17 Aug 2007

A Supply Chain Design Model with Unreliable Supply

Lian Qi

Missouri University of Science and Technology

Zuo-Jun Max Shen

Recommended Citation

Qi, L., & Shen, Z. M. (2007). A Supply Chain Design Model with Unreliable Supply. *Naval Research Logistics* Wiley.

The definitive version is available at https://doi.org/10.1002/nav.20255

WILEY InterScience® DISCOVER SOMETHING GREAT

A Supply Chain Design Model With Unreliable Supply

Lian Qi,[1] Zuo-Jun Max Shen[2]

[1] *Department of Business Administration, University of Missouri, Rolla, Missouri*

[2] *Department of Industrial Engineering and Operations Research, University of California, Berkeley, California*

Received 6 August 2006; revised 19 June 2007; accepted 30 June 2007
DOI 10.1002/nav.20255
Published online 17 August 2007 in Wiley InterScience (www.interscience.wiley.com).

Abstract: Uncertainties abound within a supply chain and have big impacts on its performance. We propose an integrated model for a three-tiered supply chain network with one supplier, one or more facilities and retailers. This model takes into consideration the unreliable aspects of a supply chain. The properties of the optimal solution to the model are analyzed to reveal the impacts of supply uncertainty on supply chain design decisions. We also propose a general solution algorithm for this model. Computational experience is presented and discussed. © 2007 Wiley Periodicals, Inc. Naval Research Logistics 54: 829–844, 2007

Keywords: supply chain design; disruption; nonlinear optimization

1. INTRODUCTION

A supply chain network is a complex system where many uncertainties exist, such as customer demands, yields of the suppliers, and defect rates of the products. Since most supply chain design decisions (for example, facility location) are very expensive to change, it is not reasonable to use deterministic parameters to describe these uncertainties when designing a supply chain.

In this article, we study a supply chain design problem with multiple retailers and unreliable supply. Our model builds upon recent developments of integrated supply chain design models that simultaneously consider location, shipment, and inventory decisions in the same model [7, 17, 19], while taking into consideration the unreliable nature of the supply side.

Specifically, we consider the following multiperiod problem: in each period, multiple retailers order a specific product from a supplier, and the supplier ships the product to some intermediate facilities selected from a set of candidate locations. Each retailer can be served by more than one facility (see Fig. 1). Because of the benefit of economies of scale in transportation associated with bulk shipments from the supplier to the facilities, some amount of inventory may be kept at these facilities. Furthermore, some assembly and packaging activities may be performed at these facilities to satisfy orders from different retailers. However, the amount of final product delivered on time to a retailer may not be exactly the amount requested. This is because of the quality issues resulting from different production/assembly capabilities in different facilities, mistakes made during the assembly/packaging operations, and damages caused by loading and transportation. Most research in the supply chain design literature ignores this portion of loss, or simply uses a constant parameter instead of a random variable to model it. We model the amount of goods delivered from a facility to a retailer using the product of the order quantity from the retailer and a random variable associated with the facility, which we call the reliability coefficient. This method is prevalent in the random yield literature (see [2] for an example). We assume that the supplier is perfectly reliable (100% reliable) to simplify the problem. We show in the conclusion of this article that the problem with unreliable suppliers can be treated in a similar manner.

The retailers sell the product at different prices, and the objective of our model is to maximize the expected annual profit for the whole supply chain. We want to point out that minimizing total cost has been the primitive objective in most of the supply chain network design models. These models typically require that every potential demand has to be satisfied. However, for a profit-maximizing business, it may not always be optimal to satisfy all potential demands, especially if the additional cost is higher than the additional revenue associated with servicing some customers. In our setting, we may not want to satisfy some retailer's demand if the facilities

Correspondence to: Zuo-Jun Max Shen (Shen@ieor.berkeley.edu)

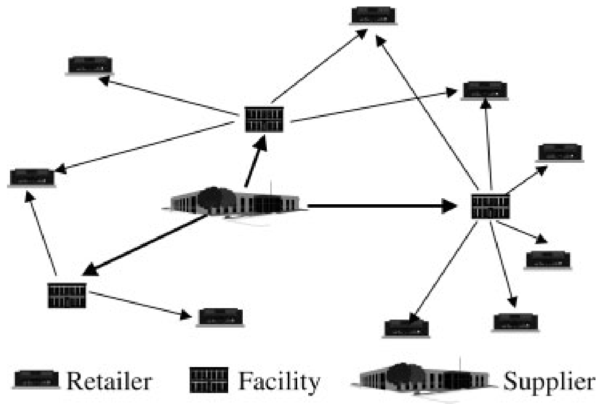

Figure 1. The three-tiered supply chain studied in the article.

that may serve this retailer either have very low reliability coefficients or very high shipping costs to the retailer. Most supply chain related costs, such as transportation and inventory costs of an item depend on total demand volume, and no clear method exists for determining a customer's profitability a priori, based solely on the characteristics of this customer. To address this problem, we formulate a model that takes the unreliability issue into consideration when answering the following questions: To maximize the total expected profit, (1) how many and which facilities should be opened from the candidate set? (2) how should retailers be served by open facilities? (3) what is the inventory policy at each open facility? Note that the facilities are heterogenous. For instance, one type of facility has very high reliability coefficient, but the related costs (setup cost, transportation cost) are high too; another type of facility has low reliability coefficient, but the related costs are low. Which type of facility is preferred? Our model can be used to answer such kinds of questions. Specifically, we study the impacts of reliability and cost factor on the optimal retailer assignments in Sections 4 and 6.

Another contribution of this article is the development of an efficient solution algorithm for a nonlinear optimization problem. Although this problem arises from the subproblem of our integrated supply chain design model, the proposed algorithm can be applied to a family of nonlinear optimization problems in which the objective function to be minimized can be written as the difference between two increasing and differentiable functions of the decision variables. The convergence of the algorithm is also proved.

The rest of this article is organized as follows. In Section 2, we review the related literature. Then in Section 3 we propose an integrated supply chain design model considering unreliable supply. We analyze the properties of the optimal solution of this model in Section 4, develop a solution algorithm for the model in Section 5, and conduct numerical studies in Section 6. We conclude the article in Section 7. The proofs for this article can be found in the Appendix.

2. LITERATURE REVIEW

The study by Parlar and Berkin [12] is among the earliest works that incorporate supply disruptions into the classical inventory models. They analyze the supply uncertainty problem for a class of EOQ models, where the supply is available only during an interval of random length, and then unavailable for another interval of random length. By using the renewal reward theorem, they construct an average cost objective function and find the optimal value of the order quantity. Their work is based on two main assumptions. First, at any time, the decision maker knows the availability status of the product; second, the retailer follows a zero-inventory ordering policy. Their cost function is shown to be incorrect in two respects by Berk and Arreola-Risa [1], who propose a corrected cost function.

Snyder [20] develops a simple but tight approximation for the model introduced by Berk and Arreola-Risa [1], whose cost function cannot be minimized in closed form, and has not been proven to be convex. Snyder shows in his work that his approximate cost function not only is convex but also yields a closed-form solution and behaves similarly to the classical EOQ cost function.

Parlar and Perry [13] relax the two assumptions made in [12]. First, their model can deal with the case where the decision maker is not aware of the ON–OFF status of the supply, and is assumed to identify the state if and only if an order is placed. Second, instead of a zero-inventory-ordering policy, the model has the reorder point as one of the decision variables. They consider both deterministic and random yields of the supplier.

Besides the uncertainty of supply, Gupta [9] assumes that the unit demands are generated according to a Poisson process and the shortages result in lost sales. A more general model is studied by Parlar [14], where the lead time is assumed to be stochastic.

The above works assume there is only one supplier. Tomlin [23] presents a dual-sourcing model in which orders may be placed with either a cheap but unreliable supplier or an expensive but reliable supplier. In this work, he also generalizes the stochastic recovery process of the supplier to be any member of the decreasing mean residual life family, instead of the exponential or constant assumption frequently made. Sheffi [15] also discusses a dual-sourcing problem. lie only presents a small illustrative example, while no analytical formulation is provided.

More generally, Dada et al. [6] consider a newsvendor problem in which the retailer is served by multiple suppliers, and any given supplier may be unreliable. They develop a modeling framework in which the newsvendor can diversify the risk of inadequate delivery amounts by spreading orders among any number and combination of available suppliers who differ in terms of cost and reliability. Then, they establish properties of the optimal solution and obtain corresponding insights into the tradeoff between cost and reliability.

All the above papers only consider problems in which there is one retailer. In recent years, the integrated supply chain design problem with multiple retailers is being widely studied. Erlebacher and Meller [8] formulate a highly nonlinear integer location/inventory model. They attack the problem by using a continuous approximation as well as a number of construction and bounding heuristics. Computation times on a 600 node problem using an exchange heuristic averaged 117 h on a Sun Ultra Sparcstation. Shen [16], Shen et al. [17], and Daskin et al. [7] study the joint location/inventory model in which location, shipment, and nonlinear safety stock inventory costs are included in the same model. They develop an integrated approach to determine the number of DCs to establish, the location of the DCs, and the magnitude of inventory to maintain at each center. Shen et al. [17] use column generation whereas Daskin et al. [7] apply Lagrangian relaxation to solve this problem. They use a low-order polynomial algorithm for the nonlinear integer programming subproblem that must be solved in either of the two approaches. Instead of using the more complex routing costs, they assume linear direct shipment costs from a DC to the customers it serves. They also assume that the variance of demand of customer i, σ_i^2, is proportional to the mean demand of customer i, μ_i, and that the proportionality constant is the same for each customer i. In other words, they assume $\mu_i = \gamma\sigma_i^2 \ \forall i \in I$. This allows them to reduce the number of nonlinear terms in the objective function from two to one. Recently, Shu et al. [19] extended the above works by relaxing this assumption on demand.

On the basis of these works, we consider an integrated supply chain design problem with unreliable supply. The following are some recent works on supply chain design with unreliable supply.

Snyder and Daskin [21] consider facility location models where some facilities will fail with a given probability. They present two reliability models that are based on the classical P-median problem and the uncapacitated fixed-charge location problem, respectively, for choosing facility locations to minimize cost. The expected increase in transportation cost after failures of facilities is also considered. They do not consider inventory cost.

Snyder et al. [22] design a scenario-based integrated supply chain design model in which they describe uncertainties by different scenarios. However, using a scenario-based model will possibly limit the size of the problem that they can solve, since the amount of decision variables in the problem increases as the number of scenarios increases.

Vidal and Goetschalckx [24] give a qualitative discussion of global supply chain design with a very general large-scale MIP that incorporates the reliability of suppliers into the constraints. They also analyze the effects of some uncertain factors through sensitivity analysis.

Bundschuh et al. [3] study a model for a multistage supply chain network, where each node in the network is the supplier of the nodes in the next stage. They also consider the reliability of suppliers by adding constraints to the model. They use the idea of redundancy to increase the robustness of the network. One disadvantage associated with [3] is the use of fixed constants to represent the reliability of suppliers.

More recently, Chopra et al. [5] study sourcing strategies when both ongoing supply uncertainty (caused by machine reliability and congestion of orders, etc.) and the disruption of supply (caused by low likelihood events such as natural disasters) are considered. By studying two single-period models, they conclude that bundling disruption and ongoing supply uncertainty into a single measure results in higher inventory than optimal, higher supply chain costs than optimal, and an underutilization of reliable supply sources.

As mentioned in the introduction section, the unreliability of facilities in our problem is treated by a similar method used in many random yield problems. For a comprehensive review of the random yield literature, we refer the readers to Yano and Lee [25]. This line of research typically involves the production plant only or a two-tiered supply chain. We study the impact of yield uncertainty/product defects in a three-echelon supply chain in this article.

3. MODEL FORMULATION

In our model, we consider facility location costs and inventory costs at facilities, as well as the penalty costs and transportation costs associated with retailers. The retail price of the product at each retailer is given, and each retailer's inventory problem is modeled as a classic newsvendor problem. The objective of our model is to maximize the expected annual profit. We use the following notation throughout the paper:

Parameters

- I: the index set of all retailers, and $|I| = n$
- J: the index set of all facility candidates, and $|J| = m$
- c: purchasing price from the supplier per unit of product

- χ: number of ordering periods per year (for retailers). We assume that the durations of all ordering periods at retailers are equal.

For each facility candidate at location $j \in J$ (facility j)

- f_j: yearly location cost for facility j
- R_j: the reliability coefficient associated with facility j, which is a random variable between 0 and 1. We define $\theta_j = E(R_j)$ and $\tau_j^2 = Var(R_j)$.
- F_j: fixed cost (such as administrative and handling costs) of placing an order at facility j
- g_j: fixed transportation cost per shipment from the supplier to facility j
- a_j: unit shipment cost from the supplier to facility j
- h_j: yearly holding cost at facility j per unit of product
- $\mathcal{N}_j$: number of shipments per year from the supplier to facility j. We assume that χ is dividable by $\mathcal{N}_j$.

For each retailer $i \in I$

- D_i: demand on retailer i in one ordering period, which is a random variable. We assume it follows a normal distribution $N(\mu_i, \sigma_i^2)$, and is mutually independent over different periods
- p_i: retail price at retailer i per unit of product
- π_i: penalty cost of lost goodwill at retailer i per unit of product
- v_i: salvage value at retailer i per unit of product
- d_{ij}: delivery cost from facility $j \in J$ to retailer i per unit of product

Decision variables

- X_j: $\begin{cases} 1 & \text{facility } j \in J \text{ is open} \\ 0 & \text{otherwise} \end{cases}$
- Q_{ij}: order quantity at facility $j \in J$ from retailer $i \in I$ in each ordering period (We assume that the order quantity from retailer i to facility j is the same in each period.)

We use $\boldsymbol{X}$ to denote the $1 \times m$ matrix $(X_j, j \in J)$, $\boldsymbol{Q}$ to denote the $n \times m$ matrix $(Q_{ij}, i \in I, j \in J)$, and $S_{ij}(Q_{ij}, R_j)$ to represent the actual quantity retailer i receives from facility j in each ordering period if retailer i orders Q_{ij} from facility j, and the reliability coefficient associated with facility j is R_j. According to the discussion in Section 1, $S_{ij}(Q_{ij}, R_j) = R_j \cdot Q_{ij}$.

3.1. Working Inventory Cost at Facility $j \in J$

We assume that facility j determines its order size from the supplier using an EOQ-like inventory policy illustrated in Fig. 2. It, differs from the EOQ model since the demand facility j faces only arrives at the beginning of each ordering period, and therefore, the demand comes in batches periodically instead of smooth and continuous arrivals. Since the supplier is perfectly reliable in our problem, facilities will always receive the exact amount they order. Hence, the frequency of shipments ($\mathcal{N}_j$ per year) and the size of one shipment from the supplier to facility j are determined by the order quantity assigned to facility j. According to our assumption that χ is dividable by $\mathcal{N}_j$, the duration between two consecutive shipments from the supplier to facility $j(1/\mathcal{N}_j)$ is a multiple of the ordering cycle length at retailers $(1/\chi)$. It is easy to see that the average size of a shipment from the supplier to facility j is $\chi \sum_{i \in I} Q_{ij}/\mathcal{N}_j$. We assume that defective and damaged products caused during packaging/assembly operations and storage at facilities are identified by the quality inspection right before delivery to the retailers, so it is easy to see from Fig. 2 that the average inventory level at facility j is $\left(\frac{\frac{\chi}{\mathcal{N}_j}+1}{2}\right) \sum_{i \in I} Q_{ij}$, which is based on the amount ordered instead of the amount actually delivered to the retailers (each facility takes the responsibility of the product loss that occurs within it). Thus, the cost of holding inventory at facility j is $h_j \left(\frac{\frac{\chi}{\mathcal{N}_j}+1}{2}\right) \sum_{i \in I} Q_{ij}$. With the widely used linear delivery cost assumption, the delivery cost from the supplier to facility j can be represented as $g_j + a_j \chi \sum_{i \in I} Q_{ij}/\mathcal{N}_j$, then the annual working inventory cost at facility $j \in J$ is $F_j \mathcal{N}_j + (g_j + a_j \chi \sum_{i \in I} Q_{ij}/\mathcal{N}_j)\mathcal{N}_j + h_j(\frac{\chi}{\mathcal{N}_j} + 1) \sum_{i \in I} Q_{ij}/2 = (F_j + g_j)\mathcal{N}_j + \frac{h_j \chi}{2\mathcal{N}_j} \sum_{i \in I} Q_{ij} + (a_j \chi + h_j/2) \sum_{i \in I} Q_{ij}$.

It is easy to show that the optimal value of $\mathcal{N}_j$ that minimizes the annual working inventory cost

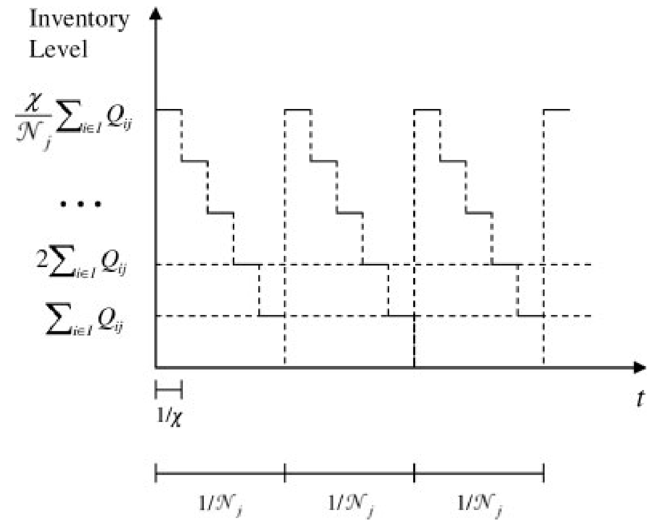

Figure 2. The inventory management policy at facility j.

at facility j is $\sqrt{h_j \chi \sum_{i\in I} Q_{ij}/(2(F_j+g_j))}$. Hence, the corresponding total optimal annual working inventory cost at facility j can be expressed as $(a_j\chi + h_j/2)\sum_{i\in I} Q_{ij} + \sqrt{2h_j(F_j+g_j)\chi \sum_{i\in I} Q_{ij}}$. We define $K_j = \sqrt{2h_j(F_j+g_j)\chi}$ to simplify the notation as $(a_j\chi + h_j/2)\sum_{i\in I} Q_{ij} + K_j\sqrt{\sum_{i\in I} Q_{ij}}$.

3.2. Profit at Retailer $i \in I$

Each retailer can place orders to several open facilities, some or all of which may be unreliable. Recall in this paper we use $S_{ij}(Q_{ij}, R_j) = R_j \cdot Q_{ij}$ to represent the actual quantity retailer i receives from facility j in each ordering period if the order quantity from retailer i to facility j is Q_{ij}, and the reliability coefficient associated with facility j is R_j. With the assumptions that all excess stock is salvaged at the end of each ordering period, and inventory shortages result in lost sales, we model retailer i's ($i \in I$) inventory problem as a newsvendor problem, as proposed in Dada et al. [6]:

$$\text{Maximize} \quad T_i(\boldsymbol{Q}) \equiv E\Bigg\{p_i D_i + v_i\Bigg[\sum_{j\in J} S_{ij}(Q_{ij},R_j) - D_i\Bigg]^+ - (p_i+\pi_i)\Bigg[D_i - \sum_{j\in J} S_{ij}(Q_{ij},R_j)\Bigg]^+ - \sum_{j\in J} d_{ij} S_{ij}(Q_{ij},R_j)\Bigg\}$$

$$\text{s.t.} \quad Q_{ij} \ge 0 \quad j \in J$$

Similar to the objective function in the classic newsvendor problem, $T_i(\boldsymbol{Q})(i \in I)$ represents the expected profit earned at retailer i in each ordering period.

3.3. Integrated Model

Based on the discussions in Sections 3.1 and 3.2, we formulate our problem as:

PROBLEM P:

$$\begin{aligned}\text{Maximize} \quad & -\sum_{j\in J}\Bigg\{f_j X_j + c\chi \sum_{i\in I} Q_{ij} + \Bigg((a_j\chi + h_j/2)\sum_{i\in I} Q_{ij} + K_j\sqrt{\sum_{i\in I} Q_{ij}}\Bigg)\Bigg\} + \chi\sum_{i\in I} T_i(\boldsymbol{Q}) \\ & = \phi(\boldsymbol{Q}) - \sum_{j\in J} f_j X_j\end{aligned}$$

$$\begin{aligned}\text{s.t.} \quad & 1 - e^{-\beta Q_{ij}} \le X_j \\ & i \in I, j \in J, \beta \text{ is a positive constant} && (1)\\ & Q_{ij} \ge 0 \qquad i \in I, j \in J && (2)\\ & X_j \in \{0,1\} \qquad j \in J && (3)\end{aligned}$$

where

$$\phi(\boldsymbol{Q}) \equiv -\sum_{j\in J}\Bigg\{[(c+a_j)\chi + h_j/2]\sum_{i\in I} Q_{ij} + k_j\sqrt{\sum_{i\in I} Q_{ij}}\Bigg\} + \chi\sum_{i\in I} T_i(\boldsymbol{Q}).$$

The objective of Problem P is to maximize the expected annual profit of the entire system including all facilities and retailers. In the objective function, the first term represents the facility location cost for opening facilities and the second term is the annual purchasing cost from the supplier. The third term represents the working inventory cost associated with each facility, and the last term is the profit earned at retailers.

The first constraint stipulates that retailers can only order from open facilities. We use an exponential function to formulate this restriction, which is a novel "trick" in this article, instead of any other commonly used method such as the big-M method, because we want to utilize the quick convergence property of the exponential function. The positive constant β in constraint (1) is used to expedite the convergence. Please refer to the remarks right after Theorem 2 for details.

We use $(\boldsymbol{X}^*, \boldsymbol{Q}^*)$ to denote an optimal solution to Problem P. Since Problem P is a highly nonlinear and mixed-integer optimization problem, with the objective function neither convex nor concave, it is very difficult to solve directly by any standard algorithm. We study the relationship between Problem P and its Lagrangian Dual problem in Sections 3.4 and 3.5, then propose an algorithm to solve Problem P in Section 5.

3.4. Lagrangian Relaxation Model

By relaxing constraint (1), we derive the Lagrangian relaxation problem of Problem P:

PROBLEM LR:

$$z(\boldsymbol{\lambda}) = \text{Maximize} \quad \phi(\boldsymbol{Q}) - \sum_{j\in J} f_j X_j + \sum_{j\in J}\sum_{i\in I} \lambda_{ij}\{X_j - 1 + e^{-\beta Q_{ij}}\}$$

$$\text{s.t.} \quad Q_{ij} \geq 0 \quad i \in I, j \in J \tag{4}$$
$$X_j \in \{0, 1\} \quad j \in J \tag{5}$$

We use $\boldsymbol{\lambda}$ to denote the $n \times m$ matrix $(\lambda_{ij}, i \in I, j \in J)$ (where $\lambda_{ij} \geq 0$), and rewrite the objective function of Problem LR as $-\sum_{j\in J}(f_j - \sum_{i\in I}\lambda_{ij})X_j + \phi(\boldsymbol{Q}) + \sum_{j\in J}\sum_{i\in I}\lambda_{ij}(e^{-\beta Q_{ij}} - 1)$, which we use $L(\boldsymbol{\lambda}, \boldsymbol{X}, \boldsymbol{Q})$ to denote. It is well known that $z(\boldsymbol{\lambda})$ provides an upper bound for the optimal objective value of Problem P, which is $\phi(\boldsymbol{Q}^*) - \sum_{j\in J} f_j X_j^*$. This is easy to show by observing that

$$z(\boldsymbol{\lambda}) \geq L(\boldsymbol{\lambda}, \boldsymbol{X}^*, \boldsymbol{Q}^*) \geq \phi(\boldsymbol{Q}^*) - \sum_{j\in J} f_j X_j^*.$$

The first inequality follows from the definition of $z(\boldsymbol{\lambda})$, and the second inequality follows from the facts that $\lambda_{ij} \geq 0$ and $1 - e^{-\beta Q_{ij}^*} \leq X_j^*$ for any $i \in I$ and $j \in J$.

Since $\boldsymbol{X}$ and $\boldsymbol{Q}$ are independent in Problem LR, we can determine the optimal solutions for $\boldsymbol{X}$ and $\boldsymbol{Q}$ separately. Since $X_j \in \{0, 1\}, \forall j \in J$, its optimal solution can be directly determined by its corresponding coefficient, $-f_j + \sum_{i\in I}\lambda_{ij}$. If the coefficient of X_j is positive, then $X_j = 1$; if it is negative, $X_j = 0$. If the coefficient of X_j is equal to 0, then X_j can be either 0 or 1. In this case, we restrict $X_j = 0$ when $Q_{ij} = 0, \forall i \in I$, and $X_j = 1$ otherwise. We propose an algorithm to search the optimal $\boldsymbol{Q}$ of Problem LR in Section 5.

3.5. Connection between Problem P and Problem LR

We use Problem DP to denote the Lagrangian dual problem of Problem P, that is,

$$\text{Problem DP} : \{\text{Minimize } z(\boldsymbol{\lambda}), \quad \text{s.t. } \lambda_{ij} \geq 0 \; i \in I, j \in J\}$$

The Lagrangian dual problem DP, which has a convex objective function of $\boldsymbol{\lambda}$, can be solved by the Subgradient Algorithm (SA). For more details on SA, please refer to [18].

Define $\boldsymbol{\lambda}^t$ as the $\boldsymbol{\lambda}$ value in the t-th iteration of SA. Let $\boldsymbol{\lambda}^1 = 0$, and $(\boldsymbol{X}^t, \boldsymbol{Q}^t)$ be the corresponding optimal solution of Problem LR in the t-th iteration of SA. Define the rule of SA to be:

$$\lambda_{ij}^{t+1} = \left[\lambda_{ij}^t + b_{ij}^t\left(1 - e^{-\beta Q_{ij}^t} - X_j^t\right)\right]^+ \quad \text{for } t \geq 1 \tag{6}$$

where b_{ij}^t is a function of t and $b_{ij}^t > 0, \lim_{t\to\infty} b_{ij}^t = 0$ and $\sum_{t=1}^{\infty} b_{ij}^t = \infty$.

Shor [18] shows that with the above rule, $\lim_{t\to\infty}\boldsymbol{\lambda}^t$ is finite. We define $\boldsymbol{\lambda}^* = \lim_{t\to\infty}\boldsymbol{\lambda}^t$ and use $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ to denote the optimal solution to Problem LR when $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$.

REMARK: Although when $b_{ij}^t > 0, \lim_{t\to\infty} b_{ij}^t = 0$ and $\sum_{t=1}^{\infty} b_{ij}^t = \infty$, SA is guaranteed to converge, it may converge very slowly under this condition. As a result, we only use this rule when we analyze model properties in Sections 3 and 4. We use a heuristic version of SA to obtain close-to-optimal solutions faster when we design the actual algorithm for computational experiments (please refer to Table 1 in Section 6).

The following two lemmas reveal some useful properties of $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$.

LEMMA 1: $\lim_{t\to\infty}\boldsymbol{Q}^t = \boldsymbol{Q}^L$.

LEMMA 2: For any pair of $i \in I$ and $j \in J$, $X_J^L = 1$ and $Q_{ij}^L = 0$ can be true only when $\lambda_{ij}^* = 0$.

We now establish the relationship between $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ and Problem P.

THEOREM 1: $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ is a feasible solution to Problem P.

The values of $\boldsymbol{\lambda}^*, \boldsymbol{X}^L$, and $\boldsymbol{Q}^L$, and hence $z(\boldsymbol{\lambda}^*)$, the optimal objective value of Problem DP, and $\phi(\boldsymbol{Q}^L) - \sum_{j\in J} f_j X_j^L$, the objective value of Problem P when $\boldsymbol{X} = \boldsymbol{X}^L$ and $\boldsymbol{Q} = \boldsymbol{Q}^L$, depend on the value of β in constraint (1). Theorem 2 illustrates the relationship between them and the value of β.

THEOREM 2: The larger the value of β in constraint (1) is, the smaller the value of $z(\boldsymbol{\lambda}^*) - (\phi(\boldsymbol{Q}^L) - \sum_{j\in J} f_j X_j^L)$. Specifically, when $\beta \to \infty$, $z(\boldsymbol{\lambda}^*) - (\phi(\boldsymbol{Q}^L) - \sum_{j\in J} f_j X_j^L = 0$, and $(\boldsymbol{X}^L, \boldsymbol{Q}^L) = (\boldsymbol{X}^*, \boldsymbol{Q}^*)$.

REMARKS:

1. From Case 4 in the proof of Theorem 2, we see the advantage of the exponential function in constraint (1). It makes $X_j^L - 1 + e^{-\beta Q_{ij}^L}$ close to 0, which is not possible with the big-M method.
2. Theorem 2 shows that we can obtain the properties of $(\boldsymbol{X}^*, \boldsymbol{Q}^*)$ by studying the properties of $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ wnen $\beta \to \infty$. For numerical experiments, we can simply set β big enough to make $z(\boldsymbol{\lambda}^*) - (\phi(\boldsymbol{Q}^L) - \sum_{j\in J} f_j X_j^L)$ small enough, so that $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ is still a "good" approximation to $(\boldsymbol{X}^*, \boldsymbol{Q}^*)$ (we show $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ is a feasible solution to Problem P in Theorem 1).

4. PROPERTIES OF THE OPTIMAL SOLUTION TO PROBLEM P

Theorem 2 shows that when $\beta \rightarrow \infty, (\boldsymbol{X}^L, \boldsymbol{Q}^L) = (\boldsymbol{X}^*, \boldsymbol{Q}^*)$. Thus, to reveal properties of $(\boldsymbol{X}^*, \boldsymbol{Q}^*)$, we set $\beta \rightarrow \infty$ and study $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$.

PROPERTY 1: For any facility $j \in J$ and retailer $i \in I$, if $d_{ij} \geq (p_i + \pi_i) - \frac{c+a_j}{\theta_j} - \frac{h_j}{2\chi\theta_j}$, then retailer i will not be served by facility j in the optimal solution to Problem P.

Property 1 can also be informally justified as follows: If the unit profit for facility j to serve retailer i, which is $p_i - d_{ij} - \frac{c+a_j}{\theta_j} - \frac{h_j}{2\chi\theta_j}$ (retail price–shipment cost–purchasing cost–inventory holding cost), is no larger than $-\pi_i$, the unit profit when retailer i is not served by any facility, then it is obvious that we should not use facility j to serve retailer i. Hence, if $d_{ij} \geq (p_i + \pi_i) - \frac{c+a_j}{\theta_j} - \frac{h_j}{2\chi\theta_j}$, retailer i will not be served by facility j.

Please refer to Fig. 3 for an illustration of Property 1. Facility j can serve retailer i only if (d_{ij}, θ_j) is within the shaded region in Fig. 3.

Property 1 yields the following conclusions (please refer to Fig. 3):

- Whether facility j will serve retailer i or not depends on both the unit transportation cost from facility j and reliability factor of facility j. A retailer is more likely to be assigned to a facility with cheaper shipment cost and bigger reliability factor.
- Facility j will not be opened if its reliability is too low, or its unit transportation cost is too high. Specifically, if facility j satisfies $d_{ij} \geq (p_i + \pi_i) - \frac{c+a_j}{\theta_j} - \frac{h_j}{2\chi\theta_j}$, for all $i \in I$, then it will not be opened in the optimal solution to Problem P, i.e. $X_j^* = 0$ or $X_j^L = 0$ when $\beta \rightarrow \infty$.

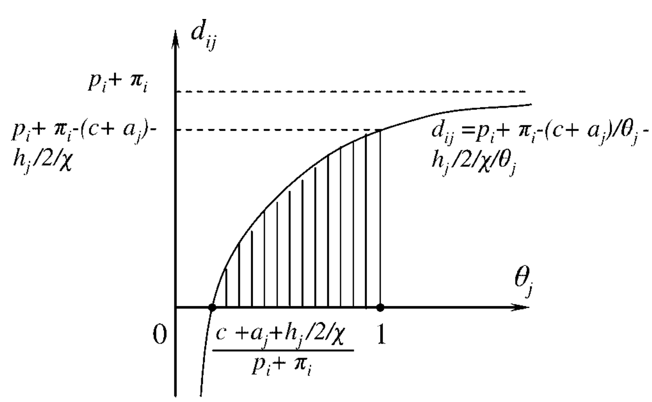

Figure 3. Unit transportation cost versus reliability.

We also discover the following property about a retailer's order fulfillment.

PROPERTY 2: If the optimal solution to Problem P suggests that facility j serves retailer i, then the expected percentage of the customer demand fulfilled at retailer i is less than $\frac{-(c+a_j) - \frac{h_j}{2\chi} + (p_i + \pi_i - d_{ij})\theta_j}{(p_i + \pi_i - v_i)\theta_j}$.

It follows from Property 2 that the expected percentage of the customer demand fulfilled at retailer $i (i \in I)$ will be no more than $\max_{j \in J} \left\{ \frac{-(c+a_j) - \frac{h_j}{2\chi} + (p_i + \pi_i - d_{ij})\theta_j}{(p_i + \pi_i - v_i)\theta_j} \right\}$, and hence we conclude that:

- A retailer with higher retail price and penalty cost is more likely to be able to fulfill the customer demand it faces.
- The expected percentage of the customer demand fulfilled at a retailer is likely to be low, if
  - facilities have low reliability, and
  - the delivery cost to this retailer is high.

5. SOLUTION ALGORITHM

5.1. Basic Approach

Section 3 shows that in order to solve Problem P, we can use the Subgradient Algorithm (SA) to solve Problem DP, and then derive $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$, which is a feasible solution to Problem P (Theorem 1). When $\beta \rightarrow \infty, (\boldsymbol{X}^L, \boldsymbol{Q}^L)$ is an optimal solution to Problem P, and for big enough β, $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ can provide a "good" approximation to $(\boldsymbol{X}^*, \boldsymbol{Q}^*)$. In the t-th iteration of SA, we need to find $(\boldsymbol{X}^t, \boldsymbol{Q}^t)$, the optimal solution to Problem LR when $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$, to determine the value of $\boldsymbol{\lambda}^{t+1}$ by (6). In Section 3.4, we have discussed how the optimal $\boldsymbol{X}$ to Problem LR can be derived. In this section, we propose an algorithm to find the optimal $\boldsymbol{Q}$ to Problem LR.

In Section 3, we show that $\boldsymbol{Q}$ and $\boldsymbol{X}$ are independent in Problem LR. Thus to find the optimal $\boldsymbol{Q}$ to Problem LR, we can focus on maximizing: $\phi(\boldsymbol{Q}) + \sum_{j \in J} \sum_{i \in I} \lambda_{ij} e^{-\beta Q_{ij}}$. Since

$$\phi(\boldsymbol{Q}) = -\sum_{j \in J} \left\{ [(c + a_j)\chi + h_j/2] \sum_{i \in I} Q_{ij} + K_j \sqrt{\sum_{i \in I} Q_{ij}} \right\} + \chi \sum_{i \in I} T_i(\boldsymbol{Q})$$

and

$$T_i(\boldsymbol{Q}) = E\left\{p_i D_i + v_i\left[\sum_{j\in J} R_j Q_{ij} - D_i\right]^+ - (p_i+\pi_i)\left[D_i - \sum_{j\in J} R_j Q_{ij}\right]^+ - \sum_{j\in J} d_{ij} R_j Q_{ij}\right\}$$
$$= E\left\{(p_i - v_i)D_i + \sum_{j\in J}(v_i - d_{ij})R_j Q_{ij} - (p_i+\pi_i-v_i)\left[D_i - \sum_{j\in J} R_j Q_{ij}\right]^+\right\}$$
$$= E\left\{(p_i - v_i)D_i + \sum_{j\in J}(p_i+\pi_i - d_{ij})R_j Q_{ij} - (p_i+\pi_i-v_i)\left(\sum_{j\in J} R_j Q_{ij} + \left[D_i - \sum_{j\in J} R_j Q_{ij}\right]^+\right)\right\}$$
$$= (p_i - v_i)\mu_i + \sum_{j\in J}(p_i+\pi_i - d_{ij})\theta_j Q_{ij} - (p_i+\pi_i-v_i)E\left\{\sum_{j\in J} R_j Q_{ij} + \left[D_i - \sum_{j\in J} R_j Q_{ij}\right]^+\right\}$$

the optimal $\boldsymbol{Q}$ to Problem LR in the t-th iteration of SA is also an optimal solution to the following problem:

PROBLEM LR′:

$$\begin{aligned}\text{Minimize}\quad & -\phi(\boldsymbol{Q}) - \sum_{j\in J}\sum_{i\in I}\lambda_{ij}^t e^{-\beta Q_{ij}} + \chi\sum_{i\in I}(p_i - v_i)\mu_i \\ & = \sum_{j\in J}\left\{\sum_{i\in I}\{[c + a_j - (p_i+\pi_i-d_{ij})\theta_j]\chi + h_j/2\}Q_{ij} + K_j\sqrt{\sum_{i\in I} Q_{ij}}\right\} - \sum_{j\in J}\sum_{i\in I}\lambda_{ij}^t e^{-\beta Q_{ij}} \\ & \quad + \chi\sum_{i\in I}(p_i+\pi_i-v_i)E\left\{\sum_{j\in J} R_j Q_{ij} + \left[D_i - \sum_{j\in J} R_j Q_{ij}\right]^+\right\} \\ \text{s.t.}\quad & Q_{ij}\ge 0 \quad i\in I, j\in J\end{aligned} \tag{7}$$

We assume that $c + a_j + h_j/2/\chi + d_{ij}\theta_j \ge v_i\theta_j$ for all $i \in I, j \in J$. This is a reasonable assumption since $c > v_i$ in most cases. With this assumption, it is easy to see that Problem LR′ has a finite optimal solution. To find the optimal $\boldsymbol{Q}$ to Problem LR, we can solve Problem LR′ instead, which can be reformulated into the following problem by introducing a new variable s:

$$\begin{aligned}\text{Minimize}\quad & s + \sum_{j\in J} K_j\sqrt{\sum_{i\in I} Q_{ij}} - \sum_{j\in J}\sum_{i\in I}\lambda_{ij}^t e^{-\beta Q_{ij}} \\ & + \chi\sum_{i\in I}(p_i+\pi_i-v_i)E\left\{\sum_{j\in J} R_j Q_{ij} + \left[D_i \text{-} \sum_{j\in J} R_j Q_{ij}\right]^+\right\} \\ \text{s.t.}\quad & \sum_{j\in J}\sum_{i\in I}\{[c + a_j - (p_i+\pi_i-d_{ij})\theta_j]\chi + h_j/2\}Q_{ij} - s \le 0 \\ & Q_{ij}\ge 0 \quad i\in I, j\in J\end{aligned} \tag{8}$$

The objective function of the above problem is an increasing and differentiable function of $\boldsymbol{Q}$. We assume that $c + a_j - (p_i + \pi_i - d_{ij})\theta_j + h_j/2/\chi < 0$ for all $i \in I$ and $j \in J$ here, since if there is a pair of $i \in I$ and $j \in J$ such that $c + a_j - (p_i + \pi_i - d_{ij})\theta_j + h_j/2/\chi \geq 0$, then the optimal Q_{ij} must be zero based on Property 1, and we can simply drop this Q_{ij} from the model. Therefore, $\sum_{j\in J}\sum_{i\in I}\{[c + a_j - (P_i + \pi_i - d_{ij})\theta_j]\chi + h_j/2\}Q_{ij}$ is a decreasing and differentiable function of $\boldsymbol{Q}$. To make our study more general, we generalize the above problem into the following version:

$$\begin{aligned} \text{Minimize} \quad & G(\boldsymbol{Q}, s) \\ \text{s.t.} \quad & H(\boldsymbol{Q}) \leq s \\ & Q_{ij} \geq 0 \quad i \in I, j \in J \\ & s \in R \end{aligned}$$

where $G(\boldsymbol{Q}, s)$ is an increasing and differentiable function of $\boldsymbol{Q}$ and s, and $H(\boldsymbol{Q})$ is a decreasing and differentiable function of $\boldsymbol{Q}$ (in our case, $H(\boldsymbol{Q})$ is linear). For any minimization problem with only nonnegative constraints, such as Problem LR′, if its objective function can be written as the difference between two increasing functions, then this problem can be transformed into the format of the above generalized problem by incorporating a new variable s as we do for Problem LR′. We next propose an algorithm to solve this generalized problem.

We use $(\boldsymbol{Q^G}, s^G)$ to denote an optimal solution to the generalized problem. From Lemma 3, we know that the optimal solution to the generalized problem is at the boundary of the feasible region.

LEMMA 3: $(\boldsymbol{Q^G}, s^G)$ satisfies $s^G = H(\boldsymbol{Q^G})$.

Since Problem LR′ has a finite solution, we only study the generalized problem when Q_{ij}^G are finite for all $i \in I, j \in J$, and use $\bar{Q}_{ij}$ to denote a big enough value that provides an upper bound of Q_{ij}^G. Thus, the optimal solution to the generalized problem is within $\mathcal{D} \equiv \{(\boldsymbol{Q}, s) : s = \boldsymbol{H}(\boldsymbol{Q}), 0 \leq Q_{ij} \leq \bar{Q}_{ij}, \forall_i \in I, j \in J\}$, which is apparently a subregion of the feasible region of the generalized problem.

We define $\bar{s} = H(\bar{\boldsymbol{Q}})$. Corollary 1 provides a lower bound to the objective value of the generalized problem.

COROLLARY 1: $G(\boldsymbol{0}, \bar{s}) \leq G(\boldsymbol{Q}, s), \quad \forall(\boldsymbol{Q}, s) \in \mathcal{D}$.

When designing an algorithm to search for the global minimum of the generalized problem, we draw insights from the bisection search algorithm and the Outer Approximation algorithm [10, 11]. We later show that our algorithm is more efficient than the Outer Approximation algorithm when being applied to solve the generalized problem above.

The basic ideas of this algorithm are:

- Partition $\mathcal{D}$, the region containing the optimal solution to the generalized problem, into several subregions. Find lower bounds for all such subregions, and put them into a set, which is denoted by V. At the beginning of the algorithm, V contains only one element $G(\boldsymbol{0}, \bar{s})$.
- In each iteration of this algorithm, we remove the smallest lower bound from the lower bound set. If this lower bound is feasible to the generalized problem, then apparently it is also optimal. If the optimal solution has not been reached, we partition the subregion corresponding to this smallest lower bound into two smaller subregions, find a lower bound for each of them, and put new lower bounds into the lower bound set. This procedure, according to Theorem 3 (presented in the later part of this section), is guaranteed to converge to a global minimum.

5.2. A Special Case

Figure 4 illustrates the algorithm procedure under a special case when the matrix $\boldsymbol{Q}$ only contains one element, i.e., the problem only has two variables, Q and s. Based on Lemma 3, the optimal solution should be on the curve $s = H(Q)$. Furthermore, the bold solid curve is $\mathcal{D}$. At the beginning of the algorithm, $(0, \bar{s})$ gives a lower bound to any feasible solution on the bold solid curve according to Corollary 1. In the first iteration of the algorithm, the initial lower bound is dropped from the lower bound set, and the bold solid curve is partitioned into two subregions by the point H^1, whose coordinates are $(\bar{Q}/2, H(\bar{Q}/2))$. It follows from Lemma 4 that the corresponding lower bounds for these two subregions are given by points $w_1^1(0, H(\bar{Q}/2))$ and $w_2^1(\bar{Q}/2, \bar{s})$ in

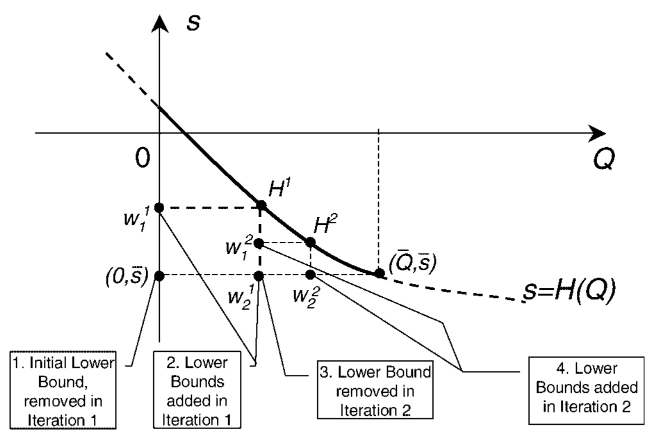

Figure 4. Illustration of the algorithm when there is only one element within the matrix $\boldsymbol{Q}$.

the figure. Therefore, after the first iteration, the lower bound set contains lower bounds given by w_1^1 and w_2^1, and both lower bounds are bigger than the initial lower bound $G(0, \bar{s})$ since $G(\boldsymbol{Q}, s)$ is an increasing function of both $\boldsymbol{Q}$ and s. In the case illustrated in Fig. 4, we suppose that $G(w_2^1) \leq G(w_1^1)$. In the second iteration of the algorithm, we drop the current lower bound provided by w_2^1 from the lower bound set, partition the feasible subregion to the right of w_2^1 into two subregions using point $H^2\left(\frac{3\bar{Q}}{4}, H\left(\frac{3\bar{Q}}{4}\right)\right)$, and find the corresponding lower bounds given by w_1^2 and w_2^2. Again, these new lower bounds are larger than the dropped lower bound $G(w_2^1)$. Thus, after the second iteration, the lower bound set contains lower bounds given by w_1^1, w_1^2, and w_2^2. In the third iteration, we choose from w_1^1, w_1^2, and w_2^2 whose $G(\bullet)$ value is the smallest. This value provides a lower bound for any feasible solution on the bold solid curve. We further partition its related feasible subregion into two smaller subregions, and repeat the above procedure till the stopping condition in the algorithm is satisfied.

On the other hand, in each iteration of the algorithm, we update the best existing feasible solution using the following procedure: for each subregion we are working on to derive a lower bound, take the middle point of the corresponding subregion as the best existing feasible solution if its objective value is smaller than the current best existing feasible solution. Apparently, the best existing feasible solution gives an upper bound to the optimal objective value of the generalized problem. We stop the algorithm if the best lower bound is close enough to the best existing feasible solution.

5.3. The General Case

Now let us see the general case when $\boldsymbol{Q}$ is a general matrix. The following Lemma suggests an efficient way to solve the general case.

LEMMA 4: For any subregion of $\mathcal{D}$ with the format of $\{(s, \boldsymbol{Q}) : s = H(\boldsymbol{Q}), \boldsymbol{L} \leq \boldsymbol{Q} \leq \boldsymbol{U}\}$, where $\boldsymbol{L}$ and $\boldsymbol{U}$ are given vectors that provide the boundary of this subregion, $(\boldsymbol{L}, H(\boldsymbol{U}))$ gives a lower bound to the objective value of any feasible solution in this subregion.

Based on Lemma 4, we propose the following steps to solve the general problem. (Superscript numbers represent the iterations, and subscript numbers are used to denote the resulted partitions.)

At the beginning of the algorithm, V contains one element, $G(\mathbf{0}, \bar{s})$. In the first iteration, $i = 1$, and $j = 1$. We drop $G(\mathbf{0}, \bar{s})$ from V, and partition $\mathcal{D}$ into two subregions, $\boldsymbol{D}_1^1 = \{(s, \boldsymbol{Q}) : (s, \boldsymbol{Q}) \in \mathcal{D}, Q_{ij} \leq \bar{Q}_{ij}/2\}$ and $D_2^1 = \{(s, \boldsymbol{Q}) : (s, \boldsymbol{Q}) \in \mathcal{D}, \bar{Q}_{ij}/2 < Q_{ij}\}$. It is easy to verify that $\boldsymbol{D}_1^1$ and $\boldsymbol{D}_2^1$ can be written in the format of $\{(s, \boldsymbol{Q}) : s = H(\boldsymbol{Q}), \boldsymbol{L} \leq \boldsymbol{Q} \leq \boldsymbol{U}\}$. Based on Lemma 4, we can easily find a lower bound for each of these subregions. We use w_1^1 and w_2^1 to denote the points that provide the lower bounds for solutions on $\boldsymbol{D}_1^1$ and $\boldsymbol{D}_2^1$, respectively, and put $G(w_1^1)$ and $G(w_2^1)$ into V. Without loss of generality, we assume $G(w_1^1) \geq G(w_2^1)$.

In the second iteration, $i = 1, j = 2$. We drop the smallest lower bound $G(w_2^1)$ from V, whose corresponding subregion is $\boldsymbol{D}_2^1$. Since $\boldsymbol{D}_2^1$ can be written in the format of $\{(s, \boldsymbol{Q}) : s = \boldsymbol{H}(\boldsymbol{Q}), \boldsymbol{L} \leq \boldsymbol{Q} \leq \boldsymbol{U}\}$, we use L_{ij} and U_{ij} to denote its lower bound and upper bound with respect to Q_{ij}, and partition it into two subregions, $\boldsymbol{D}_1^2 = \{(s, \boldsymbol{Q}) : (s, \boldsymbol{Q}) \in D_2^1, Q_{ij} \leq (L_{ij} + U_{ij})/2\}$ and $\boldsymbol{D}_2^2 = \{(s, \boldsymbol{Q}) : (s, \boldsymbol{Q}) \in D_2^1, (L_{ij} + U_{ij})/2 < Q_{ij}\}$. Again, it is easy to verify that $\boldsymbol{D}_1^2$ and $\boldsymbol{D}_2^2$ are subregions of $\mathcal{D}$ with the format of $\{(s, \boldsymbol{Q}) : s = H(\boldsymbol{Q}), \boldsymbol{L} \leq \boldsymbol{Q} \leq \boldsymbol{U}\}$, and we can easily find a lower bound for each of these subregions. We use w_1^2 and w_2^2 to denote the points giving the lower bounds for solutions on $\boldsymbol{D}_1^2$ and $\boldsymbol{D}_2^2$, respectively. We put $G(w_1^2)$ and $G(w_2^2)$ into V. Therefore, after the second iteration, there are three elements in V, which are $G(w_1^1), G(w_1^2)$, and $G(w_2^2)$.

In the third iteration, $i = 1, j = 3$. We choose the smallest one from $G(w_1^1)$, $G(w_1^2)$, and $G(w_2^2)$. This value provides a lower bound for any feasible solution on $\mathcal{D}$. We further drop it from V and partition its related subregion into two smaller subregions, and repeat the above procedure until the best lower bound is close enough to the best existing feasible solution. We may reset $i = 1$ and $j = 1$, and repeat the above process if the best lower bound after $i = n$ and $j = m$ is not good enough.

An important advantage shown in Lemma 4 is that the lower bound of each subregion of $\mathcal{D}$ can be immediately identified, while most Outer Approximation algorithms are exponential-time algorithms with respect to the dimensions of the problem when searching for lower bounds.

5.4. The Convergence Result

We use G^k to represent the smallest lower bound in the lower bound set after iteration k. From the above procedure, it is very easy to verify the following corollary.

COROLLARY 2: a. $G^k \leq G^{k+1}$; b. If the solution corresponding to G^k is a feasible solution to the generalized problem, then G^k is a global minimum; otherwise, G^k is less than the lower bounds generated in iteration $k + 1$.

We now show the convergence result of the algorithm.

THEOREM 3: The sequence of $\{G^k\}, k = 1, 2, 3, \ldots$, is guaranteed to converge to a global minimum.

It is possible that as we continue the algorithm, $\mathcal{D}$, the region containing the optimal solution to the generalized problem, may be partitioned into numerous subregions, and the size of the lower bound set may explode. To deal with this problem, we check all newly generated subregions in each iteration of the above algorithm, and eliminate some subregions that do not contain the optimal solutions. Specifically, for some $i, i' \in I$, and $j, j' \in J$, if all solutions in a given newly generated subregion satisfy $\left\{\frac{\partial}{\partial Q_{ij}} G(\boldsymbol{Q}, H(\boldsymbol{Q})) - \frac{\partial}{\partial Q_{i'j'}} G(\boldsymbol{Q}, H(\boldsymbol{Q}))\right\} < 0$, and this subregion's upper bound in direction Q_{ij} is less than its lower bound in direction $Q_{i'j'}$, we can drop this newly generated subregion within which no solution could be optimal according to the following property.

PROPERTY 3: In the generalized problem, for $i, i' \in I$, and $j, j' \in J$, if $Q_{ij}^G < Q_{i'j'}^G$, then

$$\left\{\frac{\partial}{\partial Q_{ij}} G(\boldsymbol{Q}, H(\boldsymbol{Q})) - \frac{\partial}{\partial Q_{i'j'}} G(\boldsymbol{Q}, H(\boldsymbol{Q}))\right\}\Bigg|_{\boldsymbol{Q}=\boldsymbol{Q}^G} \geq 0.$$

6. COMPUTATIONAL EXPERIMENTS

In our computational experiments, we assume that $R_j \sim N(\theta_j, \tau_j^2)$, $j \in J$, and hence $S_{ij}(Q_{ij}, R_j) = R_j \cdot Q_{ij} \sim N(Q_{ij}\theta_j, Q_{ij}^2\tau_j^2)$.

6.1. Data Sets Used in the Experiments

All computational results reported in this section are based on four different types of data sets, with sizes $m = 6, n = 14$; $m = 10$, $n = 40$; $m = 20$, $n = 40$; and $m = 20$, $n = 80$, respectively. Though most data in these data sets are randomly generated, we try to make the intervals from which we draw data as reasonable as possible, so that the data randomly drawn from these intervals are consistent with reality.

The locations of the retailers and the facility candidates are uniformly generated within a 1000×1000 region, and the values of all the other parameters are generated as follows:

- $c = 5$;
- $\chi = 52$;
- $L_j = 4$;
- d_{ij} : 0.02× (Euclidean) distance between facility j and retailer i;
- Parameters in the following table are uniformly drawn from the specified intervals:

$j \in J$		$i \in I$	
f_j: [2000, 3000]	g_j: [0, 5]	μ_i: [10, 20]	v_i: [0, 4]
θ_j: [0.6, 1]	a_j: [0, 3]	σ_i^2: [3, 6]	
τ_j: [0.01, 0.04]	h_j: [0, 2]	p_i: [30, 40]	
F_j: [0, 5]		π_i: [0, 3]	

Computational results are presented and discussed in the remaining part of this section. The algorithm is coded in C++, and all computational times are obtained on a DELL PC with a P4 2.8 GHz CPU running Windows XP.

6.2. Performance of the Algorithm

The parameter set for SA used in our computational experiments is shown in Table 1. In the t-th iteration of SA, we derive the optimal $\boldsymbol{X}^t$ to Problem LR using the method introduced in Section 3.4, and search for $\boldsymbol{Q}^t$ using the algorithm proposed in Section 5. We set $\hat{X}_j^t = 1$ if $Q_{ij}^t > 0$ ($\exists i \in I$), otherwise, $\hat{X}_j^t = X_j^t$. It is easy to see that $(\hat{\boldsymbol{X}}^t, \boldsymbol{Q}^t)$ is a feasible solution to Problem P, and hence $\phi(\boldsymbol{Q}^t) - \sum_{j \in J} f_j \hat{X}_j^t$ provides a lower bound to the optimal objective value of Problem P. On the other hand, as we defined in Section 3.4, $L(\boldsymbol{\lambda}, \boldsymbol{X}, \boldsymbol{Q})$ is the objective function of Problem LR, which is the Lagrangian relaxation problem of Problem P. Therefore, $L(\boldsymbol{\lambda}^t, \boldsymbol{X}^t, \boldsymbol{Q}^t)$ provides an upper bound to the objective value of Problem P. We stop SA when the difference between $L(\boldsymbol{\lambda}^t, \boldsymbol{X}^t, \boldsymbol{Q}^t)$ and $\phi(\boldsymbol{Q}^t) - \sum_{j \in J} f_j \hat{X}_j^t$ is small enough (please refer to the stopping condition in Table 1), and use $(\hat{\boldsymbol{X}}^t, \boldsymbol{Q}^t)$ to estimate the optimal solution to Problem P. From the stopping condition we are using, $\frac{L(\boldsymbol{\lambda}^t, \boldsymbol{X}^t, \boldsymbol{Q}^t) - (\phi(\boldsymbol{Q}^t) - \sum_{j \in J} f_j \hat{X}_j^t)}{\phi(\boldsymbol{Q}^t) - \sum_{j \in J} f_j \hat{X}_j^t} < 0.005$, we see if SA stops, current $(\hat{\boldsymbol{X}}^t, \boldsymbol{Q}^t)$ is a "good" solution to Problem P.

When β is small, the stopping condition in Table 1 sometimes may never be satisfied. We have to increase the value of β gradually in this case based on Theorem 2. According to our computational experiments, when $\beta \geq 1$, the stopping condition in Table 1 will eventually be satisfied in most cases.

Table 2 shows the average CPU times for solving problems with different sizes. Each CPU time is measured according to the stopping condition given in Table 1, and every average time reported is obtained from 10 runs. As we will discuss in Section 6.4, the value of β has an impact on the performance of the algorithm. Hence, when we study the average

Table 1. Parameters for the subgradient algorithm.

Parameter	Value
Initial λ_i	0
Initial value of scalar v	1.45
Number of iterations before halving v	4
b_{ij}^t in (6)	$\frac{v\left\{L(\boldsymbol{\lambda}^t, \boldsymbol{X}^t, \boldsymbol{Q}^t) - \left(\phi(\boldsymbol{Q}^t) - \sum_{j \in J} f_j X_j^t\right)\right\}}{\left\| 1 - X_j^t - e^{-\beta Q_{ij}^t} \right\|^2}$
Condition to stop SA	$\frac{L(\boldsymbol{\lambda}^t, \boldsymbol{X}^t, \boldsymbol{Q}^t) - \left(\phi(\boldsymbol{Q}^t) - \sum_{j \in J} f_j \hat{X}_j^t\right)}{\phi(\boldsymbol{Q}^t) - \sum_{j \in J} f_j \hat{X}_j^t} < 0.005$

Table 2. Average computational times for problems with different sizes when $\beta = 2$.

Problem size	Average CPU time (s)
$m = 6, n = 14$	125.464
$m = 10, n = 40$	286.554
$m = 20, n = 40$	473.418
$m = 10, n = 80$	586.916

CPU time of our program, we always fix $\beta = 2$, which is a good choice according to our computational experiments. Properties 1 and 3 derived in this article are applied in our program to make it more efficient.

6.3. Impacts of the Reliability of Facilities and Transportation Costs on the Optimal Solution

In this section, we numerically study the impacts of the reliability of facilities and transportation costs on the optimal decisions.

We first use a data set with $m = 20$ and $n = 80$ to study the impacts of the reliability of facilities and transportation costs on the optimal retailer assignment decisions. We pick facility $\hat{j}$, who has the largest mean value of reliability coefficient among all facilities in the data set, and vary the values of $\theta_{\hat{j}}$ and $d_{i\hat{j}}$ as shown in Table 3. Table 3 shows the optimal retailer assignments to facility $\hat{j}$ as well as the order quantity assigned to it in each period as $\theta_{\hat{j}}$ and $d_{i\hat{j}}$ vary.

From Table 3 we observe that if a facility candidate reduces its transportation cost or increases its reliability, more retailers will be assigned to it. But if its reliability is too low, then it will not be opened even if its transportation cost is 0. Similarly, if its transportation cost is too high, then it will never be opened even if it is perfectly reliable. These observations are consistent with Property 1 derived in Section 4.

Table 3. Impacts of the reliability of facilities and transportation costs on retailer assignments.

$\theta_{\hat{j}}$	$d_{i\hat{j}}, \forall i \in I$	# of retailers assigned to facility $\hat{j}$	Order quantity assigned to facility $\hat{j}$ in each period
0.98	0.15× distance	75	1813.329
0.98	0	80	3819.534
0.88	0.15× distance	58	862.684
0.88	0	80	2534.601
0.78	0.15× distance	29	251.062
0.78	0	80	1238.626
0.68	0.15× distance	0	0
0.68	0	28	97.984
0.58	0	0	0
1	15× distance	0	0

Table 4. Impacts of the reliability of facilities on facility location and retailer assignment decisions.

Times $\theta_j, \forall j \in J$	# of facilities opened	# of retailers served
1.00	12	80
0.90	13	71
0.80	16	52
0.70	12	27

With the same data set we have conducted some further tests. Instead of varying the value of $\theta_{\hat{j}}$ that corresponds to facility $\hat{j}$ only, we now adjust the values of θ_j for all facilities proportionally (as shown in the first column of Table 4).

We observe from Table 4 that when the reliability of facilities are low, few retailers will be served, though we do not see any monotone relationship between the reliability of facilities and the number of opened facilities in the optimal solutions. Furthermore, according to our extensive computational experiments, when the reliability of facilities is low, even if a retailer is served by one or more facilities, the served quantity will be low (in other words, the value of $\sum_{j \in J} Q_{ij}$ will be small).

6.4. Effect of β on the Performance of the Algorithm

Based on Theorem 2, theoretically, the larger the value of β, the better the approximation if we use $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ to estimate $(\boldsymbol{X}^*, \boldsymbol{Q}^*)$. But, according to our computational experiments, the values of β influence the performance of our algorithm.

A data set of size $m = 6$, $n = 14$ is used as an example to show the effect of β on the performance of our algorithm. Table 5 describes the CPU times needed to solve this example problem under different β values (all the CPU times shown in Table 5 are measured based on the stopping condition given in Table 1). From Table 5, we can see that when β is small, it is difficult for the stopping condition to be satisfied (in this example, when we set $\beta = 0.01$, the program does not stop after running 12 h); on the other hand, when β is big, it also becomes difficult for the program to converge quickly, and it has to spend a long time to find a "good" solution.

According to our numerous computational experiments, if we define the stopping condition for SA to be

Table 5. Influence of β on the performance of the algorithm.

Value of β	CPU time (s)	Value of β	CPU time (s)
0.01	—	5	127.725
1	137.358	6	147.703
2	139.969	7	284.981
3	132.065	8	319.724
4	133.250	10	295.570

$\frac{L(\lambda^t, X^t, Q^t) - (\phi(Q^t) - \sum_{j\in J} f_j \hat{X}_j^t)}{\phi(Q^t) - \sum_{j\in J} f_j \hat{X}_j^t} < 0.005$, the best choice of the value for β is between 2 and 5.

7. CONCLUSIONS

In this article, we introduce an integrated supply chain design model that considers unreliable supply. We show that the proposed model has the same optimal solution as its Lagrangian dual problem when the value of β approaches infinity. Utilizing this observation, we analyze the properties of the optimal solution and propose an efficient solution algorithm. The algorithm can be applied to a general class of nonlinear optimization problems with the following format:

$$\begin{aligned} \text{Minimize} \quad & G(\boldsymbol{Q}) \\ \text{s.t.} \quad & Q_{ij} \geq 0 \quad i \in I, j \in J, \end{aligned}$$

where $G(\boldsymbol{Q})$ can be written as the difference between two increasing and differentiable functions. We also prove the convergence of the algorithm.

In our model, we assume that the supplier is perfectly reliable. In fact, it is not difficult to relax this assumption. By considering the unreliability of the supplier, the total working inventory cost at each candidate facility needs to be reformulated. But as long as the working inventory cost at each facility is still an increasing function of the order quantity assigned, all conclusions in Section 3 are still correct, and we can still apply the solution algorithm proposed in this paper to this new problem.

APPENDIX

PROOF OF LEMMA 1: Since $\boldsymbol{Q}^L$ is an optimal solution to Problem LR when $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, for any $t \geq 1$

$$\phi(\boldsymbol{Q}^L) \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^* e^{-\beta Q_{ij}^L} \geq \phi(\boldsymbol{Q}^t) + \sum_{j\in J} \sum_{j\in I} \lambda_{ij}^* e^{-\beta Q_{ij}^t} \quad (9)$$

Similarly, from the fact that $\boldsymbol{Q}^t$ is an optimal solution to Problem LR when $\boldsymbol{\lambda} = \boldsymbol{\lambda}^t$, we have

$$\phi(\boldsymbol{Q}^t) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^t e^{-\beta Q_{ij}^t} \geq \phi(\boldsymbol{Q}^L) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^t e^{-\beta Q_{ij}^L} \quad (10)$$

Therefore,

$$\begin{aligned} \phi(\boldsymbol{Q}^t) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^* e^{-\beta Q_{ij}^t} &= \phi(\boldsymbol{Q}^t) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^* e^{-\beta Q_{ij}^t} - \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^t e^{-\beta Q_{ij}^t} + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^t e^{-\beta Q_{ij}^t} \\ &\geq \phi(\boldsymbol{Q}^L) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^t e^{-\beta Q_{ij}^L} + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^* e^{-\beta Q_{ij}^t} - \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^t e^{-\beta Q_{ij}^t} \\ &= \phi(\boldsymbol{Q}^L) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^* e^{-\beta Q_{ij}^L} - \sum_{j\in J} \sum_{i\in I} \left(\lambda_{ij}^* - \lambda_{ij}^t\right)\left(e^{-\beta Q_{ij}^L} - e^{-\beta Q_{ij}^t}\right) \\ &> \phi(\boldsymbol{Q}^L) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^* e^{-\beta Q_{ij}^L} - \sum_{j\in J} \sum_{i\in I} \left|\lambda_{ij}^* - \lambda_{ij}^t\right| \end{aligned} \quad (11)$$

The first inequality in (11) holds because of (10). The second inequality in (11) holds from the fact that $|e^{-\beta Q_{ij}^L} - e^{-\beta Q_{ij}^t}| < 1$.

Since $\boldsymbol{\lambda}^t \to \boldsymbol{\lambda}^*$, it follows that $\lim_{t\to\infty} \sum_{j\in J} \sum_{i\in I} |\lambda_{ij}^t - \lambda_{ij}^*| = 0$. So, from (9) and (11), we see $\phi(\boldsymbol{Q}^t) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^* e^{-\beta Q_{ij}^t} \to \phi(\boldsymbol{Q}^L) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^* e^{-\beta Q_{ij}^L}$ as $t \to \infty$. Since $\phi(\boldsymbol{Q}) + \sum_{j\in J} \sum_{i\in I} \lambda_{ij}^* e^{-\beta Q_{ij}}$ is a continuous function, we conclude $\lim_{t\to\infty} \boldsymbol{Q}^t = \boldsymbol{Q}^L$ (Note: In the case where there are more than one optimal $\boldsymbol{Q}$ for Problem LR when $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, $\boldsymbol{Q}^t$ approaches one of the optimal solutions.). ■

PROOF OF LEMMA 2: Since $\boldsymbol{\lambda}^*$, the optimal solution to Problem DP, is fixed for a given problem, no matter which solution algorithm we are using, we specify the algorithm parameter for SA to be $b_{ij}^t = 1/t$ ($\forall i \in I, j \in J$) here to simplify our proof.

Suppose there exist $i \in I$ and $j \in J$ such that $X_j^L = 1$ and $Q_{ij}^L = 0$ when $\lambda_{ij}^* > 0$. Assume that it takes SA φ iterations to converge. Then, we know in any iteration $t (t \geq \varphi)$ of SA, $\lambda_{ij}^t = \lambda_{ij}^* > 0, Q_{ij}^t = Q_{il}^L = 0$, and $X_j^t = X_j^L = 1$. Therefore, it follows from (6) that $\lambda_{ij}^{t+1} = \lambda_{ij}^t - 1/t$ for any $t \geq \varphi$. By applying this equation recursively from iteration φ to iteration $\rho\varphi$ (where ρ is a positive integer), we have $\lambda_{ij}^{\rho\varphi+1} = \lambda_{ij}^{\varphi} - \sum_{t=\varphi}^{\rho\varphi} 1/t$. Since $\lambda_{ij}^{\rho\varphi+1} = \lambda_{ij}^{\varphi} = \lambda_{ij}^*$, and it is easy to see that $\sum_{t=\varphi}^{\rho\varphi} 1/t \to \infty$ when $\rho \to \infty$, a contradiction is reached. Thus, $X_j^L = 1$ and $Q_{ij}^L = 0$ can be true only when $\lambda_{ij}^* = 0$. ■

PROOF OF THEOREM 1: Since all constraints of Problem P except constraint (1) are also the constraints of Problem LR, to prove this theorem, we only need to show that $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ satisfies constraint (1) of Problem P, i.e., $1 - e^{-\beta Q_{ij}^L} \leq X_j^L\ \forall i \in I, \forall j \in J$.

When $X_j^L = 1, 1 - e^{-\beta Q_{ij}^L} \leq 1 = X_j^L$. So we only need to consider the case when $X_j^L = 0$. Since $\sum_{i \in I} \lambda_{ij}^* \leq f_j$ at this time (otherwise, $X_j^L = 1$), based on the relative values of f_j and $\sum_{i \in I} \lambda_{ij}^*$, we discuss the following two cases:

- $\sum_{i \in I} \lambda_{ij}^* < f_j$: then there exists a constant k, so that $\sum_{i \in I} \lambda_{ij}^t < f_j$, and hence $X_j^t = 0$ for any $t > k$. If there exists an $i \in I$ such that $Q_{ij}^L > 0$, it follows from Lemma 1 that there exists an s, $Q_{ij}^t > 0$ for any $t > s$. Then for any $t > \max\{s,k\}$, $Q_{ij}^t > 0$ and $X_j^t = 0$. By recursively applying (6), we derive

$$\lambda_{ij}^{t+1} \leq \lambda_{ij}^{\max(s,k)} + \inf\{1 - e^{-\beta Q_{ij}^l},\quad l = \max\{s,k\}, \ldots, t\} \sum_{l=\max\{s,k\}}^{t} b_{ij}^l.$$

  When $t \to \infty$, $\sum_{l=\max\{s,k\}}^{t} b_{ij}^l \to \infty$, but $\inf\{1 - e^{-\beta Q_{ij}^l}, l = \max\{s,k\}, \ldots, t\}$ is finite and positive, so $\lambda_{ij}^{t+1} \to \infty$ and hence $\lambda_{ij}^* \to \infty$, which contradicts $\sum_{i \in I} \lambda_{ij}^* < f_j$. Therefore, in this case, $Q_{ij}^L = 0$ for all $i \in I$. Hence, $1 - e^{\beta Q_{ij}^L} = 1 - 1 = X_j^L$.
- $\sum_{i \in I} \lambda_{ij}^* = f_j$: As discussed in Section 3.4, when $\sum_{i \in I} \lambda_{ij}^* = f_j$, we set X_j^L according to Q_{ij}^L, and $X_j^L = 0$ means $Q_{ij}^L = 0, \forall i \in I$. Thus, $1 - e^{-\beta Q_{ij}^L} = 1 - 1 = X_j^L \forall i \in I$. ∎

PROOF OF THEOREM 2: From Theorem 1, we know that $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ is a feasible solution to Problem P. Since Problem DP is the dual problem of Problem P, the smaller the value of $z(\boldsymbol{\lambda}^*) - (\phi(\boldsymbol{Q}^L) - \sum_{j \in J} f_j X_j^L)$, the better the approximation if we use $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ to estimate the optimal solution to Problem P. If the difference is equal to 0, $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ is an optimal solution to Problem P. Since $z(\boldsymbol{\lambda}^*) - (\phi(\boldsymbol{Q}^L) - \sum_{j \in J} f_j X_j^L) = \sum_{j \in J} \sum_{i \in I} \lambda_{ij}^* \{X_j^L - 1 + e^{-\beta Q_{ij}^L}\}$, we consider the following four cases,

1. $X_j^L = 0$ and $Q_{ij}^L = 0 : X_j^L - 1 + e^{-\beta Q_{ij}^L} = 0$;
2. $X_j^L = 0$ and $Q_{ij}^L > 0$: impossible according to Theorem 1;
3. $X_j^L = 1$ and $Q_{ij}^L = 0 : \lambda_{ij}^* = 0$ by Lemma 2;
4. $X_j^L = 1$ and $Q_{ij}^L > 0 : X_j^L - 1 + e^{-\beta Q_{ij}^L}$ is very close to 0, especially for large β.

Therefore,

$$z(\boldsymbol{\lambda}^*) - \Big(\phi(Q^L) - \sum_{j \in J} f_j X_j^L\Big) = \sum_{\{j \in J, X_j^L = 1\}} \sum_{\{i \in I, Q_{ij}^L > 0\}} \lambda_{ij}^* \cdot e^{-\beta Q_{ij}^L}$$
$$\leq e^{-\beta \min_{i \in I, j \in J, Q_{ij}^L > 0}\{Q_{ij}^L\}} \sum_{\{j \in J, X_j^L = 1\}} \sum_{\{i \in I, Q_{ij}^L > 0\}} \lambda_{ij}^* \quad (12)$$

Since $\lambda_{ij}^*\ \forall i \in I, j \in J$ is finite, it follows from (12) that by adjusting the value of β, $z(\boldsymbol{\lambda}^*) - (\phi(\boldsymbol{Q}^L) - \sum_{j \in J} f_j X_j^L)$ can be as close to 0 as possible to guarantee that $(\boldsymbol{X}^L, \boldsymbol{Q}^L)$ is a good approximation to (X^*, Q^*). In the special case when $\beta \to \infty$, $z(\boldsymbol{\lambda}^*) - (\phi(\boldsymbol{Q}^L) - \sum_{j \in J} f_j X_j^L) = 0$, and hence $(\boldsymbol{X}^L, \boldsymbol{Q}^L) = X^*, Q^*)$. ∎

PROOF OF PROPERTY 1: When $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, the first partial derivative of the objective function of Problem LR with respect to Q_{ij} is

$$\frac{\partial}{\partial Q_{ij}} L(\boldsymbol{\lambda}^*, \boldsymbol{X}, \boldsymbol{Q}) = -[(c + a_j)\chi + h_j/2] - \frac{K_j}{2\sqrt{\sum_{i \in I} Q_{ij}}} - \beta \lambda_{ij}^* e^{-\beta Q_{ij}} + \chi \frac{\partial}{\partial Q_{ij}} T_i(\boldsymbol{Q})$$

where

$$\frac{\partial}{\partial Q_{ij}} T_i(\boldsymbol{Q}) = (p_i + \pi_i - d_{ij}) E_{R_j}[R_j] - (p_i + \pi_i - v_i) E_{R_j}\left[R_j Pr\left(D_i \leq \sum_{j \in J} S_{ij}(Q_{ij}, R_j) | R_j\right)\right] \quad \text{(Dada et al. [6])}$$
$$= (p_i + \pi_i - d_{ij})\theta_j - (p_i + \pi_i - v_i) E_{R_j}\left[R_j Pr\left(D_i - \sum_{k \neq j} R_k Q_{ik} \leq R_j Q_{ij} | R_j\right)\right] \quad (13)$$

We derive the following KKT condition for $\boldsymbol{Q}^L$ from Problem LR:

$$0 = \left(\frac{\partial}{\partial Q_{ij}} L(\boldsymbol{\lambda}^*, \boldsymbol{X}, \boldsymbol{Q}) + u_{ij} \frac{\partial}{\partial Q_{ij}} Q_{ij}\right)\Bigg|_{Q = Q^L}$$
$$= u_{ij} - [(c + a_j)\chi + h_j/2] - \frac{K_j}{2\sqrt{\sum_{i \in I} Q_{ij}^L}} - \beta \lambda_{ij}^* e^{-\beta Q_{ij}^L} + \chi (p_i + \pi_i - d_{ij})\theta_j - \chi (p_i + \pi_i - v_i) \times E_{R_j}\left[R_j Pr\left(D_i - \sum_{k \neq j} R_k Q_{ik}^L \leq R_j Q_{ij}^L | R_j\right)\right] \quad i \in I, j \in J \quad (14)$$

$$u_{ij} Q_{ij}^L = 0 \quad i \in I, j \in J \tag{15}$$

$$Q_{ij}^L \geq 0 \quad i \in I, j \in J \tag{16}$$

$$u_{ij} \geq 0 \quad i \in I, j \in J \tag{17}$$

It follows from (14) that when $\beta \to \infty, u_{ij} - [(c+a_j)\chi + h_j/2] + \chi(p_i + \pi_i - d_{ij})\theta_j > 0$. If $Q_{ij}^L > 0$, it is easy to see from (15) that $u_{ij} = 0$, and hence

$$-(c + a_j) - h_j/2/\chi + (P_i + \pi_i - d_{ij})\theta_j > 0 \tag{18}$$

Since $\boldsymbol{Q}^L = \boldsymbol{Q}^*$ when $\beta \to \infty$, it follows that if $Q_{ij}^* > 0$ (that is, $Q_{ij}^L > 0$ when $\beta \to \infty$), then (18) must be satisfied. Thus, for any facility $j \in J$ and retailer $i \in I$, if $d_{ij} \geq (p_i + \pi_i) - (c + a_j)/\theta_j - h_j/2/\chi/\theta_j$, then retailer i will never be served by facility j in the optimal solution to Problem P. ■

PROOF OF PROPERTY 2: It follows from (14) that when $\beta \to \infty$ and $Q_{ij}^L > 0$ (i.e. $Q_{ij}^* > 0$), we have

$$-(c + a_j) - h_j/2/\chi + (p_i + \pi_i - d_{ij})\theta_j > (p_i + \pi_i - v_i) \times E_{R_j}\left[R_j Pr\left(D_i - \sum_{k \neq j} R_k Q_{ik}^L \leq R_j Q_{ij}^L | R_j\right)\right] \tag{19}$$

Since both R_j and $Pr\left(D_i - \sum_{k \neq j} R_k Q_{ik}^L \leq R_j Q_{ij}^L | R_j\right)$ are increasing functions of R_j, it follows from Theorem 4.7.9 in [4] that

$$E_{R_j}\left[R_j Pr\left(D_i - \sum_{k \neq j} R_k Q_{ik}^L \leq R_j Q_{ij}^L | R_j\right)\right] \geq \theta_j \times E_{R_j}\left[Pr\left(D_i - \sum_{k \neq j} R_k Q_{ik}^L \leq R_j Q_{ij}^L | R_j\right)\right] \tag{20}$$

With the above two inequalities, we derive that $-(c+a_j) - h_j/2/\chi + (P_i + \pi_i - d_{ij})\theta_j > (p_i + \pi_i - v_i)\theta_j E_{R_j}[Pr(D_i - \sum_{k \neq j} R_k Q_{ik}^L \leq R_j Q_{ij}^L | R_j)]$. On the other hand, it is easy to see that in the case when $\beta \to \infty, E_{R_j}[Pr(D_i - \sum_{k \neq j} R_k Q_{ik}^L \leq R_j Q_{ij}^L | R_j)]$ represents the expected percentage of the customer demand fulfilled at retailer i. Therefore we conclude that when $Q_{ij}^* > 0$ ($\beta \to \infty$ and $Q_{ij}^L > 0$), the expected percentage of the customer demands fulfilled at retailer i is less than $\frac{-(c+a_j) - \frac{h_j}{2\chi} + (p_i + \pi_i - d_{ij})\theta_j}{(p_i + \pi_i - v_i)\theta_j}$. ■

PROOF OF LEMMA 3: Assume there exists an optimal solution to the generalized problem, say $(\boldsymbol{Q}', s')$, with $s' > H(\boldsymbol{Q}')$. Consider a feasible solution $(\boldsymbol{Q}', s'')$, where $s'' = H(\boldsymbol{Q}') < s'$. Since $G(\boldsymbol{Q}, s)$ is an increasing function of both $\boldsymbol{Q}$ and s, $G(\boldsymbol{Q}', s'') < G(\boldsymbol{Q}', s')$, which contradicts the assumption that $(\boldsymbol{Q}', s')$ is an optimal solution to the problem. Thus, we conclude that if $(\boldsymbol{Q}^G, s^G)$ is an optimal solution to the generalized problem, it must satisfy $s^G = H(\boldsymbol{Q}^G)$. ■

PROOF OF COROLLARY 1: Corollary 1 follows from the fact that $G(\boldsymbol{Q}, s)$ is an increasing function of both $\boldsymbol{Q}$ and s, and $H(\boldsymbol{Q})$ is a decreasing function of $\boldsymbol{Q}$. ■

PROOF OF LEMMA 4: This lemma naturally follows from the fact that $G(\boldsymbol{Q}, s)$ is an increasing function of both $\boldsymbol{Q}$ and s, and $H(\boldsymbol{Q})$ is a decreasing function of $\boldsymbol{Q}$. ■

PROOF OF THEOREM 3: Since the sequence of $\{G^k\}$ is nondecreasing (Corollary 2 part a) and upper bounded (because $G^k \leq G(\boldsymbol{Q}, s)\ \forall(\boldsymbol{Q}, s) \in \mathcal{D}$), it must converge to a certain value when $k \to \infty$. We use G^* to denote this value, i.e., $\lim_{k\to\infty} G^k = G^*$. To prove this theorem, we need to show that G^* is a global minimum.

Supposing G^* is not a global minimum. Therefore, G^* must be less than the global minimum (since G^* is a smallest lower bound in the lower bound set), and the solutions corresponding to G^* are infeasible. We define set V_{G^*} that contains all elements in V equal to G^* in the iteration when G^k converges. Let $\rho = |V_{G^*}|$. It is easy to see that G^* is less than any element in set $V \backslash V_{G^*}$ at this time. On the other hand, it follows from Corollary 2 part b that all new lower bounds added into the lower bound set during the ρ iterations after G^k converges are bigger than G^*. Therefore, after these ρ iterations, G^k is bigger than G^*, which contradicts $\lim_{k\to\infty} G^k = G^*$. Thus, G^* is a global minimum. ■

PROOF OF PROPERTY 3: Based on Lemma 3, the generalized problem has the same optimal solution as the problem {Minimize $G(\boldsymbol{Q}, H(\boldsymbol{Q}))$, s.t. $\boldsymbol{Q}_{ij} \geq 0\ i \in I,\ j \in J$}, whose KKT conditions

$$\frac{\partial}{\partial Q_{ij}} G(\boldsymbol{Q}, H(\boldsymbol{Q})) - u_{ij} = 0 \quad i \in I, j \in J$$

$$u_{ij} Q_{ij} = 0 \quad i \in I, j \in J$$

$$Q_{ij} \geq 0 \quad i \in I, j \in J$$

$$u_{ij} \geq 0 \quad i \in I, j \in J$$

Therefore, we have

$$0 = u_{i'j'} - u_{ij} + \left\{\frac{\partial}{\partial Q_{ij}} G(\boldsymbol{Q}, H(\boldsymbol{Q})) - \frac{\partial}{\partial Q_{i'j'}} G(\boldsymbol{Q}, H(\boldsymbol{Q}))\right\}\bigg|_{\boldsymbol{Q}=\boldsymbol{Q}^G}.$$

Since $0 \leq Q_{ij}^G < Q_{i'j'}^G$ and $u_{i'j'} Q_{i'j'}^G = 0$, we know that $u_{i'j'} = 0$, and

$$\left\{\frac{\partial}{\partial Q_{ij}} G(\boldsymbol{Q}, H(\boldsymbol{Q})) - \frac{\partial}{\partial Q_{i'j'}} G(\boldsymbol{Q}, H(\boldsymbol{Q}))\right\}\bigg|_{\boldsymbol{Q}=\boldsymbol{Q}^G} \geq 0.$$

■

ACKNOWLEDGMENTS

We want to thank Joe Geunes, Richard Francis, and Larry Snyder for constructive comments that led to this improved version. This work was supported in part by NSF CAREER award DMI-0348209. This support is gratefully acknowledged.

REFERENCES

[1] E. Berk and A. Arreola-Risa, Note on "future supply uncertainty in EOQ models", Nav Res Logist 41 (1994), 129–132.

[2] S. Bollapragada and T.E. Morton, Myopic heuristics for the random yield problem, Oper Res 47 (1999), 713–722.

[3] M. Bundschuh, D. Klabjan, and D.L. Thurston, Modeling robust and reliable supply chains, Working paper, University of Illionis at Urbana-Champaign, 2006.

[4] G. Casella and R. Berger, Statistical inference, 2nd ed., Duxbury/Thomason Learing, Pacific Grove, CA, 2002.

[5] S. Chopra, U. Mohan, and G. Reinhardt, The importance of decoupling recurrent and disruption risks in a supply chain, Working paper, Kellogg School of Management, Northwestern University, 2006.

[6] M. Dada, N. Petruzzi, and L.B. Schwarz, A newsvendor model with unreliable suppliers, Working paper, University of Illinois at Urbana-Champaign, 2003.

[7] M. Daskin, C. Coullard, and Z.-J. Shen, An inventory-location model: Formulation, solution algorithm and computational results, Ann Oper Res 110 (2002), 83–106.

[8] S.J. Erlebacher and R.D. Meller, the interaction of location and inventory in designing distribution systems, IIE Trans 32 (2000), 155–166.

[9] D. Gupta, The (q, r) Inventory System with an Unreliable Supplier, *INFOR* 34 (1996), 59–76.

[10] R. Horst, P.M. Pardalos, and N.V. Thoai, Introduction to global optimization, Kluwer Academic Publishers, Boston, 2000.

[11] R. Horst and H. Tuy, Global optimization, Springer-Verlag, New York, 1990.

[12] M. Parlar and D. Berkin, Future supply uncertainty in EOQ models, Nav Res Logist 38 (1991), 107–121.

[13] M. Parlar and D. Perry, Analysis of a (q, r, t) inventory policy with deterministic and random yields when future supply is uncertain, Eur J Oper Res 84 (1995), 431–443.

[14] M. Parlar, Continuous-review inventory problem with random supply interruptions, Eur J Oper Res 99 (1997), 366–385.

[15] Y. Sheffi, Supply chain management under the threat of international terrorism, Int J Logist Manag 12 (2001), 1–11.

[16] Z.-J. Shen, Efficient algorithms for various supply chain problems, Ph.D. Thesis, Department of Industrial Engineering and Management Sciences, Northwestern University, Evansto, IL, 2000.

[17] Z.-J. Shen, C. Coullard, and M. Daskin, A joint location-inventory model, Transport Sci 37 (2003), 40–55.

[18] N.Z. Shor, Minimization methods for non-differentiable functions, Springer-Verlag, New York, 1985.

[19] J. Shu, C.-P. Teo, and Z.-J. Shen, Stochastic transportation-inventory network design problem, Oper Res 53 (2005), 48–60.

[20] L.V. Snyder, A tight approximation for a continuous-review inventory model with supplier disruptions, Working paper, Lehigh University, 2006.

[21] L.V. Snyder and M. Daskin, Reliability models for facility location: The expected failure cost case, Transport Sci 39 (2005), 400–416.

[22] L.V. Snyder, M. Daskin, and C.-P. Teo, The stochastic location model with risk pooling, Eur J Oper Res 179 (2007), 1221–1238.

[23] B.T. Tomlin, On the value of mitigation and contingency strategies for managing supply-chain disruption risks, Manag Sci 52 (2006), 639–657.

[24] C.J. Vidal and M. Goetschalckx, Modeling the effect of uncertainties on global logistics systems, J Business Logist 21 (2000), 95–120.

[25] C.A. Yano and H.L. Lee, Lot sizing with random yields: A review, Oper Res 43 (1994), 311–334.