Mitigating Disruption Risks in Delivery Supply Chains to Serve Contracted Customers

Mert Hakan Hekimoğlu
hekimm@rpi.edu
The Lally School of Management
Rensselaer Polytechnic Institute
Troy, NY 12180

John Park
jhpark03@syr.edu
Whitman School of Management
Syracuse University
Syracuse, NY 13244

Burak Kazaz
bkazaz@syr.edu
Whitman School of Management
Syracuse University
Syracuse, NY 13244

August 30, 2016

Mitigating Disruption Risks in Delivery Supply Chains to Serve Contracted Customers

Motivated by an implementation in a Fortune 150 company, this paper helps a firm determine its capacity expansion decisions as a mitigation strategy against disruptions in a delivery supply chain. The delivery supply chain involves fulfillment centers that are responsible for delivering orders within the next day. We formulate the firm's capacity planning problem using a two-stage stochastic model. The firm determines the capacity expansion amount in each fulfillment center in stage 1. If a disruption occurs in stage 2, then the firm determines the optimal allocation of the backup capacity in order to satisfy the orders arriving at the disrupted fulfillment center. We consider the length of disruption as random. The firm restricts the probability of the number of late deliveries exceeding a tolerable threshold under risk aversion.

This paper makes five main contributions. First, we use capacity planning, rather than inventory planning, as a proactive measure against supply chain disruptions. Unlike inventory planning, capacity planning adds agility and flexibility to a delivery supply chain. Second, our work incorporates two types of disruptions that are (1) low-impact and high-likelihood disruptions, and (2) high-impact and low-likelihood disruptions. This provides a better representation of the set of disruptions a firm would face in its daily operations. Third, we show that geographic proximity does not necessarily serve as an anchor when determining the location of capacity expansion, i.e., the firm may be economically better off by adding capacity at a remotely located facility, even though providing backup from that facility would cost more than providing from a closer facility in case of a disruption. Fourth, while risk aversion generally leads to an increase in capacity investment, we find a surprising result that capacity may decrease with risk aversion, and identify the conditions that lead to this finding. The non-monotone behavior stems from incorporating high-impact disruption events that can influence nearby fulfillment centers into the analysis and the flexibility of the remote facility to serve multiple fulfillment centers. We supplement our analytical investigation with numerical analyses with data corresponding to a comprehensive set of potential disruptions (provided by the firm and collected from various national sources). Fifth, we demonstrate with numerical analysis that capacity expansion should take place at facilities that are exposed to lower risks. Our capacity expansion model is projected to make a 48% savings in the total expected operating costs stemming from disruptions under risk aversion.

1. Introduction

This paper examines a firm's capacity planning decisions as a mitigation strategy against supply chain disruptions. Our work is motivated by a risk assessment project conducted at a Fortune 150 company, an online retailer who serves contracted (business) customers. Business customers constitute the largest portion of the firm's revenues, and the firm operates a delivery supply chain to serve its contracted customers. Figure 1 illustrates the firm's delivery supply chain. There are 16 fulfillment centers located in the US, and they are responsible for delivering orders within the next business day (the firm actually operates 31 fulfillment centers. However, 15 fulfillment centers serve different set of customers with significant product assortment mismatch, and therefore, are not included in our analysis). Specifically, orders placed before 5:00pm (in regional time) are delivered the next day before 5:00pm. Thus, the firm utilizes quick delivery as its winning criterion in competition with other retailers. Fulfillment centers carry approximately 80,000 different products (SKUs), but deliver a total of more than 2 million products through its vendor shipments. These products are sorted, bundled, and wrapped at the fulfillment centers before being shipped out to the customers. If a disruption affects operations at a fulfillment center, that facility temporarily loses its capability to serve its customers until it recovers from the disruption. As a consequence, the firm might fail to deliver the orders the next day.

Figure 1. Illustration of the delivery supply chain.

For a firm standing out with the next-day delivery promise, late deliveries may cause significant consequences. Therefore, the firm needs to react quickly, and divert the orders of the disrupted facility to the functional facilities. However, this kind of a reactive approach proves useful in preventing late deliveries only if the functional facilities have sufficient excess capacity to serve as the backup facility. Therefore, the firm should take a proactive approach by determining its capacity needs in advance before a disruption occurs.

Our work helps the motivating firm (and other companies) comply with the International Standards Organization's (ISO) international business continuity standards ISO 22301. It also prepares the firm for two critical sections of the FEMA Business Continuity Guidance developed in the US: business impact analysis, and the business continuity strategies and requirements. Many firms in the US, as well as our motivating firm, are obliged to comply with the guidelines in ISO 22301 and the FEMA Business Continuity Guidance. Specifically, it is beneficial to focus on Clause 8 of ISO 22301 related with operations:

"*The organization must undertake business impact analysis to understand how its business is affected by disruption and how this changes over time. Risk assessment seeks to understand the risks to the business in a structured way and these inform the development of business continuity strategy… As it is impossible to completely predict and prevent all incidents, the approach of balancing risk reduction and planning for all eventualities is complementary. It might be said, "hope for the best and plan for the worst*".

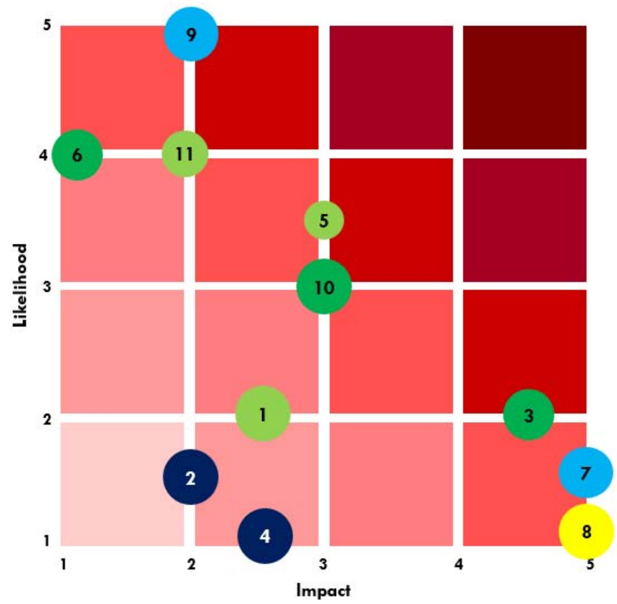
Our work identifies all potential disruptions, high-impact and/or low-impact events, in fulfillment center operations. It develops both a proactive and a reactive risk mitigation approach so that the firm can continue to serve its contracted customers in the event of a disruption. From a proactive perspective, it develops a capacity planning decision model that enables the firm to divert the orders incoming to the disrupted facility to other operational fulfillment centers. From a reactive perspective, our work determines the best re-routing plan on all orders to be fulfilled between other fulfillment centers and the firm's vendors. In sum, our work enables the firm to (1) plan ahead, (2) mitigate risk by reducing the negative financial consequences of disruptions, and (3) enhance its business continuity.

We formulate the capacity planning problem for the delivery supply chain of the above-mentioned retailer using a two-stage model. The firm determines the capacity expansion amount in each fulfillment center (FC) in stage 1. After observing the disruption, corresponding to stage 2 of our model, the firm determines how best to allocate backup capacity in order to deliver the orders arriving at the disrupted FC. In stage 2, our model considers the length of disruption as random, and the firm complies with a chance constraint that limits the probability of the number of late deliveries exceeding a threshold to be less than a tolerable probability.

Our study utilizes capacity planning, rather than inventory planning, as a proactive measure because of the characteristics of a delivery supply chain. The operations at a fulfillment center (e.g., sorting, bundling, and wrapping) require agility and flexibility since each customer order consists of a unique combination of multiple products. Therefore, satisfying those unique combinations through excess inventory is not practically possible when the operations at a fulfillment center are disrupted. As a result, the firm needs to utilize the excess capacity of its functional facilities, in addition to inventory, in order to satisfy the business customer orders of the disrupted facility. Thus, capacity planning at other fulfillment centers serves as a buffer against disruptions in a delivery supply chain.

Our analysis integrates a comprehensive set of disruptions in the delivery supply chain of the firm. The data for these disruptions are either provided by the firm (a string of eight years' information) or collected from national sources. Our work focuses on two types of disruptions: (1) Low impact and high likelihood (in the rest of the document, we refer to these disruptions shortly as low-impact disruptions) and (2) high impact and low likelihood (we refer to these disruptions as high-impact disruptions). Figure

2 provides the heat map for a comprehensive set of disruptions at the motivating firm (details of data are provided in Section 5). This figure enables us to classify bomb threat, break in, fire, flood, gas leak, tornado, power outage, and weather as low-impact disruption events, and earthquake, hurricane, and chemical and nuclear plant failures as high-impact disruption events.



ID	Risk	I	L	V	S
1	Bomb Threat	2.5	2	4	5
2	Break In	2	1.5	1	5
3	Earthquake	4.5	2	3	5
4	Fire	2.5	1	1	5
5	Flood	3	3.5	3.5	2
6	Gas Leak	1	4	3	4
7	Hurricane	5	1.5	1.5	4
8	Chemical and Nuclear	5	1	5	5
9	Power Outage	2	5	2	5
10	Tornado	3	3	3	5
11	Weather	2	4	3.5	3.5

I = Impact L = Likelihood V = Vulnerability S = Speed of onset

- Dot color reflects vulnerability: Very Low, Low, Medium, High, Very High
- Dot size reflects speed of onset

Figure 2. Heat map for the disruption categorization.

From conversations with the executives at the firm, we understand that the contingency backup plans primarily account for low-impact disruptions (e.g., power outage, gas leak, etc.). Typically, fulfillment centers are paired based on geographic distance to serve each other as primary backup in case of a disruption. However, several of those paired facilities are located close to each other, thus, it is possible that they both can be affected by a high-impact disruption occurring in the region (e.g., earthquake, hurricane, etc.). When such nearby facilities become nonfunctional at the same time, our model utilizes a third fulfillment center, which is located far away from the paired (nearby) facilities, to satisfy the orders. Thus, our work sheds light on the commonly ignored effects of high-impact disruptions.

One might expect that geographic proximity should anchor the decision on where to add capacity, i.e., the firm should be economically better off by adding more capacity at the paired facilities, which are located in closer proximity, rather than adding capacity at the distant facility. However, our work suggests the opposite under several conditions, and characterizes those conditions. This is an important result because it would motivate establishing an omni-channel backup system for a firm operating multiple channels that are not linked to each other. For example, our motivating firm operates a second distribution network that is called the retail supply chain where the distribution centers are responsible for serving the retail stores alone. Even though the distribution centers have greater amounts of excess capacity, they currently do not communicate with the fulfillment centers. However, our work provides the motivation for establishing a backup link between these two channels by justifying that the readily available excess capacity at a distribution center can be used to back up a fulfillment center even if these facilities are not located close to each other.

Our work shows how capacity planning is influenced by risk aversion. Since the firm has a next-day delivery promise, late deliveries may have a greater impact in the long-run than the immediate financial loss observed. Thus, we incorporate a chance constraint to capture the likelihood of late deliveries exceeding a tolerable amount under the presence of disruption length uncertainty. When a disruption occurs, its duration is typically uncertain, and a disruption lasting longer may lead to more late deliveries. One might intuit that, as risk aversion increases, the firm should increase its total capacity expansion. However, our work shows that there can be a substitution effect between the capacity decisions as risk aversion increases such that increasing capacity at the distant facility may lead to a decrease in the capacity at one or both of the paired (nearby) facilities. As a further consequence of the substitution effect, we find that the firm's total capacity expansion may decrease as risk aversion increases. This rather surprising result stems from two reasons: (1) Incorporating the high-impact disruptions into the analysis, and (2) the flexibility of the distant facility to serve both of the paired (nearby) facilities.

In sum, this paper makes five main contributions. First, our work utilizes capacity planning (rather than inventory planning) as a buffer against disruptions in a delivery supply chain. Second, our work examines both low-impact and high-impact disruptions together where the latter is commonly overlooked. Third, we find that geographic proximity may not anchor the decision on where to expand capacity. Fourth, we show that the firm's total capacity expansion may decrease as risk aversion increases. Fifth, our numerical illustration recommend that capacity expansion should take place in facilities that are exposed to lower risks; the consequence of this finding is that capacity expansion should not take place in facilities that are close to metropolitan areas but rather in facilities that can serve as a backup to these paired FCs located in densely populated regions. Our proposed modeling approach leads to significant savings in the firm's operational costs stemming from disruptions.

The reminder of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 presents the two-stage model, and it is analyzed in Section 4. Section 5 presents our numerical illustrations using company data. Section 6 presents our conclusions and managerial insights. All proofs and derivations are provided in the appendix.

2. Literature Review

Earlier literature mainly focuses on two types of levers against supply chain disruptions: (1) Inventory, and (2) flexible sourcing. There are many papers that examine the use of inventory as a mitigation strategy (Xia et al. 2004, Qi et al. 2009, Yang et al. 2009 and 2012, Atan and Snyder 2012, Dong and Tomlin 2012, DeCroix 2013, Tang et al. 2014, Dong et al. 2015). These studies consider manufacturing supply chains. The firm typically determines the inventory level, and then utilizes the safety stock if operations at the manufacturing facility are disrupted. Inventory planning proves useful in a manufacturing setting because manufacturing operations are identical across multiple customer orders for the same product. This enables the firm to accumulate some safety stock to be used in case of a disruption. Schmidt and Mehrotra (2016) examine the impact of inaccurate disruption duration estimates for firms that use inventory as a buffer for disruptions. We differ from these papers as our work focuses on a delivery supply chain where the operations (e.g. sorting, bundling, and wrapping) are rather unique across customer orders. Thus, increasing inventory levels at a fulfillment center does not protect the firm if the disruption halts the operations.

Flexible sourcing is another strategy in order to mitigate the supply chain disruptions. Sourcing from multiple locations are examined in various settings by Gurler and Parlar (1997), Berger et al. (2004), Tomlin and Wang (2005), Berger and Zeng (2006), Tomlin (2006), Tang (2006), Ruiz-Torres and Mahmoodi (2007), Yan and Liu (2009), Meena et al. (2011), Qi (2013). Ang et al. (2016) add a multi-tier structure to aforementioned papers. Their work studies the sourcing decisions in a supply network where the tier 2 suppliers are prone to disruption. These papers often focus on the coordination of multiple vendors. Our work also features an example of flexible sourcing which is the vendor shipment that is exercised when the backup capacity is not sufficient to recover the entire demand. However, vendor shipment is not a preferable alternative for our motivating firm as it puts emphasis on the next-day delivery. In addition to being highly costly, vendor shipments cannot satisfy the delivery commitments of the firm. Chen and Graves (2014) and Acimovic and Graves (2015) examine the decisions made at the fulfillment centers of an online retailer. The former focuses on a transportation problem where a sparsity constraint exists. The latter focuses on minimizing the outbound transportation cost through a proactive approach. However, neither paper features any sort of disruption risk. Our work employs a proactive capacity planning approach at fulfillment centers in the presence of disruption risk, and these decisions are coupled with a reactive contingency transportation planning.

Serpa and Krishnan (2016) study the strategic role of insurance in a multi-firm setting in the presence of operational failures. The likelihood of these operational failures can be controlled by the firms' efforts. They show that insurance can be used as a commitment mechanism to improve the firms' efforts, reducing the likelihood of the operational failures. In our work, disruptions are not stemming from operational failures at the firm, thus, the disruption likelihood is considered to be exogenous.

Simchi-Levi et al. (2015) make an important practical contribution in addition to the aforementioned analytical papers. Their work examines the financial impact of a generic disruption (i.e., low likelihood and high impact) at different nodes of an automotive supply chain. However, the scope of mitigation strategy is the use of inventory like the previous studies. We depart from this paper in several aspects: (1) Our work focuses on a delivery supply chain (where the nodes can serve each other as backup without featuring a precedence relationship) instead of a manufacturing supply chain (where the nodes cannot backup each other due to precedence relationships); (2) our work uses mitigation through capacity planning instead of inventory planning; (3) our model addresses a high-impact disruption in addition to a low-impact disruption; (4) Length of disruption is a static parameter in Simchi-Levi et al. (2015) whereas our model introduces randomness into the length of disrupted operations; (6) Simchi-Levi et al. (2015) study the impact when the disruption occurs, and thus, focuses on reactive operations; our work, on the other hand, examines the expected impact before the disruption actually occurs, and can be perceived as a proactive approach.

3. Model

The firm's problem is formulated using a two-stage stochastic program under disruption risk. The objective in stage 1 is to minimize the sum of the two costs: The capacity expansion cost and the expected cost from executing a contingency plan over the next one-year period (by determining the capacity expansion decisions). A contingency plan is implemented in stage 2 if a disruption occurs; the objective in stage 2 is to minimize the cost of executing the contingency plan (by determining the backup allocation decisions that are capped by stage-1 decisions).

Stage-1 capacity decisions are made under disruption risk that would halt operations at FCs. The disruption risks are classified as low impact and high impact disruptions. We develop a stylized model that examines three FCs such that FC1 and FC2 are located close to each other, and FC3 is located far away from FC1 and FC2. Later in Section 5, we relax this assumption and consider more than three facilities in the network. Because FC1 and FC2 are located in closer proximity, one serves the other as the primary backup facility in case of a low-impact disruption (e.g., gas leak). The implication of serving as a backup facility is that the demand at the nonfunctional FC is diverted to the functional FC. The probability of a low-impact disruption occurring at FC1 (FC2) over the one-year period is denoted by $p_{L,1}$ ($p_{L,2}$). The probability of a low-impact disruption occurring at FC1 and FC2 at the same time is assumed to be negligible. However, a high-impact disruption (e.g., earthquake, hurricane) may affect both FC1 and FC2 because of their geographic proximity; thus, we assume that both FC1 and FC2 become nonfunctional at the same time due to a high-impact disruption with a probability of p_H over the one-year period. It is important to note that $\{p_{L,1}, p_{L,2}\} > p_H$ in light of the disruption definitions (i.e., high vs. low likelihood). In case of a high-impact disruption, the demand of FC1 and FC2 are diverted to FC3, which is assumed to be out of the impact region due to the fact that it is sufficiently far away from FC1 and FC2. Furthermore, FC3 is capable of serving FC1 (FC2) as the secondary backup facility in case of a low-impact disruption if the primary backup FC2 (FC1) does not have sufficient excess capacity to recover the entire demand of FC1 (FC2). For orders that cannot be fulfilled through FCs, the firm engages vendor shipments.

In stage 1, the firm determines the amount of capacity expansion, denoted K_i, at fulfillment center i (shortly, FC i) where $i \in \{1, 2, 3\}$. The unit cost of capacity expansion (amortized per year) is denoted by c_K. Each FC is primarily responsible for serving its own customer demand, denoted D_i, and has a beginning capacity K_i^0 such that $D_i \le K_i^0$ (i.e., each FC has a sufficient initial capacity to satisfy its own customer demand). In order to eliminate several trivial scenarios, we assume that the primary backup facility does not have sufficient excess capacity at the beginning (i.e., the beginning capacity net of its own demand) to recover the entire demand of the nearby facility, i.e., $\{K_1^0, K_2^0\} < D_1 + D_2$. Similarly, we assume that FC3 does not have sufficient excess capacity at the beginning to recover the entire demand of FC1 and FC2 at the same time, i.e., $K_3^0 < D_1 + D_2 + D_3$. Note that $\{K_i , K_i^0, D_i\}$ represent the daily amounts.

In stage 2, one of the following four events occurs, and the firm determines the allocation of daily backup capacity, limited with the capacity expansion decisions made in stage 1, at the functional FC(s) in order to fulfill the demand at the nonfunctional FC(s):

Event 1: A low-impact disruption hits FC1 for a random duration $\tilde{t}_L$ (in days) with an expectation of $E[\tilde{t}_L] = \bar{t}_L$. The firm determines the daily backup amount $B_{i,1}$ from FC i to recover the daily demand at FC1 at a unit cost of $c_{i,1}$ where $i \in \{2, 3\}$. Note that $c_{2,1} < c_{3,1}$ due to the fact that FC2 is closer than FC3. The goal is to minimize the expected cost of implementing the contingency plan, denoted $\psi_{L,1}(\cdot)$, subject to a chance constraint that limits the probability of the number of late deliveries exceeding a threshold to be less than a tolerable probability.

Event 2: A low-impact disruption hits FC2 for a random duration $\tilde{t}_L$ (in days) with an expectation of $E[\tilde{t}_L] = \bar{t}_L$. The firm determines the daily backup amount $B_{i,2}$ from FC i to recover the daily demand at FC2 at a unit cost of $c_{i,2}$ where $i \in \{1, 3\}$. Note that $c_{1,2} < c_{3,2}$ since FC1 is closer than FC3. The goal is to minimize the expected cost of implementing the contingency plan, denoted $\psi_{L,2}(\cdot)$, subject to a chance constraint that limits the probability of the number of late deliveries exceeding a threshold to be less than a tolerable probability.

Event 3: A high-impact disruption hits both FC1 and FC2 for a random duration $\tilde{t}_H$ (in days) with an expectation of $E[\tilde{t}_H] = \bar{t}_H$. Note that $\tilde{t}_H$ has first-order stochastic dominance over $\tilde{t}_L$ in light of the disruption definitions (i.e., high vs. low impact). The firm determines the daily backup amount $B_{3,j}$ from FC3 to recover the daily demand at FC j at a unit cost of $c_{3,j}$ where $j \in \{1, 2\}$. The goal is to minimize the expected cost of implementing the contingency plan, denoted $\psi_H(\cdot)$, subject to a chance constraint that limits the probability of the number of late deliveries exceeding a threshold to be less than a tolerable probability.

Event 4: No disruption occurs with a probability of $1 - p_{L,1} - p_{L,2} - p_H$. A contingency plan is not needed; thus, it has zero cost in stage 2. As a result, this event is simply excluded from the analysis.

The cost of implementing the contingency plan involves three types of costs: (1) Additional transportation cost stemming from on-time delivery through the use of backup fulfillment centers; (2) the cost of late deliveries associated with the firm's promise of delivering within the next business day; and, (3) the cost of satisfying demand through vendor shipments. We next describe each cost individually.

A firm's capability of executing a contingency backup plan may be restricted by its level of preparedness. If a firm is not well prepared in advance for a disruption (e.g., absence of a rigorous plan to execute the contingency actions), the backup capacity may not be effectively utilized. This may cause late deliveries even if the firm has sufficient excess capacity. We denote the firm's level of preparedness as T. The preparedness affects the on-time delivery performance of the backup actions, i.e., T portion of the backup allocation is delivered on-time, and $1 - T$ portion is delivered late where $0 \leq T \leq 1$. For every unit of late delivery, the firm incurs an additional cost of c_L. The cost of on-time deliveries is denoted by $OC(\cdot)$, and the cost of late deliveries is denoted by $LC(\cdot)$.

If the firm's maximum backup capacity, which is capped by stage-1 decisions, is not sufficient to recover the entire demand of the nonfunctional FC(s), then the remaining demand is fulfilled through a vendor. Vendor deliveries are late, and incur a unit cost of $c_V + c_L$. The cost of vendor deliveries is denoted by $VC(\cdot)$. As a result, the cost of contingency plan is composed of three terms: (1) the cost of on-time deliveries $OC(\cdot)$, (2) the cost of late deliveries $LC(\cdot)$, and (3) the cost of vendor deliveries $VC(\cdot)$. It is defined that $\{c_{2,1}, c_{1,2}, c_{3,1}, c_{3,2}\} < c_V$ since vendor shipment is the most costly backup alternative.

The firm's risk consideration is modeled using a chance constraint in stage 2. According to this risk constraint, the firm limits the probability of late deliveries exceeding a tolerable threshold when a disruption occurs. The tolerable threshold is denoted by β, and the tolerable probability is denoted by α. In the firm motivating our problem, the values of β and α come strictly from the firm's promises in terms of its overall delivery performance used in attracting business customers.

The model is mathematically expressed as follows:

Stage 1:

$$\min_{K_1, K_2, K_3} \Psi(K_1, K_2, K_3) = c_K(K_1 + K_2 + K_3) + p_{L,1}\psi_{L,1}(B_{2,1}, B_{3,1} \mid K_2, K_3) + p_{L,2}\psi_{L,2}(B_{1,2}, B_{3,2} \mid K_1, K_3) + p_H\psi_H(B_{3,1}, B_{3,2} \mid K_3) \quad (1)$$

subject to $\{K_1, K_2, K_3\} \geq 0.$ (2)

Stage 2:

Event 1: A low-impact disruption occurs at FC1

$$\min_{\{B_{2,1}, B_{3,1}\} \geq 0} \psi_{L,1}(B_{2,1}, B_{3,1} \mid K_2, K_3) = [OC(B_{2,1}, B_{3,1}) + LC(B_{2,1}, B_{3,1}) + VC(B_{2,1}, B_{3,1})]E[\tilde{t}_L] \tag{3}$$

subject to

$$B_{2,1} \leq \min\{K_2^0 + K_2 - D_2, D_1\} \tag{4}$$

$$B_{3,1} \leq \min\{K_3^0 + K_3 - D_3, D_1\} \tag{5}$$

$$P[[(B_{2,1} + B_{3,1})(1 - T) + (D_1 - B_{2,1} - B_{3,1})^+]\tilde{t}_L > \beta] \leq \alpha \tag{6}$$

where

$$OC(B_{2,1}, B_{3,1}) = (c_{2,1}B_{2,1} + c_{3,1}B_{3,1})T \tag{7}$$

$$LC(B_{2,1}, B_{3,1}) = [(c_{2,1} + c_L)B_{2,1} + (c_{3,1} + c_L)B_{3,1}](1 - T) \tag{8}$$

$$VC(B_{2,1}, B_{3,1}) = (c_V + c_L)(D_1 - B_{2,1} - B_{3,1})^+. \tag{9}$$

Event 2: A low-impact disruption occurs at FC2

$$\min_{\{B_{1,2}, B_{3,2}\} \geq 0} \psi_{L,2}(B_{1,2}, B_{3,2} \mid K_1, K_3) = [OC(B_{1,2}, B_{3,2}) + LC(B_{1,2}, B_{3,2}) + VC(B_{1,2}, B_{3,2})]E[\tilde{t}_L] \tag{10}$$

subject to

$$B_{1,2} \leq \min\{K_1^0 + K_1 - D_1, D_2\} \tag{11}$$

$$B_{3,2} \leq \min\{K_3^0 + K_3 - D_3, D_2\} \tag{12}$$

$$P[[(B_{1,2} + B_{3,2})(1 - T) + (D_2 - B_{1,2} - B_{3,2})^+]\tilde{t}_L > \beta] \leq \alpha \tag{13}$$

where

$$OC(B_{1,2}, B_{3,2}) = (c_{1,2}B_{1,2} + c_{3,2}B_{3,2})T \tag{14}$$

$$LC(B_{1,2}, B_{3,2}) = [(c_{1,2} + c_L)B_{1,2} + (c_{3,2} + c_L)B_{3,2}](1 - T) \tag{15}$$

$$VC(B_{1,2}, B_{3,2}) = (c_V + c_L)(D_2 - B_{1,2} - B_{3,2})^+. \tag{16}$$

Event 3: A high-impact disruption occurs at FC1 and FC2

$$\min_{\{B_{3,1}, B_{3,2}\} \geq 0} \psi_H(B_{3,1}, B_{3,2} \mid K_3) = [OC(B_{3,1}, B_{3,2}) + LC(B_{3,1}, B_{3,2}) + VC(B_{3,1}, B_{3,2})]E[\tilde{t}_H] \tag{17}$$

subject to

$$B_{3,1} \leq \min\{K_3^0 + K_3 - D_3, D_1\} \tag{18}$$

$$B_{3,2} \leq \min\{K_3^0 + K_3 - D_3, D_2\} \tag{19}$$

$$B_{3,1} + B_{3,2} \leq K_3^0 + K_3 - D_3 \tag{20}$$

$$P[[(B_{3,1} + B_{3,2})(1 - T) + (D_1 + D_2 - B_{3,1} - B_{3,2})^+]\tilde{t}_H > \beta] \leq \alpha \tag{21}$$

where

$$OC(B_{3,1}, B_{3,2}) = (c_{3,1}B_{3,1} + c_{3,2}B_{3,2})T \tag{22}$$

$$LC(B_{3,1}, B_{3,2}) = [(c_{3,1} + c_L)B_{3,1} + (c_{3,2} + c_L)B_{3,2}](1 - T) \tag{23}$$

$$VC(B_{3,1}, B_{3,2}) = (c_V + c_L)(D_1 + D_2 - B_{3,1} - B_{3,2})^+. \quad (24)$$

The unit backup cost between the nearby facilities (i.e., FC1 and FC2) is assumed to be equal in both directions, and are therefore, relabeled as c_{12}, i.e., $c_{12} = c_{1,2} = c_{2,1}$ in the remaining part of the analysis. The unit backup cost from the distant facility (i.e., FC3) to FC1 is assumed to be equal to that to FC2, and is relabeled as c_3, i.e., $c_3 = c_{3,1} = c_{3,2}$ in the remaining part of the analysis. Furthermore, it is assumed that $c_3 - c_{12} < c_V - c_3$; this implies that the marginal benefit of recovering one order from the vendor to the distant FC (i.e., FC3) is greater than that from the distant FC to the nearby FC (i.e., FC1 or FC2). Finally, it is assumed that the probability of a low-impact disruption at FC1 is equal to that at FC2; thus, the probability of a low-impact disruption is relabeled as p_L, i.e., $p_L = p_{L,1} = p_{L,2}$ in the remaining part of the analysis. This assumption is relaxed in the numerical illustrations of Section 5 where the probabilities of disruptions vary at each FC. It is important to note that $c_3 - c_{12} < c_V - c_3$ and $p_{L,1} = p_{L,2}$ are useful for eliminating several redundant scenarios in the mathematical analysis that do not bring additional insight.

4. Analysis

This section presents the analytical results obtained from the model presented in Section 3.

4.1. Optimal Stage 2 Policies

The cost structure defined in the model section (i.e., $c_{12} < c_3 < c_V$) prioritizes the backup shipment alternatives in the following order: (1) the nearby FC, (2) the distant FC, and (3) the vendor. Note that the nearby FC alternative is not available in case of a high-impact disruption (Event 3). On the other hand, the distant FC and the vendor alternatives are available in each disruption (Events 1, 2, and 3).

Stage 2 is composed of three events, and the optimal policy for each event is presented in the following proposition.

Proposition 1. *For a given* (K_1, K_2, K_3),

(*a*) *if Event* 1 *occurs in stage* 2, *then*

(*i*) *if* $K_2^0 + K_2 - D_2 > D_1$, *then* $(B_{2,1}^*, B_{3,1}^*) = (D_1, 0)$;

(*ii*) *if* $K_2^0 + K_2 - D_2 \le D_1$, *then* $B_{2,1}^* = K_2^0 + K_2 - D_2$, *and*

(1) *if* $K_3^0 + K_3 - D_3 > D_1 - K_2^0 - K_2 + D_2$, *then* $B_{3,1}^* = D_1 - K_2^0 - K_2 + D_2$;

(2) *if* $K_3^0 + K_3 - D_3 \le D_1 - K_2^0 - K_2 + D_2$, *then* $B_{3,1}^* = K_3^0 + K_3 - D_3$;

(*b*) *if Event* 2 *occurs in stage* 2, *then*

(*i*) *if* $K_1^0 + K_1 - D_1 > D_2$, *then* $(B_{1,2}^*, B_{3,2}^*) = (D_2, 0)$;

(*ii*) *if* $K_1^0 + K_1 - D_1 \le D_2$, *then* $B_{1,2}^* = K_1^0 + K_1 - D_1$, *and*

(1) *if* $K_3^0 + K_3 - D_3 > D_2 - K_1^0 - K_1 + D_1$, *then* $B_{3,2}^* = D_2 - K_1^0 - K_1 + D_1$;

(2) *if* $K_3^0 + K_3 - D_3 \le D_2 - K_1^0 - K_1 + D_1$, *then* $B_{3,2}^* = K_3^0 + K_3 - D_3$;

(*c*) *if Event* 3 *occurs in stage* 2, *then*

(*i*) *if* $K_3^0 + K_3 - D_3 > D_1 + D_2$, *then* $(B_{3,1}^*, B_{3,2}^*) = (D_1, D_2)$;

(ii) *if* $K_3^0 + K_3 - D_3 \le D_1 + D_2$, *then* $(B_{3,1}^*, B_{3,2}^*) \in \{(B_{3,1}, B_{3,2}): B_{3,1} + B_{3,2} = K_3^0 + K_3 - D_3\}$.

In Proposition 1(a)(ii)(2), 1(b)(ii)(2), and 1(c)(ii), the total backup capacity is not sufficient to recover the entire demand at the nonfunctional FC(s). Therefore, the remaining portion of the demand is fulfilled through the vendor. However, in the other conditions, the entire demand can be backed up without any vendor shipment.

4.2. Optimal Stage 1 Policies – Risk Neutral

We begin the analysis by characterizing the first-order conditions. The next proposition makes use of the following notation to denote the expected marginal benefits of capacity expansion:

$$N_S = p_L \bar{t}_L \, (c_3 - c_{12})$$

$$N_L = p_L \bar{t}_L \, (c_V - c_{12} + c_L T)$$

$$F_S = p_H \bar{t}_H \, (c_V - c_3 + c_L T)$$

$$F_M = (p_L \bar{t}_L + p_H \bar{t}_H)(c_V - c_3 + c_L T)$$

$$F_L = (2p_L \bar{t}_L + p_H \bar{t}_H)(c_V - c_3 + c_L T)$$

where N_S and N_L (such that $N_S < N_L$) represent the small and large benefit, respectively, of capacity expansion at the nearby functional FC in case of a low-impact disruption (either Event 1 or Event 2). The value of N_S is the expected benefit of recovering one order from the distant FC to the nearby FC. The value of N_L is the expected benefit of recovering one order from the vendor to the nearby FC. On the other hand, F_S, F_M, and F_L (such that $F_S < F_M < F_L$) represent the small, moderate, and large benefit, respectively, of capacity expansion at the distant FC. Expanding capacity at the distant FC recovers the orders from the vendor. The value of F_S is the expected benefit in Event 3 only. The value of F_M is the expected benefit in either Event 1 or Event 2 in addition to Event 3. The value of F_L is the expected benefit in all three events.

Proposition 2. *The first-order conditions for* K_1, K_2, *and* K_3 *are as follows*:

$$\partial\Psi(K_1, K_2, K_3)/\partial K_1 = \begin{cases} c_K & \text{if } D_1 + D_2 - K_1^0 < K_1 \\ c_K - N_S & \text{if } D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_3 < K_1 \le D_1 + D_2 - K_1^0 \\ c_K - N_L & \text{if } K_1 \le D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_3 \end{cases}$$

$$\partial\Psi(K_1, K_2, K_3)/\partial K_2 = \begin{cases} c_K & \text{if } D_1 + D_2 - K_2^0 < K_2 \\ c_K - N_S & \text{if } D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_3 < K_2 \le D_1 + D_2 - K_2^0 \\ c_K - N_L & \text{if } K_2 \le D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_3 \end{cases}$$

$$\partial\Psi(K_1, K_2, K_3)/\partial K_3 = \begin{cases} c_K & \text{if } D_1 + D_2 + D_3 - K_3^0 < K_3 \\ c_K - F_S & \text{if } \max\{D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_2\} < K_3 \le D_1 + D_2 + D_3 - K_3^0 \\ c_K - F_M & \text{if } \min\{D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_2\} < K_3 \le \max\{D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_2\} \\ c_K - F_L & \text{if } K_3 \le \min\{D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_2\}. \end{cases}$$

The above proposition states that the capacity expansion decisions have a piecewise linear impact on the stage-1 objective function. Let us first consider the effect of K_1. FC1 serves as a primary backup if FC2 gets disrupted (Event 2). In case of Event 2, FC3 can also serve FC2 as a secondary backup. If the combined excess capacity at FC1 and FC3 is not sufficient to fulfill the entire demand at FC2, then increasing K_1 has a marginal benefit of N_L (i.e., the third region of $\partial\Psi(\cdot)/\partial K_1$) as it recovers from the vendor. Otherwise, increasing K_1 has a marginal benefit of N_S (i.e., the second region of $\partial\Psi(\cdot)/\partial K_1$) as it recovers from the distant FC. However, if K_1 is very large, then FC1 can back up the entire demand by itself (i.e., the first region of $\partial\Psi(\cdot)/\partial K_1$); thus, the marginal benefit of any further expansion becomes zero.

Similar interpretations apply to the effect of K_2; we omit the details. Let us now consider the effect of K_3. FC3 can serve as a primary backup (in case of Event 3) as well as a secondary backup (in cases of Event 1 and Event 2). If the combined capacity at FC3 and the nearby functional facility is not sufficient to fulfill the entire demand at the nonfunctional FC in both Event 1 and Event 2, then increasing K_3 has a marginal benefit of F_L (i.e., the fourth region of $\partial\Psi(\cdot)/\partial K_3$) as it recovers from the vendor in both events in addition to Event 3. If the combined capacity at FC3 and the nearby functional facility is not sufficient to fulfill the entire demand at the nonfunctional FC in either Event 1 or Event 2, then increasing K_3 has a marginal benefit of F_M (i.e., the third region of $\partial\Psi(\cdot)/\partial K_3$) as it recovers from the vendor in one of those events in addition to Event 3. Otherwise, increasing K_3 has a marginal benefit of F_S (i.e., the second region of $\partial\Psi(\cdot)/\partial K_3$) as it recovers from the vendor in Event 3 only. However, if K_3 is very large, then FC3 can back up the entire demand ($D_1 + D_2$) in Event 3 (i.e., the first region of $\partial\Psi(\cdot)/\partial K_3$); thus, the marginal benefit of any further expansion becomes zero.

It is important to note that the second and third conditions for both $\partial\Psi(\cdot)/\partial K_1$ and $\partial\Psi(\cdot)/\partial K_2$ depend on K_3. Similarly, the second, third, and fourth conditions for $\partial\Psi(\cdot)/\partial K_3$ depend on both K_1 and K_2. Therefore, the optimal stage-1 decisions depend on the ranking of $\{N_S, N_L, F_S, F_M, F_L, c_K\}$.

Lemma 1. *The marginal benefits* $\{N_S, N_L, F_S, F_M, F_L\}$ *can be ranked as follows*:

(*a*) $F_S \le N_S < F_M \le N_L < F_L$;

(b) $N_S < F_S \leq N_L < F_M < F_L$;

(c) $N_S < N_L < F_S < F_M < F_L$.

In order to solve the generalized version of this problem, we introduce two new conditions: $(K_1^0 - D_1) + (K_3^0 - D_3) < D_2$ and $(K_2^0 - D_2) + (K_3^0 - D_3) < D_1$. These conditions imply that, at the beginning, e.g., $(K_1, K_2, K_3) = (0, 0, 0)$, the functional FCs do not have sufficient excess capacity to fulfill the entire demand at the nonfunctional FC in either Event 1 or Event 2. A similar condition for Event 3 is already introduced in Section 3, i.e., $K_3^0 - D_3 < D_1 + D_2$. In the absence of these conditions, the regions where we observe $\{N_L, F_M, F_L\}$ may disappear; thus, the problem would become a sub-problem of the current version of our main problem.

The following proposition characterizes the optimal stage-1 decisions.

Proposition 3. *For a risk-neutral firm, the optimal stage-1 decisions* (K_1^N, K_2^N, K_3^N) *are*:

(a) *if* $\{N_S, N_L, F_S, F_M, F_L\} < c_K$, *then*

$(K_1^N, K_2^N, K_3^N) = (0, 0, 0)$;

(b) *if* $\{N_S, N_L, F_S, F_M\} < c_K \leq F_L$, *then*

$(K_1^N, K_2^N, K_3^N) = (0, 0, \min\{D_1 + D_2 + D_3 - K_3^0 - K_1^0, D_1 + D_2 + D_3 - K_3^0 - K_2^0\})$;

(c) *if* $F_S \leq N_S < \{F_M, c_K\} \leq N_L < F_L$, *then*

(i) *if* $F_L + c_K > 2N_L$ *and* $K_1^0 > K_2^0$, *then*

$(K_1^N, K_2^N, K_3^N) = (0, K_1^0 - K_2^0, D_1 + D_2 + D_3 - K_3^0 - K_1^0)$;

(ii) *if* $F_L + c_K > 2N_L$ *and* $K_1^0 \leq K_2^0$, *then*

$(K_1^N, K_2^N, K_3^N) = (K_2^0 - K_1^0, 0, D_1 + D_2 + D_3 - K_3^0 - K_2^0)$;

(iii) *if* $F_L + c_K \leq 2N_L$, *then*

$(K_1^N, K_2^N, K_3^N) = (D_1 + D_2 + D_3 - K_3^0 - K_1^0, D_1 + D_2 + D_3 - K_3^0 - K_2^0, 0)$;

(d) *if* $N_S < \{N_L, F_S\} < c_K \leq F_M < F_L$ *or* $N_S < F_S \leq c_K \leq N_L < F_M < F_L$, *then*

$(K_1^N, K_2^N, K_3^N) = (0, 0, \max\{D_1 + D_2 + D_3 - K_3^0 - K_1^0, D_1 + D_2 + D_3 - K_3^0 - K_2^0\})$;

(e) *if* $N_S < N_L < c_K \leq F_S < F_M < F_L$ *or* $N_S < c_K \leq \{N_L, F_S\} < F_M < F_L$, *then*

$(K_1^N, K_2^N, K_3^N) = (0, 0, D_1 + D_2 + D_3 - K_3^0)$;

(f) *if* $F_S \leq c_K \leq N_S < F_M \leq N_L < F_L$, *then*

$(K_1^N, K_2^N, K_3^N) = (D_1 + D_2 - K_1^0, D_1 + D_2 - K_2^0, 0)$;

(g) *if* $c_K \leq \{N_S, N_L, F_S, F_M, F_L\}$, *then*

$(K_1^N, K_2^N, K_3^N) = (D_1 + D_2 - K_1^0, D_1 + D_2 - K_2^0, D_1 + D_2 + D_3 - K_3^0)$.

Proposition 3(a) indicates that the firm does not invest in any additional capacity when its marginal cost is greater than the marginal benefits. Thus, vendor shipment is needed in all three events.

Proposition 3(b) indicates that the firm buys additional capacity at FC3 up to an amount such that the functional FCs have sufficient total excess capacity to fulfill the entire demand at the nonfunctional FC in either Event 1 or Event 2 (i.e., vendor shipment is not needed in one of these two events). Note that vendor shipment is still needed in Event 3.

Before proceeding with the results given in Proposition 3(c), we explain the implication of the condition $F_L + c_K > 2N_L$. Recall that the excess capacities at FC2 and FC3 are utilized in Event 1, the excess capacities at FC1 and FC3 are utilized in Event 2, and the excess capacity at FC3 is utilized in Event 3. Thus, the first unit invested in K_3 recovers one order from the vendor to the distant FC in all three events with a net benefit of $F_L – c_K$. Alternatively, the firm may invest one unit in K_1 and one unit in K_2; this investment recovers one order from the vendor to the nearby FC in events 1 and 2 with a net benefit of $2(N_L – c_K)$. Thus, the tradeoff is between (1) recovering one order from the vendor to the distant FC in three events, and (2) recovering one order from the vendor to the nearby FC in two events. If $F_L + c_K > 2N_L$ holds, it means that the firms is better off with the first alternative, thus, invests in K_3 up to the point where the region of F_L disappears. Otherwise, the firm is better off with the second alternative, thus, invests in K_1 and K_2 up to the point where the region of N_L disappears.

When $F_L + c_K > 2N_L$ holds, propositions 3(c)(i) and 3(c)(ii) show that the firm invests in the same amount of capacity at FC3 as in Proposition 3(b); furthermore, the firm adds capacity at either FC1 or FC2 up to an amount such that the functional FCs have sufficient total excess capacity to fulfill the entire demand at the nonfunctional FC in both Event 1 and Event 2 (i.e., vendor shipment is not needed in these two events). However, when $F_L + c_K > 2N_L$ does not hold, the firm invests in K_1 and K_2 as seen in Proposition 3(c)(iii). Similar to Proposition 3(c)(i) and 3(c)(ii), vendor shipment is not needed in Event 1 or Event 2. However, the total capacity expansion given in Proposition 3.3(c)(iii) is greater than that in propositions 3(c)(i) and 3(c)(ii).

Proposition 3(d) presents the same amount of total capacity expansion as in propositions 3(c)(i) and 3(c)(ii). Thus, vendor shipment is not needed in Event 1 or Event 2. However, unlike proposition 3(c)(i) and 3(c)(ii), the firm invests in capacity only at FC3. Note that vendor shipment is still needed in Event 3.

Proposition 3(e) states that the firm buys capacity at FC3 up to an amount such that FC3 has sufficient excess capacity to fulfill the total demand at FC1 and FC2 in Event 3 (i.e., vendor shipment is not needed in Event 3). This also implies that vendor shipment is not needed in Event 1 or Event 2.

Proposition 3(f) states that the firm adds capacity at FC1 and FC2 up to amounts such that one can completely recover the demand at the other FC. Thus, vendor shipment is not needed in Event 1 or Event 2. However, vendor shipment is still needed in Event 3 since $K_3^N = 0$.

Proposition 3(g) indicates that the firm makes a maximum amount of capacity investment when the cost is less than the marginal benefits. This means that FC1 and FC2 can recover each other completely in Event 1 and Event 2, whereas FC3 can recover the total demand at FC1 and FC2 completely in Event 3. Thus, no vendor shipment is needed in any event under this condition.

Proposition 3 implies that geographic proximity does not anchor the decision on where to add capacity when $K_3^N > 0$. In those conditions, the firm is economically better off by adding capacity at the distant facilities rather than at a nearby facility.

4.3. Optimal Stage 1 Policies under Risk Aversion

The risk constraints in stage 2 (see equations (6), (13), and (21)) measure the probability of the number of late deliveries exceeding the tolerable amount β in events 1, 2, and 3. Late delivery is caused by two factors: (1) the $(1 - T)$ portion of the backup allocation from the functional FC(s); and, (2) the vendor shipments. Vendor shipments can be eliminated in all events through capacity expansion decisions (i.e., $K_3 = D_1 + D_2 + D_3 - K_3^0$ as in propositions 3(e) and 3(g)). However, the $(1 - T)$ portion of the backup allocation remains to be late, thus, cannot be eliminated through capacity expansion decisions unless $T = 1$. As a consequence, depending on the values of several parameters (e.g. low values of T), the problem may be infeasible. The following lemma shows the necessary conditions to guarantee that the problem is feasible. The conditions given in this lemma are assumed to hold in the rest of the analysis. The values of $t_{L,1-\alpha}$ and $t_{H,1-\alpha}$ denote the realizations of $\tilde{t}_L$ and $\tilde{t}_H$ at fractile $1 - \alpha$, i.e., $P[\tilde{t}_L > t_{L,1-\alpha}] = P[\tilde{t}_H > t_{H,1-\alpha}] = \alpha$.

Lemma 2. *When* $K_3 = D_1 + D_2 + D_3 - K_3^0$, (*a*) *Equation* (6) *in Event 1 is satisfied if and only if* $D_1(1 - T)t_{L,1-\alpha}/\beta \le 1$ *holds*; (*b*) *Equation* (13) *in Event 2 is satisfied if and only if* $D_2(1 - T)t_{L,1-\alpha}/\beta \le 1$ *holds*; (*c*) *Equation* (21) *in Event 3 is satisfied if and only if* $(D_1 + D_2)(1 - T)t_{H,1-\alpha}/\beta \le 1$ *holds*. *Therefore, when these three conditions hold, the problem is feasible*.

From the above lemma, it follows that the risk constraints are never binding when the risk-neutral stage-1 decisions (K_1^N, K_2^N, K_3^N) can recover all the vendor shipments in all events as in propositions 3(e) and 3(g). Thus, the optimal risk-averse stage-1 decisions (K_1^A, K_2^A, K_3^A) are the same as the risk-neutral decisions for the conditions presented in propositions 3(e) and 3(g). The remark below summarizes this result.

Remark 1. $(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N)$ *if* $N_S < N_L < c_K \le F_S < F_M < F_L$ *or* $N_S < c_K \le \{N_L, F_S\} < F_M < F_L$ *or* $c_K \le \{N_S, N_L, F_S, F_M, F_L\}$.

For the remaining risk-neutral decisions, at least one of the risk constraints may be violated. The following proposition identifies the conditions under which each risk constraint is violated. Note that $R^{i,j}$ denotes j^{th} condition for Event i.

Proposition 4. *Given the optimal risk-neutral stage-1 decisions* (K_1^N, K_2^N, K_3^N):

(*a*) *The risk constraint* (6) *in Event 1 is violated when*

(*i*) $(K_2^N, K_3^N) = (0, 0)$ *and* $R^{1,1} = [D_1 - (K_2^0 - D_2 + K_3^0 - D_3)T]t_{L,1-\alpha}/\beta > 1$; *or*

(*ii*) $(K_2^N, K_3^N) = (0, D_1 + D_2 + D_3 - K_3^0 - K_1^0)$ *and* $K_1^0 > K_2^0$ *and*

$R^{1,2} = [D_1(1 - T) + (K_1^0 - K_2^0)T]t_{L,1-\alpha}/\beta > 1$;

(*b*) *The risk constraint* (13) *in Event 2 is violated when*

(i) $(K_1^N, K_3^N) = (0, 0)$ *and* $R^{2,1} = [D_2 - (K_1^0 - D_1 + K_3^0 - D_3)T]t_{L,1-\alpha}/\beta > 1$; *or*

(ii) $(K_1^N, K_3^N) = (0, D_1 + D_2 + D_3 - K_3^0 - K_2^0)$ *and* $K_1^0 \leq K_2^0$ *and*

$R^{2,2} = [D_2(1 - T) + (K_2^0 - K_1^0)T]t_{L,1-\alpha}/\beta > 1$;

(c) *The risk constraint* (21) *in Event 3 is violated when*

(i) $K_3^N = 0$ *and* $R^{3,1} = [D_1 + D_2 - (K_3^0 - D_3)T]t_{H,1-\alpha}/\beta > 1$; *or*

(ii) $K_3^N = D_1 + D_2 + D_3 - K_3^0 - K_1^0$ *and* $R^{3,2} = [(D_1 + D_2)(1 - T) + K_1^0T]t_{H,1-\alpha}/\beta > 1$; *or*

(iii) $K_3^N = D_1 + D_2 + D_3 - K_3^0 - K_2^0$ *and* $R^{3,3} = [(D_1 + D_2)(1 - T) + K_2^0T]t_{H,1-\alpha}/\beta > 1$.

If the risk-neutral decisions violate at least one of the risk constraints, then the firm should increase at least one of the capacity decisions compared to the risk-neutral benchmark in order to comply with the violated risk constraint(s). The following proposition characterizes the capacity expansion decisions that are required to comply with the risk constraints. We denote $K^{R,i}$ as the minimum total capacity expansion required to comply with the risk constraint in Event i.

Proposition 5. (a) *If the risk constraint* (6) *in Event 1 is violated*, *then the optimal risk-averse stage-1 decisions* (K_2^A, K_3^A) *must satisfy that* $K_2^A + K_3^A \geq K^{R,1} = [1/T][D_1 - \beta/t_{L,1-\alpha}] - (K_2^0 - D_2 + K_3^0 - D_3)$; ($b$) *If the risk constraint* (13) *in Event 2 is violated*, *then the optimal risk-averse stage-1 decisions* (K_1^A, K_3^A) *must satisfy that* $K_1^A + K_3^A \geq K^{R,2} = [1/T][D_2 - \beta/t_{L,1-\alpha}] - (K_1^0 - D_1 + K_3^0 - D_3)$; ($c$) *If the risk constraint* (21) *in Event 3 is violated*, *then the optimal risk-averse stage-1 decision* K_3^A *must satisfy that* $K_3^A \geq K^{R,3} = [1/T][D_1 + D_2 - \beta/t_{H,1-\alpha}] - (K_3^0 - D_3)$.

The following lemma explains that the risk constraint (21) in Event 3 is more restrictive than the risk constraints (6) and (13) in events 1 and 2, respectively. This is a consequence of high-impact disruption occurring at two facilities in Event 3 as opposed to low-impact disruption occurring at one facility in events 1 or 2.

Lemma 3. (a) $\{R^{1,1}, R^{2,1}\} < R^{3,1}$, $R^{1,2} < R^{3,2}$, *and* $R^{2,2} < R^{3,3}$; (b) $\{K^{R,1}, K^{R,2}\} < K^{R,3}$.

Lemma 3(a) states that the risk constraint (21) in Event 3 is the governing risk constraint. Lemma 3(b) states that the firm's minimum total capacity expansion required to comply with the risk constraint (21) in Event 3 is greater than that to comply with the risk constraints in events 1 and 2. Thus, the risk-averse firm should determine the capacity expansion decisions based on the governing risk constraint, which is (21) in Event 3. The following proposition characterizes the optimal risk-averse stage-1 decisions.

Proposition 6. *For a risk-averse firm*, *the optimal stage-1 decisions* (K_1^A, K_2^A, K_3^A) *are*:

(a) *if* $\{N_S, N_L, F_S, F_M, F_L\} < c_K$, *then*

(i) *if* $R^{3,1} \leq 1$, *then* $(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N) = (0, 0, 0)$;

(ii) *if* $R^{3,1} > 1$, *then* $(K_1^A, K_2^A, K_3^A) = (0, 0, K^{R,3})$;

(b) *if* $\{N_S, N_L, F_S, F_M\} < c_K \leq F_L$, *then*

(i) *if* $\max\{R^{3,2}, R^{3,3}\} \le 1$, *then*

$(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N)$

$= (0, 0, \min\{D_1 + D_2 + D_3 - K_3^0 - K_1^0, D_1 + D_2 + D_3 - K_3^0 - K_2^0\})$;

(ii) *if* $\max\{R^{3,2}, R^{3,3}\} > 1$, *then* $(K_1^A, K_2^A, K_3^A) = (0, 0, K^{R,3})$;

(c) *if* $F_S \le N_S < \{F_M, c_K\} \le N_L < F_L$, *then*

(i) *if* $F_L + c_K > 2N_L$ *and* $K_1^0 > K_2^0$, *then*

(1) *if* $R^{3,2} \le 1$, *then* $(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N)$

$= (0, K_1^0 - K_2^0, D_1 + D_2 + D_3 - K_3^0 - K_1^0)$;

(2) *if* $R^{3,2} > 1$, *then*

(2.1) *if* $K^{R,3} \le D_1 + D_2 + D_3 - K_3^0 - K_2^0$, *then*

$(K_1^A, K_2^A, K_3^A) = (0, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K^{R,3}, K^{R,3})$;

(2.2) *if* $K^{R,3} > D_1 + D_2 + D_3 - K_3^0 - K_2^0$, *then* $(K_1^A, K_2^A, K_3^A) = (0, 0, K^{R,3})$;

(ii) *if* $F_L + c_K > 2N_L$ *and* $K_1^0 \le K_2^0$, *then*

(1) *if* $R^{3,3} \le 1$, *then* $(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N)$

$= (K_2^0 - K_1^0, 0, D_1 + D_2 + D_3 - K_3^0 - K_2^0)$;

(2) *if* $R^{3,3} > 1$, *then*

(2.1) *if* $K^{R,3} \le D_1 + D_2 + D_3 - K_3^0 - K_1^0$, *then*

$(K_1^A, K_2^A, K_3^A) = (D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K^{R,3}, 0, K^{R,3})$;

(2.2) *if* $K^{R,3} > D_1 + D_2 + D_3 - K_3^0 - K_1^0$, *then* $(K_1^A, K_2^A, K_3^A) = (0, 0, K^{R,3})$;

(iii) *if* $F_L + c_K \le 2N_L$, *then*

(1) *if* $R^{3,1} \le 1$, *then*

$(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N)$

$= (D_1 + D_2 + D_3 - K_3^0 - K_1^0, D_1 + D_2 + D_3 - K_3^0 - K_2^0, 0)$;

(2) *if* $R^{3,1} > 1$, *then*

(2.1) *if* $K^{R,3} \le \{D_1 + D_2 + D_3 - K_3^0 - K_1^0, D_1 + D_2 + D_3 - K_3^0 - K_2^0\}$, *then*

$(K_1^A, K_2^A, K_3^A) =$

$(D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K^{R,3}, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K^{R,3}, K^{R,3})$;

(2.2) *if* $D_1 + D_2 + D_3 - K_3^0 - K_1^0 < K^{R,3} \le D_1 + D_2 + D_3 - K_3^0 - K_2^0$, *then*

$(K_1^A, K_2^A, K_3^A) = (0, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K^{R,3}, K^{R,3})$;

(2.3) *if* $D_1 + D_2 + D_3 - K_3^0 - K_2^0 < K^{R,3} \le D_1 + D_2 + D_3 - K_3^0 - K_1^0$, *then*

$(K_1^A, K_2^A, K_3^A) = (D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K^{R,3}, 0, K^{R,3})$;

(2.4) *if* $K^{R,3} > \{D_1 + D_2 + D_3 - K_3^0 - K_1^0, D_1 + D_2 + D_3 - K_3^0 - K_2^0\}$, *then*

$(K_1^A, K_2^A, K_3^A) = (0, 0, K^{R,3})$;

(d) *if* $N_S < \{N_L, F_S\} < c_K \le F_M < F_L$ *or* $N_S < F_S \le c_K \le N_L < F_M < F_L$, *then*

(i) *if* $\min\{R^{3,2}, R^{3,3}\} \le 1$, *then*

$(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N)$

$= (0, 0, \max\{D_1 + D_2 + D_3 - K_3^0 - K_1^0, D_1 + D_2 + D_3 - K_3^0 - K_2^0\})$;

(ii) *if* $\min\{R^{3,2}, R^{3,3}\} > 1$, *then* $(K_1^A, K_2^A, K_3^A) = (0, 0, K^{R,3})$;

(e) *if* $N_S < N_L < c_K \le F_S < F_M < F_L$ *or* $N_S < c_K \le \{N_L, F_S\} < F_M < F_L$, *then*

$(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N) = (0, 0, D_1 + D_2 + D_3 - K_3^0)$;

(f) *if* $F_S \le c_K \le N_S < F_M \le N_L < F_L$, *then*

(i) *if* $R^{3,1} \le 1$, *then* $(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N) = (D_1 + D_2 - K_1^0, D_1 + D_2 - K_2^0, 0)$;

(ii) *if* $R^{3,1} > 1$, *then* $(K_1^A, K_2^A, K_3^A) = (D_1 + D_2 - K_1^0, D_1 + D_2 - K_2^0, K^{R,3})$;

(g) *if* $c_K \le \{N_S, N_L, F_S, F_M, F_L\}$, *then*

$(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N)$

$= (D_1 + D_2 - K_1^0, D_1 + D_2 - K_2^0, D_1 + D_2 + D_3 - K_3^0)$.

In propositions 6(e) and 6(g), the governing risk constraint (21) never becomes binding, and thus, the optimal decisions are the same as the risk-neutral optimal decisions (recall Remark 1).

When $R^{3,j} \le 1$ for $j \in \{1, 2, 3\}$, the governing risk constraint (21) is not violated by the risk-neutral decisions, and the risk-neutral optimal decisions remain to be optimal under risk aversion (see Proposition 4(c)). This corresponds to the conditions given in propositions 6(a)(i), 6(b)(i), 6(c)(i)(1), 6(c)(ii)(1), 6(c)(iii)(1), 6(d)(i), and 6(f)(i).

In the remaining parts of this proposition, the risk-neutral decisions violate the governing risk constraint (21), and therefore, the firm needs to set $K_3^A = K^{R,3}$ ($> K_3^N$) in order to comply with the risk aversion constraint. Note that, as risk aversion increases (i.e., lower β and/or α), the value of $K^{R,3}$ increases. This implies that the firm increases the capacity expansion at FC3 as risk aversion increases. One might intuit that the firm should not change the capacity expansion decisions at FC1 and FC2 as they are not directly affected by the governing risk constraint (see Proposition 5(c)), and thus, the total capacity expansion should increase as the risk aversion increases. However, the following proposition presents the conditions where this intuition is not valid.

While risk aversion generally leads to an increase in the optimal capacity investment, there are conditions under which the optimal capacity can be decreasing with risk aversion. Thus, the optimal capacity investment is not monotonic in risk aversion. Using the unit cost of capacity expansion, the following proposition identifies the conditions that lead to this rather surprising finding. As will be seen in Section 5, this non-monotone behavior is confirmed in with numerical analyses using company data.

Proposition 7. *Let* $F_M \le N_L$ *hold. As risk aversion increases, for every unit of increase in FC3 capacity expansion, the firm decreases the capacity expansion* (a) *by one unit at both FC1 and FC2 when the unit* *cost of capacity expansion is such that* $N_S < c_K \leq 2N_L - F_L$; (*b*) *by one unit at either FC1 or FC2 when the unit cost of capacity expansion is such that* $2N_L - F_L < c_K \leq N_L$.

The above proposition identifies the conditions where we observe a substitution effect, i.e., increasing capacity investment at the farther FC (i.e., FC3) leads to a decrease in capacity investment at the nearby FC(s) (i.e., FC1 and FC2). Increasing capacity at FC3 by one unit brings a benefit through recovering one unit of vendor shipment by FC3. However, this diminishes the benefit provided by the last unit of capacity added at FC1 and FC2, resulting in an overinvestment. Therefore, the firm is economically better off by taking back that expansion at FC1 and/or FC2. Mathematically, increasing K_3 by one unit decreases the marginal benefit of the last unit in K_1 and K_2 from N_L to N_S (see Proposition 2). Since $N_S < \{F_M, c_K\} \leq N_L$ holds in Proposition 6(c), the firm is better off by taking back the last unit in K_1 and K_2 for every unit of increase in K_3 until K_1 and K_2 reach zero. The rate of substitution is two units in Proposition 7(a), and it is one unit in Proposition 7(b).

Proposition 7(a) describes an intriguing result that increasing the degree of risk aversion may lead to a decrease in total capacity expansion. Let us consider the case when the firm prefers making a capacity expansion at both FC1 and FC2 ($K_1^N > 0$ and $K_2^N > 0$) but not at FC3 ($K_3^N = 0$) in the risk neutral setting. At a low degree of risk aversion, the firm has to make a small amount of investment at FC3 ($K_3^A = K^{R,3}$) in order to comply with the governing risk constraint. However, the firm has to deduct the same amount from the capacity investment at both FC1 and FC2 due to the substitution effect. As a consequence, the total capacity expansion decreases as risk aversion increases until the investment at either FC1 or FC2 drops to zero (either $K_1^A = 0$ or $K_2^A = 0$).

5. Numerical Analysis

This section presents illustrations from numerical analyses. These numerical analyses confirm that our analytical findings apply in practical settings. The firm's 16 FCs are divided among four groups using a variety of criteria including proximity and trucking distances possible through the firm's transportation vehicles. Group 1 has 4 FCs, Group 2 has 6 FCs, Group 3 has 4 FCs, and Group 4 has 2 FCs that can serve each other as a backup facility in the event of a disruption. Our analysis enforces the risk constraints in (6), (13), and (21) to apply in each group independently rather than as a single set of constraints for the entire country. Incorporating the risk constraints in each group enables the firm's delivery service standards to apply in every geographic region uniformly. It also eliminates the potentially poorer service in one of the regions while satisfying the delivery service requirements in other regions.

5.1. Data regarding Disruption Risks and Costs

Our analysis examines the influence of eleven different disruption possibilities. The data regarding potential disruptions has two sources: The firm has provided eight-year long detailed information about the frequency and length of disruptions for seven of the eleven potential disruptions in Figure 2: Bomb threat, break in, fire, flood, gas leak, power outage, and weather. The data regarding the remaining four disruptions are collected from national sources using the most granular data available.

Earthquake data is collected through the US Geological Survey. The US Geological Survey describes that earthquakes that have a magnitude less than 5.0 result in minor or no impact. We consider the impact range of the average earthquake with a scale greater than or equal to 5.0 in order to determine the probability of earthquake impacting each FC.

The data for hurricanes and tornadoes are obtained from the National Oceanic and Atmospheric Administration. Facilities are influenced by hurricanes geographically and seasonally. Hurricane season predominantly occurs in the East Coast of the US from June 1st through November 30th. We derive the probability of a FC being impacted by a hurricane by using the county-level data. Tornadoes occur sporadically, and the data is only available at the state level. We take into account the impact region for an average tornado in order to compute the likelihood for each FC.

For chemical and nuclear disruptions, we first identify the ten most influential chemicals (ethylene oxide, oleum, sulfur dioxide, chlorine, furan, bromine, chlorine dioxide, hydrofluoric acid, toluene 2,6-diisocyanate, ammonia) that can cause significant disruptions at chemical production facilities. We then locate all plants that produce these ten chemicals in the US. Next, we examine the history of disruptions (33 years of data) at these facilities through the data collected from the Right-to-Know Network (www.rtknet.org). Nuclear disruptions, while limited in number in the US, are collected from the data available at the Nuclear Regulatory Commission for all of the operational nuclear power reactors in the US. We consider the proximity of the company's FCs to these nuclear power reactors and their impact radius in order to determine the frequency and length of nuclear disruptions.

We next describe the process used for estimating the probability and length of disruptions. For the company provided data, the frequency at each FC leads to the estimation of the probability at each FC independently; thus, each FC has a different probability of disruption for the seven disruption events with company data. The mean and standard deviations regarding the length of disruption are provided in the same data set. We assume Normal Distribution for these set of disruptions. For the disruption events whose data is collected from national sources, our granular data regarding the frequency of disruptions leads to different probabilities at each FC. Similarly, we assume that these events also follow Normal Distribution where the mean and standard deviation values are derived from the collected data. Our analysis employs a variety of potential disruptions at each FC. It is important to note that these random disruption events can be combined in the analysis due to the fact that the random variable representing the sum of random variables that follow the Normal Distribution also follows the Normal Distribution.

As described in Section 1, the eleven disruptions are classified in two sets: Low-impact and high-impact events. Table 1 provides the mean and standard deviation of each disruption event, and their corresponding probabilities at each FC.

		Low-Impact Disruptions								High-Impact Disruptions		
		Gas Leak	Fire	Power Outage	Break In	Weather	Bomb Threat	Tornado	Flood	Hurricane	Earthquake	Chemical and Nuclear
avg hours		0.9	4.3	5.7	7.9	8.0	30.2	49.3	64.6	7.6	13.8	2920
std. dev.		1.1	9.0	49.4	10.7	27.3	50.6	19.4	181.4	19.4	28.5	100.0
	268	2.01%	0.00%	15.95%	0.29%	0.43%	0.14%	2.23%	0.72%	0.63%	0.00%	0.08%
	368	1.55%	0.31%	3.31%	0.41%	1.14%	0.21%	1.06%	0.62%	3.75%	0.00%	0.06%
	983	0.72%	0.00%	17.31%	0.24%	0.48%	0.00%	5.23%	0.24%	9.38%	0.00%	0.02%
	269	1.00%	0.13%	7.55%	0.77%	0.00%	0.13%	0.26%	0.27%	0.00%	0.21%	0.02%
	398	2.53%	0.00%	20.73%	0.16%	0.47%	0.00%	4.02%	0.32%	3.75%	0.00%	0.06%
	948	1.30%	0.00%	13.54%	0.00%	5.47%	0.00%	2.87%	0.52%	0.00%	0.00%	0.04%
	497	1.00%	0.13%	7.55%	0.07%	0.00%	0.13%	0.26%	0.27%	0.00%	0.21%	0.02%
FC	297	0.72%	0.00%	17.31%	0.24%	0.48%	0.00%	5.23%	0.24%	9.38%	0.00%	0.02%
ID	469	1.30%	0.00%	13.54%	0.00%	5.47%	0.00%	2.87%	0.52%	0.00%	0.00%	0.04%
	598	2.50%	0.00%	12.50%	0.00%	3.75%	0.00%	3.86%	0.00%	0.00%	0.00%	0.00%
	198	2.44%	0.00%	18.29%	0.30%	0.61%	0.00%	3.05%	0.30%	0.63%	0.00%	0.08%
	292	1.07%	0.00%	7.50%	0.00%	0.36%	0.00%	3.61%	0.00%	0.00%	0.00%	0.00%
	868	0.00%	0.00%	8.33%	0.00%	0.00%	0.00%	2.39%	4.17%	0.00%	0.00%	0.02%
	281	1.25%	0.00%	10.63%	0.00%	0.00%	0.00%	2.91%	0.00%	0.00%	0.00%	0.00%
	397	0.42%	0.00%	12.92%	0.00%	0.00%	0.00%	0.15%	0.83%	0.00%	0.04%	0.00%
	697	1.39%	0.00%	1.39%	0.00%	0.00%	0.00%	0.24%	0.00%	0.00%	0.03%	0.00%

Table 1. The mean, standard deviation, and probability of each disruption at fulfillment centers.

The company has provided the initial capacity levels at each FC, and the associated cost data. In the event of a disruption at a FC, we incur three types of costs. If there is available capacity at the backup FC, the delivery would take place using the firm's own transportation vehicles. This on-time delivery would incur a transportation cost, $c_{i,j}$, that is based on mileage and the firm's standard transportation costs. The backup FC may not be fully prepared to make on-time deliveries; this occurs when $T < 1$. These late deliveries are more expensive than on-time deliveries by an additional per-delivery late fee c_L. The third possibility is to engage vendors for shipments, however, the shipments from vendors are also late. The cost of delivery using vendors, $c_V + c_L$, is 7.8 times the cost of late delivery using the firm's own transportation vehicles, i.e., $(c_V + c_L)/c_L = 7.8$. The unit cost of capacity expansion, c_K, is estimated from the construction cost of expanding conveyor belts and packaging facility in addition to the space for inventory incurring in each delivery; thus, it is expressed in terms of per delivery.

Before proceeding with the numerical analysis, let us provide some insight into the disruption risks at the 16 FCs in the firm's delivery supply chain. Using the definition of the "risk exposure index" (REI) score established by Simchi-Levi et al. (2014), we examine the level of monetary risk at each FC. Table 2 presents the REI scores (on a scale of 0 to 100) using the same cost terms provided above without making any additional capacity investment. These REI scores have been instrumental in the firm's decision to implement the recommendations from our study.

FC ID	983	268	398	948	368	297	469	598	269	868	292	281	497	397	198	697
REI	100.00	72.67	59.54	30.33	22.64	22.16	17.38	15.73	7.59	3.76	2.96	2.93	2.58	2.44	1.40	0.03

Table 2. REI scores of each FC using the same cost data.

According to the REI scores in Table 2, the highest risk appears in FCs identified as 983, 268, and 398 with REI scores greater than 50%. FCs 697, 198, 397, 497, 281, 292, 868, and 269 appear to have less than 10% of the maximum monetary risk corresponding to FC 983. Restricting the analysis to REI scores, one might intuit that the firm should invest in additional capacity at facilities with the highest REI scores. However, our results demonstrate that this is not the recommended strategy for mitigating disruption risks. As will be seen later, our optimal capacity investment does not occur at the highest REI score facilities; rather, the optimal capacity expansion recommendations take place at facilities with the lowest REI scores. Thus, our work contributes to the literature on mitigating supply chain disruption risks by extending the REI score analysis with capacity decisions and risk aversion in delivery supply chains. We expand on this observation and provide a managerially insightful discussion in Section 6.

5.2. Findings under the Risk-Neutral Setting

Table 3 provides the results of capacity investment decisions for the risk-neutral analysis at various unit costs of capacity investment (c_K) and at the level of preparedness involving $T = 1$. We use the expected total cost at $c_K = 1$ as the benchmark case in Table 3, and show the increase in expected total cost at higher values of c_K.

We make four observations from the results pertaining to the risk-neutral analysis. First, the optimal level of capacity investment decreases with higher levels of c_K. This can be seen in the optimal values corresponding to the row representing the sum of capacity investment decisions with $\sum K_i^*$ in Table 3. Beyond $c_K = 7$, additional capacity becomes too expensive and the firm does not engage in any capacity expansion.

Second, the firm does not need to make additional capacity investments in Group 1. This is a consequence of the fact that FCs in Group 1 have sufficient level of backup capacity to fulfill the demand in the event of disruption.

Third, the firm invests in additional capacity in some of its FCs in groups 2, 3, and 4. However, the selection of FC for the investment decision varies depending on the cost of capacity expansion c_K.

Fourth, FCs with the highest REI scores get limited capacity expansion, while FCs with the lowest REI score get the highest capacity investment. The facility with the highest and third highest REI scores, FCs 983 and 398, receive capacity investments that can make an additional 99 and 376 deliveries per day, respectively. The facility with the second highest REI score, FC 268, receives no additional capacity. On the other hand, facilities that are tagged with the lowest REI scores receive additional capacity investment. FC 198 with the REI score of only 1.40 is recommended for the highest level optimal capacity expansion, receiving an additional 1,443 units of daily delivery capability when $c_K \leq 2$. Similarly, FCs 281, 292, and 868, receive a positive capacity expansion when $c_K \leq 2$.

		c_K									
	FC ID	1	2	3	4	5	6	7	8	9	10
Group 1	269	0	0	0	0	0	0	0	0	0	0
	497	0	0	0	0	0	0	0	0	0	0
	697	0	0	0	0	0	0	0	0	0	0
	397	0	0	0	0	0	0	0	0	0	0
Group 2	983	99	0	0	0	0	0	0	0	0	0
	297	1,232	1,133	1,133	0	0	0	0	0	0	0
	281	92	190	190	190	190	190	92	0	0	0
	948	0	0	0	0	0	0	0	0	0	0
	469	281	281	281	0	0	0	0	0	0	0
	292	334	334	334	334	334	102	0	0	0	0
Group 3	198	1,443	1,443	1,819	761	761	761	761	0	0	0
	268	0	0	0	0	0	0	0	0	0	0
	398	376	376	0	0	0	0	0	0	0	0
	368	0	0	0	0	0	0	0	0	0	0
Group 4	868	344	344	0	0	0	0	0	0	0	0
	598	23	23	0	0	0	0	0	0	0	0
ΣK_i^*		4,224	4,124	3,757	1,285	1,285	1,053	853	0	0	0
Expected Total Cost		100%	131%	158%	177%	187%	196%	203%	206%	206%	206%

Table 3. The impact of the unit capacity cost (c_K) on the optimal capacity expansion decisions under the risk-neutral setting with $T = 1$.

Table 4 presents the comparison of REI scores in the risk-neutral setting between two scenarios: When the firm does not invest in capacity, i.e., $K_i = 0$, and when the firm makes the optimal capacity investment, i.e., $K_i = K_i^N$ for all i. We continue to use the expected total cost of FC 983 with no capacity investment as the underlying benchmark for all REI scores. The REI scores for the facilities that had the highest risk exposure prior to capacity expansion (FCs 983, 268, and 398 with REI scores of 100, 72.67, and 59.54, respectively) have significant reductions in their risk exposure (37.35, 9.85, and 22.03, respectively) after the optimal capacity investments. This leads to the conclusion that planning capacity in an optimal manner results in reducing risk exposure, and leads to a more resilient supply chain.

FC ID	983	268	398	948	368	297	469	598	269	868	292	281	497	397	198	697
REI ($K_i = 0$)	100.00	72.67	59.54	30.33	22.64	22.16	17.38	15.73	7.59	3.76	2.96	2.93	2.58	2.44	1.40	0.03
REI ($K_i = K_i^N$)	37.35	9.85	22.03	6.67	8.65	29.85	8.71	4.90	7.59	7.56	1.74	2.66	2.58	2.44	20.25	0.03

Table 4. The comparison of REI scores at each FC using the optimal capacity expansion decisions under the risk-neutral setting.

Next, we report the cost savings from capacity planning through the comparison of the total expected cost between the two supply chain designs with no capacity investment, denoted $\Psi(0)$, and with the optimal capacity expansion levels with $K = K_i^N$ for all i, denoted $\Psi(K^N)$. The cost savings from capacity expansion is denoted $\Delta\Psi_{\text{capacity}}$, and is equal to

$$\Delta\Psi_{\text{capacity}} = (\Psi(0) - \Psi(K^N))/\ \Psi(0) = 53\%$$

in the firm motivating our study. In conclusion, despite the additional capacity investment costs, the firm is projected to have a 53% savings in its total operating costs stemming from disruptions.

5.3. Findings under the Risk-Averse Setting

The data for the analysis of the risk-averse setting is also provided by the firm. The value of $\alpha = 3\%$ is derived from the firm's premise to provide an on-time delivery performance at 97%. While this delivery performance is not promised to each customer individually in the contractual agreements, the firm advertises this 97% on-time delivery performance to attract business customers. The senior administration is keen on complying with this on-time delivery premise, and tracks the performance in various regions. The number of tolerated late deliveries provides the value of β in each region. Our analysis examines the optimal capacity decisions under various values of β.

Table 5 provides the results pertaining to the optimal capacity investment decisions under increasing risk aversion with the unit capacity expansion cost set at $c_K = 1$ and the firm's preparedness at $T = 1$. We use the total expected cost of the risk-neutral setting as the benchmark cost, and show how the expected total cost increases under risk aversion.

		Risk						β					
	FC ID	Neutral	1000	900	800	700	600	500	400	300	200	100	0
Group 1	269	0	0	0	0	0	0	0	0	0	0	0	0
	497	0	0	0	0	0	0	0	0	0	0	0	0
	697	0	0	0	0	0	0	0	0	0	13	1101	2194
	397	0	0	0	0	0	0	0	0	0	0	0	0
Group 2	983	99	0	0	0	0	0	0	0	0	0	0	0
	297	1,232	843	674	504	334	165	0	0	0	0	0	0
	281	92	480	650	819	989	1,158	1,328	1,497	1,667	1,836	1,962	1,962
	948	0	0	0	0	0	0	0	0	0	0	0	483
	469	281	281	281	281	281	281	281	281	281	281	113	0
	292	334	334	334	334	334	334	334	334	334	334	501	1,144
Group 3	198	1,443	2,675	2,909	3,144	3,153	3,153	3,153	3,153	3,153	3,153	3,153	3,153
	268	0	0	0	0	0	0	0	0	0	0	0	0
	398	376	0	0	0	290	590	891	1,192	1,493	1,794	2,095	2,396
	368	0	0	0	0	0	0	0	0	0	0	0	0
Group 4	868	344	344	344	344	344	344	344	344	344	344	344	344
	598	23	23	23	23	23	23	23	23	23	23	23	23
ΣK_i^*		4,224	4,979	5,213	5,448	5,746	6,047	6,353	6,823	7,294	7,778	9,291	11,698
Expected Total Cost		100%	104%	105%	107%	109%	112%	114%	117%	120%	122%	131%	148%

Table 5. The impact of risk aversion (with decreasing values of β) on the optimal capacity decisions when $c_K = 1$ and $T = 1$.

We make three observations from this analysis pertaining to risk aversion. First, the firm generally invests in a higher total capacity (designated with $\sum K_i^*$ in Table 5) with higher degrees of risk aversion (designated with lower values of β) as a consequence of the preference to add capacity in the distant backup FC. This is an outcome of high-impact disruptions that can influence both FCs in close proximity. In Group 1, the firm does not make any additional investment in the risk-neutral setting and at lower degrees of risk aversion. However, as the degree of risk aversion increases to $\beta \leq 200$, the firm begins to increase capacity in FC 697, which serves as the distant backup facility to other FCs that are in close proximity. The result here demonstrates that it is beneficial to invest in additional capacity in the distant FC rather than nearby facilities. We observe no additional capacity investment in Group 4 when compared with the risk-neutral setting. The additional capacity investment in the risk-neutral setting is sufficient to satisfy the risk constraints (6), (13), and (21) at all degrees of risk aversion in Group 4, including the extreme risk aversion scenario of $\beta = 0$.

Our second finding relates to the rather surprising result of decreasing capacity investment with higher risk aversion. As pointed out in Proposition 7(a), the total capacity investment can decrease under risk aversion. We observe this phenomenon in our numerical results in Group 2 where the optimal capacity investment initially decreases when compared to the risk-neutral setting. In this group, the cost saving from the additional investment in nearby FCs justify the capacity expansion in the distant backup FC, yielding a decrease in the total capacity investment. Figure 3 demonstrates this surprising result by showing the optimal total capacity investment in Group 2 under the risk-neutral and risk-averse settings. When the firm switches from the risk neutral setting (with $\beta = \infty$) to the risk averse setting with $\beta = 1000$, the total capacity investment in Group 2 decreases. While substitution continues between FCs 297 and 281, the total capacity investment in Group 2 remains constant until $\beta = 500$. The total capacity investment in Group 2 begins to increase as the degree of risk aversion increases with $\beta \leq 400$.

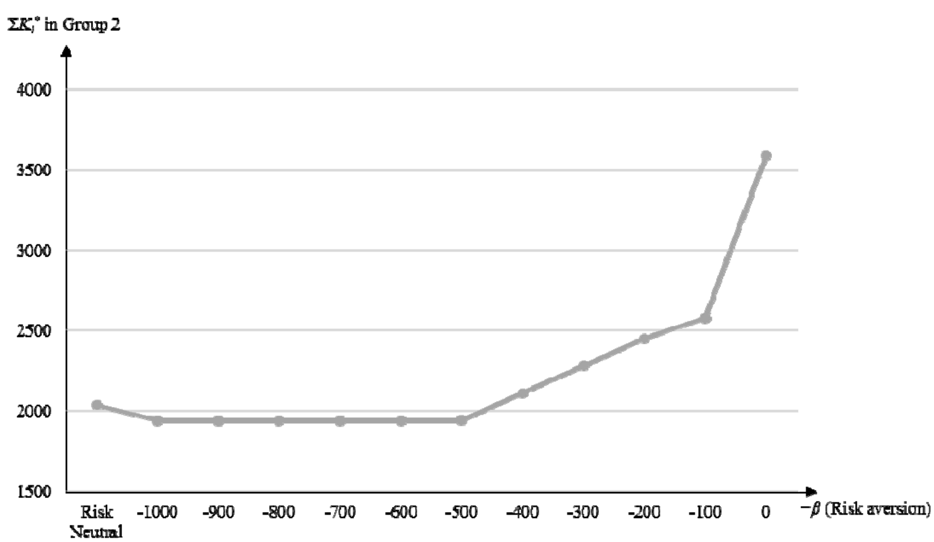


Figure 3. The impact of risk aversion on the optimal capacity expansion decisions of Group 2.

Before proceeding with the results pertaining to REI scores, it is important to highlight that the firm cannot satisfy its delivery performance constraint without capacity expansion under risk aversion. Using the firm's preference in the tolerable number of late deliveries as 800 in each group, the firm cannot satisfy the risk constraint without capacity expansion in groups 2 and 3; the realized number of late deliveries when $K_i = 0$ for all i are 1,335 and 2,140 in groups 2 and 3, respectively. Thus, we do not report on the total expected cost under $K_i = 0$ for all i as the problem becomes infeasible under risk aversion. Considering that the firm's winning criterion is on-time delivery within 24 hours, the firm would not be capable of delivering its promise to potential customers without additional capacity at its FCs in these two regions.

Third, FCs with the highest REI scores continue get limited capacity expansion, while FCs with the lowest REI score receive the highest additional capacity investment under risk aversion. Using the firm's tolerated number of late deliveries to be equal to $\beta = 800$, we observe that the facilities with the highest REI scores under no capacity expansion (FCs 983, 268, 398) receive no capacity expansion. On the other hand, facilities that are tagged with the lowest REI scores receive additional capacity investment. Once again, FC 198 with the REI score of only 1.40 is recommended for the highest level optimal capacity expansion, receiving an additional 3,144 units of daily delivery capability. Similarly, FCs 281, 292, and 868, receive a positive capacity expansion under risk aversion.

Optimal capacity planning continues to lead to significant disruption risk mitigation in delivery supply chains under risk aversion. Table 6 presents REI scores for each facility when the firm does not invest in capacity i.e., $K_i = 0$ under the risk-neutral setting and when the firm makes optimal capacity investments under risk-aversion, i.e., $K_i = K_i^A$ for all i. We continue to use the total expected cost of FC 983 with the highest risk exposure to be the benchmark for all REI scores in Table 6. Under risk aversion, the REI scores for facilities that had the highest risk exposure prior to capacity expansion (FCs 983, 268, and 398) continue to have significant reductions in their risk exposure: 40.92, 8.01, and 14.32, respectively. Thus, the facility with the highest risk exposure reduces its REI score by 59%. Thus, the adjustment in REI scores in each facility accounts for both capacity adjustment and risk aversion. It is important to highlight that incorporating the risk constraint into the problem leads to a smaller number of late deliveries, but it also increases the expected total cost due to a higher commitment in capacity expansion. As a consequence, FC 983, for example, has a higher REI score (equal to 40.92) than its REI score under the risk-neutral setting with optimal capacity expansion (equal to 37.35). However, the REI score of 40.92 under the optimal capacity planning in the risk-averse setting continues to represent a significantly smaller risk exposure for this facility when compared with the present level of capacity even in the absence of the risk constraint.

FC ID	983	268	398	948	368	297	469	598	269	868	292	281	497	397	198	697
REI ($K_i = 0$)	100.00	72.67	59.54	30.33	22.64	22.16	17.38	15.73	7.59	3.76	2.96	2.93	2.58	2.44	1.40	0.03
REI ($K_i = K_i^A$)	40.92	8.01	14.32	6.67	4.70	20.48	8.71	4.90	7.60	7.56	6.12	12.18	2.58	2.44	42.58	0.03

Table 6. The comparison of REI scores at each FC using the optimal capacity expansion decisions under risk-averse setting with no capacity expansion under the risk-neutral setting.

Finally, we report the cost savings from our capacity planning model under risk aversion. We compare the total expected cost between the two supply chain designs with no capacity investment, denoted $\Psi(0)$, and with the optimal capacity expansion levels with $K = K_i^A$ for all i, denoted $\Psi(K^A)$ under risk aversion. The cost savings from capacity expansion is denoted $\Delta\Psi_{\text{capacity\&RA}}$, and is equal to

$$\Delta\Psi_{\text{capacity\&RA}} = (\Psi(0) - \Psi(K^A))/\ \Psi(0) = 48\%$$

in the firm motivating our study. In conclusion, despite the additional cost capacity investment costs due to risk aversion, the firm is projected to have a 48% savings in its total operating costs stemming from disruptions without sacrificing its on-time delivery performance.

Earlier, Lemma 2 provided the feasibility conditions to satisfy the governing risk constraint. While not presented here, the firm's level of preparedness, designated with T, also plays a role in satisfying the risk constraint. As T decreases, the firm cannot always satisfy the risk constraint at higher degrees of risk aversion. Thus, when $T < 1$, there exists a threshold of β where the governing risk constraint (21) cannot be satisfied. For example, when $T = 0.75$ under the same cost parameters, the problem becomes infeasible as the firm cannot satisfy the risk constraint in Group 2 for $\beta \leq 300$ and in Group 3 for $\beta \leq 500$. As the firm's preparedness level decreases with lower values of T, the threshold for β increases implying that the problem becomes infeasible even at lower degrees of risk aversion.

6. Conclusions and Managerial Insights

We have examined a firm's capacity expansion decisions in a delivery supply chain in order to mitigate the negative effects of disruptions. Our work helps the firm comply with guidelines established through the ISO 22301 and the FEMA Business Continuity Guidance that are essential for operations in the US. The firm's risk preference is modeled with a chance constraint in the presence of disruption length uncertainty.

Our work makes five main contributions. First, this study uses capacity planning as a proactive measure against supply chain disruptions. Unlike inventory planning which may help in overcoming some production failures in a manufacturing chain, capacity planning brings agility and flexibility to delivery supply chains. Therefore, it serves as buffer against disruptions.

Second, our work captures different disruption characteristics by incorporating (1) low-impact and high-likelihood disruptions, and (2) high-impact and low-likelihood disruptions. This provides a more comprehensive analysis of supply chain disruptions, and more importantly, it leads to findings that cannot be obtained in the absence of high-impact disruptions.

Third, we show that geographic proximity does not necessarily serve as an anchor when determining the location of capacity expansion. This is an important result because it would motivate establishing an omni-channel backup system for a firm operating multiple channels that are not linked to each other.

Fourth, our work shows that the firm's total capacity expansion may decrease under risk aversion when compared to the risk-neutral setting. After an initial decrease, the total capacity expansion may remain the same at moderate levels of risk aversion that is followed by an increase at higher degrees of risk aversion. Our work identifies that this type of non-monotone behavior stems from two reasons: (1) Incorporating high-impact disruptions into the analysis, and (2) the flexibility of the distant facility to serve multiple nearby facilities.

Fifth, our numerical analysis recommends that capacity expansion should take place at facilities that are exposed to lower risks. In the absence of high-impact disruptions, one might intuit that the facilities with higher risk exposure deserve the attention for capacity expansion. However, our numerical analysis shows that the delivery supply chain performance is improved with lower total expected cost with capacity investment decisions made at the facilities that feature lower risk exposure index scores.

Our numerical analysis demonstrates that our proposed capacity expansion model leads to a 53% reduction in the total expected cost in the risk-neutral setting. Under risk aversion, our proposed capacity expansion results in a 48% savings in the expected total operational costs stemming from disruptions. By increasing capacity in the fulfillment centers, the firm is not simply reducing its expected costs; rather, the risk exposure at each facility decreases significantly, leading to a more resilient supply chain under disruption risk.

There are multiple managerial insights from our analytical investigation with numerical illustrations. High-impact disruption events are often neglected in supply chain risk models due to the lower frequency of occurrences. Our work, however, shows that, when high-impact disruptions are included in the analysis, the capacity investment takes place not in nearby facilities, but rather, at distant facilities. Considering the fulfillment centers in the supply chain of our motivating firm, there are more fulfillment centers in metropolitan areas that are in close proximity (primarily located in the East Coast of the US). It is important to note that facilities in metropolitan areas are exposed to higher risks with break-in, theft, and bomb threats, and they are often in closer distance of chemical and nuclear facilities to provide energy to densely populated regions. Thus, an analysis that identifies the high risk exposure would initially make managers increase capacities at these metropolitan nearby facilities. This is particularly the norm under an investigation that focuses only on low-impact and high-frequency disruption events, and one that ignores the influence of high-impact and low-frequency disruptions. Our numerical analysis, however, demonstrates that capacity expansion should take place at facilities that are not in metropolitan areas, but rather at facilities that are located in areas of smaller risk exposure. This conclusion becomes more prevalent under increasing degrees of risk aversion as the amount of capacity investment increases in fulfillment centers with the lowest risk exposure index scores. Our findings have significant implications for our motivating firm as our contingency backup recommendations based on an extended version of our stylized model are presently being implemented at the firm.

References

Acimovic, J., S.C. Graves. 2015. Making better fulfillment decisions on the fly in an online retail environment. *Manufacturing & Service Operations Management* **17**(1), 34-51.

Ang, E., D.A. Iancu, R. Swinney. 2016. Disruption risk and optimal sourcing in multitier supply networks. Forthcoming at *Management Science*.

Atan, Z., L.V. Snyder. 2012. Disruptions in one-warehouse multiple-retailer systems. Working Paper.

Berger, P.D., A. Gerstenfeld, A.Z. Zeng. 2004. How many suppliers are best? A decision-analysis approach. *Omega* **32**, 9-15.

Berger, P.D., A.Z. Zeng. 2006. Single versus multiple sourcing in the presence of risks. *The Journal of the Operational Research Society* **57**(3), 250-261.

Chen, A.I., S.C. Graves. 2014. Sparsity-constrained transportation problem. Working Paper.

DeCroix, G.A. 2013. Inventory management for an assembly system subject to supply disruptions. *Management Science* **59**(9), 2079-2092.

Dong, L., S.Y. Tang, B. Tomlin. 2015. Production chain disruptions: Inventory, preparedness, and insurance. Working Paper.

Dong, L., B. Tomlin. 2012. Managing disruption risk: The interplay between operations and insurance. *Management Science* **58**(10), 1898-1915.

Gurler, U., M. Parlar. 1997. An inventory problem with two randomly available suppliers. *Operations Research* **45**(6), 904-918.

Meena, P.L., S.P. Sarmah, A. Sarkar. 2011. Sourcing decisions under risks of catastrophic event disruptions. *Transportation Research Part E* **46**, 1058-1074.

Qi, L. 2013. A continuous-review inventory model with random disruptions at the primary supplier. *European Journal of Operational Research* **225**(1), 59-74.

Qi, L., Z.J.M. Shen, L.V. Snyder. 2009. A continuous-review inventory model with disruptions at both supplier and retailer. *Production and Operations Management* **18**(5), 516-532.

Ruiz-Torres, A.J., F. Mahmoodi. 2007. The optimal number of suppliers considering the costs of individual supplier failures. *Omega* **35**, 104-115.

Schmidt, W., M. Mehrotra. 2016. The impact of inaccurate disruption duration estimates. *Working Paper*. Cornell University.

Serpa, J.C., H. Krishnan. 2016. The strategic role of business insurance. Forthcoming at *Management Science*.

Simchi-Levi, D., W. Schmidt, Y. Wei. 2014. From superstorms to factory fires: Managing unpredictable supply chain disruptions. *Harvard Business Review* **92**(1-2), 96-101.

Simchi-Levi, D., W. Schmidt, Y. Wei, P.Y. Zhang, K. Combs, Y. Ge, O. Gusikhin, M. Sanders, D. Zhang. 2015. Identifying risks and mitigating disruptions in the automotive supply chain. *Interfaces* **45**(5), 375-390.

Tang, C.S. 2006. Robust strategies for mitigating supply chain disruptions. *International Journal of Logistics* **9**(1), 33-45.

Tang, S.Y., H. Gurnani, D. Gupta. 2014. Managing disruptions in decentralized supply chains with endogenous supply process reliability. *Production and Operations Management* **23**(7), 1198-1211.

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

Tomlin, B., Y. Wang. 2005. On the value of mix flexibility and dual sourcing in unreliable newsvendor networks. *Manufacturing and Service Operations Management* **7**(1), 37-57.

Xia, Y., M.H. Yang, B. Golany, S. M. Gilbert, G. Yu. 2004. Real-time disruption management in a two-stage production and inventory system. *IIE Transactions* **36**(2), 111-125.

Yan, X., K. Liu. 2009. An inventory system with two suppliers and default risk. *Operations Research Letters* **37**, 322-326.

Yang, Z., G. Aydin, V. Babich, D. Beil. 2009. Supply disruptions, asymmetric information, and a backup production option. *Management Science* **55**(2), 192-209.

Yang, Z., G. Aydin, V. Babich, D. Beil. 2012. Using a dual-sourcing option in the presence of asymmetric information about supplier reliability: Competition vs. diversification. *Manufacturing & Service Operations Management* **14**(2), 202-217.

Appendix

Proof of Proposition 1. (a) $\partial\psi_{L,1}(B_{2,1}, B_{3,1} \mid K_2, K_3)/\partial B_{2,1} \geq 0$ and $\partial\psi_{L,1}(B_{2,1}, B_{3,1} \mid K_2, K_3)/\partial B_{3,1} \geq 0$ when $(D_1 - B_{2,1} - B_{3,1})^+ = 0$. However, when $(D_1 - B_{2,1} - B_{3,1})^+ > 0$, we have $\partial\psi_{L,1}(B_{2,1}, B_{3,1} \mid K_2, K_3)/\partial B_{2,1} < \partial\psi_{L,1}(B_{2,1}, B_{3,1} \mid K_2, K_3)/\partial B_{3,1} < 0$ following from $c_{12} < c_3 < c_V$. Thus, the firm first uses $B_{2,1}$, and then $B_{3,1}$ until $(D_1 - B_{2,1} - B_{3,1})^+ = 0$ or until the constraints (4) and (5) become binding.

(b) $\partial\psi_{L,2}(B_{1,2}, B_{3,2} \mid K_1, K_3)/\partial B_{1,2} \geq 0$ and $\partial\psi_{L,2}(B_{1,2}, B_{3,2} \mid K_1, K_3)/\partial B_{3,2} \geq 0$ when $(D_2 - B_{1,2} - B_{3,2})^+ = 0$. However, when $(D_2 - B_{1,2} - B_{3,2})^+ > 0$, we have $\partial\psi_{L,2}(B_{1,2}, B_{3,2} \mid K_1, K_3)/\partial B_{1,2} < \partial\psi_{L,2}(B_{1,2}, B_{3,2} \mid K_1, K_3)/\partial B_{3,2} < 0$ following from $c_{12} < c_3 < c_V$. Thus, the firm first uses $B_{1,2}$, and then $B_{3,2}$ until $(D_2 - B_{1,2} - B_{3,2})^+ = 0$ or until the constraints (11) and (12) become binding.

(c) $\partial\psi_H(B_{3,1}, B_{3,2} \mid K_3)/\partial B_{3,1} \geq 0$ and $\partial\psi_H(B_{3,1}, B_{3,2} \mid K_3)/\partial B_{3,2} \geq 0$ when $(D_1 + D_2 - B_{3,1} - B_{3,2})^+ = 0$. However, when $(D_1 + D_2 - B_{3,1} - B_{3,2})^+ > 0$, we have $\partial\psi_H(B_{3,1}, B_{3,2} \mid K_3)/\partial B_{3,1} = \partial\psi_H(B_{3,1}, B_{3,2} \mid K_3)/\partial B_{3,2} < 0$ following from $c_3 = c_{3,1} = c_{3,2} < c_V$. Thus, the firm uses $B_{3,1}$ and $B_{3,2}$ indifferently until $(D_1 + D_2 - B_{3,1} - B_{3,2})^+ = 0$ or until the constraints (18), (19), and (20) become binding. ❑

Proof of Proposition 2. The first-order conditions are developed using the optimal stage-2 decisions given in Proposition 1. $\partial\Psi(K_1, K_2, K_3)/\partial K_1 = c_K$ follows from Proposition 1(b)(i). $\partial\Psi(K_1, K_2, K_3)/\partial K_1 = c_K - N_S$ follows from Proposition 1(b)(ii)(1). $\partial\Psi(K_1, K_2, K_3)/\partial K_1 = c_K - N_L$ follows from Proposition 1(b)(ii)(2).

$\partial\Psi(K_1, K_2, K_3)/\partial K_2 = c_K$ follows from Proposition 1(a)(i). $\partial\Psi(K_1, K_2, K_3)/\partial K_2 = c_K - N_S$ follows from Proposition 1(a)(ii)(1). $\partial\Psi(K_1, K_2, K_3)/\partial K_2 = c_K - N_L$ follows from Proposition 1(a)(ii)(2).

$\partial\Psi(K_1, K_2, K_3)/\partial K_3 = c_K$ follows from Proposition 1(c)(i). $\partial\Psi(K_1, K_2, K_3)/\partial K_3 = c_K - F_S$ follows from propositions 1(c)(ii), 1(a)(ii)(1) and 1(b)(ii)(1). $\partial\Psi(K_1, K_2, K_3)/\partial K_3 = c_K - F_M$ follows from one of the following two combinations: (1) Propositions 1(a)(ii)(1) and 1(b)(ii)(2); or (2) propositions 1(a)(ii)(2) and 1(b)(ii)(1). $\partial\Psi(K_1, K_2, K_3)/\partial K_3 = c_K - F_L$ follows from propositions 1(a)(ii)(2) and 1(b)(ii)(2). ❑

Proof of Lemma 1. It is trivial that $N_S < N_L$ and $F_S < F_M < F_L$. Furthermore, $N_L < F_L$ and $N_S < F_M$ following from $c_3 - c_{12} < c_V - c_3$. Note that $N_S \lozenge F_S$ if and only if $N_L \lozenge F_M$ where $\lozenge \in \{>, =, <\}$. Thus, if $N_S \geq F_S$, then we have $F_S \leq N_S < F_M \leq N_L < F_L$ as presented in part (a). Otherwise, we have $N_S < \{F_S, N_L\} < F_M < F_L$ as presented in parts (b) and (c). ❑

Proof of Proposition 3. (a) $\{\partial\Psi(K_1, K_2, K_3)/\partial K_1, \partial\Psi(K_1, K_2, K_3)/\partial K_2, \partial\Psi(K_1, K_2, K_3)/\partial K_3\} > 0$. Thus, $(K_1^N, K_2^N, K_3^N) = (0, 0, 0)$.

(b) Following from Proposition 2, for $K_3 \leq \min\{D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_2\}$, we have $\partial\Psi(K_1, K_2, K_3)/\partial K_3 = c_K - F_L \leq 0$. Thus, $K_3^N = \min\{D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_2\}$ whereas $(K_1^N, K_2^N) = (0, 0)$.

(c) Since we have $c_K \leq N_L < F_L$ in this condition, the optimal decision depends on whether one unit of K_3 with a net benefit of $F_L - c_K$ is more beneficial than that of one unit of K_1 and K_2 together with a net benefit of $2(N_L - c_K)$. Let us first consider the case when $F_L + c_K > 2N_L$ as in parts (i) and (ii), i.e., the firm prefers one unit of K_3 over two units of K_1 and K_2.

(c)(i) Following from Proposition 2, for $K_3 \leq \min\{D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_2\}$, we have $\partial\Psi(K_1, K_2, K_3)/\partial K_3 = c_K - F_L < 0$. Thus, $K_3^N = D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1$ since $K_1^0 > K_2^0$. Similarly, following from Proposition 2, for $K_2 \leq D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_3^N = K_1^0 - K_2^0$, we have we have $\partial\Psi(K_1, K_2, K_3)/\partial K_2 = c_K - N_L \leq 0$. Thus, $K_2^N = K_1^0 - K_2^0$. Consequently, $K_1^N = 0$.

(c)(ii) Symmetric to part (c)(i) with $K_1^0 \leq K_2^0$.

(c)(iii) We now consider the case when $F_L + c_K \leq 2N_L$, i.e., the firm prefers two units of K_1 and K_2 over one unit of K_3. Following from Proposition 2, for $K_1 \leq D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_3^N = D_1 + D_2 + D_3 - K_3^0 - K_1^0$, we have $\partial\Psi(K_1, K_2, K_3)/\partial K_1 = c_K - N_L \leq 0$. Similarly, for $K_2 \leq D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_3^N = D_1 + D_2 + D_3 - K_3^0 - K_2^0$, we have $\partial\Psi(K_1, K_2, K_3)/\partial K_2 = c_K - N_L \leq 0$. Thus, $(K_1^N, K_2^N) = (D_1 + D_2 + D_3 - K_3^0 - K_1^0, D_1 + D_2 + D_3 - K_3^0 - K_2^0)$ whereas $K_3^N = 0$.

(d) First, note that $F_L + c_K > 2N_L$ is always satisfied due to the ranking given in this part. Furthermore, we have $N_L < F_M$ in this condition. This implies that the firm prefers one unit of K_3 over one unit of K_1 or K_2. Following from $c_K \leq F_M < F_L$ and Proposition 2, for $K_3 \leq \max\{D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_2\}$, we have $\partial\Psi(K_1, K_2, K_3)/\partial K_3 \in \{c_K - F_M, c_K - F_L\} \leq 0$. Thus, $K_3^N = \max\{D_1 + D_2 + D_3 - K_3^0 - K_1^0 - K_1, D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_2\}$ whereas $(K_1^N, K_2^N) = (0, 0)$.

(e) First, note that $F_L + c_K > 2N_L$ is always satisfied due to the ranking given in this part. Furthermore, we have $N_L < F_M$ in this condition. This implies that the firm prefers one unit of K_3 over one unit of K_1 or K_2. Following from $c_K \leq F_S < F_M < F_L$ and Proposition 2, for $K_3 \leq D_1 + D_2 + D_3 - K_3^0$, we have $\partial\Psi(K_1, K_2, K_3)/\partial K_3 \in \{c_K - F_S, c_K - F_M, c_K - F_L\} \leq 0$. Thus, $K_3^N = D_1 + D_2 + D_3 - K_3^0$ whereas $(K_1^N, K_2^N) = (0, 0)$.

(f) We have $F_M \leq N_L$ in this condition. This implies that the firm prefers one unit of K_1 or K_2 over one unit of K_3. Following from $c_K \leq N_S < N_L$ and Proposition 2, for $K_1 \leq D_1 + D_2 - K_1^0$, we have $\partial\Psi(K_1, K_2, K_3)/\partial K_1 \in \{c_K - N_S, c_K - N_L\} \leq 0$. Similarly, for $K_2 \leq D_1 + D_2 - K_2^0$, we have $\partial\Psi(K_1, K_2, K_3)/\partial K_2 \in \{c_K - N_S, c_K - N_L\} \leq 0$. Thus, $(K_1^N, K_2^N) = (D_1 + D_2 - K_1^0, D_1 + D_2 - K_2^0)$ whereas $K_3^N = 0$.

(g) Following from $c_K \leq \{N_S, N_L, F_S, F_M, F_L\}$ and Proposition 2, we have $\partial\Psi(K_1, K_2, K_3)/\partial K_1 \leq 0$ for $K_1 \leq D_1 + D_2 - K_1^0$; $\partial\Psi(K_1, K_2, K_3)/\partial K_2 \leq 0$ for $K_2 \leq D_1 + D_2 - K_2^0$; and $\partial\Psi(K_1, K_2, K_3)/\partial K_3 \leq 0$ for $K_1 \leq D_1 + D_2 + D_3 - K_3^0$. Thus, $(K_1^N, K_2^N, K_3^N) = (D_1 + D_2 - K_1^0, D_1 + D_2 - K_2^0, D_1 + D_2 + D_3 - K_3^0)$. ❑

Proof of Lemma 2.

(a) When $K_3 = D_1 + D_2 + D_3 - K_3^0$, we have $B_{2,1}^* + B_{3,1}^* = D_1$ following from Proposition 1. As a consequence, Equation (6) becomes $P[D_1(1 - T)\tilde{t}_L > \beta] \leq \alpha$. This is satisfied if and only if $D_1(1 - T)t_{L,1-\alpha}/\beta \leq 1$ holds.

(b) When $K_3 = D_1 + D_2 + D_3 - K_3^0$, we have $B_{1,2}^* + B_{3,2}^* = D_2$ following from Proposition 1. As a consequence, Equation (13) becomes $P[D_2(1 - T)\tilde{t}_L > \beta] \leq \alpha$. This is satisfied if and only if $D_2(1 - T)t_{L,1-\alpha}/\beta \leq 1$ holds.

(c) When $K_3 = D_1 + D_2 + D_3 - K_3^0$, we have $B_{3,1}^* + B_{3,2}^* = D_1 + D_2$ following from Proposition 1. As a consequence, Equation (21) becomes $P[(D_1 + D_2)(1 - T)\tilde{t}_H > \beta] \leq \alpha$. This is satisfied if and only if $(D_1 + D_2)(1 - T)t_{H,1-\alpha}/\beta \leq 1$ holds. ❑

Proof of Remark 1. When $N_S < N_L < c_K \leq F_S < F_M < F_L$ or $N_S < c_K \leq \{N_L, F_S\} < F_M < F_L$ holds, it follows from Proposition 3(e) that $K_3^N = D_1 + D_2 + D_3 - K_3^0$. Similarly, when $c_K \leq \{N_S, N_L, F_S, F_M, F_L\}$ holds, it follows from Proposition 3(g) that $K_3^N = D_1 + D_2 + D_3 - K_3^0$. The rest follows from the proof of Lemma 2. ❑

Proof of Proposition 4. (a)(i) When $(K_2^N, K_3^N) = (0, 0)$, we have $B_{2,1}^* = K_2^0 - D_2$ and $B_{3,1}^* = K_3^0 - D_3$ following from Proposition 1. As a consequence, Equation (6) is violated if $P[[(K_2^0 - D_2 + K_3^0 - D_3)(1 - T) + (D_1 + D_2 + D_3 - K_3^0 - K_2^0)]\tilde{t}_L > \beta] > \alpha$. This is equivalent to $[D_1 - (K_2^0 - D_2 + K_3^0 - D_3)T]t_{L,1-\alpha}/\beta > 1$.

(a)(ii) When $(K_2^N, K_3^N) = (0, D_1 + D_2 + D_3 - K_3^0 - K_1^0)$ and $K_1^0 > K_2^0$, we have $B_{2,1}^* = K_2^0 - D_2$ and $B_{3,1}^* = D_1 + D_2 - K_1^0$ following from Proposition 1. As a consequence, Equation (6) is violated if $P[[(D_1 - K_1^0 + K_2^0)(1 - T) + (K_1^0 - K_2^0)]\tilde{t}_L > \beta] > \alpha$. This is equivalent to $[D_1(1 - T) + (K_1^0 - K_2^0)T]t_{L,1-\alpha}/\beta > 1$.

(b) This is symmetric to Part (a).

(c)(i) When $K_3^N = 0$, we have $B_{3,1}^* + B_{3,2}^* = K_3^0 - D_3$ following from Proposition 1. As a consequence, Equation (21) is violated if $P[[(K_3^0 - D_3)(1 - T) + (D_1 + D_2 + D_3 - K_3^0)]\tilde{t}_H > \beta] > \alpha$. This is equivalent to $[D_1 + D_2 - (K_3^0 - D_3)T]t_{H,1-\alpha}/\beta > 1$.

(c)(ii) When $K_3^N = D_1 + D_2 + D_3 - K_3^0 - K_1^0$, we have $B_{3,1}^* + B_{3,2}^* = D_1 + D_2 - K_1^0$ following from Proposition 1. As a consequence, Equation (21) is violated if $P[[(D_1 + D_2 - K_1^0)(1 - T) + K_1^0]\tilde{t}_H > \beta] > \alpha$. This is equivalent to $[(D_1 + D_2)(1 - T) + K_1^0T]t_{H,1-\alpha}/\beta > 1$.

(c)(iii) This is symmetric to part (c)(ii). ❑

Proof of Proposition 5. (a) Violating the risk constraint (6) implies that $D_1 - B_{2,1}^* - B_{3,1}^* > 0$. Thus, following from Proposition 1, we have $B_{2,1}^* = K_2^0 + K_2 - D_2$ and $B_{3,1}^* = K_3^0 + K_3 - D_3$. In order to comply with the risk constraint (6), (K_2^A, K_3^A) must satisfy that

$$P[[(K_2^0 + K_2^A - D_2 + K_3^0 + K_3^A - D_3)(1 - T) + (D_1 + D_2 + D_3 - K_3^0 - K_2^0 - K_3^A - K_2^A)]\tilde{t}_L > \beta] \leq \alpha$$

which is equivalent to $K_2^A + K_3^A \geq K^{R,1} = [1/T][D_1 - \beta/t_{L,1-\alpha}] - (K_2^0 - D_2 + K_3^0 - D_3)$.

(b) This is symmetric to part (a).

(c) Violating the risk constraint (21) implies that $D_1 + D_2 - B_{3,1}^* - B_{3,2}^* > 0$. Thus, following from Proposition 1, we have $B_{3,1}^* + B_{3,2}^* = K_3^0 + K_3 - D_3$. In order to comply with the risk constraint (21), K_3^A must satisfy that $P[[(K_3^0 + K_3^A - D_3)(1 - T) + (D_1 + D_2 + D_3 - K_3^0 - K_3^A)]\,\tilde{t}_H > \beta] \leq \alpha$ which is equivalent to $K_3^A \geq K^{R,3} = [1/T][D_1 + D_2 - \beta/t_{H,1-\alpha}] - (K_3^0 - D_3)$. ❑

Proof of Lemma 3. The proof follows immediately from $t_{H,1-\alpha} > t_{L,1-\alpha}$ due to the first-order stochastic dominance. ❑

Proof of Proposition 6. In parts (e) and (g), $(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N)$ following from 1. In parts (a)(i), (b)(i), (c)(i)(1), (c)(ii)(1), (c)(iii)(1), (d)(i), and (f)(i), $(K_1^A, K_2^A, K_3^A) = (K_1^N, K_2^N, K_3^N)$ following from Proposition 4(c) and Lemma 3.

In parts (a)(ii) and (b)(ii), $K_3^A = K^{R,3} > K_3^N$ following from propositions 4(c), 5(c) and Lemma 3. Furthermore, $(K_1^A, K_2^A) = (K_1^N, K_2^N) = (0, 0)$ since $\{N_S, N_L\} < c_K$.

In part (c)(i)(2), $K_3^A = K^{R,3} > K_3^N$ following propositions 4(c), 5(c) and Lemma 3. However, $K_2^A < K_2^N$ because increasing K_3 by one unit shifts $\partial\Psi(K_1, K_2, K_3)/\partial K_2$ from $c_K - N_L \leq 0$ to $c_K - N_S > 0$ for one unit of K_2 (see Proposition 2). Thus, in part (c)(i)(2.1), K_2^A is such that $K_2^A + K_3^A = K_2^N + K_3^N$. In part (c)(i)(2.2), however, K_2^A reaches zero, and thus, $K_2^A + K_3^A > K_2^N + K_3^N$. Note that $K_1^A = K_1^N = 0$. Part (c)(ii)(2) is symmetric to part (c)(i)(2).

In part (c)(iii)(2), $K_3^A = K^{R,3} > K_3^N$ following propositions 4(c), 5(c) and Lemma 3. However, $K_1^A < K_1^N$ and $K_2^A < K_2^N$ because increasing K_3 by one unit shifts $\partial\Psi(K_1, K_2, K_3)/\partial K_1$ from $c_K - N_L \leq 0$ to $c_K - N_S > 0$ for one unit of K_1, and shifts $\partial\Psi(K_1, K_2, K_3)/\partial K_2$ from $c_K - N_L \leq 0$ to $c_K - N_S > 0$ for one unit of K_2 (see Proposition 2). Thus, in part (c)(iii)(2.1), K_1^A and K_2^A are such that $K_1^A + K_2^A + K_3^A < K_1^N + K_2^N + K_3^N$. In parts (c)(iii)(2.2) and (c)(iii)(2.3), either K_1^A or K_2^A reaches zero. In part (c)(iii)(2.4), both K_1^A and K_2^A reach zero.

In part (d)(ii), $K_3^A = K^{R,3} > K_3^N$ following propositions 4(c), 5(c) and Lemma 3. Furthermore, $(K_1^A, K_2^A) = (K_1^N, K_2^N) = (0, 0)$ due to $N_L < F_M$.

In part (f)(ii), $K_3^A = K^{R,3} > K_3^N$ following propositions 4(c), 5(c) and Lemma 3. Furthermore, $(K_1^A, K_2^A) = (K_1^N, K_2^N) = (D_1 + D_2 - K_1^0, D_1 + D_2 - K_2^0)$ due to $c_K \leq N_S < N_L$. ❑

Proof of Proposition 7. Note that we have $N_L < F_L$ and $N_S < F_M$ following from $c_3 - c_{12} < c_V - c_3$. Furthermore, assuming $F_M \leq N_L$ implies that $F_S \leq N_S$. Thus, we have $F_S \leq N_S < F_M \leq N_L < F_L$. This order further implies that $N_S < 2N_L - F_L \leq N_L$.

Risk aversion increases by decreasing α and/or β. Decreasing α and/or β lead to an increase in $K_3^A = K^{R,3}$. When we have $N_S < c_K \leq N_L$, increasing K_3 by one unit shifts $\partial\Psi(K_1, K_2, K_3)/\partial K_1$ from $c_K - N_L \leq 0$ to $c_K - N_S > 0$ for one unit of K_1, and shifts $\partial\Psi(K_1, K_2, K_3)/\partial K_2$ from $c_K - N_L \leq 0$ to $c_K - N_S > 0$ for one unit of

K_2 (see Proposition 2). Thus, as risk aversion increases, the firm is better off decreasing each of K_1^A and K_2^A in the same amount as the increase in K_3^A.

(a) Note that $F_L + c_K \leq 2N_L$ implies that both $K_1^N > 0$ and $K_2^N > 0$ because a total of two units of investment in capacity at FC1 and FC2 has a greater net benefit than that of one unit of investment in capacity at FC3, i.e., $2(N_L - c_K) \geq F_L - c_K$. Following from the logic above, increasing risk aversion leads to an increase in K_3 such that $K_3^N < K_3^A$ and a decrease in both K_1 and K_2 such that $K_1^N > K_1^A$ and $K_2^N > K_2^A$ where $K_1^N - K_1^A = K_2^N - K_2^A = K_3^A - K_3^N$. This further implies that $K_1^N + K_2^N + K_3^N > K_1^A + K_2^A + K_3^A$.

(b) Note that $F_L + c_K > 2N_L$ and $F_M \leq N_L$ imply that either $K_1^N > 0$ or $K_2^N > 0$. Following from the logic above, increasing risk aversion leads to an increase in K_3 such that $K_3^N < K_3^A$ and a decrease in either K_1 or K_2 such that $K_1^N > K_1^A$ or $K_2^N > K_2^A$ where either $K_1^N - K_1^A = K_3^A - K_3^N$ or $K_2^N - K_2^A = K_3^A - K_3^N$. This further implies that $K_1^N + K_2^N + K_3^N = K_1^A + K_2^A + K_3^A$. ❑