

This article was downloaded by: [HKUST Library]

On: 11 September 2014, At: 21:01

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



IIE Transactions

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/uiie20>

Risk diversification and risk pooling in supply chain design

Ho-Yin Mak^a & Zuo-Jun (Max) Shen^b

^a Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

^b Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, CA, 94720, USA

Published online: 01 Jun 2012.

To cite this article: Ho-Yin Mak & Zuo-Jun (Max) Shen (2012) Risk diversification and risk pooling in supply chain design, IIE Transactions, 44:8, 603-621, DOI: [10.1080/0740817X.2011.635178](https://doi.org/10.1080/0740817X.2011.635178)

To link to this article: <http://dx.doi.org/10.1080/0740817X.2011.635178>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Risk diversification and risk pooling in supply chain design

HO-YIN MAK^{1,*} and ZUO-JUN (MAX) SHEN²

¹*Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong*
E-mail: hymak@ust.hk

²*Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, CA 94720, USA*

Received July 2010 and accepted October 2011

Recent research has pointed out that the optimal strategies to mitigate supply disruptions and demand uncertainty are often mirror images of each other. In particular, risk diversification is favorable under the threat of disruptions and risk pooling is favorable under demand uncertainty. This article studies how dynamic sourcing in supply chain design provides partial benefits of both strategies. Optimization models are formulated for supply chain network design with dynamic sourcing under the risk of temporally dependent and temporally independent disruptions of facilities. Using computational experiments, it is shown that supply chain networks that allow small to moderate degrees of dynamic sourcing can be very robust against both disruptions and demand uncertainty. Insights are attained on the optimal degree of dynamic sourcing under different conditions.

Keywords: Supply chain design, facility location, disruption management, inventory sharing, stochastic programming

1. Introduction

Supply chain design is a critical and difficult process, involving long-term decisions such as which products to manufacture, which markets to serve, what technology to acquire or develop, which suppliers to source from, where to locate plants and warehouses, and which facilities to use to serve different markets. One major difficulty encountered in the process is that such long-term strategic decisions are often costly or even impossible to reverse, while they must be made under uncertainty in the future business environment, including the uncertainty of future demand and supply, competitive environment, tax rates, and government policy, among others. It is critical that all of these factors and the associated impacts on profits and costs are assessed carefully in the strategic supply chain design phase. Unfortunately, the unavailability of accurate forecasts for the tactical and operational phases (e.g., demand information, fuel costs) makes this task a challenge.

Researchers have studied the use of various strategies to mitigate the effects of uncertainty. One of the most well-known and widely adopted practices to temper the effects of demand uncertainty is centralization or risk pooling (Eppen, 1979). By aggregating demand, it is possible to reduce the coefficient of variation of demand and, therefore, safety stock costs. On the other hand, Snyder and Shen (2008)

show that risk diversification should be favored under the threat of disruptions. When both types of uncertainty are present, the optimal strategy should be based on a trade-off between these two opposing forces. Classical risk pooling requires centralization of inventory at a central facility handling a large demand volume. This lowers inventory holding costs while intensifying the risk of losing (or making temporarily unavailable) all inventory when there is a disruption. In contrast, risk diversification suggests stocking at multiple warehouses such that the impact of a disruption at any single warehouse is not catastrophic. However, since each warehouse handles a smaller volume of demand, inventory costs will be higher.

Mak and Shen (2010) introduce the dynamic multiple sourcing strategy in integrated supply chain network design. Instead of being a strategic decision, the assignment of retailers to Distribution Centers (DCs) is deferred to be an operational decision, leading to inventory sharing benefits. In this article, we show that the dynamic sourcing strategy can be even more beneficial under both demand uncertainty and the threat of disruptions. Since dynamic sourcing allows inventory sharing while operating multiple DCs, it provides both risk pooling and risk diversification. Therefore, a supply chain network design with dynamic sourcing can perform much better than one with a static sourcing strategy when both types of uncertainty are present.

The remainder of this article is organized as follows. We first review some related literature in Section 2 and provide a detailed problem description and the basic formulation

*Corresponding author

in Section 3. Then, in Section 4, we focus on the case where disruptions are temporally independent. Note that this is also the standard assumption made in all existing location models in the literature with disruption considerations. In Section 5, we discuss how it is possible to modify our model to relax this constraint. Finally, in Sections 6 and 7, we present our computational results and draw conclusions.

2. Related literature

In recent years, researchers have begun taking an integrated approach to supply chain design. They have developed models that integrate facility location models with tactical and operational-level decisions such as inventory management and vehicle routing. Shen *et al.* (2003) and Daskin *et al.* (2002) study a facility location problem in which inventory is controlled with a continuous (r, Q) policy. The model captures the economies-of-scale effect (Eppen, 1979) since safety stock can be reduced by pooling demand from multiple retailers at a single DC. Their computational results show that if location and inventory decisions are optimized separately, the number of facilities opened will be more than is optimal.

Extending the Shen *et al.* (2003) study, Ozsen *et al.* (2008, 2009) consider the effect of storage capacity at facilities under single and multiple sourcing. Unlike the traditional approach in which capacity is defined as the maximum demand that can be assigned to a facility (e.g., Geoffrion and McBride (1978)), the authors impose an upper bound on the amount of space that can be occupied by inventory at a facility.

While the models discussed above point out the importance of the integrated planning approach, only first-order inventory sharing has been studied. Besides, all of the mentioned papers assume that each retailer is served by at most one DC, with the exception of Ozsen *et al.* (2009), which allows multiple sourcing in static proportions to achieve better capacity utilization. Mak and Shen (2010) study a dynamic sourcing strategy, under which sourcing proportions can be adjusted dynamically over time, subject to realized demand levels. They discuss how such a strategy allows second-order inventory sharing; i.e., the virtual sharing of inventory across multiple stocking locations. In their terminology, first-order inventory sharing refers to classic risk pooling at a single stocking location.

Recall from Shen *et al.* (2003) that first-order inventory sharing encourages the location of fewer, but larger, facilities, due to economies of scale. Although inventory costs can be reduced by such a location strategy, transportation costs will inevitably increase, because each facility has to serve demand in a larger geographical region. Second-order inventory sharing allows one to better balance the two costs, because one may still locate a large number of small facilities and utilize dynamic sourcing to reduce inventory costs. It is clear that network design considerations

must be considered differently under second-order inventory sharing than under first-order sharing. Mak and Shen (2011a) further study such differences using an analytical approach to complement the computational studies in Mak and Shen (2010). For a review of this stream of literature, readers may refer to Shen (2007) and Mak and Shen (2011b).

All of the papers previously mentioned assume that facilities never fail and focus only on demand uncertainty. Snyder and Shen (2008) explore the differences between the supply and demand uncertainty in supply chains and show that the optimal strategies for coping with the two types of uncertainty might be exactly opposite to each other. Therefore, there is a need to study the optimal supply chain design under the threat of disruptions.

Snyder (2006) provides an excellent review of the earlier works on facility location under uncertainty. More recently, Snyder and Daskin (2005) studied the two-objective facility location model considering both normal operating costs and the expected cost of disruptions. In the model, each customer is assigned to facilities as backups at different levels. The customer is served by an assigned facility if all of the facilities assigned at lower levels have been disrupted. This backup assignment can be viewed as a contingency strategy that is only used when a disruption happens. While the expected failure cost objective reflects the operating cost incurred by this contingency strategy, the model does not consider the design cost required to enable such a strategy.

There have been several extensions to Snyder and Daskin's model. Shen *et al.* (2011) relax the assumption of uniform failure probabilities, formulate the stochastic fixed-charged problem as a non linear mixed integer problem, and provide several heuristic solution algorithms. Cui *et al.* (2010) provide a linear integer programming formulation and an efficient algorithm for the problem with unequal disruption probabilities. They also compare the integer programming model to a continuum approximation model for the same problem.

Berman *et al.* (2007) study the network facility location problem under disruption risks where facilities can have different failure probabilities. They show conditions under which the Hakimi property (i.e., there exists an optimal solution in which all facilities are located at nodes but not along arcs) holds and study asymptotic properties of the model. Heuristics and worst-case analysis are also provided. Lim *et al.* (2009, 2010) study a facility hardening problem where there is an option to invest to reduce the probability of failure.

None of the previously discussed work on disruptions considers inventory costs. As previously mentioned, risk pooling and risk diversification are "opposite" strategies that work well under demand and supply disruptions, respectively. Models without inventory cost considerations do not capture risk pooling effects. For a one-warehouse, multiple-retailer setting, Schmitt *et al.* (2011) show that it is typically optimal for a risk-averse manager to choose

a decentralized (risk diversification) configuration over a centralized (risk pooling) one.

Snyder *et al.* (2007) extend the location–inventory model of Shen *et al.* (2003) to a scenario-based formulation to handle uncertainty. They assume that cost and demand parameters, including the mean and standard deviation of demand, are uncertain and can be characterized by a set of discrete scenarios with known probability. The facility location decisions are made before observing the realization of the random parameters, and a (Q, r) inventory policy is optimized after the uncertainty is resolved. This is a rather favorable assumption, since it allows tactical inventory control decisions to be made under complete knowledge of the realization of uncertain parameters. Qi and Shen (2007) consider a supply chain design model where the distribution centers' yields are random. Qi *et al.* (2010) further study a problem in which both suppliers and distribution centers are subject to disruption risks. Although the inventory control decisions are made with considerations of disruption threats, the model does not include the possibility of rerouting demand once disruption occurs. As shown by Mak and Shen (2010), dynamic rerouting of demand in a supply chain, or dynamic sourcing, can be effective in coping with demand uncertainty. In this article, we study how the same strategy helps under the threat of disruptions and how the supply chain network should be designed to allow dynamic sourcing.

3. Problem description and basic formulation

Our goal is to design a supply chain to serve a number of retailers with known locations and random demand. We are given a set of candidate sites, a subset of which is to be selected for locating DCs. Following the standard assumption in location theory, we assume that a (site-specific) fixed cost is incurred for locating each DC. The problem can be divided into two phases. In the first stage (we shall refer to it as the *design* phase throughout this article), the network structure—i.e., locations of DCs and the allocation of retailers to DCs—are to be determined. Then, in the second (*management*) phase, the inventory and shipment decisions are made for a number of periods given the network structure determined in the design stage. In the *management* phase, each candidate site can be disrupted in any period of the operating horizon with known probability, and the demand at retailers follows some known probability distribution.

In addition to the number and locations of DCs, the *design* decisions also include connecting the opened DCs to retailers. An arc must be constructed in the *design* phase before it can be used in the *management* phase. Therefore, the set of arcs constructed determines the flexibility of the network for dynamic sourcing. Following the assumption made by Mak and Shen (2010), we use a fixed cost to capture the physical expenses and managerial disutility of making the network more complex.

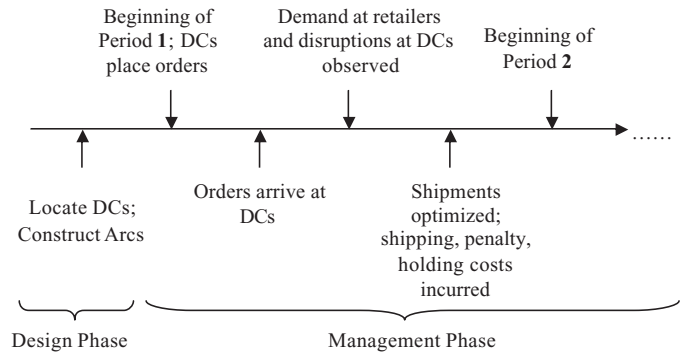


Fig. 1. Sequence of events.

After the *design* decisions are made, we move on to the *management* phase, which consists of a number of periods (the operating horizon). Each period begins with the DCs placing replenishment orders, which have zero lead time. Since orders are placed before observing whether or not disruptions occur, it is possible that a DC gets disrupted after receiving a shipment. After the shipments arrive, demand at the retailers and the subset of DCs disrupted (if any) are observed. Disruptions do not damage inventory, but rather make it temporarily unavailable for an entire period. Then, the shipments from DCs to retailers are optimized over the constructed network, minimizing the sum of shipping, shortage, and holding cost of any leftover inventory. Any unmet demand is assumed to be lost, incurring some penalty cost. This sequence of events is summarized in Fig. 1.

We define the following notation:

Sets

I : set of retailers;

J : set of candidate DC locations.

Cost and other parameters

f_j : fixed (annualized) cost of opening a DC at j ;

c_{ji} : fixed (annualized) cost of creating an arc connecting DC j and retailer i ;

C_j : maximum storage capacity of candidate facility location j ;

T : number of periods in management stage;

$\mathcal{D}_{t,i}$: demand in period t at retailer i , assumed to be independent and identically distributed across periods;

$\mathcal{D}_t \equiv (\mathcal{D}_{t,1}, \dots, \mathcal{D}_{t,|I|})$: demand vector in period t ;

$\mathfrak{d}_t \equiv (\mathfrak{d}_{t,1}, \dots, \mathfrak{d}_{t,|I|})$: realization of demand vector \mathcal{D}_t in period t ;

$\mathcal{S}_{t,j}$: binary disruption indicator in period t : a value of one indicates that the site j is working and a value of zero indicates that the site is disrupted, assumed to be random;

$\mathcal{S}_t \equiv (\mathcal{S}_{t,1}, \dots, \mathcal{S}_{t,|J|})$: disruption vector in period t ;

$\mathfrak{s}_t \equiv (\mathfrak{s}_{t,1}, \dots, \mathfrak{s}_{t,|J|})$: realization of disruption vector \mathcal{S}_t in period t .

Decision variables

X_j : = 1 if we open DC at j , 0 otherwise;

\mathbf{X} : the vector of X_j variables;

Z_{ji} : = 1 if we build an arc between DC j and retailer i , 0 otherwise;

\mathbf{Z} : the matrix of Z_{ji} variables.

The supply chain design problem with dynamic sourcing under disruptions (SCDDSD) can be formulated as follows:

$$(\text{SCDDSD}) : V = \min \sum_{j \in J} f_j X_j + \sum_{i \in I} \sum_{j \in J} c_{ij} Z_{ij} + v(\mathbf{X}, \mathbf{Z}),$$

subject to:

$$\begin{aligned} X_j &\geq Z_{ij} && \text{for each } i \in I, j \in J, \\ X_j &\in \{0, 1\} && \text{for each } j \in J, \\ Z_{ij} &\in \{0, 1\} && \text{for each } i \in I, j \in J. \end{aligned}$$

In this formulation, $v(\cdot)$ is defined to be the expected recourse function that includes the inventory, shipping, and shortage costs in the management stage. The formulation and properties of the recourse problem will be discussed in detail in the next section. The objective of the two-stage problem is to minimize the design-stage fixed costs of locating facilities and constructing arcs plus the expected value of the management-stage inventory, shipping, and shortage costs.

The only constraints in the design stage are the requirements that arcs cannot be constructed unless the corresponding DC is located and the binary constraints. Note that we do not impose the restriction that every retailer must be connected to at least one DC but rather allow the optimal subset of open retailers be determined by the force of shortage costs. This treatment allows the possibility of maximizing supply chain profit by selecting the subset of profitable retailers, rather than minimizing the cost of serving all retailers. For a more thorough discussion of these contrasting settings, we refer the reader to Shen (2006).

The above formulation of (SCDDSD) is generic. By modifying the function $v(\cdot)$, it is possible to capture different types of disruptions. In the upcoming sections, we discuss two possibilities. In Section 4, we first follow the standard assumption in the existing literature that disruptions are independent over time. Then, in Section 5, we discuss how it is possible to generalize the model to deal with temporally dependent disruptions.

4. Temporally independent disruptions

4.1. Model formulation

In this section, we analyze the case where disruptions are temporally independent. In this setting, the occurrence of disruptions in one period does not affect the probability of disruptions in future periods. This assumption is valid if the

planning periods are long enough (e.g., a month) or if disruptions are short-term (e.g., adverse weather conditions like snowstorms). Under this setting, the disruption indicators $\mathfrak{S}_{t,j}$ follow Bernoulli distributions. Many models in the literature assume the probability of disruption to be identical for all sites (e.g., Snyder and Daskin (2005)), but we allow the probability to vary across different locations. To summarize:

Assumption 1. The sequence of random vectors $(\mathfrak{S}_1, \dots, \mathfrak{S}_T)$ is independent and identically distributed.

To formulate the management phase cost function, we need the following additional notation.

Sets

$J(\mathbf{X})$: set of open DCs given \mathbf{X} as determined in design phase;

$A(\mathbf{Z})$: set of constructed arcs connecting retailers and DCs given \mathbf{Z} .

Cost and demand parameters

α : discount factor per period;

c : purchasing cost for one unit of stock;

$\hat{p}_i(s)$: lost sales cost for losing s units of retailer i 's demand;

$p_i(s) \equiv \hat{p}_i(s) - cs$: effective penalty cost for losing s units of retailer i 's demand;

\hat{h}_j : holding cost per unit per period;

$h_j \equiv \hat{h}_j + (1 - \alpha)c$: effective holding cost per unit per period;

d_{ji} : shipping cost per unit from DC j to retailer i ;

q_j : probability that facility $j \in J(\mathbf{X})$ is disrupted.

Decision variables in each period before observing demand

$y_{t,j}$: order-up-to level at DC j at the beginning of period t ;

$\mathbf{y}_t \equiv (y_{t,1}, \dots, y_{t,|J(\mathbf{X})|})$: Vector of order-up-to levels at opened DCs in period t .

Recourse decision variables after observing demand

$w_{t,ji}$: shipment from DC j to retailer i in period t ;

$s_{t,i}$: shortage to retailer i in period t ;

$x_{t,j}$: ending inventory in period $t - 1$ at DC j ; i.e., beginning inventory in period t ;

$\mathbf{x}_t \equiv (x_{t,1}, \dots, x_{t,|J(\mathbf{X})|})$: vector of ending inventory in period $t - 1$ at DC j ; i.e., beginning inventory in period t .

Facing the risk of disruption events with low probability but devastating impact, it is natural to consider risk-averse objectives. Therefore, we assume that the lost sales cost at each retailer is increasing and convex.

Assumption 2. The lost sales cost function is $\hat{p}_i(s) = \tilde{p}_i(s) + \underline{p}_i s$. In addition, $\tilde{p}_i(s)$ is increasing, convex, equal to zero at $s = 0$, and differentiable for $s \geq 0$, and $\underline{p}_i > c$.

Minimizing the expected value of a convex cost function is equivalent to maximizing a risk-averse (concave) utility

function. In the inventory literature, linear shortage cost functions are common. A disadvantage of a linear shortage cost function in our setting is that it does not distinguish between losing tiny amounts of demand at every retailer (which is often acceptable) and losing all of the demand at a single retailer (which definitely should be avoided). In contrast, an increasing convex shortage cost function implies that the marginal cost of shortage is increasing and thus discourages large shortage quantities at individual retailers. The requirement that $\underline{p}_i > c$ ensures that it is never beneficial to intentionally incur shortages to save purchase costs.

To simplify the optimal inventory policy, we retain the following assumption made by Mak and Shen (2010):

Assumption 3. It is possible to sell any remaining inventory at the end of the planning horizon at purchase cost c .

In each earlier period t where $t \leq T$, beginning with inventory level \mathbf{x}_t , each DC makes an order to minimize its expected (discounted) costs in the current and remaining periods. Using the inventory level vector \mathbf{x}_t at the end of each period as the state variable, we may formulate the optimality equations as follows:

$$\begin{aligned} v(\mathbf{X}, \mathbf{Z}) &= \hat{f}_1(\mathbf{0}), \\ \hat{f}_t(\mathbf{x}_t) &\equiv \min_{\mathbf{y}_t \geq \mathbf{x}_t} \hat{G}_t(\mathbf{y}_t | \mathbf{x}_t), \\ \hat{G}_t(\mathbf{y}_t | \mathbf{x}_t) &\equiv E[\hat{H}_t(\mathbf{y}_t, \mathcal{D}_t, \mathcal{S}_t | \mathbf{x}_t)], \\ \hat{H}_t(\mathbf{y}_t, \mathcal{D}_t, \mathcal{S}_t | \mathbf{x}_t) &\equiv \min_{\mathbf{x}_{t+1}, \mathbf{w}_t, \mathbf{s}_t \geq 0} \sum_{j \in J(\mathbf{X})} c(y_{t,j} - x_{t,j}) \\ &\quad + \sum_{(i,j) \in A(\mathbf{Z})} d_{ij} w_{t,ij} + \sum_{i \in I} \hat{p}_i(s_{t,i}) \\ &\quad + \sum_{j \in J(\mathbf{X})} \hat{h}_j x_{t+1,j} + \alpha f_{t+1}(\mathbf{x}_{t+1}), \end{aligned}$$

subject to:

$$x_{t+1,j} + \sum_{i \in I} \mathfrak{s}_{t,j} w_{t,ji} = y_{t,j} \text{ for each } j \in J(\mathbf{X}), \quad (1)$$

$$\sum_{j \in J(\mathbf{X})} \mathfrak{s}_{t,j} w_{t,ji} + s_{t,i} = \mathfrak{d}_{t,i} \text{ for each } i \in I. \quad (2)$$

In the presented recourse problem defining $\hat{H}_t(\cdot)$, Equations (1) and (2) are the flow balance constraints. Constraints (1) require the sum of all shipments out of every DC and the leftovers to equal the inventory level before sending shipments. Constraints (2) require that the sum of all shipments received by a retailer and the shortage to equal the realized demand. The shipment variables $w_{t,ji}$ in our disruption model are multiplied by the corresponding disruption indicator $\mathfrak{s}_{t,j}$. Therefore, if a DC is disrupted, it cannot make any shipments to reduce the inventory level or satisfy any demand. Note that with the positive cost coefficient for $w_{t,ji}$, all of the shipment variables corresponding to disrupted DCs will equal zero at optimality.

The function $\hat{f}_t(\mathbf{x}_t)$ denotes the cost-to-go from period t to the end of the horizon, given a starting inventory level vector \mathbf{x}_t , assuming that optimal stocking decisions are made from period t onwards. Function $\hat{G}_t(\mathbf{y}_t | \mathbf{x}_t)$ represents the cost-to-go from period t to the end of the horizon, assuming that we order up to the vector \mathbf{y}_t in period t and behave optimally from period $t+1$ onwards. $\hat{H}_t(\mathbf{y}_t, \mathcal{D}_t, \mathcal{S}_t | \mathbf{x}_t)$ represents the cost-to-go given that we have ordered up to \mathbf{y}_t in period t and the demand-disruption scenario $(\mathcal{D}_t, \mathcal{S}_t)$ has been realized.

The presented inventory problem formulation considers the actual costs incurred by stocking decisions. It is possible to convert it into an equivalent formulation that considers the minimization of incremental costs over purchase costs to satisfy demand. In particular, note that in every period, $\sum_{j \in J(\mathbf{X})} y_{t,j} = \sum_{i \in I} (\mathfrak{d}_{t,i} - s_{t,i}) + \sum_{j \in J(\mathbf{X})} x_{t+1,j}$. Therefore, the purchase cost component can be rewritten as

$$\begin{aligned} \sum_{j \in J(\mathbf{X})} c(y_{t,j} - x_{t,j}) &= c \sum_{i \in I} \mathfrak{d}_{t,i} - \sum_{i \in I} c s_{t,i} \\ &\quad + \sum_{j \in J(\mathbf{X})} c x_{t+1,j} - \sum_{j \in J(\mathbf{X})} c x_{t,j}. \end{aligned}$$

Note that the first term on the right of the equality sign does not depend on our decision variables. The second and third terms can be combined with the penalty cost and holding cost terms, respectively. The last term is independent of decisions in period t and should be considered in period $t-1$. With Assumption 3, we may then obtain the following equivalent formulation in which the incremental costs are minimized:

$$v(\mathbf{X}, T) = f_1(0) + c \sum_{t=1}^T \sum_{j \in J(\mathbf{X})} \alpha^{t-1} E[\mathcal{D}_{t,j}],$$

$$f_t(\mathbf{x}_t) \equiv \min_{\mathbf{y}_t \geq \mathbf{x}_t} G_t(\mathbf{y}_t),$$

$$G_t(\mathbf{y}_t) \equiv E[H_t(\mathbf{y}_t, \mathcal{D}_t, \mathcal{S}_t)],$$

$$\begin{aligned} H_t(\mathbf{y}_t, \mathcal{D}_t, \mathcal{S}_t) &\equiv \min_{\mathbf{x}_{t+1}, \mathbf{w}_t, \mathbf{s}_t \geq 0} \sum_{(i,j) \in A(\mathbf{Z})} d_{ij} w_{t,ij} + \sum_{i \in I} p_i(s_{t,i}) \\ &\quad + \sum_{j \in J(\mathbf{X})} h_j x_{t+1,j} + \alpha f_{t+1}(\mathbf{x}_{t+1}), \end{aligned}$$

subject to:

$$x_{t+1,j} + \sum_{i \in I} \mathfrak{s}_{t,j} w_{t,ji} = y_{t,j} \text{ for each } j \in J(\mathbf{X}),$$

$$\sum_{j \in J(\mathbf{X})} \mathfrak{s}_{t,j} w_{t,ji} + s_{t,i} = \mathfrak{d}_{t,i} \text{ for each } i \in I,$$

where $h_j \equiv \hat{h}_j + (1 - \alpha)c$ is the effective holding cost; i.e., actual holding cost plus the opportunity cost of purchasing unused inventory one period early, and $p_i(s) \equiv \hat{p}_i(s) + \hat{p}_i s = \tilde{p}_i(s) + (\underline{p}_i - c)s$ is the effective penalty cost; i.e., actual penalty cost minus the savings in purchase cost due

to losing sales. Note that $H_1(\mathbf{y}_t, \mathfrak{D}_t, \mathfrak{s}_t) = \hat{H}_1(\mathbf{y}_t, \mathfrak{D}_t, \mathfrak{s}_t|0) - c \sum_{t=1}^T \sum_{j \in J(\mathbf{X})} \alpha^{t-1} E[\mathfrak{D}_{t,j}]$.

Mak and Shen (2010) show that a stationary base stock policy is optimal for the inventory control problem given any (\mathbf{X}, \mathbf{Z}) , without considering the possibility of disruptions. One may check that all steps of the proof will still go through even assuming non-zero disruption probabilities at candidate facility locations. Therefore:

Proposition 1. *Given any (\mathbf{X}, \mathbf{Z}) , there exists a base stock-level vector \mathbf{y}^* that is optimal for all $t = 1, \dots, T$; i.e., a stationary base stock policy is optimal.*

This result suggests the sufficiency of considering one single period in the management phase, since the cost of multiple periods can be obtained by scaling. In particular, the pre-discount expected (incremental) cost incurred in every period will be equal to $G_T(\mathbf{y}^*)$. Accounting for the discount factor, we know that $G_t(\mathbf{y}^*) = \sum_{l=t}^T \alpha^{l-t} G_T(\mathbf{y}^*) = (1 - \alpha^{T-t})/(1 - \alpha)$. Therefore, the cost of a T -period problem is $(1 - \alpha^{T-1})/(1 - \alpha)$ times that of a one-period problem. Note that the result does not hold in the absence of the assumption of temporal independence of disruptions. This is the major source of difficulty of solving the problem with temporally dependent disruptions.

By Proposition 1, we may consider the case where there is a single period in the management phase for the remainder of Section 4. For notational convenience, the subscript t will be temporarily dropped. Before discussing the proposed solution algorithm, we first provide an equivalent formulation with only one period in the management phase. For notational convenience, let Ω denote the set of possible realizations of the random vectors \mathfrak{D} and \mathfrak{S} in the single period. Furthermore, for realization $\omega \in \Omega$, let $\mathfrak{D}(\omega)$ and $\mathfrak{S}(\omega)$ denote the values of demand and disruption vectors corresponding to the realization. Finally, the recourse decision variables now depend on ω as well, as their values are determined after observing $\mathfrak{D}(\omega)$ and $\mathfrak{S}(\omega)$.

(SCDDSD'): min

$$\sum_{j \in J} f_j X_j + \sum_{i \in I} \sum_{j \in J} c_{ji} Z_{ji} + E \left[\sum_{i \in I} \sum_{j \in J} (d_{ji} - h_j) w_{ji}(\omega) + \sum_{i \in I} p_i \left(\mathfrak{D}_i(\omega) - \sum_{j \in J} w_{ji}(\omega) \right) + \sum_{j \in J} h_j y_j \right] + c \sum_{j \in J(\mathbf{X})} E[\mathfrak{D}_j(\omega)],$$

subject to:

$$\begin{aligned} X_j &\geq Z_{ji} \quad \text{for each } i \in I, j \in J, \\ y_j &\leq C_j X_j \quad \text{for each } j \in J, \\ w_{ji}(\omega) &\leq \mathfrak{D}_i(\omega) Z_{ji} \quad \text{for each } i \in I, j \in J, \omega \in \Omega, \end{aligned}$$

$$\begin{aligned} \sum_{i \in I} w_{ji}(\omega) &\leq y_j \quad \text{for each } j \in J, \omega \in \Omega, \\ \sum_{j \in J} w_{ji}(\omega) &\leq \mathfrak{D}_i(\omega) \quad \text{for each } i \in I, \omega \in \Omega, \\ X_j &\in \{0, 1\} \quad \text{for each } j \in J, \\ Z_{ji} &\in \{0, 1\} \quad \text{for each } i \in I, j \in J, \\ y_j &\geq 0 \quad \text{for each } j \in J, \\ 0 \leq w_{ji}(\omega) &\leq \mathfrak{D}_i(\omega) \mathfrak{S}_j(\omega) \quad \text{for each } i \in I, j \in J, \omega \in \Omega. \end{aligned}$$

4.2. Solution approach

We next present an algorithm to solve the problem (SCDDSD') approximately. The algorithm is based on Lagrangian relaxation and variable splitting. We first introduce auxiliary variables $u_{ji}(\omega)$ and set them equal to $w_{ji}(\omega)$. Then, we replace the w variables by u variables in some terms of the objective function and some constraints. Finally, we relax the constraints that require u and w to be equal and impose penalties for violations. More specifically, we modify the formulation as follows:

(SCDDSD''): min

$$\sum_{j \in J} f_j X_j + \sum_{i \in I} \sum_{j \in J} c_{ji} Z_{ji} + E \left[\sum_{i \in I} \sum_{j \in J} (d_{ji} - h_j) w_{ji}(\omega) + \sum_{i \in I} p_i \left(\mathfrak{D}_i(\omega) - \sum_{j \in J} u_{ji}(\omega) \right) + \sum_{j \in J} h_j y_j \right],$$

subject to:

$$\begin{aligned} X_j &\geq Z_{ji} \quad \text{for each } i \in I, j \in J, \\ y_j &\leq C_j X_j \quad \text{for each } j \in J, \\ w_{ji}(\omega) &\leq \mathfrak{D}_i(\omega) Z_{ji} \quad \text{for each } i \in I, j \in J, \omega \in \Omega, \quad (3) \\ \sum_{i \in I} w_{ji}(\omega) &\leq y_j \quad \text{for each } j \in J, \omega \in \Omega, \\ \sum_{j \in J} u_{ji}(\omega) &\leq \mathfrak{D}_i(\omega) \quad \text{for each } i \in I, \omega \in \Omega, \quad (4) \\ w_{ji}(\omega) &= u_{ji}(\omega) \quad \text{for each } i \in I, j \in J, \omega \in \Omega, \quad (5) \\ X_j &\in \{0, 1\} \quad \text{for each } j \in J, \\ Z_{ji} &\in \{0, 1\} \quad \text{for each } i \in I, j \in J, \\ y_j &\geq 0 \quad \text{for each } j \in J, \\ 0 \leq w_{ji}(\omega) &\leq \mathfrak{D}_i(\omega) \mathfrak{S}_j(\omega) \quad \text{for each } i \in I, j \in J, \omega \in \Omega, \\ 0 \leq u_{ji}(\omega) &\leq \mathfrak{D}_i(\omega) \mathfrak{S}_j(\omega) \quad \text{for each } i \in I, j \in J, \omega \in \Omega. \end{aligned}$$

Notice that the penalty cost term in the objective function now depends on the variables u_{ji} instead of w_{ji} . Moreover, the Constraints (4) that require the sum of shipments from all DCs not to exceed demand is also rewritten with the u_{ji} variables. A minor point to note is that we have dropped the expected purchase cost term for the brevity of notation, because it is a constant and does not depend on any of our decision variables. To solve the problem, we dualize

Constraints (3) and (5) with Lagrangian multipliers η_{ji} and τ_{ji} , respectively. Upon relaxation of these constraints, we may separate the problems into a subproblem involving variables \mathbf{X} , \mathbf{Z} , \mathbf{y} , \mathbf{w} and one involving the \mathbf{u} variables. Moreover, the first subproblem is further separable by candidate location j and the second one is separable by retailer i .

In our algorithm, we impose state-independent Lagrangian multipliers; i.e., ones that take on the same values across all demand and disruption realizations. Given fixed values of (\mathbf{X}, \mathbf{Z}) , the exact recourse value for our convex stochastic program can be obtained by using state-dependent multipliers. However, the number of such multipliers would equal the number of possible states of the system, which is infinite with a continuous demand distribution. Therefore, for practical reasons, we use state-independent Lagrangian multipliers. Similar techniques based on Lagrangian relaxation of weakly coupled recourse problems have been utilized in the literature to approximate recourse functions of stochastic programming problems. One example is the network recourse decomposition method introduced by Powell and Cheung (1994). Recent papers by Topaloglu and Kunnumkal (2006) and Kunnumkal and Topaloglu (2008) apply similar techniques to approximate the value functions in stochastic dynamic programming. Such techniques provide computationally efficient approximations for the recourse function values that would otherwise require computationally expensive sampling or discretization methods.

Given a set of dual multipliers $(\boldsymbol{\eta}, \boldsymbol{\tau})$, the subproblem in $(\mathbf{X}, \mathbf{Z}, \mathbf{y}, \mathbf{w})$ corresponding to a particular $j \in J$ has the following form:

$$\min_{\mathbf{X}} \sum_{j \in J} (f_j - W_j(\boldsymbol{\eta}, \boldsymbol{\tau})) X_j,$$

where $W_j(\boldsymbol{\eta}, \boldsymbol{\tau})$ is defined as the optimal objective value in the following subproblem:

$$\begin{aligned} (\text{SP}_1) : W_j(\boldsymbol{\eta}, \boldsymbol{\tau}) = & \min \sum_{i \in I} (c_{ji} - \eta_{ji} E[\mathcal{D}_i(\omega)]) Z_{ji} \\ & + E \left[\sum_{i \in I} (d_{ji} + \eta_{ji} - \tau_{ji} - h_j) w_{ji}(\omega) + h_j y_j \right], \quad (6) \end{aligned}$$

subject to:

$$\begin{aligned} \sum_{i \in I} w_{ji}(\omega) & \leq \mathcal{D}_i(\omega) \quad \text{for each } \omega \in \omega, \\ 0 \leq w_{ji}(\omega) & \leq \mathcal{D}_i(\omega) \mathfrak{S}_j(\omega) \quad \text{for each } i \in I, \omega \in \Omega. \end{aligned}$$

As pointed out by Mak and Shen (2010), the subproblem (SP₁) can be solved efficiently using a convex optimization procedure. In particular, if the demands follow a multivariate normal distribution, we can obtain the optimal objective value to (SP₁) as follows. We first define: $\hat{d}_{ji} = (1 - q_j)(d_{ji} + \eta_{ji} - \tau_{ji} - h_j)$. Then, we have the following result.

Proposition 2. Suppose the demands \mathcal{D}_i values follow a multivariate normal distribution. Let $\hat{I} = \{i \in I | \hat{d}_{ji} \leq h_j\}$ and let $m = |\hat{I}|$. Sort the retailers in set \hat{I} in increasing order of \hat{d}_{ji} ; i.e., $\hat{d}_{1j} \leq \hat{d}_{2j} \leq \dots \leq \hat{d}_{mj} \leq h_j$, and let $\hat{d}_{m+1} = h_j$. Then, the optimal objective value of (SP₁) is given by

$$W_j(\boldsymbol{\tau}, \boldsymbol{\eta}) = \sum_{i \in I} \min\{c_{ji} - \eta_{ji} E[\mathcal{D}_i(\omega)], 0\} + Q_j(y_j^*), \quad (7)$$

where:

$$Q_j(y_j) = h_j y_j - \sum_{i=1}^m (\hat{d}_{i+1,j} - \hat{d}_{ji}) \int_0^{y_j} \bar{\Phi} \left(\frac{s - \mu(i)}{\sigma(i)} \right) ds, \quad (8)$$

$$\mu(i) = E \left[\sum_{k=1}^i \mathcal{D}_k \right],$$

$$\sigma(i) = \sqrt{\text{Var} \left(\sum_{k=1}^i \mathcal{D}_k \right)},$$

and y_j^* is the solution to

$$h_j = \sum_{i=1}^m (\hat{d}_{i+1,j} - \hat{d}_{ji}) \bar{\Phi} \left(\frac{y_j^* - \mu(i)}{\sigma(i)} \right).$$

Proof. Let $\rho(s, i)$ be the probability that the s th unit of stock at DC j is shipped to retailer i and $\hat{\rho}(s)$ be the probability that the unit is not shipped out, conditional on facility j not being disrupted. Note that the cost rate corresponding to such a shipment is given by $\hat{d}_{ji}/(1 - q_j)$. The unconditional cost rate is given by \hat{d}_{ji} . In the case where the unit is not shipped, no matter whether there is a disruption or not, the cost rate is h_j . Then, the expected recourse function—i.e., the value of the expectation in Equation (6) given y_j —is given by

$$Q_j(y_j) = \int_0^{y_j} \sum_{i \in I} [\hat{d}_{ji} \rho(s, i) + h_j \hat{\rho}(s)] ds.$$

The optimal value is then given by $\min_{y_j \geq 0} Q_j(y_j)$.

Note that for $i \in I$ where $\hat{d}_{ji} > h_j$, $\rho(s, i) = 0$ for all s . Sort the remaining retailers (denoted by \hat{I}) in increasing order of \hat{d}_{ji} ; i.e., $\hat{d}_{1j} \leq \hat{d}_{2j} \leq \dots \leq \hat{d}_{mj} \leq h_j$. Then, the probabilities are given by

$$\begin{aligned} \rho(s, i) & = P \left(\sum_{k=1}^i \mathcal{D}_k \geq s \right) - P \left(\sum_{k=1}^{i-1} \mathcal{D}_k \geq s \right) \\ \hat{\rho}(s) & = P \left(\sum_{k=1}^m \mathcal{D}_k < s \right). \end{aligned}$$

Because the demand distributions at retailers follow a multivariate normal distribution, the sum of demands at different retailers follows a univariate normal distribution (see, for example, Kale (1970)). Then, the probabilities are

given by

$$P\left(\sum_{k=1}^i \mathcal{D}_k \geq s\right) = 1 - \Phi\left(\frac{s - \mu(i)}{\sigma(i)}\right) = \bar{\Phi}\left(\frac{s - \mu(i)}{\sigma(i)}\right)$$

where:

$$\mu(i) = E\left[\sum_{k=1}^i \mathcal{D}_k\right],$$

$$\sigma(i) = \sqrt{\text{Var}\left(\sum_{k=1}^i \mathcal{D}_k\right)}.$$

Knowing this and letting $\hat{d}_{m+1} = h_j$, we may express $Q_j(y_j)$ as follows:

$$Q_j(y_j) = h_j y_j - \sum_{i=1}^m (\hat{d}_{i+1,j} - \hat{d}_{ji}) \int_0^{y_j} \bar{\Phi}\left(\frac{s - \mu(i)}{\sigma(i)}\right) ds.$$

It is easy to show that $Q_j(y_j)$ is convex and the optimal y_j^* can be found easily by solving for the first-order condition:

$$h_j = \sum_{i=1}^m (\hat{d}_{i+1,j} - \hat{d}_{ji}) \bar{\Phi}\left(\frac{y_j^* - \mu(i)}{\sigma(i)}\right). \quad (9)$$

Equation (9) is a generalization of the critical fractile condition for the standard newsvendor problem. Since the (\mathbf{w}, \mathbf{y}) terms and the \mathbf{Z} terms in Equation (6) are not linked, we can solve (SP_1) by finding their optimal values separately. Let the optimal inventory level y_j found by solving Equation (9) be y_j^* , then the optimal value of the expected value term in Equation (6) is given by $Q_j(y_j^*)$. Moreover, the optimal \mathbf{Z} values can be easily found by setting Z_{ji} to one if its cost coefficient is negative and zero otherwise. To summarize, the optimal objective value of (SP_1) is given by

$$W_j(\boldsymbol{\tau}, \boldsymbol{\eta}) = \sum_{i \in I} \min\{c_{ji} - \eta_{ji} E[\mathcal{D}_i(\omega)] - \tau_i, 0\} + Q_j(y_j^*).$$

The subproblem involving the \mathbf{u} variables corresponding to a particular retailer $i \in I$ has the following form for a given realization (ω) :

$$(SP_2) : \hat{U}_i(\omega) = \min p_i \left(\mathcal{D}_i(\omega) - \sum_{j \in J} u_{ji}(\omega) \right) + \sum_{j \in J} \tau_{ji} u_{ji}(\omega),$$

subject to:

$$\sum_{j \in J} u_{ji}(\omega) \leq \mathcal{D}_i(\omega),$$

$$0 \leq u_{ji}(\omega) \leq \mathcal{D}_i(\omega) \mathcal{S}_j(\omega) \quad \text{for each } j \in J.$$

We need to solve (SP_2) for any realization (ω) and take the expected value $U_i = E[\hat{U}_i(\omega)]$. Given any realization (ω) , the optimal solution to (SP_2) can be obtained using the following results.

Lemma 1. *Given any realization (ω) , there exists an optimal solution to (SP_2) where $u_{ji}(\omega) = 0$ for all $j \neq j^* = \text{argmin}\{\tau_{ik} | k \in J(\omega)\}$, where $J(\omega) \equiv \{j \in J | \mathcal{S}_j(\omega) = 1\}$. Therefore, $u_{ji}(\omega)$ can be positive only for the j that has the minimum τ_{ji} value among the set that is not disrupted.*

Proof. Suppose in an optimal solution u^* , there is a $k \neq j^*$ where $u_{ik}^*(\omega) > 0$. Then, we may produce another feasible solution u' by setting $u'_{ik}(\omega) = u_{ik}^*(\omega) - \epsilon$ and $u'_{j^*i}(\omega) = u_{j^*i}^*(\omega) + \epsilon$ for some $\epsilon > 0$. The new solution does not change the sum of the u_{ji} values; i.e., it does not affect the $p_i(\cdot)$ term. The objective value changes by $(\tau_{j^*i} - \tau_{ik})\epsilon \geq 0$ by the definition of j^* , and thus u' is also optimal. By repeated application of this argument, we can obtain a solution in which only $u_{j^*i}(\omega)$ can be strictly positive. ■

Lemma 2. *The solution to (SP_2) is given by*

$$u_{ji}(\omega) = \begin{cases} u^*(\omega) & \text{if } j = \text{argmin}\{\tau_{ik} | k \in J(\omega)\}, 0 \leq u^*(\omega) \leq \mathcal{D}_i(\omega), \\ 0 & \text{if } j = \text{argmin}\{\tau_{ik} | k \in J(\omega)\}, u^*(\omega) < 0, \\ \mathcal{D}_i(\omega) & \text{if } j = \text{argmin}\{\tau_{ik} | k \in J(\omega)\}, u^*(\omega) > \mathcal{D}_i(\omega), \\ 0 & \text{if } j \neq \text{argmin}\{\tau_{ik} | k \in J(\omega)\}, \end{cases}$$

where $u^*(\omega)$ is the solution to

$$p'_i(\mathcal{D}_i(\omega) - u^*) = \min\{\tau_{ik} | k \in J(\omega)\}.$$

Proof. By Lemma 1, we can reduce (SP_2) to a univariate minimization problem. Since $p_i(\cdot)$ is a convex function, the first-order sufficient condition gives rise to the stated optimal solution. ■

For the remainder of this article, we consider the quadratic penalty cost function $p_i(x) = 1/2 \tilde{p}_i x^2 + \hat{p}_i x$ for simplicity. However, we note that more complex penalty cost functions can be accommodated. With the quadratic penalty function, $p'_i(x) = \tilde{p}_i x + \hat{p}_i$. Then,

$$u^*(\omega) = \mathcal{D}_i(\omega) - \frac{\min\{\tau_{ik} | k \in J(\omega)\} - \hat{p}_i}{\tilde{p}_i}.$$

Proposition 3. *Consider the case of a quadratic penalty cost function; i.e., $p_i(x) = 1/2 \tilde{p}_i x^2 + \hat{p}_i x$. Suppose the set of DCs is sorted such that $\tau_{i1} \leq \tau_{i2} \leq \dots \leq \tau_{ij} < 0 \leq \tau_{i(j+1)} \leq \dots$. Then, the expected optimal objective value of (SP_2) and*

the expected value of the optimal u_{ji} variables are given by

$$\begin{aligned}
 U_i &= \sum_{j=1}^J \tau_{ji} E[\mathcal{D}_i] \hat{q}_j \\
 &+ \sum_{j=\hat{j}+1}^{|J|} \left(\left\{ \frac{\tilde{p}_i}{2} E \left[\mathcal{D}_i^2 \middle| \mathcal{D}_i < \frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \right] \right. \right. \\
 &\left. \left. + \hat{p}_i E \left[\mathcal{D}_i \middle| \mathcal{D}_i < \frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \right] \right\} \right. \\
 &P \left(\mathcal{D}_i < \frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \right) \hat{q}_j + P \left(\mathcal{D}_i \geq \frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \right) \hat{q}_j \\
 &\times \left\{ \frac{\tilde{p}_i}{2} \left(\frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \right)^2 + \hat{p}_i \left(\frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \right) \right. \\
 &\left. + \tau_{ji} E \left[\mathcal{D}_i - \frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \middle| \mathcal{D}_i \geq \frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \right] \right\} \Bigg). \quad (10)
 \end{aligned}$$

$$\begin{aligned}
 E[u_{ji}] &= \begin{cases} E \left[\mathcal{D}_i - \frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \middle| \mathcal{D}_i \geq \frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \right] P \left(\mathcal{D}_i \geq \frac{\tau_{ji} - \hat{p}_i}{\tilde{p}_i} \right) \hat{q}_j & \text{if } \tau_{ji} \geq 0 \\ E[\mathcal{D}_i] \hat{q}_j & \text{if } \tau_{ji} < 0 \end{cases} \quad (11)
 \end{aligned}$$

where:

$$\hat{q}_j = P(j \text{ not disrupted and } 1, \dots, j-1 \text{ disrupted}).$$

Proof. We condition on the event $E_j = \{j = \operatorname{argmin}\{\tau_{ik} \in J(\omega)\}\}$, which has probability \hat{q}_j . Note that E_j for $j = 1, \dots, |J|$ are mutually exclusive and exhaustive. Now we can consider two cases:

Case 1: $\tau_{ji} < 0$, then Lemma 2 suggests that the optimal $u_{ji}(\omega) = \mathcal{D}_i(\omega)$. Since demand is independent of the availability of facilities (i.e., the event E_j), the conditional expectations are given by: $E[u_{ji}|E_j \text{ and Case 1}] = E[\mathcal{D}_i]$ and $E[\text{shortage}|E_j \text{ and Case 1}] = 0$.

Case 2: $\tau_{ji} \geq 0$, then Lemma 2 suggests that $u^*(\omega) = \mathcal{D}_i(\omega) - (\tau_{ji} - \hat{p}_i)/(\tilde{p}_i)$. Then, there are two subcases:

Case 2a: $\tau_{ji} \geq 0$ and $\mathcal{D}_i < (\tau_{ji} - \hat{p}_i)/(\tilde{p}_i)$. The conditional expectations are $E[u_{ji}|E_j \text{ and Case 2a}] = 0$ and $E[\text{shortage}|E_j \text{ and Case 2a}] = E[\mathcal{D}_i | \mathcal{D}_i < (\tau_{ji} - \hat{p}_i)/(\tilde{p}_i)]$.

Case 2b: $\tau_{ji} \geq 0$ and $\mathcal{D}_i \geq (\tau_{ji} - \hat{p}_i)/(\tilde{p}_i)$. The conditional expectations are $E[u_{ji}|E_j \text{ and Case 2b}] = E[\mathcal{D}_i | \mathcal{D}_i \geq (\tau_{ji} - \hat{p}_i)/(\tilde{p}_i)] - (\tau_{ji} - \hat{p}_i)/(\tilde{p}_i)$ and $E[\text{shortage}|E_j \text{ and Case 2b}] = (\tau_{ji} - \hat{p}_i)/(\tilde{p}_i)$.

Unconditioning and combining all cases, we obtain Equations (10) and (11). ■

In the above, the probability \hat{q}_j is the probability that $j = \operatorname{argmin}\{\tau_{ik} \in J(\omega)\}$; i.e., the probability that τ_{ji} is the smallest among all j values that are not disrupted. These probabilities, as well as Equations (10) and (11), are straightforward to compute. Combining these procedures, we can obtain a lower bound for (SCDDSD'') given Lagrangian multipliers (η, τ) as follows:

$$L(\eta, \tau) = \sum_{j \in J} \min\{W_j(\eta, \tau) + f_j, 0\} + \sum_{i \in I} U_i(\eta, \tau). \quad (12)$$

Using the presented procedure, it is possible to obtain a lower bound to (SCDDSD'') given a set of dual multipliers. The standard subgradient optimization approach (see, for example, Fisher (1985)) requires a subroutine to estimate tight upper bounds given information from the lower bound solutions. However, for our problem, the recourse function value—i.e., the expected value term in (SCDDSD)—needs to be evaluated in order to find a feasible solution or upper bound. While approximate methods for doing so exist in the literature (e.g., modifications of the algorithms proposed by Cheung and Powell (1996) and Harrison and Van Mieghem (1999)), it is not practical to embed any of these as a subroutine and run it for a large number of iterations (in the order of thousands) since doing so would significantly increase computation time. There are alternative subgradient procedures that guarantee convergence of the lower bound to its maximal value that do not require upper bounding. An example is the Variable Target Value Method (VTVM) proposed by Sherali *et al.* (2000). In our computational experiments, we apply the VTVM to update the dual multipliers in each iteration.

5. Temporally dependent disruptions

5.1. Model formulation

When decision periods are short or disruptions have long-term effects, it is important to consider temporal dependence of disruptions. For example, a labor strike can make a facility unavailable for several weeks. In many cases like this, whether a facility is disrupted or not is dependent on the state in the previous period. In this section, we assume that the state of facility disruptions—i.e., the sequence of $\{\mathcal{S}_t\}$ vectors—is a discrete time Markov process. This implies that given the knowledge of which current locations are disrupted, the future availability of facility locations is independent of the past history of events.

The Markov chain model of disruptions is a common assumption in the disruption management literature. In single facility inventory problems, it is also known as the (α, β) model (see, for example, Snyder and Shen (2008)). If the facility is currently working, there is a certain probability

α that it gets disrupted in the next period. If it is currently down, the probability of recovering in the next period is β . In our multi-location problem, the state of the system is a vector \mathfrak{S}_t . Since there are $2^{|J|}$ states, defining the Markov chain would require an exponential number of transition probabilities in general.

Assumption 4. The Markov process that characterizes the availability of facility locations is ergodic.

Following the notation introduced in Sections 3 and 4, we formulate the management phase problem as follows, assuming any initial state of disruptions \mathfrak{s}_0 :

$$\begin{aligned} v(\mathfrak{D}_1, \dots, \mathfrak{D}_T, \mathfrak{S}_1, \dots, \mathfrak{S}_T) &= f_0(\mathbf{0}, \mathfrak{s}_0) \\ &= \min_{y_0 \geq 0} G_T(y_0 | \mathfrak{s}_0), \\ f_t(\mathbf{x}_t, \mathfrak{s}_{t-1}) &= \min_{y_t \geq x_t} \{G_t(y_t | \mathfrak{s}_{t-1})\} \\ &= \min_{y_t \geq x_t} E\{H_t(y_t, \mathfrak{D}_t, \mathfrak{S}_t) | \mathfrak{S}_{t-1} = \mathfrak{s}_{t-1}\}. \end{aligned}$$

Because the disruption process follows a Markov chain and demand is temporally independent, $G_t(\cdot)$ is now the conditional expectation of $H_t(\cdot)$ on the disruption vector realized in the previous period, \mathfrak{S}_{t-1} . As before, the value of $H_t(\cdot)$ after observing the demand and disruption realizations in the current period t (i.e., $\mathfrak{D}_t = \mathfrak{d}_t$, $\mathfrak{S}_t = \mathfrak{s}_t$) can be obtained by solving the recourse problem:

$$\begin{aligned} H(y_t, \mathfrak{d}_t, \mathfrak{s}_t) &= \min_{\mathbf{x}_{t+1}, \mathbf{w}_t, \mathfrak{s}_t} \sum_{(i,j) \in A(\mathbf{Z})} d_{ji} w_{t,ji} + \sum_{i \in I} p_i(s_{t,i}) \\ &\quad + \sum_{j \in J(\mathbf{X})} h_j x_{t+1,j} + \alpha f_{t+1}(\mathbf{x}_{t+1}, \mathfrak{s}_t), \end{aligned}$$

subject to:

$$\begin{aligned} x_{t+1,j} + \sum_{i \in I} \mathfrak{s}_{t,j} w_{t,ji} &= y_{t,j} \quad \text{for each } j \in J(\mathbf{X}), \\ \sum_{j \in J(\mathbf{X})} \mathfrak{s}_{t,j} w_{t,ji} + s_{t,i} &= \mathfrak{d}_{t,i} \quad \text{for each } i \in I, \\ x_{t+1,j} &\geq 0 \quad \text{for each } j \in J(\mathbf{X}), \\ w_{t,ji} &\geq 0 \quad \text{for each } i \in I, j \in J(\mathbf{X}), \\ &\quad \text{such that } (i, j) \in A(\mathbf{Z}), \\ s_{t,i} &\geq 0 \quad \text{for each } i \in I. \end{aligned}$$

As with the case with temporally independent disruptions, we shall show that the optimal inventory policy allows us to simplify the problem significantly. However, there are extra complications due to the following result.

Proposition 4. *Given any (\mathbf{X}, \mathbf{Z}) , \mathbf{x}_t and \mathfrak{s}_{t-1} , it is optimal to order up to \mathbf{y}_t^* , the value of which depends on \mathbf{x}_t , \mathfrak{s}_{t-1} as well as t . In other words, a non-stationary state-dependent base stock policy is optimal where the state is the set of sites disrupted in the previous period.*

Proof. We may prove that an optimal \mathbf{y}_t^* exists for every period given \mathbf{x}_t and \mathfrak{s}_{t-1} using arguments similar to those in the proof of Proposition 1. We next show that the optimal vector may depend on t using a simple example.

Suppose that there is only one DC and one retailer and the shipping cost is zero. The marginal holding and penalty (assume to be constant) costs are equal to one, demand is deterministic and equal to one every period, and there is no discounting. Furthermore, $P(\mathfrak{S}_t = 0 | \mathfrak{S}_{t-1} = 1) = 0.49$ and $P(\mathfrak{S}_t = 0 | \mathfrak{S}_{t-1} = 0) = 1$. Therefore, if the DC was not disrupted in the previous period, then it will be disrupted with probability 0.49. If it was disrupted in the previous period, then it will be disrupted with probability one. Then, for the final period $t = T$, if the DC is not disrupted in the previous period, the optimal stocking level \mathbf{y}_T^* can be found by solving:

$$\min 0.49y + 0.51[(y-1)^+ + (1-y)^+].$$

The optimal solution is $\mathbf{y}_T^* = 1$ and the expected cost is 0.49. Now consider period $t = T-1$ and suppose that the DC was not disrupted in period $T-2$. The optimal stocking level \mathbf{y}_{T-1}^* can be found by solving:

$$\begin{aligned} \min 0.49y + 0.51[(y-1)^+ + (1-y)^+] \\ + E[f_T(y-1)^+ | \mathfrak{S}_{T-2} = 1]. \end{aligned}$$

Now we need to determine the $f_T(\mathbf{x}_T)$ function. It is clear that we will never order more than one unit; therefore, the starting inventory \mathbf{x}_T in period T will not exceed one. If the DC is not disrupted in period $T-1$, then the problem in period T can be solved as discussed before and the conditional expected cost is 0.49. If the DC is disrupted in period $T-1$, then since the DC can never recover by assumption, all starting inventory in period T will be held until the end of the period and the conditional expected cost is \mathbf{x}_T , which is \mathbf{y}_{T-1} because the DC is disrupted in period $T-1$. Therefore,

$$E[f_T(y-1)^+ | \mathfrak{S}_{T-2} = 1] = 0.51(0.49) + 0.49y.$$

Then, the optimal stocking level in period $T-1$ can be found by solving:

$$\min 0.98y + 0.51[(y-1)^+ + (1-y)^+ + 0.49].$$

Now the optimal stocking level \mathbf{y}_{T-1}^* is zero. Note that this is different from the optimal solution in period $t = T$ after observing no disruption in the previous period. Therefore, the optimal base stock level is both state and time dependent. ■

Since the optimal policy is not time-stationary, we cannot collapse the management phase into a single period as we have done for the temporally independent disruptions case. Indeed, the multiple-location, multiple-period dynamic sourcing inventory problem is very challenging and deserves further study. However, in this article, our focus is on the network design issues rather than the optimal inventory policy. Under our integrated supply chain design framework, we attempt to include the most important trade-offs regarding inventory management into the integrated network design model to optimize the network design using a more accurate objective function. In complex

problems like this, we resort to making relatively strong assumptions in order to keep the network design model tractable.

Assumption 5. It is possible to buy or sell on-hand inventory at each DC at the beginning of every period at cost c .

This assumption allows the possibility of rebalancing starting inventory at DCs without any extra cost. Therefore, we replace the constraint of $\mathbf{y}_t \geq \mathbf{x}_t$ in minimizing $G_t(\cdot)$ by $\mathbf{y}_t \geq \mathbf{0}$. While this may appear restrictive and unrealistic, we argue that its impact on the problem is rather small. Under the assumption that demand is stationary over time, rebalancing of stock is needed only when some DCs have very high ending inventory levels. When the probability of disruptions is small, the optimal inventory levels at DCs should have low variability and the ending inventory level after satisfying demand in one period is rarely higher than the optimal inventory level in the subsequent period. This should only happen if the DC has stocked up too much due to disruptions of other DCs, which should be rare in practice. Furthermore, when considering disruptions, we are taking a risk-averse standpoint and worrying more about stock-outs than about having more inventory than needed. Therefore, we maintain that it is not too rough an approximation to ignore the possibility that some starting inventory levels are higher than optimal and thus need rebalancing.

With Assumption 5, the optimal inventory policy becomes much simpler.

Proposition 5. *The optimal replenishment policy under Assumption 5 is state dependent and time-stationary. Therefore, \mathbf{y}_t^* depends on \mathbf{s}_{t-1} but not on t .*

Proof. Again, we can follow the same arguments in the proof of Proposition 1 (Mak and Shen, 2010) to show that $G_t(\cdot)$ is convex in every period and a base stock policy is optimal. Next we want to show that the order-up-to points in any two periods following the same state of disruptions are the same. Consider two sample paths and periods t_1 and t_2 in the two sample paths, respectively, with $t_1 \neq t_2$. Suppose $\mathbf{s}_{t_1+1} = \mathbf{s}_{t_2+1}$, and we want to show that $\mathbf{y}_{t_1}^* = \mathbf{y}_{t_2}^*$. The objective functions of the recourse problems in these two sample paths are given by

$$\begin{aligned} I_{t_1}(\mathbf{x}_{t_1+1}, \mathbf{w}_{t_1}, \mathbf{s}_{t_1}) &= \sum_{(i,j) \in A(\mathbf{Z})} d_{ji} w_{t_1,ji} + \sum_{i \in I} p_i(s_{t_1,i}) \\ &+ \sum_{j \in J(\mathbf{X})} h_j x_{t_1-1,j} + \alpha f_{t_1+1}(\mathbf{x}_{t_1+1}, \mathbf{s}_{t_1}), \\ I_{t_2}(\mathbf{x}_{t_2-1}, \mathbf{w}_{t_2}, \mathbf{s}_{t_2}) &= \sum_{(i,j) \in A(\mathbf{Z})} d_{ji} w_{t_2,ji} + \sum_{i \in I} p_i(s_{t_2,i}) \\ &+ \sum_{j \in J(\mathbf{X})} h_j x_{t_2+1,j} + \alpha f_{t_2+1}(\mathbf{x}_{t_2+1}, \mathbf{s}_{t_2}). \end{aligned}$$

Under Assumption 5, we have $f_t(\mathbf{x}_t, \mathbf{s}_{t-1}) = f_t(\mathbf{0}, \mathbf{s}_{t-1})$ since inventory can be bought or sold at the beginning of the period. This implies that given the same realization—i.e., $\mathbf{s}_{t_1} = \mathbf{s}_{t_2}$, $f_{t_1+1}(\mathbf{x}_{t_1+1}, \mathbf{s}_{t_1})$ and $f_{t_2-1}(\mathbf{x}_{t_2-1}, \mathbf{s}_{t_2})$ —only differs by a constant and thus the optimal solution $(\mathbf{w}, \mathbf{x}, \mathbf{s})$ is the same in both cases. Moreover, from the definition of the Markov chain, we know that \mathfrak{S}_{t_1} follows the same distribution as \mathfrak{S}_{t_2} conditioning on the same previous state. Therefore, the expected values of the recourse problems, $G_{t_1}(\mathbf{y}_{t_1}, \mathbf{s}_{t_1-1})$ and $G_{t_2}(\mathbf{y}_{t_2}, \mathbf{s}_{t_2-1})$ differ by only a constant. This implies that their minimizers are equal. ■

5.2. Solution approach

Proposition 5 states that the optimal inventory control policy under Assumption 5 is state dependent and time-stationary. Therefore, we may approximate the problem by assuming that the Markov chain of disruptions is in steady state. This approximation is accurate when the disruptions are caused by external environmental factors that do not depend on whether or not a DC is located at the site. For example, if the occurrence of earthquakes can be modeled by a time-stationary Markov chain, we may safely assume that the process has reached its steady state.

A major difficulty of the Markov chain representation is that the number of possible states is exponential in the number of candidate facility locations. To simplify the problem, we solve the problem based on a subset of states, denoted by $\bar{\Theta}$. Denote the probability of disruption scenario $\bar{\theta}$ happening by $P(\bar{\theta})$. It is possible to obtain subset $\bar{\Theta}$ by sampling over the stationary distribution of \mathfrak{S}_t , by simulating the evolution of the Markov chain until a steady state is reached. An alternative method is to define such a subset based on what the management believe to be “important” scenarios. For example, the management may be concerned about the risk of earthquakes disrupting sites in California or that of hurricanes hitting the Gulf states. Note that the probabilities of each $\omega \in \Omega$ realizing is now dependent on $\bar{\theta}$. Let the probability of scenario $\bar{\theta}$ happening be $P(\bar{\theta})$. We now present a formulation of the resulting supply chain design problem:

$$\begin{aligned} (\text{SCDDSDMC}): \min & \sum_{j \in J} f_j X_j + \sum_{i \in I} \sum_{j \in J} c_{ji} Z_{ji} \\ & + \sum_{\bar{\theta} \in \bar{\Theta}} E \left[\sum_{i \in I} \sum_{j \in J} (d_{ji} - h_j) w_{ji}(\omega) \right. \\ & \left. + \sum_{i \in I} p_i \left(\mathcal{D}_i(\omega) - \sum_{j \in J} w_{ji}(\omega) \right) + \sum_{j \in J} h_j y_j \middle| \bar{\theta} \right] P(\bar{\theta}), \end{aligned}$$

subject to:

$$\begin{aligned}
 X_j &\geq Z_{ji} \quad \text{for each } i \in I, j \in J, \\
 y_j &\leq C_j X_j \quad \text{for each } j \in J, \\
 w_{ji}(\omega) &\leq \mathcal{D}_i(\omega) Z_{ji} \quad \text{for each } i \in I, j \in J, \omega \in \Omega, \\
 \sum_{i \in I} w_{ji}(\omega) &\leq y_j \quad \text{for each } j \in J, \omega \in \Omega, \\
 \sum_{j \in J} w_{ji}(\omega) &\leq \mathcal{D}_i(\omega) \quad \text{for each } i \in I, \omega \in \Omega, \\
 X_j &\in \{0, 1\} \quad \text{for each } j \in J, \\
 Z_{ji} &\in \{0, 1\} \quad \text{for each } i \in I, j \in J, \\
 y_j &\geq 0 \quad \text{for each } j \in J, \\
 0 \leq w_{ji}(\omega) &\leq \mathcal{D}_i(\omega) \mathcal{S}_j(\omega) \quad \text{for each } i \in I, j \in J, \omega \in \Omega.
 \end{aligned}$$

This formulation is very similar to (SCDDSD'). The management phase cost, instead of the simple expectation used in (SCDDSD'), is now approximated as follows. We first condition on the disruption state of the previous period from a sampled set of states ($\bar{\theta} \in \bar{\Theta}$). The conditional expectation of the one-period management phase cost is structurally identical to that management phase cost formulation in the temporally independent case, except that the expectation is taken over the conditional expectation given $\bar{\theta}$.

The structural similarity suggests that we may apply a similar solution approach. Applying variable splitting and Lagrangian relaxation, it is possible to decompose the problem into subproblems that are structurally identical to (SP₁) and (SP₂). Note that now we have $|\bar{\Theta}|$ sets of such subproblems. Each subproblem can then be solved using the procedure discussed in Section 4.2.

6. Computational results

We now present the computational results based on the algorithm proposed in Section 4.3. We use the same 30-city dataset as in Mak and Shen (2010). This dataset is obtained by selecting the 30 cities with the largest population out of the 150-city dataset provided by Daskin (1995). The unit shipment cost is equal to the great circle distance between two cities (in miles) divided by 100. The fixed cost of constructing an arc between two cities is equal to the distance multiplied by a factor of 0.05. The fixed cost of opening a facility, the unit penalty cost, and the unit holding cost are 1000 and 5, respectively, for all locations. The penalty cost function for shortage is set to $p_i(s) = 10bs$, where $b = 1, 2, 3$, in Sections 6.1 and 6.3, and $p_i(s) = ks^2 + 30s$ for $k = 2, 4, 6, 8, 10$ in Section 6.2. Demand follows multivariate normal distributions with mean equal to the city population divided by 20 000, standard deviation equal to 0.6 times the mean, and correlation equal to zero for any pair of cities. Finally, capacities at warehouses are assumed to be large numbers; i.e., we consider the uncapacitated problem. The maximum number of Lagrangian iterations is set at 3500.

Mak and Shen (2010) show that the state-independent Lagrangian relaxation approach produces high-quality solutions with short computation times, by cross-validating solutions produced by the Lagrangian relaxation with a distributionally robust optimization approach. Figure 2 shows an example of computational performance of our approach for the case with disruptions. In the figure, we plot, for each instance solved with unit shortage penalty

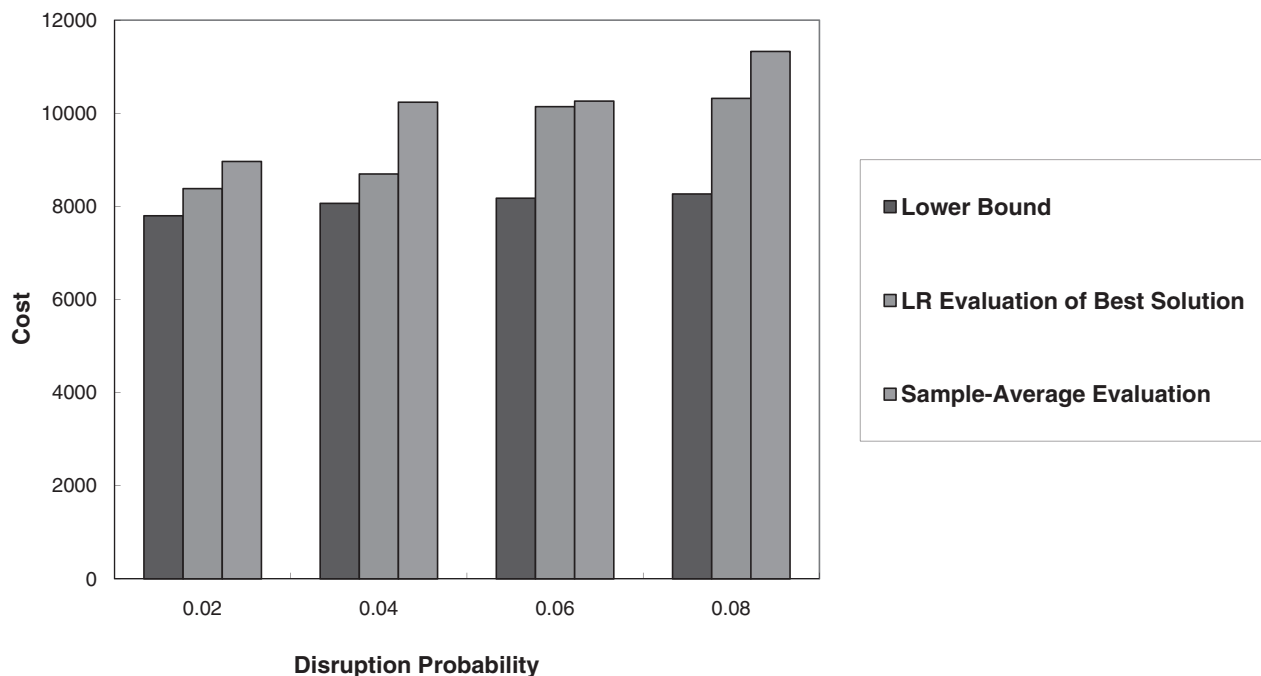


Fig. 2. Example of computational performances.

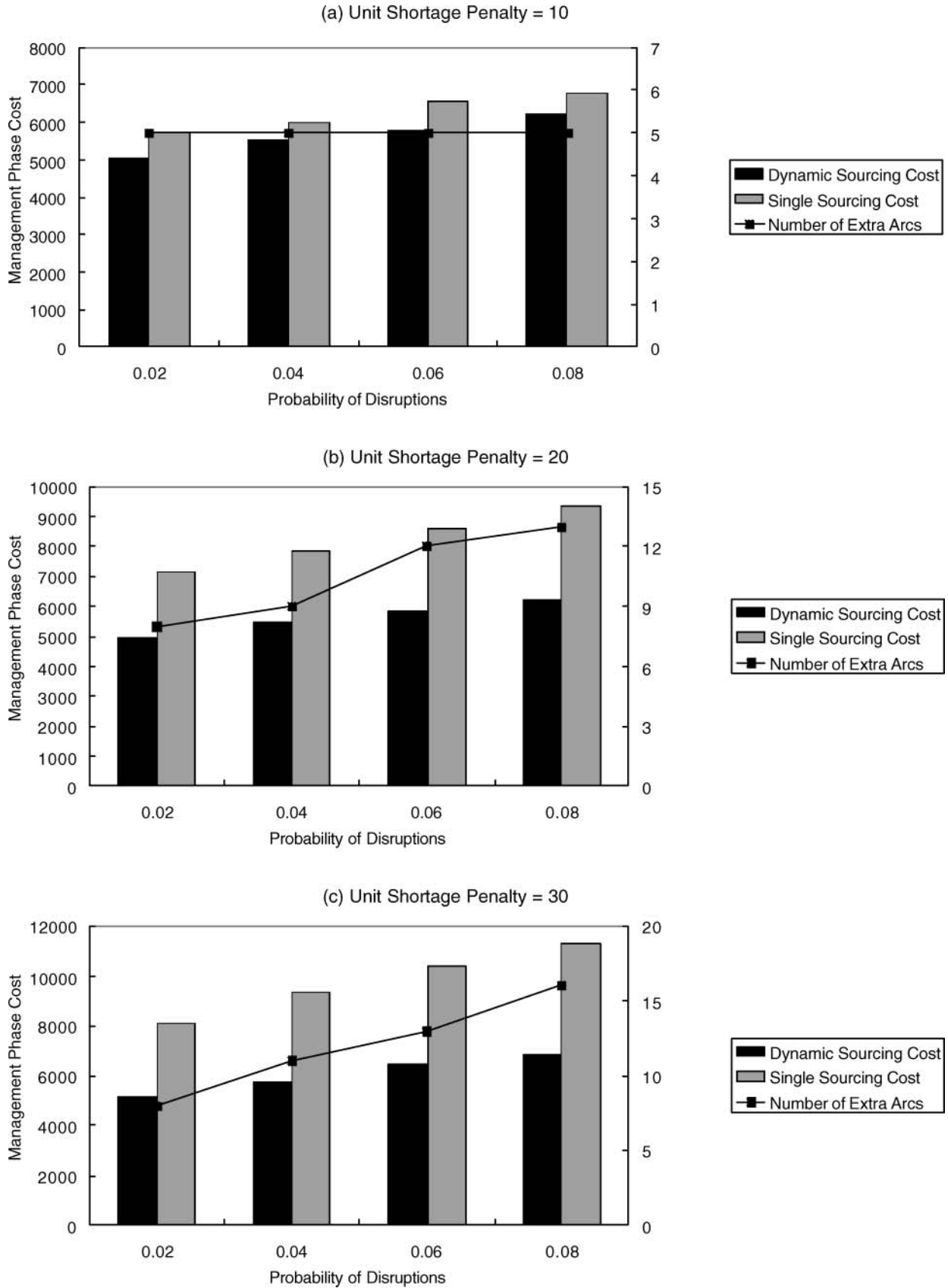


Fig. 3. Effect of disruption probability for unit shortage penalty costs of (a) 10, (b) 20, and (c) 30.

equal to 10, the lower bound obtained in our Lagrangian relaxation algorithm (“Lower Bound”), the expected cost estimate obtained by applying our Lagrangian relaxation procedure with $\mathbf{X}, \mathbf{Z}, \mathbf{y}$ set to the best solution obtained (“LR Evaluation of Best Solution”), and the expected cost evaluation of the same given solution ($\mathbf{X}, \mathbf{Z}, \mathbf{y}$ variables) using a sample-average approach with 2000 samples (“Sample-Average Evaluation”). The sample-average computations require solving 2000 linear programs for the recourse problem for each instance. We can see that the sample-average expected cost of the best solution obtained is typically within about 20% of the lower bound. For a combinatorial stochastic optimization with high complexity, we believe that such performances are reasonably good. More important, the expected cost evaluations using our Lagrangian relaxation approach and using the sample-average approach, for the same solution, are typically within a few percent. This suggests that the Lagrangian relaxation approach does provide reasonable approximations of the expected costs given a solution.

6.1. Effect of disruption probabilities

In this experiment, we examine the effect of different probabilities of disruptions. We vary the disruption probabilities, q_j , between 0.02 and 0.08. The lost sales penalty function is assumed to be linear (i.e., risk neutral) in this experiment.

The results for the test regarding the effects of disruption probabilities are shown in Fig. 3. In the figure, we report three performance measures. “Dynamic Sourcing Cost” refers to the expected cost in the management phase costs (i.e., excluding fixed costs) under the best solution obtained, allowing dynamic sourcing. “Single Sourcing Cost” refers to the expected management phase costs with the same facilities, but only single sourcing; i.e., each retailer is solely served by the nearest open DC. The number of extra arcs refer to the difference between the numbers of arcs constructed (i.e., (i, j) pairs for which $Z_{ji} = 1$) in the dynamic sourcing and single sourcing solutions. Because one arc is constructed per retailer in the single sourcing solution, this number reflects the degree of dynamic sourcing in the network. Inventory is more freely shared across the network if the number of extra arcs is larger.

With a low unit penalty cost (Fig. 3(a)), there are only a few dynamic sourcing arcs in the network. The savings from dynamic sourcing—i.e., the difference between the management phase costs of dynamic and single sourcing—are relatively small at about 10–15% (of the dynamic sourcing cost). When the unit penalty becomes large (see Figs. 3(b) and 3(c)), the savings become more significant and show an upward trend with the probability of disruptions. The cost of single sourcing increases rapidly with the probability of disruptions, whereas the cost of dynamic sourcing stays relatively robust. To enable these savings, more dynamic sourcing arcs are added to the network as the probability of disruptions increases.

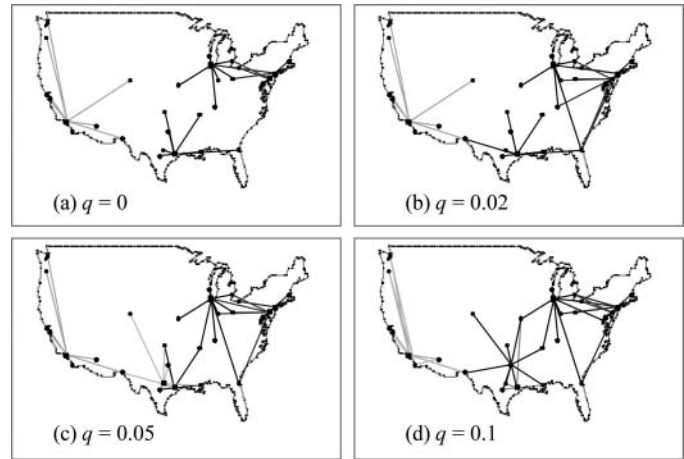


Fig. 4. Example solution for 30-city problem with disruptions.

An example set of solutions is displayed in Fig. 4. When the probability of disruptions is zero, the five DCs are located in New York, Philadelphia, Chicago, Houston, and Los Angeles, respectively. These cities serve high demand volumes nearby. The first three DCs are joined together by dynamic sourcing arcs such that second-order inventory sharing is possible. When the probability of disruptions increases slightly to 0.02, the set of open DCs remains the same. However, the number of dynamic sourcing arcs increases and now all five DCs are joined together, allowing inventory sharing across the entire network. The number of multi-sourcing retailers increases because of the new disruption threat that can make one of the connected DCs unavailable.

When the probability of disruptions increases further to 0.05, the number of dynamic sourcing arcs further increases. In addition, there is an additional DC located at Austin, Texas. This DC is close to the one in Houston and, unlike the cases of Philadelphia and New York, the local demand at Austin is relatively small. The main reason for opening this DC is for risk diversification. Since the DC at Houston is now subject to a 5% chance of being disrupted and there is no other DC nearby, it is beneficial to open a new neighboring DC at Austin such that the risk is diversified. Similarly, when the probability of disruptions increases further to 0.1, there is an extra DC opened at San Diego as a backup for the Los Angeles DC.

6.2. Effect of risk aversion

Recall that one of the reasons to allow the lost sales penalty cost function to be an increasing convex function is to reflect risk aversion. In this experiment, we evaluate the performance of the dynamic sourcing solution under different risk preferences. The lost sales cost for each retailer is $p_i(s) = ks^2 + 30s$ for s units of lost sales. A larger value of k (i.e., the marginal penalty cost is increasing more rapidly) indicates a higher degree of risk aversion and $k = 0$ gives

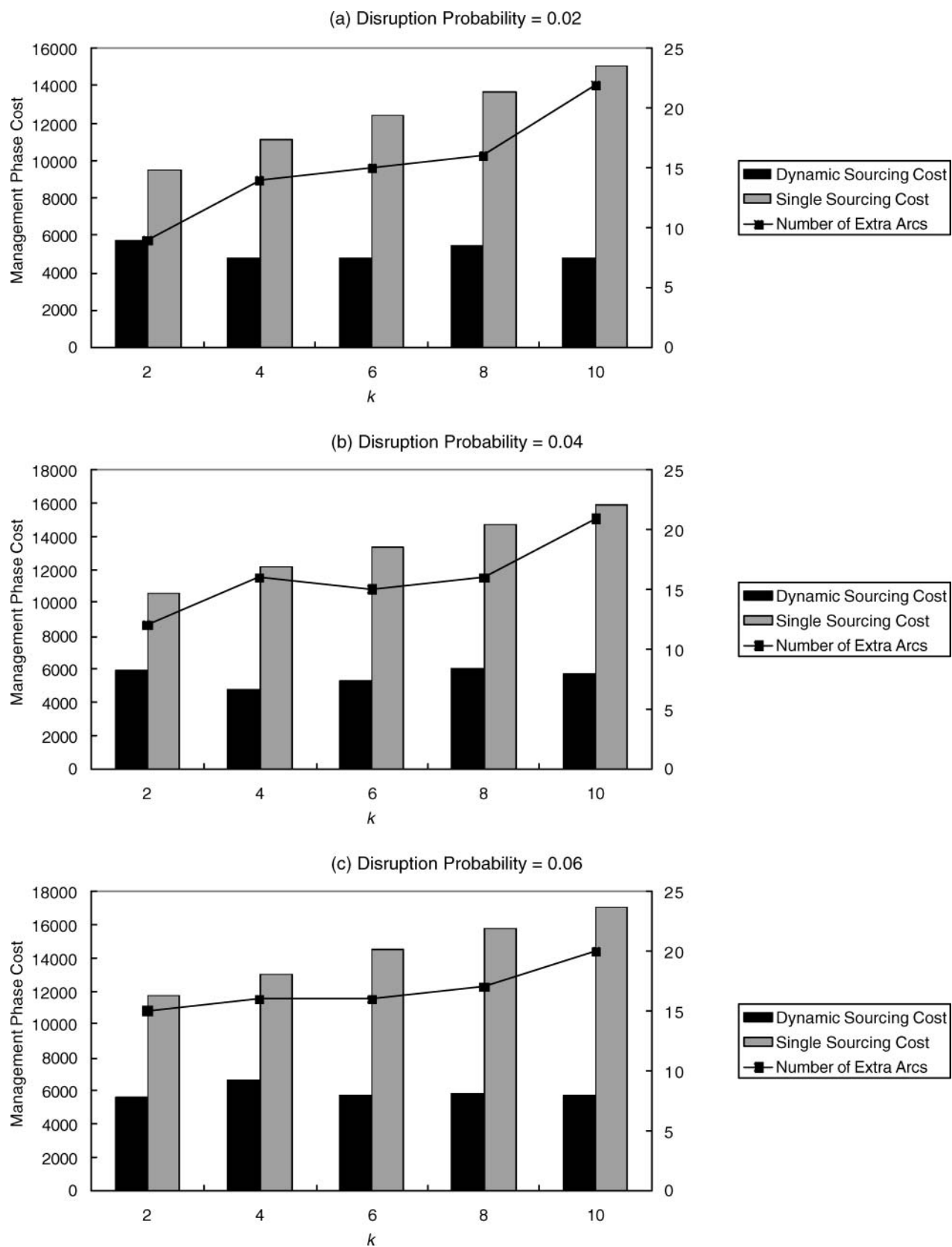


Fig. 5. Effect of risk aversion for disruption probability values of (a) 0.02, (b) 0.04, and (c) 0.06.

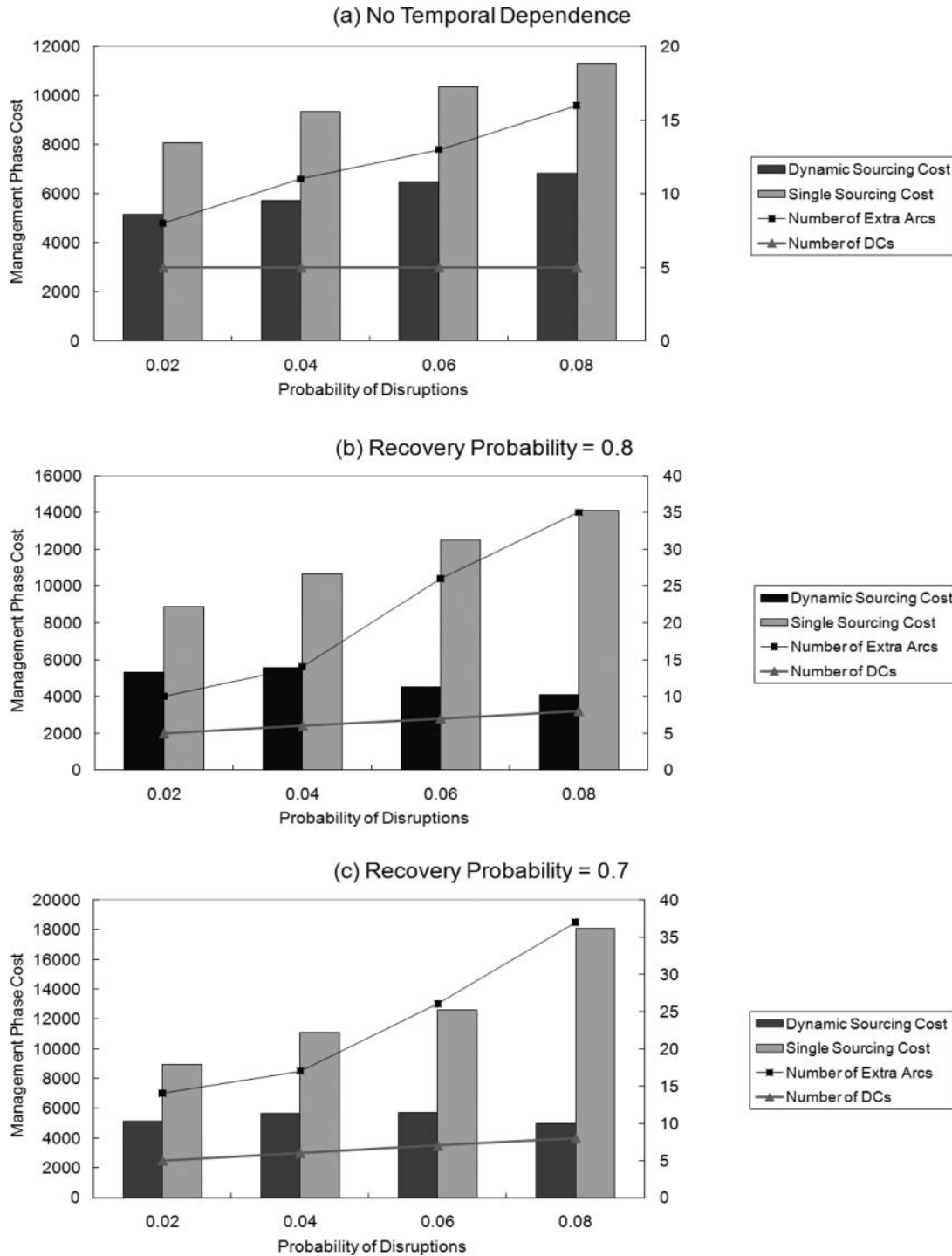


Fig. 6. Effect of temporal dependence of disruptions (a) no temporal dependence, (b) recovery probability = 0.8, and (c) recovery probability = 0.70.

a risk-neutral objective function. We let the probability of disruptions equal 0.02, 0.04, and 0.06 and k equal 2, 4, 6, 8, 10.

The results are shown in Fig. 5. At each level of disruption probability, the management phase cost of single sourcing increases rapidly with the degree of risk aversion. This indicates that single sourcing is not effective

when risk aversion is high. The requirement that each retailer can source from only one (unreliable) DC limits the possibility of risk diversification. Therefore, when the decision maker becomes more risk averse, single sourcing becomes less favorable.

The management phase cost of dynamic sourcing stays roughly constant as the degree of risk aversion increases.

This suggests that supply chain networks designed for dynamic multiple sourcing can work well under different degrees of risk aversion. Compared with the single sourcing case, dynamic sourcing is relatively more favorable when the decision maker is more risk averse. Moreover, the number of dynamic sourcing arcs tends to increase with the degree of risk aversion. A sparse network with less than one extra arc per retailer can achieve very dramatic savings under a risk averse objective.

6.3. Effect of temporal dependence

In the next set of experiments, we test the impact of temporal dependence of disruptions using the (SCDDSDMC) model discussed in Section 5. As done in Section 6.1, we assume the penalty cost to be $p_i(s) = 30s$ for s units of lost sales. The disruption process follows the typical (α, β) model assumed in the disruption literature (e.g., Snyder and Shen (2008)). In particular, we assume that, if a facility is currently working, we let the probability of it being disrupted in the next period is $\alpha = 0.02, 0.04, 0.06, 0.08$. If the facility is down currently, we let the probability that it recovers in the next period be $\beta = 1 - \alpha, 0.8, 0.7$. Note that if $\beta = 1 - \alpha$, then the conditional probability that a facility cannot recover (i.e., remains disrupted) in the next period, given that it is currently disrupted, will be the same as the probability of a working facility becoming disrupted. This reflects the case that there is no temporal correlation. A smaller β value implies that disruption durations tend to be longer. The remaining problem parameters are set to the same values as in the previous experiments. To solve the problem (SCDDSDMC), we simulate a sample path of disruptions following the (α, β) model with a sample path of 1000 000 periods. Then, we randomly draw samples from this sample path, ignoring the first 100 000 warm-up periods, until we obtain 200 distinct scenarios. Then, we use this set of 200 scenarios as our set Θ and the relative frequencies of the scenarios as their respective probabilities $P(\theta)$.

The results are summarized in Fig. 6. We may observe that the management phase cost of single sourcing increases very rapidly in both the disruption probability (α) and the duration of disruptions ($1 - \beta$). This is an expected result, because as the proportions of time that facilities are unavailable increase, shortages will inevitably happen more frequently, leading to increased shortage costs. In response, it is beneficial for the firm to include more extra arcs in the network design; i.e., increase the degree of dynamic sourcing flexibility. After doing so, the management phase cost can be significantly reduced, even with the same set of facilities. We may also observe that the differences between the management phase costs with and without dynamic sourcing tends to increase in both likelihoods of disruptions, as discussed in previous experiments, and their durations. This indicates that, if disruptions tend to last longer, dynamic sourcing gives more significant benefits. This further confirms the effectiveness of dynamic sourcing as a favorable network design strategy under disruption threats.

Finally, we also observe that the management phase cost under dynamic sourcing does not always increase monotonically in disruption probability if durations of disruptions are long. In cases in which disruptions become more and more likely and long lasting, both the number of facilities and the number of extra arcs increase significantly. The location of more and smaller facilities is consistent with the conventional wisdom of risk diversification. The non-monotonicity of management phase cost in disruption probability suggests that the risk diversification effect alters the fundamental trade-off between fixed costs of location and transportation costs in the management phase. As a large number of smaller facilities are located, each retailer is closer to the nearest DC, and therefore transportation costs in the management phase cost may decrease. This effect counteracts the increase of shortage costs due to higher disruption probabilities, as demonstrated in the previous discussion on the case with no temporal dependence, leading to the non-monotonic relationship. Observe that, if we consider the huge cost differences between single and dynamic sourcing in these instances, we can see that risk diversification alone, with only single sourcing, is not effective enough. If dynamic sourcing is coupled with the conventional wisdom of risk diversification, cost savings can be very significant.

7. Conclusion and future research

In this article, we study a dynamic sourcing supply chain design model under both demand uncertainty and disruption risks. This allows us to consider the other benefit of dynamic sourcing: since inventory can be shared without pooling all demand at a single location, it is possible to spread out the risk of disruptions compared to the case where only first-order inventory sharing is possible. Therefore, dynamic multiple sourcing allows the supply chain network to capture benefits of both risk pooling and risk diversification.

We show that the case where disruptions are short and temporally independent can be handled by modifying the formulation in Mak and Shen (2010). We propose a similar Lagrangian relaxation algorithm for this new formulation. For the case where disruptions are temporally dependent, the analysis is more difficult because the inventory control policy is not time-stationary without further assumptions. We discuss such an assumption and propose a formulation based on sampling the stationary distribution of the Markov chain characterizing the disruption process. Such a formulation can be solved using an algorithm similar to that proposed for the temporally independent case.

Our computational results show that dynamic multiple sourcing is indeed beneficial in mitigating the risk of disruptions. By adding just a limited number of dynamic sourcing arcs, it is possible to significantly reduce the expected cost under the threat of disruptions. The savings tend to increase with the lost sales penalty cost and the probability of disruptions.

The tactical and operational aspects of this problem form an important area for future research. We adopted several simplifying assumptions such that the design problem can be tractably formulated and solved. When determining how to operate the supply chain, we have to relax some of these assumptions. For example, backorders and delivery lead times are critical issues in this problem. To see why, consider the case where a retailer dynamically sources from two DCs. If the closer DC is disrupted, it is optimal for the retailer to receive as close to its realized demand as possible from the second DC in the case of lost sales and zero delivery lead time. However, if shortages can be backordered, it may be beneficial to accept backorders if the disrupted DC is expected to recover soon. It may also be possible that by the time the retailer receives shipment from the further DC (with non-zero lead time), the closer DC has already recovered from the disruption. In view of issues like this, more detailed models are needed to analyze the tactical and operational issues of this problem.

Acknowledgements

The authors would like to thank the Associate Editor and two anonymous referees for their helpful comments and suggestions. This research was supported by Hong Kong Research Grants Council grant DAG09/10.EG07 (Mak), National Science Foundation grant CMMI 1068862 (Shen) and National Natural Science Foundation of China grants 70731003, 90924012, 71128001 (Shen).

References

- Berman, O., Krass, D. and Menezes, M.B.C. (2007) Facility reliability issues in network p -median problems: strategic centralization and co-location effects. *Operations Research*, **55**(2), 332–350.
- Cheung, R.K.M. and Powell, W.B. (1996) Models and algorithms for distribution problems with uncertain demands. *Transportation Science*, **30**(1), 43–59.
- Cui, T., Ouyang, Y. and Shen, Z.J.M. (2010), Reliable facility location design under the risk of disruptions. *Operations Research*, **58**(4), 998–1011.
- Daskin, M. (1995) *Network and Discrete Location: Models, Algorithms, and Applications*, John Wiley, New York, NY.
- Daskin, M.S., Coullard, C. and Shen, Z.J.M. (2002) An inventory-location model: formulation, solution algorithm and computational results. *Annals of Operations Research*, **110**, 83–106.
- Eppen, G.D. (1979) Effects of centralization on expected costs on the multi-location newsboy problem. *Management Science*, **25**(5), 498–501.
- Fisher, M. (1985) An applications oriented guide to Lagrangian relaxation. *Interfaces*, **15**(2), 2–21.
- Geoffrion, M. and McBride, R. (1978) Lagrangian relaxation applied to capacitated facility location problem. *IIE Transactions*, **10**, 40–47.
- Harrison, J.M. and Van Mieghem, J.A. (1999) Multi-resource investment strategies: operational hedging under demand uncertainty. *European Journal of Operations Research*, **113**, 17–29.
- Kale, B.K. (1970) Normality of linear combinations of non-normal random variables. *The American Mathematical Monthly*, **77**, 992–995.
- Kunnumkal, S. and Topaloglu, H. (2008) A duality-based relaxation and decomposition approach for inventory distribution systems. *Naval Research Logistics*, **55**(7), 612–631.
- Lim, M., Daskin, M., Bassamboo, A. and Chopra, S. (2009) Facility location decisions in supply chain networks with random disruption and imperfect information. *Naval Research Logistics*, **57**(1), 58–70.
- Lim, M., Daskin, M., Bassamboo, A. and Chopra, S. (2010) A facility reliability problem: formulation, properties and algorithm. Working Paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.
- Mak, H.Y. and Shen, Z.J.M. (2010) Supply chain network design with dynamic sourcing Working Paper, Hong Kong University of Science and Technology, Hong Kong.
- Mak, H.Y. and Shen, Z.J.M. (2011a) Agility and proximity considerations in supply chain design. Working Paper, Hong Kong University of Science and Technology, Hong Kong.
- Mak, H.Y. and Shen, Z.J.M. (2011b) Integrated Supply Chain Design Models, in *Wiley Encyclopedia of Operations Research and Management Science*.
- Ozsen, L., Coullard, C. and Daskin, M. (2008) Capacitated facility location model with risk pooling. *Naval Research Logistics*, **55**(4), 295–312.
- Ozsen, L., Daskin, M. and Coullard, C. (2009) Facility location modeling and inventory management with multisourcing. *Transportation Science*, **43**(4), 455–472.
- Powell, W.B. and Cheung, R.K.M. (1994) A network recourse decomposition method for dynamic networks with random arc capacities. *Networks*, **24**, 369–384.
- Qi, L. and Shen, Z.J.M. (2007) A supply chain design model with unreliable supply. *Naval Research Logistics*, **54**(8), 829–844.
- Qi, L., Shen, Z.J.M. and Snyder, L.V. (2010) The effect of supply disruptions on supply chain design decisions. *Transportation Science*, **44**(2), 274–289.
- Schmitt, A.J., Snyder, L.V. and Shen, Z.J.M. Centralization versus decentralization: Risk pooling, risk diversification, and supply uncertainty in a one-warehouse multiple-retailer system. Submitted for publication. (Download from SSRN.) 7/11.
- Shen, Z.J.M. (2006). A profit-maximizing supply chain network design model with demand choice flexibility. *Operations Research Letters*, **34**, 673–682.
- Shen, Z.J.M. (2007) Integrated supply chain design models: a survey and future research directions. *Journal of Industrial and Management Optimization*, **3**(1), 1–27.
- Shen, Z.J.M., Coullard, C. and Daskin, M.S. (2003) A joint location-inventory model. *Transportation Science*, **37**, 40–55.
- Shen, Z.J.M., Zhan, R.L. and Zhang, J. (2011). The reliable facility location problem: formulations, heuristics, and approximation algorithms. *Informatics Journal on Computing*, **23**(3), 470–482.
- Sherali, H.D., Choi, G. and Tuncbilek, C.H. (2000) A variable target value method for nondifferentiable optimization. *Operations Research Letters*, **26**(1), 1–8.
- Snyder, L. (2006) Facility location under uncertainty: a review. *IIE Transactions*, **38**, 537–554.
- Snyder, L.V. and Daskin, M.S. (2005) Reliability models for facility location: the expected failure cost case. *Transportation Science*, **39**(3), 400–416.
- Snyder, L., Daskin, M.S. and Teo, C.P. (2007) The stochastic location model with risk pooling. *European Journal of Operational Research*, **179**, 1221–1238.
- Snyder, L.V. and Shen, Z.J.M. (2008) Supply and demand uncertainty in multi-echelon supply chains. Working Paper, Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA.
- Topaloglu, H. and Kunnumkal, S. (2006) Approximate dynamic programming methods for an inventory allocation problem under uncertainty. *Naval Research Logistics*, **53**, 822–841.
- Zhan, R.L., Shen, Z.J.M. and Daskin, M.S. (2007) System reliability with location-specific failure probabilities. Working Paper, Department of Industrial Engineering and Operations Research, University of California, Berkeley.

Biographies

Ho-Yin Mak is an Assistant Professor in the Department of Industrial Engineering and Logistics Management at the Hong Kong University of Science and Technology. His research interests span the areas of supply chain design, green transportation, sustainable operations management, and applied optimization.

Zuo-Jun (Max) Shen is the Chancellor's Professor in the Department of Industrial Engineering and Operations Research at UC Berkeley.

He received his Ph.D. from Northwestern University in 2000. He has been active in the following research areas: integrated supply chain design and management, market mechanism design, applied optimization, and decision making with limited information. He is currently on the editorial/advisory board for several leading journals. He received the CAREER award from the National Science Foundation in 2003.