Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Facility Location Decisions with Random Disruptions and Imperfect Estimation

Michael K. Lim, Achal Bassamboo, Sunil Chopra, Mark S. Daskin,

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 15, No. 2, Spring 2013, pp. 239–249
ISSN 1523-4614 (print) | ISSN 1526-5498 (online)
http://dx.doi.org/10.1287/msom.1120.0413

Facility Location Decisions with Random Disruptions and Imperfect Estimation

Michael K. Lim
Department of Business Administration, University of Illinois at Urbana–Champaign, Champaign, Illinois 61820, mlim@illinois.edu
Achal Bassamboo, Sunil Chopra
Department of Managerial Economics and Decision Sciences, Northwestern University, Evanston, Illinois 60208 {a-bassamboo@kellogg.northwestern.edu, s-chopra@kellogg.northwestern.edu}
Mark S. Daskin
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan 48109, msdaskin@umich.edu

Supply chain disruptions come with catastrophic consequences in spite of their low probability of occurrence. In this paper, we consider a facility location problem in the presence of random facility disruptions where facilities can be protected with additional investments. Whereas most existing models in the literature implicitly assume that the disruption probability estimate is perfectly accurate, we investigate the impact of misestimating the disruption probability. Using a stylized continuous location model, we show that underestimation in disruption probability results in greater increase in the expected total cost than overestimation. In addition, we show that, when planned properly, the cost of mitigating the misestimation risk is not too high. Under a more generalized setting incorporating correlated disruptions and finite capacity, we numerically show that underestimation in both disruption probability and correlation degree result in greater increase in the expected total cost compared to overestimation. We, however, find that the impact of misestimating the correlation degree is much less significant relative to that of misestimating the disruption probability. Thus, managers should focus more on accurately estimating the disruption probability than the correlation.

Key words: logistics and transportation; supply chain disruptions; facility network design; estimation error; correlated disruptions; continuous approximation
History: Received: September 3, 2011; accepted: August 10, 2012. Published online in *Articles in Advance* December 14, 2012.

1. Introduction

Over the past few decades, significant effort has been expended in making supply chains leaner and cheaper. However, recent studies indicate that although such efforts have successfully reduced supply chain operational costs, they have concomitantly increased supply chain risks (Rice and Caniato 2003, Sheffi and Rice 2005). Among the various types of supply chain risks, we focus on *supply chain disruptions* in this study. Supply chain disruptions are fundamentally different from the risks arising from machine failures or demand uncertainties because they completely stop the production flow and typically persist longer (Kleindorfer and Saad 2005); thus, the impact of supply chain disruptions can be much more catastrophic, although their likelihood of occurrence is very low. Hendricks and Singhal (2005) point out that supply chain disruptions expose a firm to negative financial impact, and the recovery from such shocks is typically very slow. Furthermore, reports from reinsurance companies show that the frequency of natural hazards is on the rise, further increasing the cost associated with supply chain disruptions (Munich Re 2008). To be better prepared against supply chain disruptions, it is important to understand how to design robust supply chain networks.

We consider a facility location problem in the presence of facility disruptions. On a distribution network, disruptions can be caused by random events such as natural disasters or by premeditated events such as adversarial attacks. In addition, disruptions may be independent of each other (such as facility contamination or a fire in the plant) or correlated across the network (such as a flood or an earthquake that affects multiple facilities in the neighborhood). In this study, we focus on *random* disruptions that are both *independent* and *correlated*. For premeditated disruptions (also referred to as man-made disruptions), refer to Church and Scaparra (2007).

We examine the optimal facility network design where the reliability of a facility is measured as the *disruption probability*; this can be viewed as the probability of a facility being down or the facility's expected fractional downtime. Most studies in the literature assume this as a known parameter. Unfortunately, the disruption probability is very difficult to estimate because such events do not occur regularly or often. Moreover, managers tend to underestimate the disruption probability (or the impact of disruptions), deceived by its low probability of occurrence, and discount it even further when such events do not occur for some period of time (Swiss Re 2009). As a result, firms often design and operate distribution networks with a significant error in their estimate of disruption probability.

The supply chain literature, however, has very limited discussion on the impact of misestimating disruption probability. The one exception is a discussion by Tomlin (2006) that provides insights on the impact of misestimation on a firm's sourcing decision. Our paper aims to address the impact of misestimation (and derive managerial insights) in the context of supply chain network design. We employ a stylized continuous model to facilitate the analysis incorporating both independent and correlated random facility disruptions. The main contributions of our paper are as follows.

(1) *Understanding the impact of imperfect estimation.* Most existing models in the literature implicitly assume that the disruption probability estimate is perfectly accurate. Consequently, it is uncertain how robust the optimal policy will be, given misestimated inputs. In contrast, we assume that the estimation of the disruption probability is imperfect (and thus comes with an estimation error) and investigate the impact of its misestimation on optimal network design.

(2) *Underestimation versus overestimation.* We contrast the impact of underestimating the disruption probability to that of overestimating. We show that underestimation results in greater increase in the expected total cost than overestimation. Further, the impact of underestimation relative to overestimation increases with estimation error.

(3) *Correlated disruptions with finite capacity.* We extend the model by incorporating correlated disruptions in the context of facilities with finite capacity. We show that underestimation in both disruption probability and correlation degree result in greater increase in the expected total cost compared to overestimation. However, the impact of misestimating the correlation degree is much less significant relative to that of misestimating the disruption probability.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 proposes a continuous facility reliability model and characterizes the optimal network structure. Section 4 analyzes the impact of misestimating the disruption probability and provides an optimal estimate to hedge against the worst-case scenario when the disruption probability is difficult to estimate. Section 5 extends the model to the capacitated case to provide further insights on the impact of correlated disruptions. Finally, §6 concludes the paper. All proofs are presented in the online appendix (available at http://dx.doi.org/10.1287/msom.1120.0413).

2. Literature Review and Synthesis

Supply chain disruptions have gained considerable attention, especially in the past few decades. Chopra and Sodhi (2004), Kleindorfer and Saad (2005), and Tang (2006) identify and categorize supply chain risks and provide mitigation strategies to deal with these risks. In the context of facility location, several papers (e.g., Pirkul and Schilling 1988, Batta and Mannur 1990, Ball and Lin 1993, Berman et al. 2007) focus on network designs that are robust when facing demand uncertainty. Baron et al. (2008) apply queueing models to capture demand fluctuation and provide optimal facility decisions. In contrast, our focus is on the supply-side uncertainty. We study the design of distribution network that maintains robust performance in the presence of supply disruption. We employ a contingency tactic similar to that in Pirkul and Schilling (1988): each demand is assigned to a primary and backup server to cope with random disruptions.

Tomlin (2006) considers a variety of contingency strategies on a firm's sourcing decision (including excess inventory or a backup supplier) in the presence of disruption risks and characterizes conditions under which different contingency strategies are more effective. The author further studies the impact of misestimating disruption probabilities and points out that overestimation is more harmful than underestimation when the backorder cost is high, and the opposite holds when the backorder cost is low. Our paper aims to draw similar insights in the context of facility location in supply chain networks where facilities may randomly fail but its probability is not perfectly known. Berman et al. (2009) also study the impact of incomplete information in facility network design where the customers do not have advance information about whether a given facility is operational or not. We consider imperfect information from the network designer's perspective.

Facility location problems have typically been modeled using discrete networks. Snyder and Daskin (2005) present two models to demonstrate the tradeoff between the expected failure cost and the total cost in the presence of facility disruptions. Snyder and Daskin (2006) introduce the concept of stochastic p-robustness in which the relative regret is always less than p for any possible scenario to guarantee a given level of system performance. Lim et al. (2010) propose a facility reliability model that allows for site-specific disruption probabilities. Although models in discrete networks reflect reality more accurately, it is very challenging to obtain generalized insights (due to the lack of analytical tractability) or good solutions for large instances within a limited time frame (because the problems are typically NP-hard).

Our paper aims to fill this gap by obtaining reliable network design insights using a *continuous approximation* (CA). The idea of CA is to concisely convert discrete data (e.g., locations on a service region or specific information on each location points) into a continuous scale (or into a differentiable function) for deriving an analytical model. With this parsimonious data representation, the CA model allows us to focus on the main issues, such as the facility size and service region under a simplified topology. This provides great analytical tractability, often in closed form, and useful managerial insights can be derived exploiting the clear relationships between the decision variables and other parameters. The CA technique was first introduced by Newell (1973) for analyzing logistics systems, and was then employed by many others (e.g., Hall 1984, Langevin et al. 1996, Dasci and Verter 2001, Mak and Shen 2011) for solving facility location problems. These papers show that for large-scale problems for which the objective function is relatively insensitive to the details of the parameters, the CA solution is a very good approximation for the exact problem. Ouyang and Daganzo (2006) provide an algorithm for translating CA solutions into discrete design and validate the accuracy of the CA model. For more details on CA, we refer the readers to Daganzo (2005), which explains the modeling methodology in great detail along with various applications.

There are only a few papers that examine facility location issues in the presence of supply disruption using the CA approach. Daganzo and Erera (1999) suggest approximation methods to systematically analyze large-scale logistics systems in the presence of demand uncertainty (as opposed to supply uncertainty). Wang et al. (2006) employ CA for evaluating the service reliability of a large-scale distribution system. Cui et al. (2010) propose both continuous and discrete models for designing a reliable distribution network incorporating independent random failures. Through a set of test instances, the authors show that the continuous model serves as a good alternative to the discrete model for solving large-scale problems. We contribute to this short literature by investigating the impact of misestimating the disruption probability on distribution network design.

Although many facility disruption instances exhibit spatial correlation in practice, to the best of our knowledge, the only paper to address the issue in the facility location literature is by Li and Ouyang (2010). They capture correlated disruptions using conditional probabilities and the beta-binomial distribution, similar to Bakkaloglu et al. (2002) in the information storage systems literature. Although both this work and ours incorporate correlated disruptions in the model, the objective of the two papers are different. Li and Ouyang (2010) provide methodologies to formulate the correlation among adjacent facility disruptions on large-scale systems using the CA approach. They present numerical experiments to illustrate how the model can be used to optimize facility location designs for various types of correlation structures. In contrast, using a parsimonious model, we primarily focus on understanding the impact of misestimating the disruption probability and the degree of its correlation on such systems.

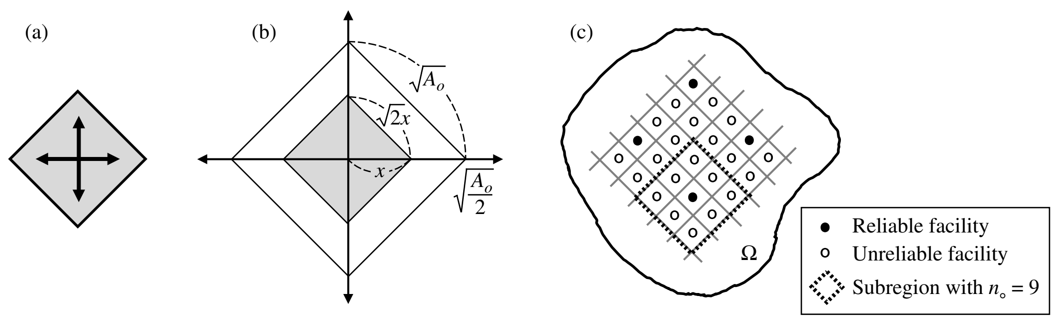
3. The Model

In this section, we introduce the basic network design model with the assumption that the disruption probability is known (§3). Then, we use these results to study the impact of misestimating the disruption probability for the later sections (§§4, 5).

Consider a service region Ω with area A on a plane that is sufficiently large. We assume that demand is uniformly distributed with a density of ρ (demand per unit time per square unit area). The network designer seeks to locate a set of facilities to satisfy all the demand while accounting for random facility disruptions. We denote the disruption probability of a facility by q (> 0). Similar to Tomlin (2006) and Chopra et al. (2007), two types of facilities are considered: (i) regular facilities, hereafter referred to as *unreliable* facilities, which are subject to random disruptions; and (ii) *reliable* facilities that are "hardened" as a result of additional investment and are thus immune to disruptions. The notion of hardening represents various protection plans ranging from physically protecting the facility to outsourcing contracts with exogenous suppliers. We assume that demand is *primarily* served by the closest (reliable or unreliable) facility. If the closest facility is down (if it was an unreliable one), demand is served from the closest reliable facility as a *backup*. Although each demand may go to its next nearest available facility (as in Cui et al. 2010), for analytical tractability, we assume that only the reliable facilities have ample excess capacity to serve other demands in case of disruptions. In §5, we extend the model to the case in which reliable facilities are capacitated.

We consider the network designer to be risk neutral; thus, the objective is to minimize the total cost of locating n_r reliable and n_u unreliable facilities and the expected transportation cost between facilities and demands. Denoting the transportation cost per unit demand per unit distance by c, the total cost (TC) can be expressed as $TC =$ fixed cost of facilities + expected distance $\times$ $\rho A c$. We denote the cost

Figure 1 **(a) Travel Directions; (b) Distance of Subregion $\Omega_\circ$; (c) Example Configuration with $n_\circ = 9$**



of each reliable and unreliable facility to be f_r and $f_u(< f_r)$, respectively. Then, the fixed cost of facilities can be expressed as $f_r n_r + f_u n_u$. To simplify the analysis, we employ the L_1 metric (Manhattan distance metric) for computing the distance between two points; i.e., $d(p_1, p_2) = \|p_1 - p_2\|_1$. The travel direction of the L_1 metric relative to the sides of a service region is shown in Figure 1(a). If the service region is infinite (i.e., a plane), the optimal configuration that minimizes the expected distance between the facility and demand is a collection of nonoverlapping identical diamond-shaped tiles with facilities located at the center of the tile because the customers always go to the nearest facility (Beckmann 1968 and Newell 1973). We adopt this optimal tiling scheme in formulating the model for a finite but sufficiently large service region. An infinite homogeneous plane (sufficiently large service area with uniform demand) is a common assumption made in the CA literature, such as in Cui et al. (2010) and Li and Ouyang (2010). Under this assumption, the number of facilities are often expressed as the density of facilities per unit area. For our case, the density of each type of facility corresponds to n_r/A and n_u/A, respectively. For further discussions and validations on CA, see Daganzo (2005) and Ouyang and Daganzo (2006).

To facilitate the derivation of the model, we first consider a subregion $\Omega_\circ$ with area $A_\circ$, as shown in Figure 1(b). The expected distance between a facility in the center and demand is given as

$$\mathbb{E}[D] = \int_{\Omega_\circ} \mathbb{P}(D > x)\, dx = \int_0^{\sqrt{A_\circ/2}} [1 - F_D(x)]\, dx = \frac{2}{3}\sqrt{\frac{A_\circ}{2}}, \qquad (1)$$

where $F_D(x) = 2x^2/A_\circ$.

Consider that $\Omega_\circ$ is covered by $n_\circ$ facilities, each serving a diamond area of $A_\circ/n_\circ$. Among the $n_\circ$ facilities in the subregion $\Omega_\circ$, assume that only one is reliable. Because the reliable facility does not fail, it is best to locate the tile with reliable facility in the *center* of $\Omega_\circ$ surrounded by $(n_\circ - 1)$ tiles with unreliable facilities. Thus, customer demands will be served by the facility from its own tile as a primary assignment and by the reliable facility from the central tile as a backup assignment. This assumption restricts the ratio between the total number of facilities and the number of reliable facilities to $n_t/n_r = n^2$, $\forall\, n \in \mathbb{N}$. An example of such a configuration is illustrated in Figure 1(c). For mathematical convenience, we obtain an approximation by dropping the integrality requirement on n. In Online Appendix A, we show that the approximation error resulting from this assumption is small. Within each tile, if its own facility is working, the expected distance between the facility and demand is given by $(2/3)\sqrt{(A_\circ/2n_\circ)}$ (using (1)). The expected distance between the reliable facility and demands from the tiles with unreliable facilities, $\mathbb{E}[D^u]$, can be obtained as follows. From (1), the expected distance between the reliable facility and demands in the subregion $\Omega_\circ$ is $(2/3)\sqrt{A_\circ/2}$ and the expected distance between the reliable facility and demands in the central tile is $(2/3)\sqrt{A_\circ/(2n_\circ)}$. Thus, we know that

$$\frac{2}{3}\sqrt{\frac{A_\circ}{2}} = \frac{2}{3}\sqrt{\frac{A_\circ}{2n_\circ}}\left(\frac{1}{n_\circ}\right) + \mathbb{E}[D^u]\left(\frac{n_\circ - 1}{n_\circ}\right).$$

This implies that

$$\mathbb{E}[D^u] = \left(\frac{n_\circ}{n_\circ - 1}\right)\left[\frac{2}{3}\sqrt{\frac{A_\circ}{2}} - \frac{2}{3n_\circ}\sqrt{\frac{A_\circ}{2n_\circ}}\right]. \qquad (2)$$

Now consider that the entire service region Ω with area A is tiled with nonoverlapping identical copies of $\Omega_\circ$ described above (a configuration that contains one reliable facility in the center surrounded by some number of unreliable facilities). Hence, the facilities are evenly spread with each reliable facility surrounded by some number of unreliable facilities. Extending the results from Beckmann (1968) and Newell (1973), one can show that this tiling scheme is optimal when the service region is an infinite plane. For a finite service region, this is an approximation because the boundary of the service region may not be tiled using symmetric diamonds. We consider that

Ω is served by n_r reliable facilities and n_u unreliable facilities (i.e., covered by $(n_r + n_u)$ tiles). We denote the total number of facilities by $n_t (= n_r + n_u)$. Because we assume Ω to be sufficiently large, we drop the integrability constraint on the number of facilities for mathematical convenience.

Within each tile, if its own facility is working, the expected distance between the facility and demand is given by $(2/3)\sqrt{A/(2n_t)}$ from (1). Hence, the expected transportation costs for serving the demands in the service region Ω is

$$\frac{2}{3}\sqrt{\frac{A}{2n_t}}\rho Ac\left(\frac{n_r}{n_t}\right) + \left[(1-q)\frac{2}{3}\sqrt{\frac{A}{2n_t}}\rho Ac\left(\frac{n_u}{n_t}\right) + q\mathbb{E}[D^u]\rho Ac\right]\left(\frac{n_u}{n_t}\right). \quad (3)$$

The first term corresponds to the expected transportation cost for the demands in the tiles with reliable facilities. The second term corresponds to the expected transportation cost for the demands in the tiles with unreliable facilities taking disruption into account. Observe that for every n_u/n_r unreliable facilities, we have one reliable facility. Again, we assume that the tiling is such that each tile with a reliable facility is surrounded by $n_\circ - 1 = n_u/n_r$ tiles with unreliable facilities. Thus, we derive $\mathbb{E}[D^u]$ by setting $A_\circ = A/n_r$, $n_\circ = n_t/n_r$ in (2):

$$\mathbb{E}[D^u] = \frac{2}{3}\sqrt{\frac{A}{2n_r}}\left(\frac{n_t}{n_u}\right) - \frac{2}{3}\sqrt{\frac{A}{2n_t}}\left(\frac{n_r}{n_u}\right). \quad (4)$$

Finally, the expected total cost can be derived by adding the facility costs and the expected transportation costs (using (3) and (4)). After algebra, we obtain the following:

$$\mathbb{E}[TC] = f_r n_r + f_u n_u + q\gamma\sqrt{\frac{1}{n_r}} + (1-q)\gamma\sqrt{\frac{1}{n_t}}, \quad (5)$$

where $\gamma = (\sqrt{2}/3)\rho A^{3/2} c$ is a constant. Therefore, the optimal solution of the problem is given as follows.

Proposition 1 (Facility Deployment Policy). *The threshold disruption probability, $q_{th} = (f_r - f_u)/f_r$, divides the supply chain network structure into two regimes: When $q \le q_{th}$, it is optimal to deploy both types of facilities using*

$$n_r^* = \left(\frac{\gamma q}{2(f_r - f_u)}\right)^{2/3} \quad \text{and} \quad n_t^* = \left(\frac{\gamma(1-q)}{2f_u}\right)^{2/3}. \quad (6)$$

Otherwise (when $q > q_{th}$), it is optimal to deploy only reliable facilities using

$$n_r^* = \left(\frac{\gamma}{2f_r}\right)^{2/3} \quad \text{and} \quad n_u^* = 0. \quad (7)$$

By substituting the optimal numbers of facilities, we obtain the optimal expected total cost as

$$\mathbb{E}[TC(n_r^*, n_u^*)] = \begin{cases} 3\left(\frac{\gamma}{2}\right)^{2/3}[q^{2/3}(f_r - f_u)^{1/3} + (1-q)^{2/3} f_u^{1/3}] & \text{if } q \le q_{th}, \\ 3\left(\frac{\gamma}{2}\right)^{2/3} f_r^{1/3} & \text{if } q > q_{th}. \end{cases} \quad (8)$$

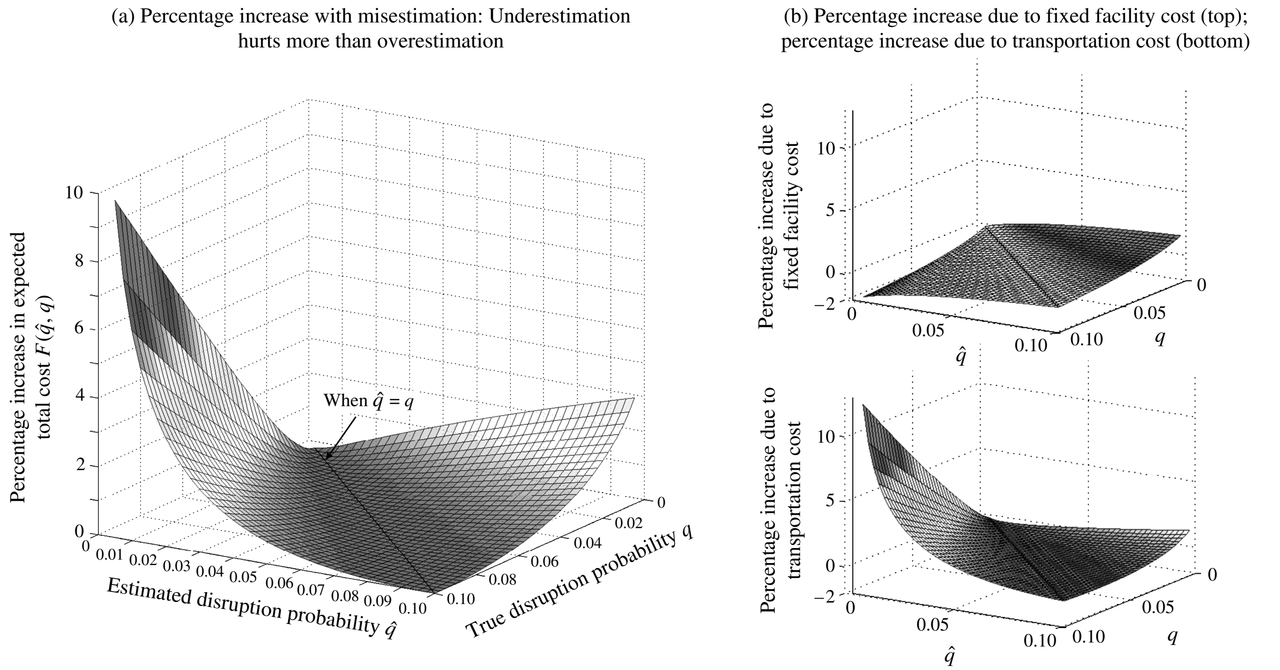
Observe that the threshold $q_{th} = (f_r - f_u)/f_r$ approaches zero as the fixed cost of a reliable facility f_r approaches f_u. This means that if the disruption probability is large or the incremental facility hardening cost is small relative to the fixed cost of an unreliable facility, it is best to locate only reliable facilities. However, if the disruption probability is low or the facility hardening cost is large, it is optimal to use both types of facilities. We do not consider the regime in which we only use unreliable facilities because $q > 0$ and at least one reliable facility must exist to satisfy this condition. Note that this problem is an extension of the analytic version of the uncapacitated fixed-charge location problem (UFLP) (Drezner and Hamacher 2002). The problem reduces to the UFLP when q approaches zero. A simpler model on a one-dimensional line is provided in Online Appendix B.

It is interesting to note that when q is small, even a slight change in q strongly impacts the number of reliable facilities $n_r^* = (\gamma q/(2(f_r - f_u)))^{2/3}$ because n_r^* is proportional to $q^{2/3}$. In contrast, the total number of facilities $n_t^* = (\gamma(1-q)/(2f_u))^{2/3}$ is proportional to $(1-q)^{2/3}$, making the entire term relatively insensitive to changes in q. Thus, for a low value of q, although the total number of facilities stays relatively stable, the fraction of reliable facilities changes significantly with q. This suggests that misestimation can have a significant impact on network design, especially when the disruption probability is small. In the next section, we study how the estimation error in disruption probability impacts the expected total cost.

4. Impact of Misestimating the Disruption Probability

To analyze the impact of misestimating the disruption probability, we define $F(q, \hat{q}, f_u, r)$ as the percentage increase (relative difference) in the expected total cost when the true disruption probability q is estimated by $\hat{q}$. We refer to the facility hardening cost factor f_r/f_u, r. For brevity of notation, we drop f_u and r and let $F(q, \hat{q})$ represent the percentage increase in the

Figure 2 Asymmetry in the Impact of Disruption Probability Misestimation



expected total cost for fixed values of f_u and r. Thus, we have

$$F(q, \hat{q}) = \frac{TC(q, n_r(\hat{q}), n_u(\hat{q}))}{TC(q, n_r(q), n_u(q))} - 1,$$

where $TC(q, n_r(\hat{q}), n_u(\hat{q}))$ denotes the expected total cost when $\hat{q}$ is the estimate used for the true disruption probability q.

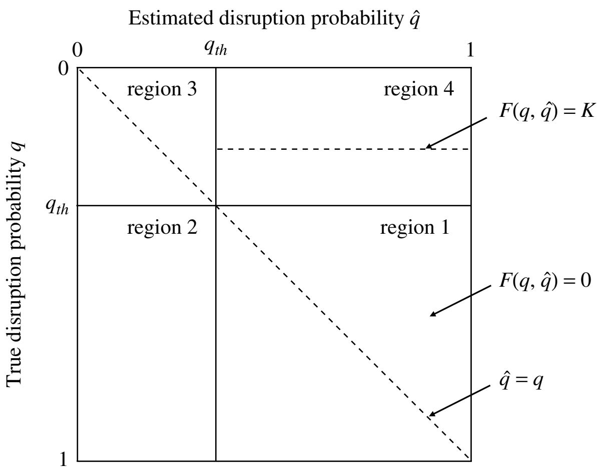
We first illustrate a numerical example on the impact of misestimating the disruption probability q in which the parameters are set to $A = 3{,}050{,}456$ (square miles), $\rho = 25.7$, $f_u = \$1{,}000{,}000$, $c = 0.005$, and $r = 1.25$, where $q \in [0.005, 0.1]$. The diagonal line represents the case in which the disruption probability is accurately estimated ($\hat{q} = q$). Figure 2(a) exhibits that the percentage increase in the expected total cost is asymmetric about this diagonal. Moreover, *the percentage increase in the expected total cost is higher when the disruption probability is underestimated compared to when it is overestimated*. The asymmetry between the impact of underestimation and overestimation is especially important given the human tendency to underestimate the probability of a disruption as the last such event recedes further into the past (Swiss Re 2009). A similar message with regards to rare events is also given by Taleb (2007), who claims that rare events tend to be underpriced by the market and can result in a disastrous consequence.

When q is overestimated, the network designer overinvests in facility cost while reducing the transportation cost. In contrast, when q is underestimated, the transportation is increased due to fewer facilities. This is captured in Figure 2(b), where the percentage increase in the expected total cost is decomposed into two components: fixed facility cost and the transportation cost. As illustrated in the figures, the transportation cost increases more steeply as the underestimation error rate increases and thus contributes to significant increase in the percentage increase.

To further study the impact of misestimation, we define four regions based on whether q and $\hat{q}$ are above or below q_{th}, as illustrated in Figure 3. In region 1, both q and $\hat{q}$ are above q_{th}, and the optimal network contains only reliable facilities. Thus, $F(q, \hat{q}) = 0$; i.e., there is no cost of misestimation. In region 4, $\hat{q} > q_{th}$, whereas $q < q_{th}$. Thus, for a given q, $F(q, \hat{q})$ is a constant because all facilities are hardened. In region 2, the disruption probability is underestimated with $\hat{q} < q_{th}$ when $q > q_{th}$. The percentage increase in the expected total cost increases with q in this region. Finally, in region 3, both q and $\hat{q}$ are below q_{th}. In this region, both underestimation and overestimation can occur. Because this is the most relevant region for many firms in practice, we focus on studying this particular region analytically.

For analytical convenience, we assume that $q < \hat{q} < 2q$. This allows us to define an estimation error rate $\delta (0 < \delta < 1)$, where an underestimation is presented as $\hat{q} = q(1-\delta)$ and an overestimation as $\hat{q} = q(1+\delta)$. Furthermore, we limit the range of disruption probability to $q \leq q_{th}/2$ so that $\hat{q} \in (0, q_{th}]$ for any δ. This is to focus on the more practical and interesting

Figure 3 **Four Regions of Possible Estimations on q-$\hat{q}$ Plane**



setting and exclude tedious cases where all facilities are hardened. We obtain the following result on the impact of misestimation.

Proposition 2 (Comparative Statics on the Impact of Misestimation). *The percentage increase in the expected total cost with fixed misestimation error rate δ increases with q; i.e., $\partial F(q, q(1 \pm \delta))/\partial q > 0$.*

This implies that the impact of misestimation (represented by the percentage increase in the expected total cost) increases with the true disruption probability. We now contrast the impact of underestimation to that of overestimation more specifically.

Proposition 3 (Underestimation vs. Overestimation). *The percentage increase in the expected total cost on underestimation with error rate δ is always greater than that from overestimation; i.e., $F(q, q(1+\delta)) < F(q, q(1-\delta))$.*

Proposition 4 (Impact of Estimation Error Rate). *The percentage increase in the expected total cost for underestimating the disruption probability by error rate δ relative to overestimation increases with δ; i.e., $F(q, q(1-\delta))/F(q, q(1+\delta))$ is 1 when $\delta = 0$ and increases with δ.*

Proposition 3 confirms the observation from Figure 2(a) that underestimating the disruption probability is more costly than overestimation by the same amount. Furthermore, Proposition 4 reveals that the impact of underestimation relative to overestimation increases as estimation error increases. This asymmetric risk structure between the impact of underestimation and overestimation suggests that managers must be particularly aware of the danger of underestimating the disruption probability. In Online Appendix C, we numerically test this under a more generalized setting (using Lim et al. 2010) incorporating heterogeneous location, demand, and facility cost on a discrete network. Although the modeling framework and some assumptions are relaxed, we observe a strong concurrence in general trends with the results of the continuous model.

Because historical data on disruptions may be sparse and new types of disruptions are constantly emerging, it may be very difficult to estimate the disruption probability in practice. Hence, it is important to consider the robustness of the planning decision when parameters (i.e., disruption probability q) are given as an interval estimate (e.g., as a result of confidence interval). See Ben-Tal et al. (2011) for applying robust optimization approach on emergency logistics and Li et al. (2011) for more general literature on risk management in supply chains. Our numerical results indicate that there are significant gains if the true disruption probability can be bounded within a reasonable range. Thus, we study a disruption probability estimate that optimally hedges against the worst-case scenario. More precisely, we obtain $\hat{q}^*$ that minimizes the maximum relative regret F, which measures the percentage increase in the expected total cost, when q lies between $\underline{q}$ and $\bar{q}$.

Proposition 5 (Optimal Estimate for the Disruption Probability). *For every $\underline{q}$ and $\bar{q}$ such that $q \in (\underline{q}, \bar{q})$, there exists $\hat{q}^*$ that satisfies*

$$\inf_{\hat{q} \in (\underline{q}, \bar{q})} \sup_{q \in (\underline{q}, \bar{q})} F(q, \hat{q}) = \sup_{q \in (\underline{q}, \bar{q})} F(q, \hat{q}^*). \tag{9}$$

Further, if $\underline{q} < q_{th}$, then $\hat{q}^*$ is unique and is strictly less than the threshold q_{th} ($\hat{q}^* < q_{th}$). Else, any $\hat{q}^* \in (\underline{q}, \bar{q})$ satisfies (9).

Given any range within which the disruption probability is bounded, Proposition 5 shows the existence of $\hat{q}^*$ that balances the error from over- and underestimation. If $\underline{q} \geq q_{th}$, $\hat{q}^*$ may take any values in $(\underline{q}, \bar{q})$ because all facilities are hardened.

To draw further insight, we numerically compute $\hat{q}^*$. We set the range of true disruption probability $q \in (\underline{q}, \bar{q})$ in relation to the median estimate $\tilde{q}$ by setting $\underline{q} = \tilde{q}/(1+\delta)$ and $\bar{q} = \tilde{q}/(1-\delta)$. To summarize our numerical results, we define $\bar{F}(q') = \sup_{q \in (\underline{q}, \bar{q})} F(q, q')$. In Table 1, we compute $\hat{q}^*$ with $\tilde{q} = 0.03$ for various levels of r. Other parameters are set identical to the earlier case. For example, when the estimation error rate is $\delta = 0.75$, the optimal estimate for the disruption probability is $\hat{q}^* = 0.0578$ for $r = 1.5$. Note that $\hat{q}^* = 0.0578$ is greater than the median estimate $\tilde{q} = 0.03$ because the impact of underestimation is greater than that of overestimation.

Although $\hat{q}^*$ is a very conservative estimate (hedging against the worst case), we observe that using this estimate results in a network whose cost is not too far from that of the optimal network with true disruption probability. For example, for $\delta = 0.75$ and $r = 1.25$, we see only a 1.00% increase in the expected total cost

Table 1 **Optimal Estimate and Percentage Increase Under the Worst-Case Scenario**

	$\tilde{q}=0.03$; $\delta=0.25$			$\tilde{q}=0.03$; $\delta=0.5$			$\tilde{q}=0.03$; $\delta=0.75$			$\tilde{q}=0.03$; $\delta=0.9$		
r	$\hat{q}^*$	$\bar{F}(\hat{q}^*)$ (%)	$\bar{F}(\tilde{q})$ (%)	$\hat{q}^*$	$\bar{F}(\hat{q}^*)$ (%)	$\bar{F}(\tilde{q})$ (%)	$\hat{q}^*$	$\bar{F}(\hat{q}^*)$ (%)	$\bar{F}(\tilde{q})$ (%)	$\hat{q}^*$	$\bar{F}(\hat{q}^*)$ (%)	$\bar{F}(\tilde{q})$ (%)
1.25	0.0315	0.04	0.06	0.0376	0.22	0.47	0.0582	0.82	2.82	0.0872	1.57	7.31
1.5	0.0315	0.05	0.08	0.0375	0.27	0.57	0.0578	1.00	3.40	0.0859	1.91	8.68
2	0.0315	0.07	0.10	0.0374	0.33	0.70	0.0573	1.22	4.75	0.0846	2.31	10.21
5	0.0315	0.10	0.14	0.0372	0.49	1.02	0.0562	1.80	5.72	0.0815	3.32	13.74

for using $\hat{q}^*$, whereas one may incur a 3.40% increase by naively using $\tilde{q}$. This suggests that, when planned properly, the cost of mitigating the misestimation risk is not too high. In the next section, we extend the model to incorporate correlated facility disruptions.

5. Facility Location Model with Correlated Disruptions

The goal of this section is to understand how correlation in disruption impacts the insights obtained in §4 and to derive relevant managerial insights. To capture the impact of correlated disruptions, we now assume that reliable facilities have a finite capacity K. Note that, without capacity, correlation does not affect the "expected" total cost. In addition to the cost components in the uncapacitated model (§4), we impose a unit penalty cost of p to the reliable facility when it serves additional demand beyond its capacity. The penalty cost at the reliable facility can be interpreted as a premium for utilizing emergency supply options such as outsourcing or overtime production. Hence, the capacitated facility reliability problem is to minimize the expected total cost that consists of facility fixed cost, expected transportation cost, and expected penalty cost. Although capacity is another important decision a firm should make, we leave it as an exogenous parameter to focus on the relationship between the impact of misestimation and correlation in disruption probability.

Following the construction of the uncapacitated model, we again consider a subregion in which one reliable facility is located in the center surrounded by $m=n_u/n_r$ unreliable facilities. With capacity K, the reliable facility first serves its own demand, $A\rho/n_t$, then serves additional demand from unreliable facilities that have failed. If facility disruptions occur independently with probability q, each disruption from m unreliable facilities, assuming m is a positive integer, can be considered as a Bernoulli trial. Furthermore, the total number of disruptions then follows a binomial distribution with parameters (m, q). If disruptions are *correlated*, we can consider each disruption from m unreliable facilities as a correlated Bernoulli (0–1) random variable Y_i, with failure probability q and covariance matrix Σ where its elements are denoted as σ_{ij}; i.e., $\mathbb{P}(Y_i=1)=q$ and $\operatorname{cov}(Y_i, Y_j)=\sigma_{ij}$ for $0<i, j<m$. Furthermore, the total number of disruptions follows a correlated binomial random variable X with parameters (m, q, Σ). Because the number of unreliable facilities in each subregion, $m=n_u/n_r$, may not be integer, we approximate $X(m, q, \Sigma)$ as follows:

$$X(m,q,\Sigma) := \begin{cases} \sum_{i=1}^{\lfloor m\rfloor} Y_i & \text{w.p. } (\lceil m\rceil - m), \\ \sum_{i=1}^{\lceil m\rceil} Y_i & \text{otherwise.} \end{cases}$$

Finally, we define the capacitated reliability model as follows:

$$\begin{aligned} \min \mathbb{E}[TC] &= f_r n_r + f_u n_u + q\gamma\sqrt{\frac{1}{n_r}} + (1-q)\gamma\sqrt{\frac{1}{n_t}} \\ &\quad + p n_r \mathbb{E}\left[\frac{A\rho}{n_t}\left(1+X\left(\frac{n_u}{n_r}, q, \Sigma\right)\right) - K\right]^+. \end{aligned} \tag{10}$$

Note that in the above formulation, we assumed that the firm uses nonoverlapping diamond-shaped tiling scheme as from the uncapacitated model. If, however, the penalty cost is high enough (and transportation cost is cheap), it is possible that the firm is better off having overlapping covering regions for each reliable facility. In contrast, if the penalty cost is low enough (in relation to the transportation cost), then the current configuration and backup assignment setting is optimal.

In practice, disruptions often exhibit local correlation, where the correlation between two facility disruptions decays as the distance between those facilities grows. For example, geographical dependence may induce such correlation structure with positive correlation. One can also envision cases in which the disruptions are negatively correlated. For example, failures due to disease-stricken workers would raise alert for the nearby facilities and may result in reduced disruption probability. Whereas the above model allows any arbitrary correlation structure among the m unreliable facilities, we let $\sigma_{ij}(d) = (d/(|i-j|+1))q(1-q)$ for $i \neq j$ and $\sigma_{ij}=1$ for $i=j$, where d represents the degree of correlation. Hence, $\sigma_{ij}(d)$ increases with d, and $d<0$ represents the negative correlation. (One may extend the model by capturing the density of the facility n_t/A in determining the degree of correlation. In this study, we limit the analysis to a simplified version because the results qualitatively remain the same.) If $d = 0$, X simply reduces to a (uncorrelated) binomial distribution with parameters (m, q). Contrasting the independent disruption case to the correlated case, we derive the following result.

Proposition 6. *The expected total cost for the capacitated facility reliability problem increases with the degree of correlation d and decreases with capacity K.*

Intuitively, a positive (negative) correlation among facility disruptions increases (decreases) the variance of the total number of disruptions while its mean remains the same. Hence, for fixed K, the expected penalty cost, and thus the expected total cost, increases with disruption correlation. In addition, the decrease in capacity for each reliable facility increases the penalty cost and thus leads to higher expected total cost. We note that the first part of this result is in line with the numerical findings of Li and Ouyang (2010), albeit with different modeling assumptions and setting.

Similar to the previous section, we conduct a numerical study to analyze the impact of misestimation in disruption probability. Figure 4 contrasts the impact of underestimation to that of overestimation with varying degrees of correlation d and capacity K. The disruption probability estimate $\hat{q}$ ranges from 0 to 0.10 when the true disruption probability q is 0.05. Parameters are set identically as the previous examples with $p = 1$, $d = 1$, and $K = 4{,}000{,}000$ unless specified otherwise. We assume that the correlation degree d is known at the moment.

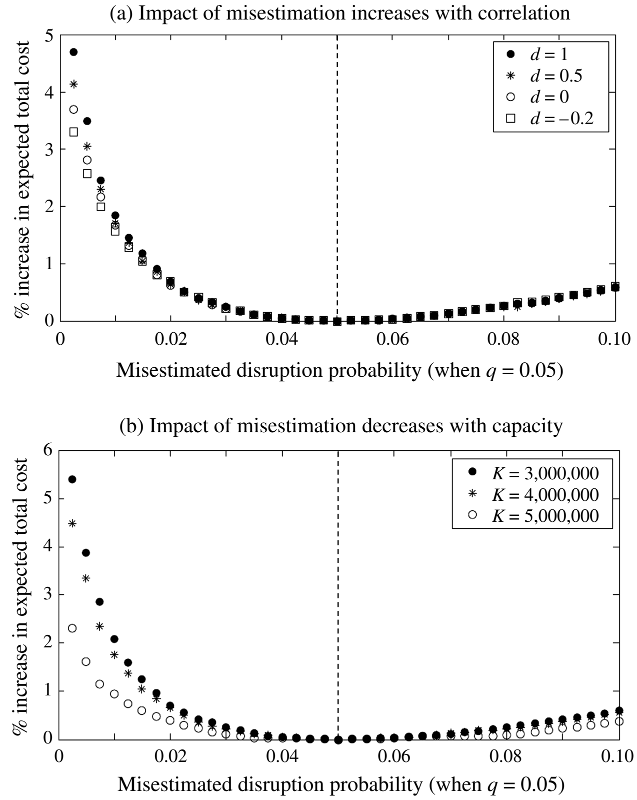
First, we observe that the impact of underestimation is greater than that of overestimation as from the uncapacitated (and independent disruption) case. In addition, Figure 4(a) shows the impact of underestimation increases with degree of correlation in particular, whereas the difference in overestimation is inconspicuous. Figure 4(b) shows that the impact of misestimation reduces as capacity increases. This suggests the value of capacity investment for hedging against the risk of disruption probability misestimation, especially against the risk of underestimation.

We now consider the case in which both the disruption probability and the correlation degree are misestimated. To facilitate the analysis, we define the percentage increase in the expected total cost to be

$$G(q, \hat{q}, d, \hat{d}) = \frac{TC(q, d, n_r(\hat{q}, \hat{d}), n_u(\hat{q}, \hat{d}))}{TC(q, d, n_r(q, d), n_u(q, d))} - 1,$$

where $TC(q, d, n_r(\hat{q}, \hat{d}), n_u(\hat{q}, \hat{d}))$ denotes the expected total cost when q and d are estimated by $\hat{q}$ and $\hat{d}$,

Figure 4 **Comparative Statics with Respect to d and K on the Impact of Misestimation**



respectively, for some fixed f_u and r. Figures 5(a) and 5(b) are contour plots of $G(q, \hat{q}, d, \hat{d})$ with respect to $\hat{q}$ (ranging from 0 to 0.1 for case (a) and from 0.04 to 0.06 for case (b)) and $\hat{d}$ (ranging from −0.2 to 1) for $q = 0.05$ and $d = 0.5$. Hence, G is 0 when both q and d are correctly estimated, and G increases as misestimation errors in $\hat{q}$ and $\hat{d}$ increase. We set all other parameters identical to the previous examples.

In Figure 5(a), it is interesting to note that the system is relatively insensitive to the misestimation in d compared to that of q. This suggests that managers should focus primarily on estimating the disruption probability. As the estimation accuracy in q increases, however, we observe that the misestimation in d starts to play a role as shown in Figure 5(b): When the disruption probability q is underestimated, overestimation in d reduces the impact of joint misestimation, whereas underestimation in d magnifies the impact even further. This is because the total number of disruption increases in q and its variance increases in d. Hence, a misestimation in opposite directions reduces the impact, whereas misestimation in the same direction magnifies the impact. In addition, we observe that the impact of joint underestimation (southwest corner) is greater than that of joint overestimation (northeast corner).

In the presence of capacity and penalty cost, we can consider the probability that a reliable facility

Figure 5 Percentage Increase in the Expected Total Cost When Both q and d are Misestimated

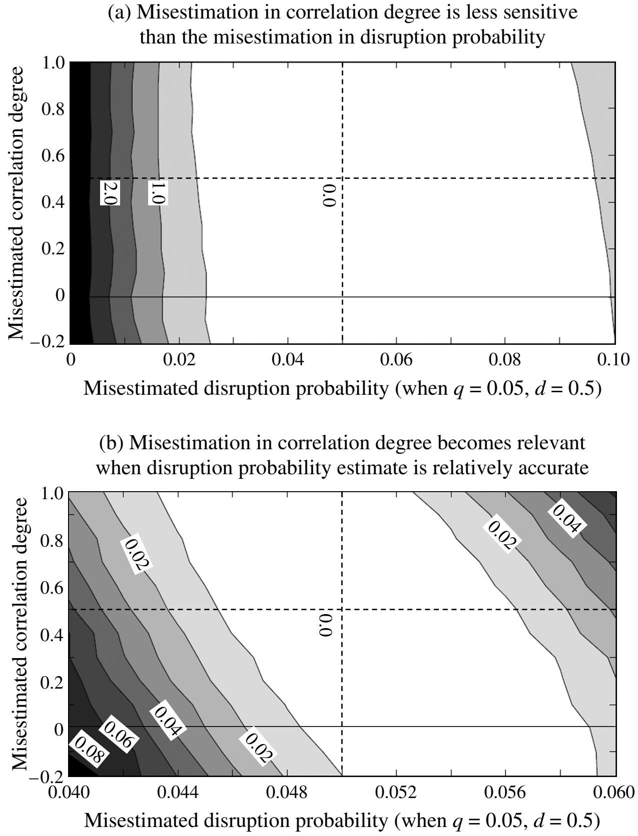
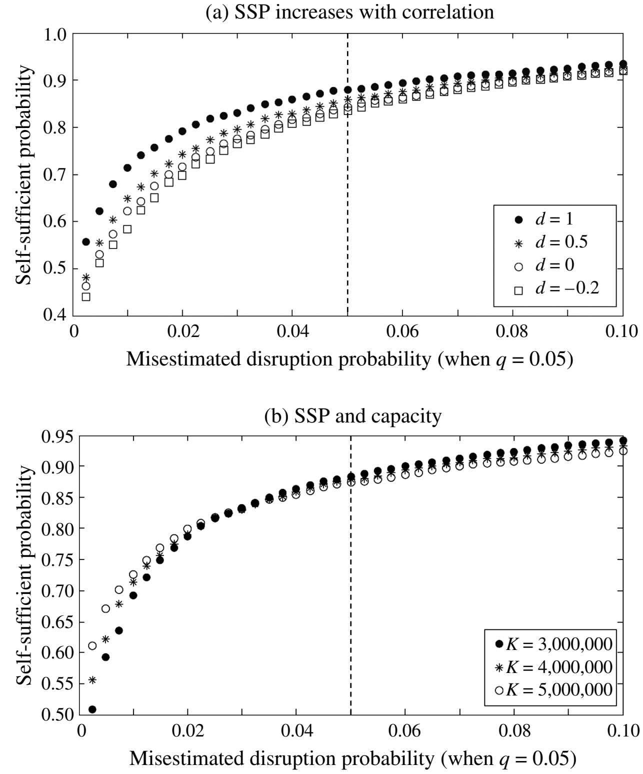


Figure 6 Comparative Statics with Respect to d and K on the Self-Sufficient Probability



does not incur penalty costs. We refer to this measure as *self-sufficient probability* (SSP) and define it as $\mathbb{P}((A\rho/n_t^*)(1 + X(n_u^*/n_r^*, q, \Sigma)) < K)$ from solving (10). Managers might be interested in SSP because low SSP represents high dependency on outsourcing or heavy stress on the system due to constant overtime production. Figure 6 illustrates the SSP for various levels of correlation and capacity. We observe that the impact of underestimation is greater than that of overestimation for the SSP; that is, the disparity level in the SSP—the difference between the target SSP (when $q = \hat{q} = 0.05$) and the resulting SSP due to misestimation—is greater when q is underestimated than overestimated. This is because the number of reliable facility reduces significantly as $\hat{q}$ decreases; thus, the system relies more on the emergency supply options (and incurs penalty cost more often). Interestingly, Figure 6(a) shows that SSP increases with correlation. This is because the positive correlation leads to a greater number of reliable facilities in the system, which in turn increases the SSP. Figure 6(b) shows that the range of SSP when q varies from 0 to 0.1 increases as capacity becomes tighter.

6. Conclusions

In this paper, we consider a facility location problem in the presence of random facility disruption where facilities can be protected with additional investments. Using a stylized continuous approximation model, we characterize the structure of the optimal distribution network and obtain managerial insights on the impact of misestimating the disruption probability. Based on our findings, we summarize the following guidelines for managers designing robust distribution networks:

1. When the estimate intervals of disruption probability are wide, misestimating the disruption probability can be expensive, especially when it is underestimated. We observe a large asymmetry in the risk structure between over- and underestimation. Never underestimate what can go wrong.

2. For a given range of disruption probability estimate, we suggest an estimate $\hat{q}^*$ that minimizes the worst-case scenario. Although this is a very conservative estimate, using this does not increase the cost too much compared to the optimal network with true disruption probability. When planned properly, the cost of mitigating the misestimation risk is not too high.

3. Underestimation in both disruption probability and correlation degree result in a greater increase in the expected total cost compared to the overestimation. However, the expected total cost is much less sensitive to the misestimation of correlation degree relative to the misestimation of disruption probability. Thus, accurately estimating the disruption probability is more critical than the correlation. Furthermore,

when the uncertainty in the disruption probability is high (and, moreover, if positive correlation is expected), investing on capacity is recommended.

Our findings pose new questions and motivate additional future research. First, we do not account for partial facility failures or investments that make a facility partially reliable. Second, capacity is not considered as a decision variable. It would be informative to investigate how insights would change given capacity as an endogenous factor. Last, our work is limited to static decisions and does not consider dynamic facility deployment policy. Taking advantage of the continuous model, one can examine the dynamics of the optimal facility decision by updating the disruption probability estimate over time. These questions are outside the scope of the current paper and are left for future research.

Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/msom.1120.0413.

Acknowledgments

The research was supported in part by the National Science Foundation [Grant DMI-0457503] and the Office of Naval Research [Grant N00014-05-1-0190]. This support is gratefully appreciated. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Office of Naval Research.

References

Bakkaloglu M, Wylie JJ, Wang C, Ganger GR (2002) On correlated failures in survival storage systems. Technical Report CMU-CS-02-129, Carnegie Mellon University, Pittsburgh.

Ball MO, Lin FL (1993) A reliability model applied to emergency service vehicle location. *Oper. Res.* 41(1):18–36.

Baron O, Berman O, Krass D (2008) Facility location with stochastic demand and constraints on waiting time. *Manufacturing Service Oper. Management* 10(3):484–505.

Batta R, Mannur NR (1990) Covering-location models for emergency situations that require multiple response units. *Management Sci.* 36(1):16–23.

Beckmann M (1968) *Location Theory* (Random House, New York).

Ben-Tal A, Chung BD, Mandala S, Yao T (2011) Robust optimization for emergency logistics planning: Risk mitigation of humanitarian relief supply chains. *Transportation Res. B.* 45(8):1177–1189.

Berman O, Krass D, Menezes M (2007) Facility reliability issues in network p-median problems: Strategic centralization and colocation effects. *Oper. Res.* 55(2):332–350.

Berman O, Krass D, Menezes M (2009) Locating facilities in the presence of disruptions and incomplete information. *Decision Sci.* 40(4):845–868.

Chopra S, Sodhi MS (2004) Managing risk to avoid supply-chain breakdown. *MIT Sloan Management Rev.* 46(1):53–61.

Chopra S, Reinhardt G, Mohan U (2007) The importance of decoupling recurrent and disruption risks in a supply chain. *Naval Res. Logist.* 54(5):544–555.

Church RL, Scaparra MP (2007) Protecting critical assets: The r-interdiction median problem with fortification. *Geographical Anal.* 39(2):129–146.

Cui T, Ouyang Y, Shen Z-JM (2010) Reliable facility location design under the risk of disruptions. *Oper. Res.* 58(4):998–1011.

Daganzo CF (2005) *Logistics System Analysis*, 4th ed. (Springer, Berlin).

Daganzo CF, Erera AL (1999) *On Planning and Design of Logistics Systems for Uncertain Environments, in New Trends in Distribution Logistics* (Springer, Berlin).

Dasci A, Verter V (2001) A continuous model for production-distribution system design. *Eur. J. Oper. Res.* 129(2):287–298.

Drezner Z, Hamacher HW (2002) *Facility Location: Applications and Theory* (Springer, Berlin).

Hall RW (1984) Travel distance through transportation terminals on a rectangular grid. *J. Oper. Res. Soc.* 35(12):1067–1078.

Hendricks KB, Singhal VR (2005) An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production Oper. Management* 14(1):35–52.

Kleindorfer PR, Saad G (2005) Managing disruption risks in supply chains. *Production Oper. Management* 14(1):53–68.

Langevin A, Mbaraga P, Campbell JF (1996) Continuous approximation models in freight distribution: An overview. *Transportation Res. B.* 30(3):163–188.

Li X, Ouyang Y (2010) A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions. *Transportation Res. B.* 44(4):535–548.

Li J, Chen J, Wang S (2011) *Risk Management of Supply and Cash Flows in Supply Chains* (Springer, New York).

Lim M, Daskin MS, Bassamboo MA, Chopra S (2010) A facility reliability problem: Formulation, properties and algorithm. *Naval Res. Logist.* 57(1):58–70.

Mak HY, Shen ZJ (2011) Agility and proximity considerations in supply chain design—An analytical approach. Working paper, Hong Kong University of Science and Technology, Hong Kong.

Munich Re Group (2008) Topics Geo natural catastrophes 2007. Report. Accessed August 2011, http://www.munichre.com/publications/302-05699_en.pdf.

Newell GF (1973) Scheduling, location, transportation and continuous mechanics: Some simple approximations to optimization problems. *SIAM J. Appl. Math.* 25(3):346–360.

Ouyang Y, Daganzo CF (2006) Discretization and validation of the continuum approximation scheme for terminal system design. *Transportation Sci.* 40(1):89–98.

Pirkul H, Schilling DA (1988) The siting of emergency service facilities with workload capacities and backup service. *Management Sci.* 34(7):896–908.

Rice JB, Caniato F (2003) Building a secure and resilient supply network. *Supply Chain Management Rev.* 7(5):22–30.

Sheffi Y, Rice J (2005) A supply chain view of the resilient enterprise. *MIT Sloan Management Rev.* 47(1):41–48.

Snyder LV, Daskin MS (2005) Reliability models for facility location: The expected failure cost case. *Transportation Sci.* 39(3):400–416.

Snyder LV, Daskin MS (2006) Stochastic p-robust location problems. *IIE Tran.* 38(11):971–985.

Swiss Re (2009) Sigma report: Natural catastrophes and Man-made disasters. Report. Accessed August 2011, http://www.swissre.com/resources/dd6346004d4e9669ac76eecedd316cf3-sigma2_2009_e.pdf.

Taleb N (2007) *The Black Swan: The Impact of the Highly Improbable* (Random House, New York).

Tang CS (2006) Perspectives in supply chain risk management. *Internat. J. Production Econom.* 103(2):451–488.

Tomlin B (2006) On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Sci.* 52(5):639–657.

Wang N, Lu JC, Kvam P (2006) Reliability modeling in spatially distributed logistics systems. *IEEE Tran. Reliability* 55(3):525–534.