. Second, we show that the Gauss-Seidel type algorithm in [15, 18] can be modified to yield a sequence converging to the user-grouping NEP. Since the bandwidth allocation is of central concern in networking systems, our results might result in the reinvestigation and reformulation of other networking issues such that more practical approaches can be developed.

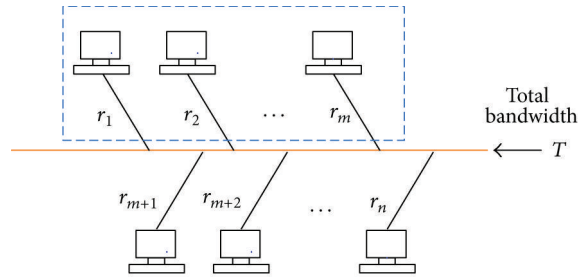

Figure 1: The system bandwidth allocation $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ subject to user-wise constraints: $\underline{r}_i \le r_i \le \overline{r}^i$ for $i \in \{1, 2, \ldots, n\}$, and a user-grouping constraint $\underline{R}_g \le \sum_{i=1}^{m} r_i \le \overline{R}_g$. The total allocation $R = \sum_{i=1}^{n} r_i$ is less than T, the total bandwidth.

2. Preliminaries

Suppose a networking system has n users and they compete for the system bandwidth. Each user is assigned with the bandwidth subject to predecided upper and lower bounds. In addition, m of the n users form a group and the total bandwidth of the m users is also subject to a predecided constraint. We would like to design the system bandwidth allocation policy that optimizes the performance index of each user in the game-theoretical sense. For convenience, we use the list of nomenclature shown at the end of the paper.

Assume the utility function for each user depends on the bandwidth of that user and the total bandwidth of other users. That is, the utility function for user i in $\mathcal{N}$ can be written as $U_i(r_i, R)$. Also, assume the utility function satisfies the following continuity and concavity properties [18].

Assumption 1. For each utility function $U_i(r_i, R)$

(a) $U_i(r_i, R)$ is continuously differentiable with respect to r_i;

(b) $(\partial/\partial r_i)U_i(r_i, R)$ is strictly decreasing with respect to r_i and nonincreasing with respect to R.

Now we define the allocation function. For user i in $\mathcal{N}\backslash\mathcal{M}$ with the available bandwidth T_i, the allocation function A_i is defined as

$$A_i(T_i) = \arg\max_{\underline{r}_i \le r \le \overline{r}_i} U_i(r, r + T - T_i). \tag{1}$$

For user i in $\mathcal{M}$ with the available bandwidth T_i and inside-group information T_i^g, the allocation function $\mathcal{A}_i$ is defined as

$$\mathcal{A}_i(T_i, T_i^g) = \arg\max_{\underline{r}_i \le r \le \overline{r}_i,\ \underline{R}_g \le T_i^g + r \le \overline{R}_g} U_i(r, r + T - T_i). \tag{2}$$

Assumption 2. The bandwidth allocation with the constraint parameters has the following properties:

(a) $T_i > \mathscr{A}_i(T_i, T_i^g)$ for each feasible T_i and T_i^g where $i \in \mathscr{M}$, and $T_i > A_i(T_i)$ for each feasible T_i where $i \in \mathscr{N} \setminus \mathscr{M}$;

(b) $\sum_{i=1}^m \overline{r}_i > \overline{R}_g > \underline{R}_g > \sum_{i=1}^m \underline{r}_i$.

In our framework, the classical Nash equilibrium point (NEP) in the allocation game is defined as

$$\mathbf{r}^* := \left(r_1^*, r_2^*, \ldots, r_n^*\right), \tag{3}$$

where

$$r_i^* \in \arg\max_{r \in C_i} U_i\left(r, r + \sum_{j \neq i} r_j^*\right). \tag{4}$$

Note that C_i is $\{r \mid \underline{r}_i \le r \le \overline{r}_i, \underline{R}_g \le r + \sum_{j \in \mathscr{M} \setminus \{i\}} r_j^* \le \overline{R}_g\}$ if $i \in \mathscr{M}$ and is $\{r \mid \underline{r}_i \le r \le \overline{r}_i\}$ otherwise. The *user-grouping* NEP is defined as

$$\mathbf{r}_g^* := \left(\frac{R_g^*}{m}, r_{m+1}^*, r_{m+2}^*, \ldots, r_n^*\right), \tag{5}$$

where

$$R_g^* := \sum_{i=1}^m r_i^*. \tag{6}$$

Remark 3. The definitions (3)-(4) reflect the central concept of the well-studied constrained NEP. That is, given the constrained strategy space C_i for each user $i \in \mathscr{N}$, r_i^* is defined as the maximizer of the utility function of user i provided that the strategy r_j^* is adopted by user j for each $j \in \mathscr{N} \setminus \{i\}$. Note that the bandwidth of users in the group $\mathscr{M}$ should satisfy the extra group constraint and thus r_i^* might lose its uniqueness in C_i as the group constraint is active. The novel concept of user-grouping NEP in (5)-(6) is thus proposed to compensate the property of the equilibrium point. We will show in Section 3.1 that our setting ensures the uniqueness of the user-grouping NEP.

Remark 4. Suppose $T_i^* := T - \sum_{j \neq i}^n r_j^*$. Also, let $T_i^{g*} := \overline{R}_g - \sum_{j \neq i}^m r_j^*$, then

$$r_i^* := \begin{cases} \mathscr{A}_i\left(T_i^*, T_i^{g*}\right) & \text{if } i \in \mathscr{M} \\ A_i\left(T_i^*\right) & \text{if } i \in \mathscr{N} \setminus \mathscr{M}. \end{cases} \tag{7}$$

By part (a) in Assumption 2 we have

$$r_i^* < T_i^* = T - \sum_{j \neq i}^n r_j^*. \tag{8}$$

Consequently,

$$R^* := \sum_{i=1}^n r_i^* < T. \tag{9}$$

This means that the NEP, or $\mathbf{r}^*$, satisfies the natural constraint $R^* \le T$ and the constraint is always inactive. Part (b) in Assumption 2 is a natural condition such that the constraint $\underline{R}_g \le \sum_{i=1}^m r_i \le \overline{R}_g$ makes sense.

The existence of the NEP in our setting is guaranteed by Rosen's result in the following.

Theorem 5 (see [22, Theorem 1]). *An equilibrium point exists for every concave n-person game.*

Theorem 5 can be obtained using the classical Kakutani fixed point theorem and in some sense generalizes Nash's setting on the strategy space of the users [23, 24]. In the next section we delve into other properties and propose an algorithm to locate the NEP.

3. Main Results

3.1. Uniqueness. Our first result is concerned with the uniqueness of the user-grouping NEP. This property as shown in [18, page 13] is not implied by the uniqueness theorem in [22]. For the NEP $\mathbf{r}^* = (r_1^*, r_2^*, \ldots, r_n^*)$ defined in (3)-(4), the Karash-Kuhn-Tucker (KKT) conditions must be satisfied. That is, for each $i \in \mathscr{N} \setminus \mathscr{M}$ there exist KKT multipliers $\overline{\lambda}_i^*$ and $\underline{\lambda}_i^*$ (see e.g., [25, page 458]) such that

$$-\frac{\partial}{\partial r_i} U_i\left(r_i^*, R^*\right) + \overline{\lambda}_i^* - \underline{\lambda}_i^* = 0, \tag{10}$$

$$\overline{\lambda}_i^*\left(r_i^* - \overline{r}_i\right) = 0, \tag{11}$$

$$\underline{\lambda}_i^*\left(\underline{r}_i - r_i^*\right) = 0, \tag{12}$$

$$\underline{r}_i \le r_i^* \le \overline{r}_i, \tag{13}$$

$$\underline{\lambda}_i^* \ge 0, \qquad \overline{\lambda}_i^* \ge 0. \tag{14}$$

In addition, for each $i \in \mathscr{M}$ there exist KKT multipliers $\overline{\lambda}_i^*$ and $\underline{\lambda}_i^*$ satisfying (11)–(14), and $\overline{\gamma}_i^*$ and $\underline{\gamma}_i^*$ satisfying

$$-\frac{\partial}{\partial r_i} U_i\left(r_i^*, R^*\right) + \overline{\lambda}_i^* - \underline{\lambda}_i^* + \overline{\gamma}_i^* - \underline{\gamma}_i^* = 0, \tag{15}$$

$$\overline{\gamma}_i^*\left(R_g^* - \overline{R}_g\right) = 0, \tag{16}$$

$$\underline{\gamma}_i^*\left(\underline{R}_g - R_g^*\right) = 0, \tag{17}$$

$$\underline{R}_g \le R_g^* \le \overline{R}_g, \tag{18}$$

$$\underline{\gamma}_i^* \ge 0, \qquad \overline{\gamma}_i^* \ge 0, \tag{19}$$

where R_g^* is defined in (6).

Lemma 6. *For each $i \in \mathscr{N} \setminus \mathscr{M}$, let $r_i^{(1)} := A_i(T_i^{(1)})$ and $r_i^{(2)} := A_i(T_i^{(2)})$ for some feasible $T_i^{(1)}$ and $T_i^{(2)}$, then*

$$r_i^{(1)} > r_i^{(2)} \Longrightarrow \overline{\lambda}_i^{(1)} - \underline{\lambda}_i^{(1)} \ge \overline{\lambda}_i^{(2)} - \underline{\lambda}_i^{(2)}, \tag{20}$$

where the nonnegative KKT multipliers $\overline{\lambda}_i^{(1)}$, $\underline{\lambda}_i^{(1)}$, $\overline{\lambda}_i^{(2)}$ *and* $\underline{\lambda}_i^{(2)}$ *satisfy*

$$\overline{\lambda}_i^{(1)}\left(r_i^{(1)}-\overline{r}_i\right)=0,\qquad \underline{\lambda}_i^{(1)}\left(\underline{r}_i-r_i^{(1)}\right)=0, \tag{21}$$

$$\overline{\lambda}_i^{(2)}\left(r_i^{(2)}-\overline{r}_i\right)=0,\qquad \underline{\lambda}_i^{(2)}\left(\underline{r}_i-r_i^{(2)}\right)=0. \tag{22}$$

Proof. Since $r_i^{(1)}$ and $r_i^{(2)}$ are both feasible, $r_i^{(1)} > r_i^{(2)}$ implies $\underline{\lambda}_i^{(1)} = 0$ and $\overline{\lambda}_i^{(2)} = 0$ by (21) and (22), respectively. Consequently, the result follows since $\overline{\lambda}_i^{(1)} \geq 0$ and $\underline{\lambda}_i^{(2)} \geq 0$. □

Theorem 7. *The user-grouping NEP defined in* (5) *is unique.*

Proof. Suppose $\mathbf{r}^{(1)} = (r_1^{(1)}, r_2^{(1)}, \ldots, r_n^{(1)})$ and $\mathbf{r}^{(2)} = (r_1^{(2)}, r_2^{(2)}, \ldots, r_n^{(2)})$ are both the equilibrium points. Let $R^{(1)} = \sum_{i=1}^n r_i^{(1)}$ and $R^{(2)} = \sum_{i=1}^n r_i^{(2)}$. We can thus write for $i \in \mathcal{N} \setminus \mathcal{M}$ that

$$\begin{aligned}-\frac{\partial}{\partial r_i}U_i\left(r_i^{(1)},R^{(1)}\right)+\overline{\lambda}_i^{(1)}-\underline{\lambda}_i^{(1)}&=0,\\ -\frac{\partial}{\partial r_i}U_i\left(r_i^{(2)},R^{(2)}\right)+\overline{\lambda}_i^{(2)}-\underline{\lambda}_i^{(2)}&=0,\end{aligned} \tag{23}$$

where $\overline{\lambda}_i^{(1)}$, $\underline{\lambda}_i^{(1)}$, $\overline{\lambda}_i^{(2)}$, $\underline{\lambda}_i^{(2)}$ are the associated KKT multipliers. Assume that $R^{(1)} > R^{(2)}$. If there exists $i \in \mathcal{N} \setminus \mathcal{M}$ such that $r_i^{(1)} > r_i^{(2)}$, by Lemma 6 and Assumption 1 we have

$$\begin{aligned}&-\frac{\partial}{\partial r_i}U_i\left(r_i^{(1)},R^{(1)}\right)+\overline{\lambda}_i^{(1)}-\underline{\lambda}_i^{(1)}\\ &\quad>-\frac{\partial}{\partial r_i}U_i\left(r_i^{(2)},R^{(2)}\right)+\overline{\lambda}_i^{(2)}-\underline{\lambda}_i^{(2)},\end{aligned} \tag{24}$$

which is a contradiction. We thus have $r_i^{(1)} \leq r_i^{(2)}$ for i in $\mathcal{N} \setminus \mathcal{M}$. This implies $R_g^{(1)} := \sum_{i=1}^m r_i^{(1)} > \sum_{i=2}^m r_i^{(2)} = R_g^{(2)}$. Note that, for $i \in \mathcal{M}$,

$$\begin{aligned}-\frac{\partial}{\partial r_i}U_i\left(r_i^{(1)},R^{(1)}\right)+\overline{\lambda}_i^{(1)}-\underline{\lambda}_i^{(1)}+\overline{\gamma}_i^{(1)}-\underline{\gamma}_i^{(1)}&=0,\\ -\frac{\partial}{\partial r_i}U_i\left(r_i^{(2)},R^{(2)}\right)+\overline{\lambda}_i^{(2)}-\underline{\lambda}_i^{(2)}+\overline{\gamma}_i^{(2)}-\underline{\gamma}_i^{(2)}&=0.\end{aligned} \tag{25}$$

With similar arguments in proving Lemma 6 we can show that

$$R_g^{(1)} > R_g^{(2)} \Longrightarrow \overline{\gamma}_i^{(1)}-\underline{\gamma}_i^{(1)} \geq \overline{\gamma}_i^{(2)}-\underline{\gamma}_i^{(2)}. \tag{26}$$

If $r_i^{(1)} > r_i^{(2)}$ for some $i \in \mathcal{M}$, we have

$$\begin{aligned}&-\frac{\partial}{\partial r_i}U_i\left(r_i^{(1)},R^{(1)}\right)+\overline{\lambda}_i^{(1)}-\underline{\lambda}_i^{(1)}+\overline{\gamma}_i^{(1)}-\underline{\gamma}_i^{(1)}\\ &\quad>-\frac{\partial}{\partial r_i}U_i\left(r_i^{(2)},R^{(2)}\right)+\overline{\lambda}_i^{(2)}-\underline{\lambda}_i^{(2)}+\overline{\gamma}_i^{(2)}-\underline{\gamma}_i^{(2)},\end{aligned} \tag{27}$$

which is a contradiction. We thus have $r_i^{(i)} \leq r_i^{(2)}$ for each $i \in \mathcal{M}$ and therefore $R_g^{(1)} \leq R_g^{(2)}$, also a contradiction. With analogous arguments we can show that assuming $R^{(1)} < R^{(2)}$ also leads to a contradiction. Therefore $R^{(1)} = R^{(2)}$ and by Lemma 6 $r_i^{(1)} = r_i^{(2)}$ for $i \in \mathcal{N} \setminus \mathcal{M}$. This implies $R_g^{(1)} = R_g^{(2)}$; therefore $\mathbf{r}_g^*$ is unique. □

3.2. Fairness. A bandwidth allocation $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ is said to be fair if for any feasible t_1 and t_2

$$\mathcal{A}_i(t_1,t_2) \geq \mathcal{A}_j(t_1,t_2) \Longrightarrow r_i \geq r_j, \tag{28}$$

where $i, j \in \mathcal{M}$, and for any feasible t

$$A_i(t) \geq A_j(t) \Longrightarrow r_i \geq r_j, \tag{29}$$

where $i, j \in \mathcal{N} \setminus \mathcal{M}$. This definition suggests that a fair allocation guarantees the user in greater need of bandwidth actually obtains more bandwidth.

Theorem 8. *The bandwidth allocation based on the NEP defined in* (3)-(4) *is fair.*

Proof. Suppose $i \in \mathcal{M}$. Let $r_i^{(1)} = \mathcal{A}_i(T_i^{(1)}, t)$ and $r_i^{(2)} = \mathcal{A}_i(T_i^{(1)} + \Delta T,\ t + \Delta T)$. We thus have

$$\begin{aligned}-\frac{\partial}{\partial r_i}U_i\left(r_i^{(1)},R^{(1)}\right)+\overline{\lambda}_i^{(1)}-\underline{\lambda}_i^{(1)}+\overline{\gamma}_i^{(1)}-\underline{\gamma}_i^{(1)}&=0,\\ -\frac{\partial}{\partial r_i}U_i\left(r_i^{(2)},R^{(2)}\right)+\overline{\lambda}_i^{(2)}-\underline{\lambda}_i^{(2)}+\overline{\gamma}_i^{(2)}-\underline{\gamma}_i^{(2)}&=0.\end{aligned} \tag{30}$$

Assume $r_i^{(2)} - \Delta T > r_i^{(1)}$, which implies $r_i^{(2)} > r_i^{(1)}$ and by Lemma 6 $\overline{\lambda}_i^{(2)} - \underline{\lambda}_i^{(2)} \geq \overline{\lambda}_i^{(1)} - \underline{\lambda}_i^{(1)}$. Also,

$$\begin{aligned}R^{(2)} &:= r_i^{(2)} + R_{-i}^{(2)} > r_i^{(1)} + \Delta T + R_{-i}^{(2)}\\ &= r_i^{(1)} + R_{-i}^{(1)} := R^{(1)}.\end{aligned} \tag{31}$$

Similarly,

$$\begin{aligned}R_g^{(2)} &:= r_i^{(2)} + \sum_{j\neq i}^m r_j^{(2)} > r_i^{(1)} + \Delta T + \sum_{j\neq i}^m r_j^{(2)}\\ &= r_i^{(1)} + \sum_{j\neq i}^m r_j^{(1)} := R_g^{(1)}.\end{aligned} \tag{32}$$

Note that (32) implies $\overline{\gamma}_i^{(2)} - \underline{\gamma}_i^{(2)} \geq \overline{\gamma}_i^{(1)} - \underline{\gamma}_i^{(1)}$. We then have

$$\begin{aligned}&-\frac{\partial}{\partial r_i}U_i\left(r_i^{(2)},r_i^{(2)}+R_{-i}^{(2)}\right)+\overline{\lambda}_i^{(2)}-\underline{\lambda}_i^{(2)}+\overline{\gamma}_i^{(2)}-\underline{\gamma}_i^{(2)}\\ &\quad>-\frac{\partial}{\partial r_i}U_i\left(r_i^{(1)},r_i^{(1)}+R_{-i}^{(1)}\right)+\overline{\lambda}_i^{(1)}-\underline{\lambda}_i^{(1)}+\overline{\gamma}_i^{(1)}-\underline{\gamma}_i^{(1)},\end{aligned} \tag{33}$$

which is a contradiction. Therefore, $r_i^{(2)} - \Delta T \leq r_i^{(1)}$, which implies

$$T_i^{(1)} + \Delta T - r_i^{(2)} \geq T_i^{(1)} - r_i^{(1)}, \tag{34}$$

namely,

$$T_i^{(1)} + \Delta T - \mathscr{A}_i\left(T_i^{(1)} + \Delta T, t + \Delta T\right) \geq T_i^{(1)} - \mathscr{A}_i\left(T_i^{(1)}, t\right). \tag{35}$$

To show the allocation based on $\mathbf{r}^*$ is fair, consider first the case that $i, j \in \mathscr{M}$. By definition $r_i^* = \mathscr{A}_i(T_i^*, T_i^{g*})$, $r_j^* := \mathscr{A}_j(T_j^*, T_j^{g*})$ where T_i^*, T_i^{g*}, T_j^* and T_j^{g*} are all feasible. With R^* defined in (9) we have

$$T - R^* = T_i^* - \mathscr{A}_i\left(T_i^*, T_i^{g*}\right) = T_j^* - \mathscr{A}_j\left(T_j^*, T_i^{g*}\right), \tag{36}$$

which equals the remaining bandwidth of the system after allocation. Given $\mathscr{A}_i(T_i^*, T_i^{g*}) > \mathscr{A}_j(T_i^*, T_i^{g*})$, we have

$$\begin{aligned} T_i^* - \mathscr{A}_j\left(T_i^*, T_i^{g*}\right) &> T_i^* - \mathscr{A}_i\left(T_i^*, T_i^{g*}\right) \\ &= T_j^* - \mathscr{A}_j\left(T_j^*, T_j^{g*}\right). \end{aligned} \tag{37}$$

Note that

$$T_i^* - T_j^* = T_i^{g*} - T_j^{g*} = r_i^* - r_j^* := \Delta T^*. \tag{38}$$

Equations (35) and (37) imply $\Delta T^* > 0$; hence $r_i^* > r_j^*$. The case for $i, j \in \mathscr{N} \setminus \mathscr{M}$ can be similarly proved and is thus ignored (see [18, Theorem 2.3] for details). □

3.3. Algorithm. In this section we analyze the scheme to identify the user-grouping Nash equilibrium point. We say an *individual update* is implemented on user i if the bandwidth of each user other than i is fixed and the bandwidth of user i is updated to maximize his or her utility function. In addition, we say a *batch update* occurs in the collection $\mathscr{K}$ of users if the bandwidth of each user not in $\mathscr{K}$ is fixed and the individual update is sequentially implemented on each user in $\mathscr{K}$ repeatedly till an equilibrium is reached. Here $\mathscr{K}$ is either $\mathscr{M}$ or $\mathscr{N} \setminus \mathscr{M}$. If $\mathscr{K} = \mathscr{M}$ the batch update is implemented assuming no group constraint, namely, $\overline{R}_g = \infty$ and $\underline{R}_g = 0$. Note that the batch update is guaranteed to reach an equilibrium (see [15] and [18, Section 2.4]). As a result, suppose $\mathbf{r}^k = (r_1^k, r_1^k, \ldots, r_n^k)$ is the system allocation at step k. If an individual update occurs at user $i \in \mathscr{N} \setminus \mathscr{M}$ at step $k+1$, that means $r_j^{k+1} = r_j^k$ for $j \in \mathscr{N} \setminus \{i\}$, and

$$-\frac{\partial}{\partial r_i} U_i\left(r_i^{k+1}, R^{k+1}\right) + \overline{\lambda}_i^{k+1} - \underline{\lambda}_i^{k+1} = 0, \tag{39}$$

for some KKT multipliers $\overline{\lambda}_i^{k+1}$ and $\overline{\lambda}_i^{k+1}$ and $R^{k+1} = \sum_{i=1}^n r_i^{k+1}$. If a batch update occurs at the group of users at step $k+1$, that means $r_j^{k+1} = r_j^k$ for each $j \in \mathscr{N} \setminus \mathscr{M}$, and we can write for each $i \in \mathscr{M}$

$$-\frac{\partial}{\partial r_i} U_i\left(r_i^{k+1}, R^{k+1}\right) + \overline{\lambda}_i^{k+1} - \underline{\lambda}_i^{k+1} + \overline{\gamma}_i^{k+1} - \underline{\gamma}_i^{k+1} = 0, \tag{40}$$

where $\overline{\lambda}_i^{k+1}$, $\overline{\lambda}_i^{k+1}$, $\overline{\gamma}_i^{k+1}$ and $\overline{\gamma}_i^{k+1}$ are the associated KKT multipliers. We now show that a repeated implementation of sequential batch updates on users in $\mathscr{M}$ and users in $\mathscr{N} \setminus \mathscr{M}$, as outline in Algorithm 1, yields a sequence converging to the user-grouping NEP $\mathbf{r}_g^*$ in (5). Define first the error measure $e(\mathbf{r}_g^k)$ between $\mathbf{r}_g^k$ and $\mathbf{r}_g^*$ as

$$e\left(\mathbf{r}_g^k\right) = \sum_{i=m+1}^n \left|r_i^k - r_i^*\right| + \left|R_g^k - R_g^*\right| + \left|R^k - R^*\right|, \tag{41}$$

where $R_g^k := \sum_{i=1}^m r_i^k$. R_g^* and R^* are defined in (6) and (9), respectively.

Lemma 9. *The error measure $e(\mathbf{r}_g^k)$ defined in* (41) *is nonincreasing, namely, $e(\mathbf{r}_g^{k+1}) \leq e(\mathbf{r}_g^k)$ for any positive integer k.*

Proof. If at step $k+1$ an individual update occurs at $i \in \mathscr{N} \setminus \mathscr{M}$, (39) is satisfied. Since (10) is also satisfied, we have by Lemma 6 and part (b) in Assumption 1 that

$$\begin{aligned} R^{k+1} \geq R^* &\Longrightarrow r_i^{k+1} \leq r_i^*, \\ R^{k+1} \leq R^* &\Longrightarrow r_i^{k+1} \geq r_i^*. \end{aligned} \tag{42}$$

Under the assumption

$$\begin{aligned} e\left(\mathbf{r}_g^{k+1}\right) - e\left(\mathbf{r}_g^k\right) &= \left|r_i^{k+1} - r^*\right| - \left|r_i^k - r^*\right| \\ &\quad + \left|R^{k+1} - R^*\right| - \left|R^k - R^*\right|. \end{aligned} \tag{43}$$

Suppose $R^{k+1} \geq R^*$. If $R^k \geq R^{k+1}$, then

$$\begin{aligned} e\left(\mathbf{r}_g^{k+1}\right) - e\left(\mathbf{r}_g^k\right) &= -\left(r_i^{k+1} - r^*\right) - \left|r_i^k - r^*\right| + R^{k+1} - R^k \\ &\Longrightarrow \begin{cases} = 0 & \text{if } r_i^k < r_i^* \\ < 0 & \text{o.w.} \end{cases} \end{aligned} \tag{44}$$

Since $R^k < R^{k+1}$ implies $r_i^k < r_i^{k+1}$, we have

$$\begin{aligned} e\left(\mathbf{r}_g^{k+1}\right) - e\left(\mathbf{r}_g^k\right) &= -r_i^{k+1} + r_i^k + R^{k+1} - R^* - \left|R^k - R^*\right| \\ &\Longrightarrow \begin{cases} = 0 & \text{if } R^k > R^* \\ < 0 & \text{o.w.} \end{cases} \end{aligned} \tag{45}$$

Using similar arguments we can analyze the case for $R^{k+1} < R^*$ and obtain also that $e(\mathbf{r}_g^{k+1}) \leq e(\mathbf{r}_g^k)$. If at step $k+1$ a batch update occurs, (15) and (40) hold for each $i \in \mathscr{M}$. Suppose $R^{k+1} \geq R^*$. If there exists an $i \in \mathscr{M}$ such that $r_i^{k+1} > r_i^*$ then part (b) in Assumption 1 implies

$$\overline{\gamma}_i^{k+1} - \underline{\gamma}_i^{k+1} < \overline{\gamma}_i^* - \underline{\gamma}_i^*; \tag{46}$$

therefore by (26) $R_g^* \geq R_g^{k+1}$. If no such i exists, namely, $r_i^{k+1} \leq r^*$ for each $i \in \mathscr{M}$, then naturally $R_g^{k+1} \leq R_g^*$. A similar result can be derived for the case $R^{k+1} \leq R^*$. We then have

$$R^{k+1} \geq R^* \Longrightarrow R_g^{k+1} \leq R_g^*, \tag{47}$$

$$R^{k+1} \leq R^* \Longrightarrow R^{k+1} \geq R_g^*. \tag{48}$$

Table 1: The parameters of α_i for $i \in \mathcal{N} = \{1, 2, \ldots, 100\}$.

i	α_i									
1–10	.6111	.7052	.0283	.5209	.2591	.5082	.9732	.8182	.8588	.8966
11–20	.2995	.8800	.2552	.1921	.1866	.8611	.4446	.3866	.8320	.2065
21–30	.8978	.4595	.1634	.2687	.5207	.7724	.3874	.0444	.9868	.5230
31–40	.8437	.8919	.6856	.5178	.4420	.3801	.0318	.2159	.0853	.7969
41–50	.4018	.0498	.3165	.8712	.9314	.4257	.8662	.4806	.9083	.2505
51–60	.5266	.2448	.8384	.5257	.3498	.2571	.1688	.5992	.8110	.8489
61–70	.8448	.2350	.5083	.6694	.0907	.5430	.2980	.3818	.5339	.4265
71–80	.2320	.0284	.0260	.2743	.6945	.3300	.7505	.7957	.0537	.0092
81–90	.7281	.6644	.5853	.9190	.1673	.9720	.2196	.0382	.7078	.2290
91–100	.3526	.7095	.5887	.1088	.6468	.7311	.8634	.0226	.5252	.1512

Now

$$e\left(\mathbf{r}_g^{k+1}\right) - e\left(\mathbf{r}_g^{k}\right) = \left|R_g^{k+1} - R_g^*\right| - \left|R_g^{k} - R_g^*\right| + \left|R^{k+1} - R^*\right| - \left|R^{k} - R^*\right|. \tag{49}$$

Assume $R^{k+1} \geq R^*$. If $R^k \geq R^{k+1}$ then

$$e\left(\mathbf{r}_g^{k+1}\right) - e\left(\mathbf{r}_g^{k}\right) = -\left(R_g^{k+1} - R_g^*\right) - \left|R_g^{k} - R_g^*\right| + R^{k+1} - R^k \Longrightarrow \begin{cases} = 0 & \text{if } R_g^k < R_g^* \\ < 0 & \text{o.w.} \end{cases} \tag{50}$$

If $R^k < R^{k+1}$, which is equivalent to $R_g^k < R_g^{k+1}$, then

$$e\left(\mathbf{r}_g^{k+1}\right) - e\left(\mathbf{r}_g^{k}\right) = -R_g^{k+1} + R_g^k + R^{k+1} - R^* - \left|R^k - R^*\right| \Longrightarrow \begin{cases} = 0 & \text{if } R^k > R^* \\ < 0 & \text{o.w.} \end{cases} \tag{51}$$

Applying similar arguments again we can analyze the case for $R^{k+1} < R^*$ and obtain also that $e(\mathbf{r}_g^{k+1}) \leq e(\mathbf{r}_g^k)$. □

Theorem 10. *Algorithm 1 yields a sequence $\{\mathbf{r}_g^k\}_{k=1}^{\infty}$ converging to $\mathbf{r}_g^*$, the user-grouping NEP of the bandwidth allocation game.*

Proof. In the light of Lemma 9, we only need to show that for each positive integer k, $\mathbf{r}_g^k \neq \mathbf{r}_g^*$ implies the existence of a finite integer k_1 such that $e(\mathbf{r}_g^{k+k_1}) < e(\mathbf{r}_g^k)$. Suppose it is not the case then there exists some k such that $e(\mathbf{r}_g^{k+k_2}) = e(\mathbf{r}_g^k)$ for any positive integer k_2, where $\mathbf{r}_g^k \neq \mathbf{r}_g^*$. Consider the update scheme that, at step $k + i - m$ for $i \in \mathcal{N} \setminus \mathcal{M}$, the individual update occurs at user i, and at step $k + n + 1 - m$ the batch update takes place for users in $\mathcal{M}$. Without loss of generality we assume $R^{k+n+1-m} \geq R^*$, then $R_g(k + n + 1 - m) \leq R_g^*$ by (47). Since $e(\mathbf{r}_g^{k+n+1-m}) = e(\mathbf{r}_g^{k+n-m})$, (50) together with (51) implies $R^{k+n-m} \geq R^*$; hence $R_g^{k+n-m} \leq R_g^*$. That is, $R^{\cdot} - R^*$ and $R_g^{\cdot} - R_g^*$ do not change their signs as the step number is increased from $k+n-m$ to $k+n+1-m$. Moreover, $R^{k+n-m} \geq R^*$ implies $r_n^{k+n-m} \leq r_n^*$. Since $e(\mathbf{r}_g^{k+n+1-m}) = e(\mathbf{r}_g^{k+i-m-1})$ for $i \in \mathcal{N} \setminus \mathcal{M}$, (44) and (45) together imply $R^{k+i-m} \geq R^*$ and thus $r_i^{k+i-m} \leq r_i^*$, for $i \in \mathcal{N} \setminus \mathcal{M}$. As a result, $r_i^{k+n-m} \leq r_i^*$ for $i \in \mathcal{N} \setminus \mathcal{M}$. Note that $R_g^{k+n-m} \leq R_g^*$ and

$$R^{k+n-m} = R_g^{k+n-m} + \sum_{i=m+1}^{n} r_i^{k+n-m} \geq R^* = R_g^* + \sum_{i=m+1}^{n} r_i^*. \tag{52}$$

We conclude that $R_g^{k+n-m} = R_g^*$ and $r_i^{k+n-m} = r_i^*$ for $i \in \mathcal{N} \setminus \mathcal{M}$, namely, $\mathbf{r}_g^k = \mathbf{r}_g^*$, a contradiction. □

4. A Numerical Example

Consider a data communication network system with 100 users. Suppose 30 of them form a group. We then have $\mathcal{N} = \{1, 2, \ldots, 100\}$ and $\mathcal{M} = \{1, 2, \ldots, 30\}$. The adopted utility function is

$$U_i\left(r_i, R\right) = r_i^{\alpha_i}\left(T - R\right), \tag{53}$$

where the parameters α_i's are listed in Table 1. Note that the utility function, known as *the generalized power function* [26], has the continuity and concavity properties required by Assumption 1. In particular, it can be shown [18, page 11] easily that the maximizer

$$\arg\max_r U_i\left(r, R\right) = \frac{\alpha_i}{1 + \alpha_i} T_i < T_i, \tag{54}$$

and thus part (a) in Assumption 2 is satisfied. Suppose the total available bandwidth $T = 6000$, and the upper and lower bounds for total bandwidth allocated to the group is $\overline{R}_g = 2000$ and $\underline{R}_g = 800$, respectively. Assume that the individual bandwidth constraint for each user in Table 2 is used. Clearly these parameters satisfy the natural requirements of part (b) in Assumption 2. Applying Algorithm 1 yields a dynamic bandwidth allocation evolving with the implementation step, as shown in Figure 2. The left part of the figure shows the evolution of total bandwidth allocated to the group, which is

(P1) Initiate the algorithm with a feasible $\mathbf{r}_g$ and set $s = 0$.
(P2) (if current step is k) Perform a batch update for users in $\mathscr{M}$ at step $k+1$
by assigning $r_j^{k+1} = r_j^k$ for all $j \in \mathscr{N} \setminus \mathscr{M}$,
to reach $r_i^{k+1} = \mathscr{A}_i\left(T_i^{k+1}, T_i^{g,k+1}\right)$ for all $i \in \mathscr{M}$.
(P3) (if current step is $\widehat{k}$) Perform an individual update at user i in $\mathscr{N} \setminus \mathscr{M}$
by assigning $r_j^{\widehat{k}+1} = r_j^{\widehat{k}}$ for all $j \in \mathscr{N} \setminus \{i\}$,
to reach $r_i^{\widehat{k}+1} = A_i(T_i^{\widehat{k}})$, and let $\widehat{k} = \widehat{k} + 1$.
Sequentially implement this whole procedure till each user in $\mathscr{N} \setminus \mathscr{M}$ is updated.
(P4) Repeat (P3) to complete a batch update at users in $\mathscr{N} \setminus \mathscr{M}$.
(P5) Set $s = s + 1$ and record the current $\mathbf{r}_g$ as $\mathscr{R}(s)$.
If $\|\mathscr{R}(s) - \mathscr{R}(s-1)\| < \varepsilon$ then stops, otherwise go to (P2).

Algorithm 1: The scheme to locate the user-grouping NEP.

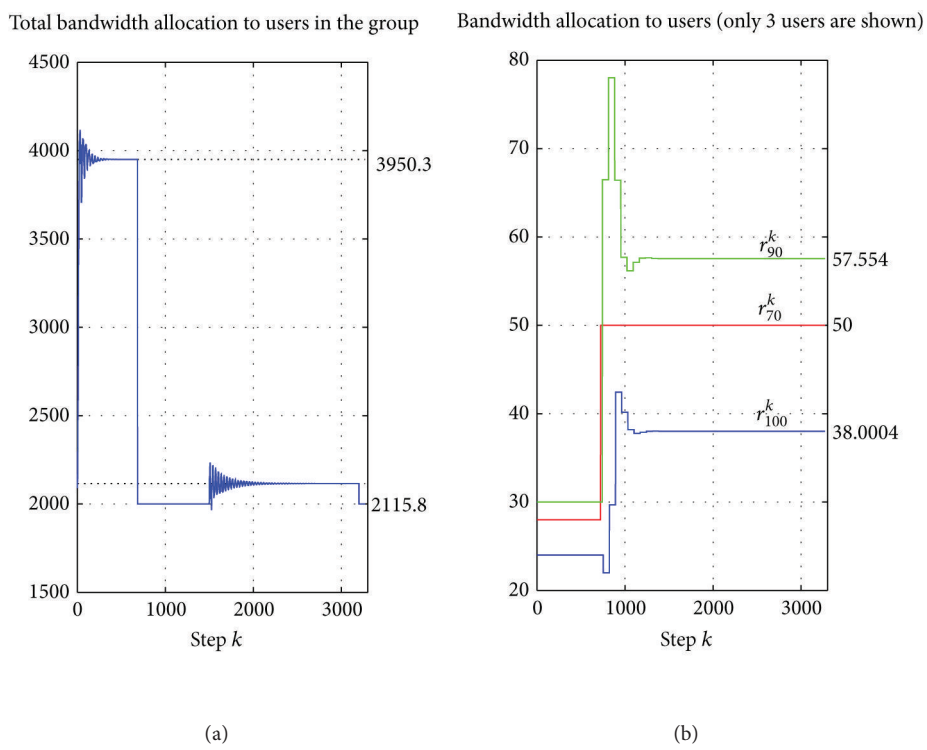

Figure 2: The sequence generated by Algorithm 1 for the example.

composed of user 1, user 2, up to user 30. At the beginning of the algorithm, an initial feasible bandwidth allocation is allocated to each user of the system. Fix the total bandwidth allocated to the users not in the group and find the optimal total bandwidth R_g. In the example R_g is 3950.3. Since this value is greater than the upper bound $\overline{R}_g = 2000$. R_g is replaced with $\overline{R}_g$. Fix this R_g we have the total available bandwidth $T - R_g = 6000 - 2000 = 4000$ for users not in the group. Based on this availability of the bandwidth we can find the equilibrium point for users not in the group. In the example we have, for instance, the bandwidth $r_{70} = 50$, $r_{90} = 57.554$, and $r_{100} = 38.0004$. Now fix the total bandwidth of the users not in the group, and find the optimal bandwidth R_g again. In the example we obtain $R_g = 2115.8$. Since this value is greater than the upper bound $\overline{R}_g = 2000$, R_g is replaced with $\overline{R}_g$ again. Note that the current optimal bandwidth allocation for each user outside the group is found based on the condition that the total bandwidth for the group is $\overline{R}_g$. The current $\overline{R}_g$ and $r_{m+1}, \ldots, r_n$ is thus the $\overline{R}_g^*$ and $r_{m+1}^*, \ldots, r_n^*$ for the user-grouping NEP in (5).

5. Conclusion

We have proposed a novel bandwidth allocation model based on game theory. The consideration of the user-grouping constraint distinguishes this model from the abundant ones

Table 2: The bandwidth constraint for each user.

User (i)	1–10	11–20	21–30	31–70	71–90	91–100
Upper bound ($\overline{r}_i$)	100	200	290	50	78	100
Lower bound ($\underline{r}_i$)	10	20	30	5	10	22

concerning similar allocation issues. Suppose each user competes for the system bandwidth resources and is granted with a constrained decision space. In particular, some users are united in one group and the total bandwidth allocated to the group is constrained as well. Given the appropriate constraint parameters and the utility function satisfying mild continuity and concavity conditions for each user, we have shown the unique existence of the user-grouping Nash equilibrium point for the allocation game. In addition, we have shown the fairness, in a proper sense, of the allocation based on this equilibrium point. Finally, we have proposed an iterative algorithm and proved that a sequence converging to the point can be generated by the algorithm. A practical example illustrating a network satisfying our settings has been given to show how the equilibrium point can be located successfully.

Nomenclature

$\mathcal{N}$: The index set for the users, that is, $\mathcal{N} := \{1, 2, \ldots, n\}$

$\mathcal{M}$: The index set for the users in the group, that is, $\mathcal{M} := \{1, 2, \ldots, m\}$

$\mathcal{N} \setminus \mathcal{M}$: The index set for the users not in the group, that is, $\mathcal{N} \setminus \mathcal{M} := \{m, m+1, \ldots, n\}$

r_i: The bandwidth allocated to user i

$\underline{r}_i$: Lower bound for r_i

$\overline{r}_i$: Upper bound for r_i

R_g: Total bandwidth allocated to the group members, that is, $R_g := r_1 + r_2 + \cdots + r_m$

$\underline{R}_g$: Lower bound for R_g

$\overline{R}_g$: Upper bound for R_g

T: Total bandwidth available in the system

R: Total bandwidth allocated, that is, $R := r_1 + r_2 + \cdots + r_n$

R_{-i}: Total bandwidth allocated, excluding to user i, that is, $R_{-i} := r_1 + \cdots + r_{i-1} + r_{i+1} + \cdots + r_n$

T_i: Total bandwidth available for user i, that is, $T_i := T - R_{-i}$

T_i^g: Inside-group information for user $i \in \mathcal{M}$, that is, $T_i^g := \overline{R}_g - \sum_{j \neq i}^{m} r_j$.

Acknowledgment

This research was supported by the National Science Council of Taiwan under Grant NSC 100-2221-E-005-071.

References

[1] J. Anselmi, U. Ayesta, and A. Wierman, "Competition yields efficiency in load balancing games," *Evaluation*, vol. 68, no. 11, pp. 986–1001, 2011.

[2] U. Ayesta, O. Brun, and B. J. Prabhu, "Price of anarchy in noncooperative load balancing games," *Performance Evaluation*, vol. 68, no. 12, pp. 1312–1332, 2011.

[3] R. J. La and V. Anantharam, "Optimal routing control: repeated game approach," *IEEE Transactions on Automatic Control*, vol. 47, no. 3, pp. 437–450, 2002.

[4] O. Richman and N. Shimkin, "Topological uniqueness of the nash equilibrium for selfish routing with atomic users," *Mathematics of Operations Research*, vol. 32, no. 1, pp. 215–232, 2007.

[5] T. Boulogne, E. Altman, H. Kameda, and O. Pourtallier, "Mixed equilibrium (ME) for multiclass routing games," *IEEE Transactions on Automatic Control*, vol. 47, no. 6, pp. 903–916, 2002.

[6] Y. A. Korilis, A. A. Lazar, and A. Orda, "Architecting noncooperative networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1241–1251, 1995.

[7] Y. A. Korilis, A. A. Lazar, and A. Orda, "Achieving network optima using Stackelberg routing strategies," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 161–173, 1997.

[8] Y. A. Korilis, A. A. Lazar, and A. Orda, "Capacity allocation under noncooperative routing," *IEEE Transactions on Automatic Control*, vol. 42, no. 3, pp. 309–325, 1997.

[9] D. Niyato and E. Hossain, "QoS-aware bandwidth allocation and admission control in IEEE 802.16 broadband wireless access networks: a non-cooperative game theoretic approach," *Computer Networks*, vol. 51, no. 11, pp. 3305–3321, 2007.

[10] A. Ganesh, K. Laevens, and R. Steinberg, "Congestion pricing and noncooperative games in communication networks," *Operations Research*, vol. 55, no. 3, pp. 430–438, 2007.

[11] H. Yaïche, R. R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 667–678, 2000.

[12] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, "A survey on networking games in telecommunications," *Computers & Operations Research*, vol. 33, no. 2, pp. 286–311, 2006.

[13] F. Bonomi and K. W. Fendick, "Rate -based flow control framework for the available bit rate ATM service," *IEEE Network*, vol. 9, no. 2, pp. 25–39, 1995.

[14] D. Niyato and E. Hossain, "Queue-aware uplink bandwidth allocation and rate control for polling service in IEEE 802.16 broadband wireless networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 6, pp. 668–679, 2006.

[15] A. A. Lazar, A. Orda, and D. E. Pendarakis, "Virtual path bandwidth allocation in multiuser networks," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 861–871, 1997.

[16] J. W. Evans and C. Filsfils, *Deploying IP and MPLS QoS for Multiservice Networks: Theory & Practice*, Morgan Kaufmann, Boston, Mass, USA, 1st edition, 2007.

[17] S. H. Rhee and T. Konstantopoulos, "Decentralized optimal flow control with constrained source rates," *IEEE Communications Letters*, vol. 3, no. 6, pp. 188–190, 1999.

[18] S. H. Rhee, *Optimal flow control and bandwidth allocation in multiservice networks: decentralized approaches [Ph.D. thesis]*, The University of Texas at Austin, 1999.

[19] C. M. Assi, Y. Ye, S. Dixit, and M. A. Ali, "Dynamic bandwidth allocation for quality-of-service over ethernet PONs," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 9, pp. 1467–1477, 2003.

[20] J. Xie, S. Jiang, and Y. Jiang, "A dynamic bandwidth allocation scheme for differentiated services in EPONs," *IEEE Communications Magazine*, vol. 42, no. 8, pp. 32–39, 2004.

[21] S. I. Choi and J. D. Huh, "Dynamic bandwidth allocation algorithm for multimedia services over ethernet PONs," *ETRI Journal*, vol. 24, no. 6, pp. 465–468, 2002.

[22] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, pp. 520–534, 1965.

[23] J. F. Nash, "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 36, pp. 48–49, 1950.

[24] J. Nash, "Non-cooperative games," *Annals of Mathematics, Second Series*, vol. 54, pp. 286–295, 1951.

[25] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*, Wiley-Interscience, Hoboken, NJ, USA, 3rd edition, 2008.

[26] A. Bovopoulos and A. Lazar, "Decentralized algorithms for optimal flow control," in *Proceedings of the 25th Allerton Conference on Communication, Control, and Computing*, pp. 979–988, 1987.