Facility Location Decisions with Random Disruptions and Imperfect Estimation

Michael K. Lim
Department of Business Administration, University of Illinois,
Urbana-Champaign, IL 61820, USA, mlim@illinois.edu

Achal Bassamboo, Sunil Chopra
Department of Managerial Economics and Decision Sciences, Northwestern University, Evanston, IL 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, MI 48109,
msdaskin@umich.edu

August 4, 2012

Abstract

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. While 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.

Keywords: logistics and transportation; supply chain disruptions; facility network design; estimation error; correlated disruptions; continuous approximation

1. Introduction

Over the past few decades, significant effort has been expended in making supply chains leaner and cheaper. However, recent studies indicate that, while such efforts have successfully reduced supply chain operational costs, they have concomitantly increased supply chain risks (Rice and Caniato 2003, and 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 persists 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. Further, 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), please 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 since 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 the estimation of the disruption probability is imperfect (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, Section 6 concludes the paper. All proofs are presented in the Online Appendix.

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, and 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 back-order cost is high, and the opposite holds when the back-order 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 trade-off 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 lack of analytical tractability) or good solutions for large instances within a limited time frame (since 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 then employed by many others (e.g. Hall 1984, Langevin et al. 1996, Dasci and Verter 2001, and 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 few papers that examine facility location issues in the presence of supply disruption using 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.

While many facility disruption instances exhibit spatial correlation in practice, to the best of our knowledge, Li and Ouyang (2010) is the only one to address the issue in the facility location literature. 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 demand is uniformly distributed with a density of ρ (demand per unit time per sq. 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 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*. While each demand may go to its next nearest available facility (as in Cui et al. (2010)), for analytical tractability, we assume 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, and 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 $\times\,\rho Ac$. We denote the cost 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 non-overlapping identical diamond-shaped tiles with facilities located at the center of the tile since 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 $\frac{n_r}{A}$ and $\frac{n_u}{A}$, respectively. For further discussions and validations on CA, please 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:

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

where $F_D(x) = \frac{2x^2}{A_\circ}$. Consider $\Omega_\circ$ is covered by $n_\circ$ facilities each serving a diamond area of $\frac{A_\circ}{n_\circ}$.

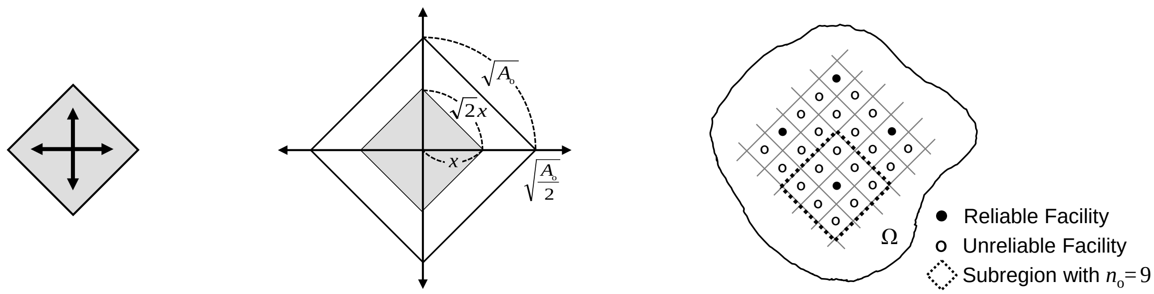


Figure 1: (a) Travel directions; (b) Distance of subregion $\Omega_\circ$; (c) Example Configuration with $n_\circ = 9$

Among the $n_\circ$ facilities in the subregion $\Omega_\circ$, assume that only one is reliable. Since 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 $\frac{n_t}{n_r} = n^2, \forall n \in \mathbb{N}$. An example of such configuration is illustrated in Figure 1(c). For mathematical convenience, we obtain an approximation by dropping the integrality requirement on n. In the 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 $\frac{2}{3}\sqrt{\frac{A_\circ}{2n_\circ}}$ (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 $\frac{2}{3}\sqrt{\frac{A_\circ}{2}}$ and the expected distance between the reliable facility and demands in the central tile is $\frac{2}{3}\sqrt{\frac{A_\circ}{2n_\circ}}$. Thus, we know $\frac{2}{3}\sqrt{\frac{A_\circ}{2}} = \frac{2}{3}\sqrt{\frac{A_\circ}{2n_\circ}}\big(\frac{1}{n_\circ}\big) + \mathbb{E}[D^u]\big(\frac{n_\circ - 1}{n_\circ}\big)$. This implies

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

Now consider the entire service region Ω with area A is tiled with non-overlapping identical copies of $\Omega_\circ$ described above (a configuration which 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 since the boundary of the service region may not be tiled using symmetric diamonds. We consider 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 $(= n_r + n_u)$. Since 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 $\frac{2}{3}\sqrt{\frac{A}{2n_t}}$ from (1). Hence, the expected transportation costs for serving the demands in the service region Ω is:

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

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 $\frac{n_u}{n_r}$ 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 $n_\circ - 1 = \frac{n_u}{n_r}$ tiles with unreliable facilities. Thus, we derive $\mathbb{E}[D^u]$ by setting $A_\circ = \frac{A}{n_r}$, $n_\circ = \frac{n_t}{n_r}$ in (2):

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

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:

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

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

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

$$n_r^* = \Big(\frac{\gamma q}{2(f_r - f_u)}\Big)^{\frac{2}{3}} \;\; \textit{and} \;\; n_t^* = \Big(\frac{\gamma(1-q)}{2f_u}\Big)^{\frac{2}{3}}. \tag{6}$$

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

$$n_r^* = \Big(\frac{\gamma}{2f_r}\Big)^{\frac{2}{3}} \;\; \textit{and} \;\; n_u^* = 0. \tag{7}$$

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

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

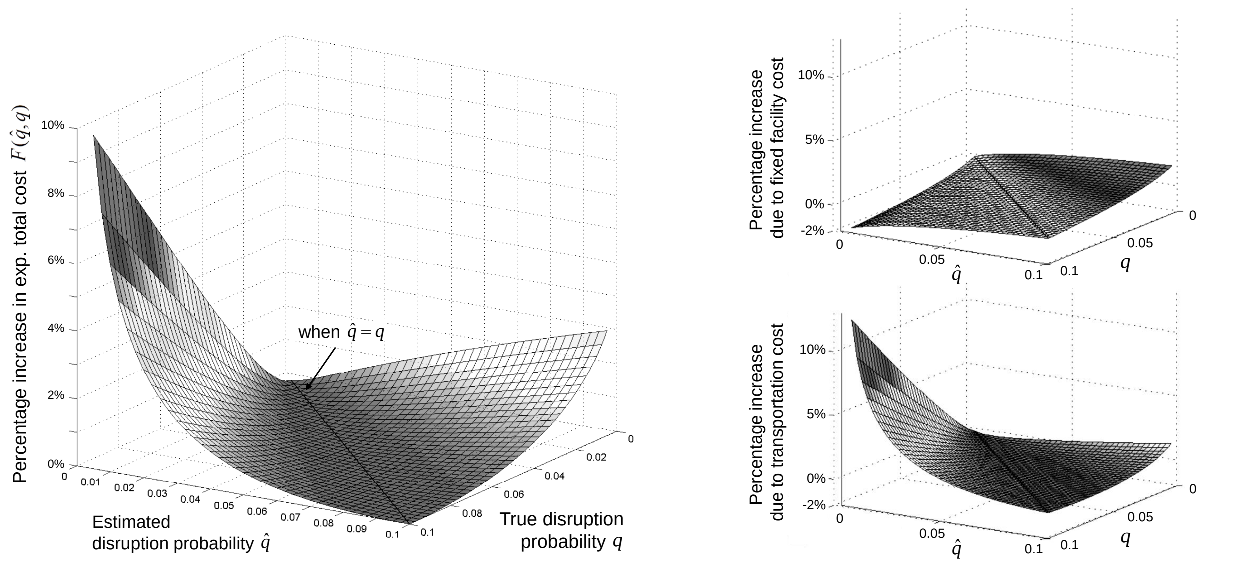
Observe that the threshold $q_{th} = \frac{f_r - f_u}{f_r}$ approaches 0 as the fixed cost of a reliable facility f_r approaches f_u. This means that if the disruption probability is large or the incremental facility hardening cost is small relative to the fixed cost of an unreliable facility, it is 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 since $q > 0$ 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 0. A simpler model on a 1-dimensional line is provided in the 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 $n_r^* = \left(\frac{\gamma q}{2(f_r - f_u)}\right)^{\frac{2}{3}}$ because n_r^* is proportional to $q^{\frac{2}{3}}$. In contrast, the total number of facilities $n_t^* = \left(\frac{\gamma(1-q)}{2f_u}\right)^{\frac{2}{3}}$ is proportional to $(1-q)^{\frac{2}{3}}$ making the entire term relatively insensitive to changes in q. Thus, for a low value of q, while 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 $(\frac{f_r}{f_u})$, r. For brevity of notation, we shall drop f_u and r and let $F(q, \hat{q})$ represent the percentage increase in the 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 $A = 3,050,456$ (sq. mi.), $\rho = 25.7$, $f_u = \$1,000,000$, $c = 0.005$, 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



(a) Percentage increase with misestimation: Underestimation hurts more than overestimation

(b) Percentage increase due to fixed facility cost (top); Percentage increase due to transportation cost (bottom)

Figure 2: Asymmetry in the impact of disruption probability misestimation

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 over-invests 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 since 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. Since this is the most relevant region for many firms in practice, we focus on studying this particular region analytically.

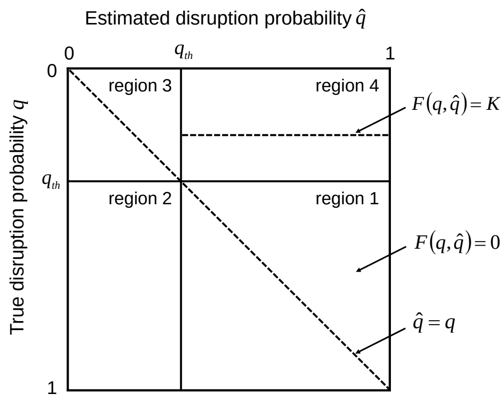


Figure 3: Four regions of possible estimations on q-$\hat{q}$ plane

For analytical convenience, we assume $q < \hat{q} < 2q$. This allow 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 $\hat{q} = q(1 + \delta)$. Further, we limit the range of disruption probability to $q \leq \frac{q_{th}}{2}$ so that $\hat{q} \in (0, q_{th}]$ for any δ. This is to focus on the more practical and interesting 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., $\frac{\partial F(q,q(1\pm\delta))}{\partial q} > 0$.*

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., $\frac{F(q,q(1-\delta))}{F(q,q(1+\delta))}$ 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. Further, 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 the Online Appendix C, we numerically test this under a more generalized setting (using the from 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.

Since 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). Please 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, that 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 $(\underline{q}, \bar{q})$ since 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 $\underline{q} = \frac{\tilde{q}}{1+\delta}$ and $\bar{q} = \frac{\tilde{q}}{1-\delta}$. 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 $\tilde{q} = 0.03$ since the impact of underestimation is greater than that of overestimation.

	$\tilde{q} = 0.03;\ \delta = 0.25$			$\tilde{q} = 0.03;\ \delta = 0.5$			$\tilde{q} = 0.03;\ \delta = 0.75$			$\tilde{q} = 0.03;\ \delta = 0.9$		
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%

Table 1: The optimal estimate and the percentage increase under the worst-case scenario

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 1.00% increase in the expected total cost for using $\hat{q}^*$, while one may incur 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 Section 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. While 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 $m = \frac{n_u}{n_r}$ unreliable facilities. With capacity K, the reliable facility first serves its own demand, $\frac{A\rho}{n_t}$, 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. Further, 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 $cov(Y_i, Y_j) = \sigma_{ij}$ for $0 < i, j < m$. Further, the total number of disruptions follows a correlated binomial random variable X with parameters (m, q, Σ). Since the number of unreliable facilities in each subregion, $m = \frac{n_u}{n_r}$, may not be integer, we approximate $X(m, q, \Sigma)$ as follows:

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

Finally, we define the capacitated reliability model as follows:

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

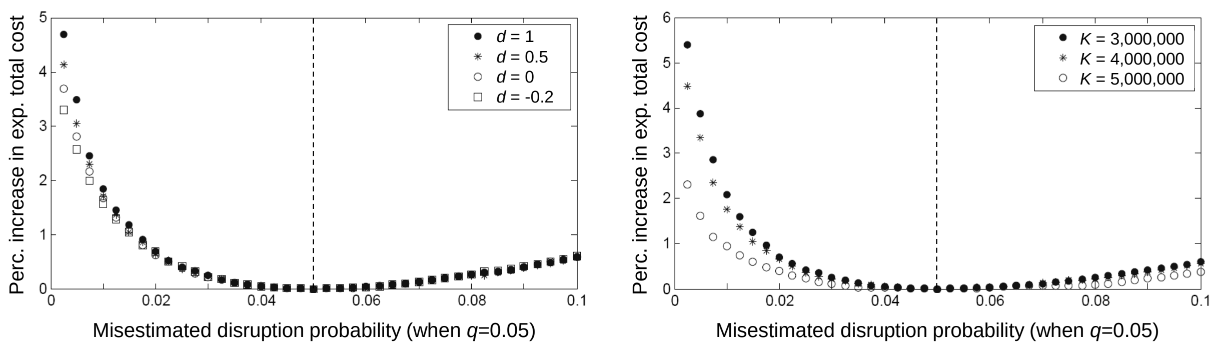
Note that in the above formulation, we assumed that the firm uses non-overlapping diamond shaped tiling scheme as from the uncapacitated model. If, however, the penalty cost is high enough (and transportation cost is cheap), it is possible that the firm is better off having overlapping covering regions 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. While the above model allows any arbitrary correlation structure among the m unreliable facilities, we let $\sigma_{ij}(d) = \frac{d}{|i-j|+1} q(1-q)$ 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 ($\frac{n_t}{A}$) in determining the degree of correlation. In this study, we limit the analysis to a simplified version as 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, 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 (p.9) in 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. Figures 4 contrast the impact of underestimation to that of overestimation with varying degree 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 identical as the previous examples with $p = 1$, $d = 1$, $K = 4,000,000$ unless specified otherwise. We assume the correlation degree d is known at the moment.

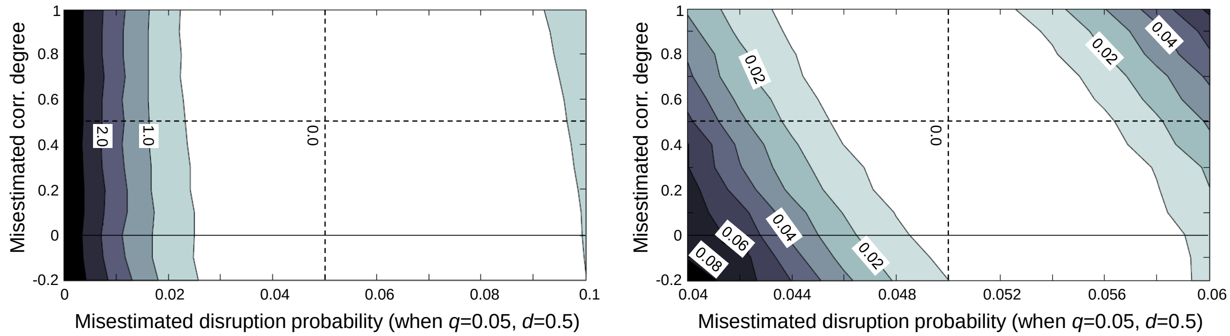


(a) Impact of misestimation increases with correlation (b) Impact of misestimation decreases with capacity

Figure 4: Comparative statics with respect to d and K on the impact of misestimation

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, while 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 $\hat{d}$, respectively, for some fixed f_u and r. Figures 5 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.



(a) Misestimation in correlation degree is less sensitive than the misestimation in disruption probability

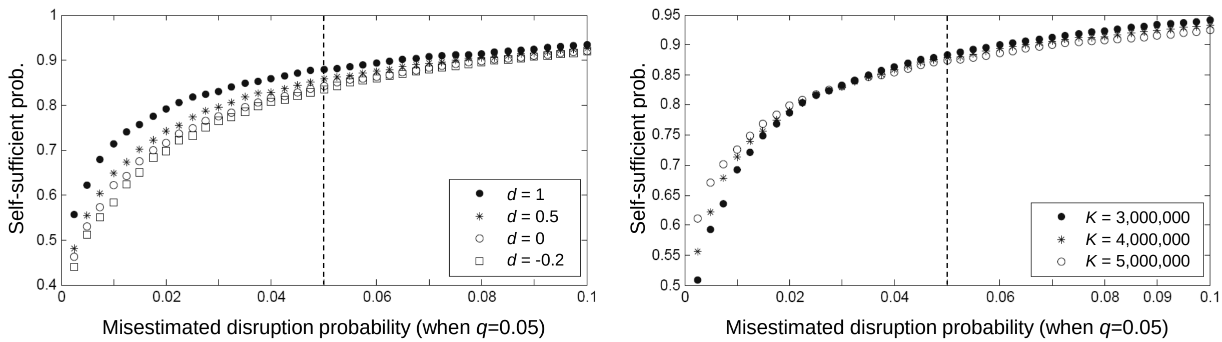
(b) Misestimation in correlation degree becomes relevant when disruption probability estimate is relatively accurate

Figure 5: Percentage increase in the expected total cost when both q and d are misestimated

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 while 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 while misestimation in the same direction magnifies the impact. In addition, we observe that the impact of joint-underestimation (south-west corner) is greater than that of joint-overestimation (north-east corner).

In the presence of capacity and penalty cost, we can consider the probability that a reliable facility does not incur penalty costs. We refer to this measure as *self-sufficient probability* (SSP) and define as $\mathbb{P}\big(\frac{A\rho}{n_t^*}\big(1+X\big(\frac{n_u^*}{n_r^*}, q, \Sigma\big)\big) < K\big)$ from solving (10). Managers might be interested in SSP since low SSP represents high dependency on outsourcing or heavy stress on the system due to constant overtime production. Figures 6 illustrate 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 (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 more 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.



(a) SSP increases with correlation (b) SSP and capacity

Figure 6: Comparative statics with respect to d and K on the self-sufficient probability

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 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. Further, 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. Lastly, 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.

References

Ball, M.O., F.L. Lin. 1993. A reliability model applied to emergency service vehicle location. *Oper. Res.* **41(1)** 18–36.

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

Bakkaloglu, M., Wylie, J.J., Wang, C., Ganger, G.R. 2002. On correlated failures in survival storage systems. Technical Report CMU-CS-02-129, Carnegie Mellon University, Pittsburgh, PA.

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

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

Beckmann, M. 1968. Location Theory. Random House, New York, NY.

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

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

Chopra, S., M.S. Sodhi. 2004. Managing risk to avoid supply-chain breakdown. *MIT Sloan Management Rev.* **46** 53–61.

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

Church, R.L., M.P. Scaparra. 2007. Protecting critical assets: The r-interdiction median problem with fortification. *Geographical Anal.* **39** 129–146.

Cui, T., Y. Ouyang, Z.-J.M. Shen. 2010. Reliable Facility Location Design under the Risk of Disruptions. *Oper. Res.* **58(4)** 998-1011.

Daganzo, C.F. 2005. Logistics system analysis. 4th ed. Springer, Berlin, Germany.

Daganzo, C.F., A.L. Erera. 1999. On planning and design of logistics systems for uncertain environments, in New Trends in Distribution Logistics. Springer, Berlin, Germany.

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

Drezner, Z., H.W. Hamacher. 2002. Facility location: Applications and theory, Springer, Berlin, Germany.

Hall, R.W. 1984. Travel distance through transportation terminals on a rectangular grid. *J. Oper. Res. Society.* **35(12)** 1067–1078.

Hendricks, K.B., V.R. Singhal. 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, P.R., G. Saad. 2005. Managing disruption risks in supply chains. *Production Oper. Management.* **14(1)** 53–68.

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

Li, J., J. Chen, S. Wang. 2011. Risk management of supply and cash flows in supply chains. Springer, New York, NY.

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

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

Mak, H.Y., Z.J. Shen. 2011. Agility and proximity considerations in supply chain design - An analytical approach. Working paper, Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Hong Kong.

Munich Re Group. 2008. Topics Geo Natural catastrophes 2007. Available online via `<http://www.munichre.com/publications/302-05699_en.pdf>`.

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

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

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

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

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

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

Snyder, L.V., M.S. Daskin. 2006. Stochastic p-robust location problems. *IIE Tran.* **38(11)** 971–985.

Swiss Re, Sigma report: Natural catastrophes and Man-made disasters, 2009. Available online via `<http://www.swissre.com/resources/dd6346004d4e9669ac76eecedd316cf3-sigma2_2009_e.pdf>`.

Tang, C.S. 2006. Perspectives in supply chain risk management. *Internat. J. Production Econom.* **103** 451–488.

Taleb, N. 2007. The black swan: The impact of the highly improbable. Random House, New York, NY.

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., J.C. Lu, P. Kvam. 2006. Reliability modeling in spatially distributed logistics systems. *IEEE Tran. Reliability.* **55(3)** 525–534.

Online Appendix to

Facility Location Decisions with Random Disruptions and Imperfect Estimation

Appendix A: Validation of the tiling scheme

The derivation of (5) is based on a diamond-shaped tiling scheme such that the facilities can be evenly spread while locating one reliable facility surrounded by unreliable facilities. Here, we test how sensitive the results are when this particular tiling scheme is relaxed. Denoting the configuration ratio between the total number of facilities and the number of reliable facilities to β (i.e., $\beta = \frac{n_t}{n_r} > 0$), the expected total cost becomes

$$\mathbb{E}[TC(\beta)] = 3\Big(\frac{\gamma}{2}\Big)^{\frac{2}{3}} f_u^{\frac{1}{3}} \Big(r + \beta - 1\Big)^{\frac{1}{3}} \Big(q + \frac{1-q}{\sqrt{\beta}}\Big)^{\frac{2}{3}} \tag{11}$$

where the optimal configuration ratio is $\beta^* = \big(\frac{(1-q)(r-1)}{q}\big)^{\frac{2}{3}}$ for $q \geq q_{th}$ and $\beta^* = 1$ otherwise. The proposed tiling scheme provides an "exact" solution only for $\beta = n^2, \forall n \in \mathbb{N}$. For other configuration ratios, $\beta \neq n^2$, finding the *actual* optimal tiling scheme may be difficult. This resulting configuration may not result in diamond-shape tiles. We assume such an actual tiling scheme exists.

Based on an infinite service region (to avoid integrality issue on the number of facilities), we investigate how large the error will be for using the proposed tiling scheme instead of actual tiling scheme. We refer to the expected total cost based on an actual optimal tiling scheme to *true expected total cost*. Denoting the true expected total cost by $\overline{TC}(\beta)$ and its actual configuration ratio by $\overline{\beta}$, we provide the following proposition.

Proposition 7. *$\mathbb{E}[\overline{TC}(\beta)]$ is bounded below by $\mathbb{E}[TC(\beta)]$. Further, $\mathbb{E}[\overline{TC}(\overline{\beta})]$ is bounded below by $\mathbb{E}[TC(\beta^*)]$.*

Proof. Proof. Since all unreliable facilities have the same disruption probability, it is sufficient for the optimal configuration to satisfy the optimal tiling scheme (non-overlapping diamond-shaped tiles) for the following two extreme cases: (a) when all unreliable facilities are working, (b) when none of the unreliable facilities are working. For a contiguous service region, the only configuration that satisfies this condition is when $\beta = n^2, n \in \mathbb{N}$ (one reliable facility in the center surrounded by unreliable facilities). For $\beta \neq n^2, n \in \mathbb{N}$, an actual tiling scheme does not satisfy the optimal tiling scheme and incurs higher total cost. Hence, $\mathbb{E}[TC(\beta)] \leq \mathbb{E}[\overline{TC}(\beta)]$ for all $\beta > 0$.

Taking the infimum on each side from Proposition 7, we have $\inf_\beta \mathbb{E}[TC(\beta)] \leq \inf_\beta \mathbb{E}[\overline{TC}(\beta)]$. Since $\mathbb{E}[\overline{TC}(\overline{\beta})] = \inf_\beta \mathbb{E}[\overline{TC}(\beta)]$ and $\mathbb{E}[TC(\beta^*)] = \inf_\beta \mathbb{E}[TC(\beta)]$, we conclude that $\mathbb{E}[TC(\beta^*)] \leq \mathbb{E}[\overline{TC}(\overline{\beta})]$. □ □

We also obtain the upper bound on the true expected total cost. By restricting the configuration ratio to $\beta = n^2$, $\forall n \in \mathbb{N}$ and using the proposed diamond tiling scheme, we can find the configuration ratio $\tilde{\beta}$ that minimizes the expected total cost; that is, $\tilde{\beta} = \mathrm{argmin}_{\beta \in \{\beta = n^2, \forall n \in \mathbb{N}\}} \mathbb{E}[TC(\beta)]$.

Proposition 8. *The expected total cost with configuration ratio $\tilde{\beta}$, $\mathbb{E}[TC(\tilde{\beta})]$, can be derived as*

$$\mathbb{E}[TC(\tilde{\beta})] = \min\left\{ \mathbb{E}[TC(\lfloor\sqrt{\beta^*}\rfloor^2)],\, \mathbb{E}[TC(\lceil\sqrt{\beta^*}\rceil^2)] \right\}.$$

Then, $\mathbb{E}[\overline{TC}(\overline{\beta})]$ is bounded above by $\mathbb{E}[TC(\tilde{\beta})]$.

Proof. Proof. We know that $\mathbb{E}[TC(\beta)]$ is strictly convex in β since $\frac{\partial^2 TC}{\partial \beta^2} = \frac{3\gamma}{4\sqrt{n_r}}(1-q)\beta^{-\frac{5}{2}} > 0$. Hence, $\mathbb{E}[TC(\beta)]$ is a unimodal function with the minimum at β^* and thus $\tilde{\beta}$ is the minimum between $\lfloor\sqrt{\beta^*}\rfloor^2$ and $\lceil\sqrt{\beta^*}\rceil^2$.

Further, we know that $\inf_\beta \mathbb{E}[\overline{TC}(\beta)] \leq \mathbb{E}[\overline{TC}(\tilde{\beta})]$ holds. Therefore, $\mathbb{E}[\overline{TC}(\overline{\beta})] = \inf_\beta \mathbb{E}[\overline{TC}(\beta)] \leq \mathbb{E}[\overline{TC}(\tilde{\beta})] = \mathbb{E}[TC(\tilde{\beta})]$. □ □

We next conduct a numerical experiment to study the performance of the current solution method. For the case example, the percentage optimality gap ($\left[\frac{\mathbb{E}[TC(\tilde{\beta})]}{\mathbb{E}[TC(\beta^*)]} - 1\right] \times 100\,(\%)$) in the expected total cost by restricting to a diamond-shaped tiling scheme is computed in Table 2. This result shows that the optimality gap is typically very small implying that our current solution method provides a very good approximation in terms of the expected total cost.

Table 2: Percentage optimality gap using the diamond-shaped tiling scheme

r	$q = 0.01$	0.05	0.10	0.20	0.30	0.40	0.50
1.25	0.002%	0.216%	0.873%	0%	0%	0%	0%
1.5	0.025%	0.030%	0.462%	1.015%	0.057%	0%	0%
2	0.025%	0.144%	0.024%	1.053%	1.893%	0.451%	0%
5	0.015%	0.049%	0.183%	0.712%	0.066%	0.228%	1.288%

Appendix B: Line analysis

Denote l to be length of the line (assumed large) and other notation remain the same (such as ρ, c, n). With n facilities on the line, the expected distance for any demand to its nearest facility is $\mathbb{E}[D] = \frac{l\rho c}{4n}$. When there is only one reliable facility (that is located at the center of the line), the expected distance to the reliable facility for the demand outside of the central region is $\mathbb{E}[D_U] = \left(\frac{n+1}{4n}\right)l\rho c$.

Given that the number of each type of facilities are n_r and n_u, the expected distance to the nearest facility is $\mathbb{E}[D] = \frac{l\rho c}{4(n_r+n_u)}$. Also, the expected distance to the reliable facility when its primary unreliable facility fails is $\mathbb{E}[D_U] = \frac{l\rho c}{4}\left(\frac{1}{n_r} + \frac{1}{n_r+n_u}\right)$. Hence, the expected total cost is $\mathbb{E}[TC] = f_r n_r + f_u n_u + q\bar{\gamma}\frac{1}{n_r} + (1-q)\bar{\gamma}\frac{1}{n_r+n_u}$ where $\bar{\gamma} = \frac{l\rho c}{4}$. Note that this has the same functional form as equation (5) for our analysis on plane. The square root term is due to the difference between a line and a plane. Using this, we obtain the optimal number of facilities as:

$$\begin{cases} n_r^* = \left(\frac{\bar{\gamma}q}{f_r - f_u}\right)^{\frac{1}{2}} \text{ and } n_r^* + n_u^* = \left(\frac{\bar{\gamma}(1-q)}{f_u}\right)^{\frac{1}{2}} & \text{where } q \leq q_{th} = \frac{f_r - f_u}{f_r}, \\ n_r^* = \left(\frac{\bar{\gamma}}{f_r}\right)^{\frac{1}{2}} \text{ and } n_u^* = 0 & \text{otherwise} \end{cases}.$$

Appendix C: Robustness test on the discrete model

To verify the robustness of the proposed model in this paper, we test one of our main results (underestimation hurts more than overestimation) in this paper under a more general setting. We employ a discrete version counterpart model introduced by Lim et al. (2010) briefly explained below.

Consider a set of demand points I where customers resides in and a set of candidate sites J where the facilities can be located. We assume that each node is a demand node and a candidate facility site, i.e., $I = J$, for this study. At each node, $j \in N$ where N is the set of all nodes, we can locate either an unreliable facility at a cost of f_j^U or a reliable facility at a cost of $f_j^R > f_j^U$. The cost of traveling to a facility j which has not failed from demand node i is given by d_{ij}^P and the cost of traveling to a backup facility (if the primary facility has failed) is given by $d_{ij}^B \geq d_{ij}^P$. We also denote d_{ij}^S $(= d_{ij}^B - d_{ij}^P \geq 0)$ as the unit savings that has to be subtracted from the objective function when demand node i is assigned to a reliable facility at j as both the primary facility at a unit cost of d_{ij}^P and the backup facility at a unit cost of d_{ij}^B since the true assignment cost of node i to the facility at j should be only d_{ij}^P, since the facility is completely reliable. Finally, we define the following decision variables:

X_j^U = 1 if an unreliable facility is located at candidate site j; 0 if not

X_j^R = 1 if a reliable facility is located at candidate site j; 0 if not

Y_{ij}^P = 1 if demands at i are assigned to a facility at j as the primary site; 0 if not

Y_{ij}^B = 1 if demands at i are assigned to a facility at j as the backup site; 0 if not

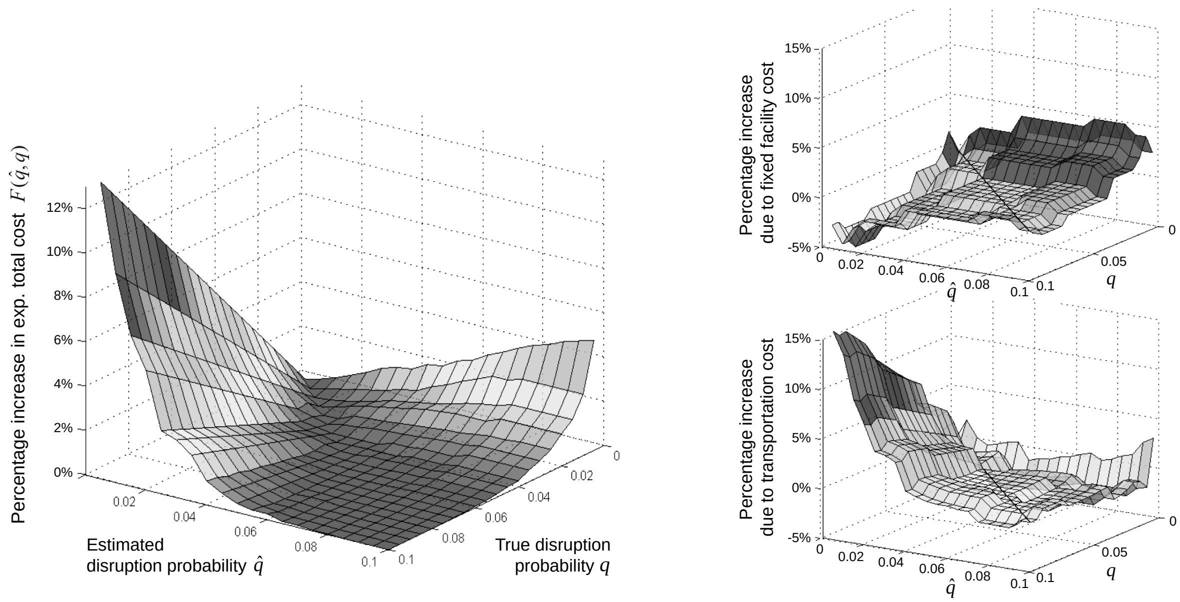
Y_{ij}^S = 1 if demands at i are assigned to a facility at j as the primary and backup site; 0 if not

With this notation, the discrete version of the model can be formulated as follows:

$$
\begin{aligned}
\text{Minimize} \quad & \sum_{j \in N} f_j^U X_j^U + \sum_{j \in N} f_j^R X_j^R \\
& + \sum_{i \in N} \sum_{j \in N} (1 - q_j) h_i d_{ij}^P Y_{ij}^P + \sum_{i \in N} \sum_{j \in N} q_j h_i d_{ij}^B Y_{ij}^B - \sum_{i \in N} \sum_{j \in N} q_j h_i d_{ij}^S Y_{ij}^S \\
\text{subject to} \quad & \\
& \sum_{j \in N} Y_{ij}^P = 1, \qquad \sum_{j \in N} Y_{ij}^B = 1 \qquad \forall i \in N \\
& Y_{ij}^P \le X_j^R + X_j^U, \qquad Y_{ij}^B \le X_j^R \qquad \forall i, j \in N \\
& Y_{ij}^S \le Y_{ij}^B, \qquad Y_{ij}^S \le Y_{ij}^P \qquad \forall i, j \in N \\
& X_j^R + X_j^U \le 1 \qquad \forall j \in N \\
& X_j^U,\, X_j^R \in \{0, 1\} \qquad \forall j \in N \\
& Y_{ij}^P,\, Y_{ij}^B,\, Y_{ij}^S \in \{0, 1\} \qquad \forall i, j \in N.
\end{aligned}
$$

To introduce heterogeneity, we relax five assumptions from the continuous model: (a) instead of demand being uniformly distributed, each demand node has its own demand level (we use the population of the city), (b) each facility is allowed to have different fixed costs (we set the facility cost proportional to the population of the city, $f_j^U = 500,000 \times (1.7h_j)$ where h_j is the population of city j), (c) each facility can have its own disruption probability (though here we use $q_j = 0.05$ to be consistent with the current work), (d) any type of distance metric is allowed (we use great circle distances), (e) transportation unit costs are allowed to differ for the backup assignments to capture the extra cost incurred when a disruption occurs (we use $d_{ij}^B/d_{ij}^P = 1.25$). The data set of the 263 largest cities in the contiguous United States was used in the numerical study. We set all other parameters identical (or as close as possible) to the numerical example in Figures 2. For the solution methods and some structural properties of this model, please see Lim et al. (2010).

The impact of misestimating the disruption probability for the discrete model is illustrated in the below figure. This is a counterpart of Figures 2 in the paper. Although the modeling framework and some key assumptions are relaxed, we observe a strong concurrence in general trends with the results of the continuous model: the impact of underestimation is much greater than the impact of overestimation for the discrete model. We believe this suggests that the insights and the analytical results from the continuous model is quite robust under a more generalized setting. Unfortunately, capacity issue in discrete facility location problems come with a huge computational challenge, we limit the analysis to the basic setting for this study.



(a) Percentage increase with misestimation

(b) Percentage increase due to fixed facility cost (top); Percentage increase due to transportation cost (bottom)

Figure 7: Impact of disruption probability misestimation in discrete model

Appendix D: Proofs of the main results

Proof. Proof of Proposition 1. Using the first order conditions (and that (5) is convex with respect to n_r and n_u), the optimal number of reliable and unreliable facilities (6) is derived. Since the number of unreliable facilities has to be non-negative, we have $n_t^* \geq n_r^*$. Hence, for (6) to be a feasible solution, we have $\big(\frac{\gamma(1-q)}{2f_u}\big)^{\frac{2}{3}} \geq \big(\frac{\gamma q}{2(f_r-f_u)}\big)^{\frac{2}{3}}$ which in turn gives us the following threshold condition on q: if $q \leq \frac{(f_r-f_u)}{f_r}$ the optimal facility configuration is given by (6). If $q > \frac{(f_r-f_u)}{f_r}$, it is optimal to locate only reliable facilities. In this case, the optimal solution becomes (7). $\square$ $\square$

Proof. Proof of Proposition 2. To show $\frac{\partial F(q,q(1\pm\delta))}{\partial q} = \frac{TC(q,q)\frac{\partial TC(q,q(1\pm\delta))}{\partial q} - TC(q,q(1\pm\delta))\frac{\partial TC(q,q)}{\partial q}}{TC(q,q)^2} > 0$, it is sufficient to show that

$$TC(q,q)\frac{\partial TC(q,q(1\pm\delta))}{\partial q} - TC(q,q(1\pm\delta))\frac{\partial TC(q,q)}{\partial q} \tag{12}$$

is strictly positive. We start with the overestimation case, $\hat{q} = q(1+\delta)$. We know $TC(q,q) = 3f^{\frac{1}{3}}(\frac{\gamma}{2})^{\frac{2}{3}}[q^{\frac{2}{3}}(r-1)^{\frac{1}{3}} + (1-q)^{\frac{2}{3}}]$ from (8). Further, we have $TC(q,q(1+\delta)) = f^{\frac{1}{3}}(\frac{\gamma}{2})^{\frac{2}{3}}\big[q^{\frac{2}{3}}(3+\delta)(\frac{r-1}{1+\delta})^{\frac{1}{3}} + (1-q(1+\delta))^{\frac{2}{3}}\big(1+\frac{2(1-q)}{1-q(1+\delta)}\big)\big]$, $\frac{\partial TC(q,q)}{\partial q} = 2f^{\frac{1}{3}}(\frac{\gamma}{2})^{\frac{2}{3}}\big[(\frac{r-1}{q})^{\frac{1}{3}} - (\frac{1}{1-q})^{\frac{1}{3}}\big]$, and $\frac{\partial TC(q,q(1+\delta))}{\partial q} = \frac{2}{3}f^{\frac{1}{3}}(\frac{\gamma}{2})^{\frac{2}{3}}(3+\delta)\big[(\frac{r-1}{q(1+\delta)})^{\frac{1}{3}} - (\frac{1}{1-q(1+\delta)})^{\frac{4}{3}}\big(1-q(1+\delta)-\frac{\delta}{3+\delta}\big)\big]$. Using this (with some algebra), (12) can be expressed as the sum of three terms as follows:

$$2f^{\frac{2}{3}}\Big(\frac{\gamma}{2}\Big)^{\frac{4}{3}}\Big\{\Big[q\delta^2\Big(\frac{1}{1-q}\Big)^{\frac{1}{3}}\Big(\frac{1}{1-q(1+\delta)}\Big)^{\frac{4}{3}}\Big]$$
$$+\ q^{\frac{2}{3}}(r-1)^{\frac{1}{3}}(3+\delta)\Big(\frac{1}{1-q(1+\delta)}\Big)^{\frac{1}{3}}\Big[-1+\frac{1}{(1-q(1+\delta))(3+\delta)}+\Big(\frac{1}{(1-q)(1+\delta)}\Big)^{\frac{1}{3}}\Big]$$
$$+\ \Big(\frac{r-1}{q}\Big)^{\frac{1}{3}}(1-q)^{\frac{2}{3}}\Big[(3+\delta)\Big(\frac{1}{1+\delta}\Big)^{\frac{1}{3}}-\Big(\frac{1-q(1+\delta)}{1-q}\Big)^{\frac{2}{3}}-2\Big(\frac{1-q}{1-q(1+\delta)}\Big)\Big]\Big\}.$$

It follows that for $0 < q \leq \frac{q_{th}}{2} = \frac{r-1}{2r}$, $r > 1$, and $0 < \delta < 1$, all the terms within each bracket are positive. More specifically, it is straightforward to see that the first term is strictly positive. For the second term, one can show that this is strictly positive by investigating: (a) $(1-q)(1+\delta) > 1$ and (b) $(1-q)(1+\delta) \leq 1$. The last term can be shown to be strictly positive by using $0 < \delta < 1$, $0 < q \leq \frac{r-1}{2r} < \frac{1}{2}$. Thus the sign of (12) is also strictly positive. The same result can be shown for the underestimation case, $\hat{q} = q(1-\delta)$, through a very similar exercise. Therefore, we conclude $\frac{\partial F(q,q(1\pm\delta))}{\partial q} > 0$ for $0 < \delta < 1$. □ □

Proof. Proof of Proposition 3. To show $F(q, q(1+\delta)) < F(q, q(1-\delta))$, we equivalently show $TC(q, q(1+\delta)) < TC(q, q(1-\delta))$. From (5), we have

$$\begin{aligned}
&TC(q, q(1+\delta)) - TC(q, q(1-\delta))\\
&= f_r\big(n_r(q(1+\delta)) - n_r(q(1-\delta))\big) + f_u\big(n_u(q(1+\delta)) - n_u(q(1-\delta))\big)\\
&\quad + q\gamma\Big(\sqrt{\tfrac{1}{n_r(q(1+\delta))}} - \sqrt{\tfrac{1}{n_r(q(1-\delta))}}\Big) + (1-q)\gamma\Big(\sqrt{\tfrac{1}{n_t(q(1+\delta))}} - \sqrt{\tfrac{1}{n_t(q(1-\delta))}}\Big)\\
&= f^{\frac{1}{3}}\Big(\frac{\gamma}{2}\Big)^{\frac{2}{3}} q^{\frac{2}{3}}(r-1)^{\frac{1}{3}}\Big((3+\delta)(1+\delta)^{-\frac{1}{3}} - (3-\delta)(1-\delta)^{-\frac{1}{3}}\Big).
\end{aligned}$$

The above quantity is strictly negative for all $0 < q \leq \frac{q_{th}}{2} = \frac{r-1}{2r}$, $r > 1$, and $0 < \delta < 1$. Therefore, we conclude $F(q, q(1+\delta)) < F(q, q(1-\delta))$. □ □

Proof. Proof of Proposition 4. Define $G(q,\delta) := \frac{F(q,q(1-\delta))}{F(q,q(1+\delta))} = \frac{TC(q,q(1-\delta))}{TC(q,q(1+\delta))}$. Then, $G(q, 0) = 1$. To show that the percentage increase due to underestimating the disruption probability by δ relative to overestimation increases with error rate δ, we show

$$\frac{\partial G(q,\delta)}{\partial\delta} = \frac{TC\big(q,q(1+\delta))\big)\frac{\partial TC(q,q(1-\delta))}{\partial\delta} - TC\big(q,q(1-\delta)\big)\frac{\partial TC(q,q(1+\delta))}{\partial\delta}}{TC\big(q,q(1+\delta)\big)^2}$$

is strictly positive for $0 < \delta < 1$. Since $\frac{\partial TC(q,q(1\pm\delta))}{\partial\delta} = \gamma\sqrt{\frac{1}{n_r(q(1\pm\delta))}} - \gamma\sqrt{\frac{1}{n_t(q(1\pm\delta))}}$, it is sufficient to examine the sign of

$$TC\big(q,q(1+\delta)\big)\Big[\sqrt{\tfrac{1}{n_r(q(1-\delta))}} - \sqrt{\tfrac{1}{n_t(q(1-\delta))}}\Big] - TC\big(q,q(1-\delta)\big)\Big[\sqrt{\tfrac{1}{n_r(q(1+\delta))}} - \sqrt{\tfrac{1}{n_t(q(1+\delta))}}\Big]. \quad (13)$$

After some algebra, (13) can be expressed as follows:

$$f^{\frac{2}{3}}\Big(\frac{\gamma}{2}\Big)^{\frac{1}{3}}\Big\{2\delta(r-1)^{\frac{2}{3}}\Big(\frac{q}{1-\delta^2}\Big)^{\frac{1}{3}} + 2(1-q)\Big(\frac{1}{1-q(1+\delta)}\Big)^{\frac{1}{3}}\Big[\Big(\frac{r-1}{q(1-\delta)}\Big)^{\frac{1}{3}}-\Big(\frac{1}{1-q(1-\delta)}\Big)^{\frac{1}{3}}\Big] - 2(1-q)\Big(\frac{1}{1-q(1-\delta)}\Big)^{\frac{1}{3}}\Big[\Big(\frac{r-1}{q(1+\delta)}\Big)^{\frac{1}{3}}-\Big(\frac{1}{1-q(1+\delta)}\Big)^{\frac{1}{3}}\Big]\Big\}.$$

The first term is strictly positive. The second and third terms are also strictly positive since $\Big[\Big(\frac{r-1}{q(1-\delta)}\Big)^{\frac{1}{3}}-\Big(\frac{1}{1-q(1-\delta)}\Big)^{\frac{1}{3}}\Big] > \Big[\Big(\frac{r-1}{q(1+\delta)}\Big)^{\frac{1}{3}}-\Big(\frac{1}{1-q(1+\delta)}\Big)^{\frac{1}{3}}\Big] > 0$ and $2(1-q)\Big(\frac{1}{1-q(1+\delta)}\Big)^{\frac{1}{3}} > 2(1-q)\Big(\frac{1}{1-q(1-\delta)}\Big)^{\frac{1}{3}} > 0$. Hence, (13) is strictly positive for all $0 < q \leq \frac{q_{th}}{2} = \frac{r-1}{2r}$, $r > 1$, and $0 < \delta < 1$; therefore $\frac{\partial G(q,\delta)}{\partial \delta} > 0$. Since $G(q,0) = 1$ and $G(q,\delta)$ is monotonically increasing in δ, we conclude that the percentage increase in expected total cost for underestimating the disruption probability by error rate δ relative to overestimation increases with error rate δ. □ □

Proof. Proof of Proposition 5. If $\underline{q} \geq q_{th}$, $F(q,\hat{q})$ is a constant regardless of q; that is, $F(q,\hat{q}) = K$ for $\{\hat{q} > q_{th}, q \leq q_{th}\}$ (region 4) and $F(q,\hat{q}) = 0$ for $\{\hat{q} > q_{th},\ q > q_{th}\}$ (region 1). Hence, any $\hat{q} \in (\underline{q},\bar{q})$ satisfies (9).

To prove the case for $\underline{q} < q_{th}$, we need the following two lemmas.

Lemma 1. $\sup\limits_{q\in(\underline{q},\bar{q})} F(q,\hat{q}) = \max\limits_{\hat{q}\leq q_{th}} [F(\underline{q},\hat{q}), F(\bar{q},\hat{q})]$.

Proof. Proof. We first show that $F(q,\hat{q})$ is unimodal in q. Note that we can reduce the region of $\hat{q}$ to $(\underline{q}, q_{th})$ since $F(q,\hat{q})$ is a constant for $\hat{q} > q_{th}$. We further note that $F(q,\hat{q}) = 0$ at $q = \hat{q}$ and $F(q,\hat{q}) \geq 0$ elsewhere. Thus, to complete the proof, we shall show that $F(q,\hat{q})$ monotonically decreases in q for $q < \hat{q}$ and monotonically increases for $q \geq \hat{q}$. We examine the sign of

$$\frac{\partial F(q,\hat{q})}{\partial q} = \frac{TC(q,q)\frac{\partial TC(q,\hat{q})}{\partial q} - TC(q,\hat{q})\frac{\partial TC(q,q)}{\partial q}}{TC(q,q)^2}$$

where $\frac{\partial TC(q,q)}{\partial q} = \gamma\sqrt{\frac{1}{n_r(q)}} - \gamma\sqrt{\frac{1}{n_t(q)}}$ and $\frac{\partial TC(q,\hat{q})}{\partial q} = \gamma\sqrt{\frac{1}{n_r(\hat{q})}} - \gamma\sqrt{\frac{1}{n_t(\hat{q})}}$. To verify the unimodality of $F(q,\hat{q})$, it is sufficient to examine the sign of

$$TC(q,q)\Big[\sqrt{\tfrac{1}{n_r(\hat{q})}} - \sqrt{\tfrac{1}{n_t(\hat{q})}}\Big] - TC(q,\hat{q})\Big[\sqrt{\tfrac{1}{n_r(q)}} - \sqrt{\tfrac{1}{n_t(q)}}\Big]\ . \tag{14}$$

We examine three cases: (a) $q < \hat{q} \leq q_{th}$, (b) $\hat{q} \leq q \leq q_{th}$, and (c) $\hat{q} \leq q_{th} < q$. After algebraic work by plugging the solutions from (6)-(7) into (14), the sign of (14) can be shown to be negative, positive and positive for each case respectively.[1] Therefore, $F(q,\hat{q})$ is unimodal in q and the $\sup_{q\in(\underline{q},\bar{q})} F(q,\hat{q})$ is achieved at the maximum between $F(\underline{q},\hat{q})$ and $F(\bar{q},\hat{q})$ for $\hat{q} \leq q_{th}$. □ □

[1]Note that $\frac{\partial F(q,\hat{q})}{\partial q}$ here is for a fixed $\hat{q}$. This is different from the $\frac{\partial F(q,\hat{q})}{\partial q}$ from the Proposition 2 since the estimate $\hat{q}$ $(= q(1\pm\delta))$ is fixed for δ but dependent on q.

Lemma 2. *For $\hat{q} \leq q_{th}$, $F(\underline{q}, \hat{q})$ is strictly increasing in $\hat{q}$ and $F(\bar{q}, \hat{q})$ is strictly decreasing in $\hat{q}$.*

Proof. Proof. (a) From $F(\underline{q}, \hat{q}) = \frac{TC(\underline{q},\hat{q})}{TC(\underline{q},\underline{q})} - 1$, we have $\frac{\partial F(\underline{q},\hat{q})}{\partial \hat{q}} = \frac{\frac{\partial TC(\underline{q},\hat{q})}{\partial \hat{q}}}{TC(\underline{q},\underline{q})}$. Hence, $\frac{\partial TC(\underline{q},\hat{q})}{\partial \hat{q}} > 0$ implies $\frac{\partial F(\underline{q},\hat{q})}{\partial \hat{q}} > 0$. After some algebra, we have

$$\frac{\partial TC(\underline{q}, \hat{q})}{\partial \hat{q}} = \frac{2}{3} f^{\frac{1}{3}} \left(\frac{\gamma}{2}\right)^{\frac{2}{3}} \left[(\hat{q} - \underline{q})\Big((r-1)^{\frac{1}{3}} \hat{q}^{-\frac{4}{3}} + (1-\hat{q})^{-\frac{4}{3}}\Big)\right] > 0.$$

Hence, $F(\underline{q}, \hat{q})$ is strictly increasing in $\hat{q}$.

(b) Similar to the above case, we have

$$\frac{\partial TC(\bar{q}, \hat{q})}{\partial \hat{q}} = \frac{2}{3} f^{\frac{1}{3}} \left(\frac{\gamma}{2}\right)^{\frac{2}{3}} \left[(\hat{q} - \bar{q})\Big((r-1)^{\frac{1}{3}} \hat{q}^{-\frac{4}{3}} + (1-\hat{q})^{-\frac{4}{3}}\Big)\right] < 0.$$

This implies $\frac{\partial F(\bar{q},\hat{q})}{\partial \hat{q}}$ is strictly negative. Hence, $F(\bar{q}, \hat{q})$ is strictly decreasing in $\hat{q}$. □ □

From Lemma 1, $\sup_{q \in (\underline{q}, \bar{q})} F(q, \hat{q}) = \max_{\hat{q} \leq q_{th}} [F(\underline{q}, \hat{q}), F(\bar{q}, \hat{q})]$. Also, since $F(\underline{q}, \hat{q})$ is strictly increasing in $\hat{q}$ and $F(\bar{q}, \hat{q})$ is strictly decreasing in $\hat{q}$ from Lemma 2, the infimum of $\sup_{\substack{q \in (\underline{q}, \hat{q}) \\ \hat{q} \leq q_{th}}} F(q, \hat{q})$ can be uniquely obtained by setting $F(\underline{q}, \hat{q}) = F(\bar{q}, \hat{q})$. Thus, if $\underline{q} < q_{th}$, an optimal estimate of the disruption probability $\hat{q}^*$ which satisfies (9) is unique and this value is strictly less than the threshold q_{th} ($\hat{q}^* < q_{th}$). □ □

Proof. Proof of Proposition 6. The variance of $X\big(\frac{n_u}{n_r}, q, \Sigma\big)$ increases with d since $\sigma_{ij}(d)$ is increasing in d. Hence, $\left[\frac{A\rho}{n_t}\big(1 + X\big(\frac{n_u}{n_r}, q, \Sigma\big)\big) - K\right]^+$ is stochastically increasing in d. Therefore, $\mathbb{E}[TC]$ increases with d.

Similarly, $\left[\frac{A\rho}{n_t}\big(1 + X\big(\frac{n_u}{n_r}, q, \Sigma\big)\big) - K\right]^+$ is stochastically decreasing in K. Therefore, $\mathbb{E}[TC]$ decreases with K. □ □