Service at Risk in Delivery Operations

Mert Hakan Hekimoğlu

hekimm@rpi.edu

Lally School of Management

Rensselaer Polytechnic Institute

Troy, NY 12180

John H. Park

john.park3@pepperdine.edu

Seaver College

Pepperdine University

Malibu, CA 90263

Burak Kazaz

bkazaz@syr.edu

Whitman School of Management

Syracuse University

Syracuse, NY 13244

May 25, 2022

Forthcoming in *Decision Sciences*.

Service at Risk in Delivery Operations

This paper examines disruption risks at fulfillment centers and develops risk mitigation strategies based on inventory stocking and delivery decisions. It considers a Fortune 150 firm whose delivery operations are designed to fulfill the orders from contracted business customers within the next day. The firm promises its customers that the probability of late deliveries exceeding a certain threshold will be limited. We coin this requirement as the Service-at-Risk (SaR) constraint. The firm proactively determines the inventory amount to be kept in each fulfillment center. If a disruption occurs, the firm determines the best way to deliver orders from its operational fulfillment centers and vendors under disruption length and demand uncertainty to minimize additional costs and satisfy the SaR constraint.

This paper makes four main contributions. First, we find a surprising result that total inventory commitment can decrease with risk aversion when there exists a disruption possibility that impacts two nearby facilities together. Using actual data from the motivating firm, the numerical analysis demonstrates that this phenomenon exists in practice. Second, we define a new metric: The Risk Dispersion Index (RDI), which measures the dispersion in risk exposure across fulfillment centers. It leads to a lower and more balanced risk exposure in the firm's delivery operations. Third, we find that a facility may elect to abandon its own customers to serve the customers of a disrupted facility; this behavior becomes more prominent under risk aversion. Fourth, the introduction of demand uncertainty leads to a smaller inventory commitment for a risk-neutral retailer.

Keywords: *disruption, service at risk, risk dispersion index, delivery operations*

1. Introduction

This paper examines disruption risks and develops risk mitigation strategies in order to sustain delivery operations. The study is motivated by a risk assessment project conducted at a Fortune 150 company, an online retailer focusing on business customers which represent the firm's largest market segment in revenues. Successfully designing and managing delivery operations has become an important *winning criterion* in today's highly competitive online retail industry. As a result, firms need to not only optimize their normal day-to-day operations, but also prepare proactively for disruptions interfering with their normal operations. In developing and executing such contingency plans, firms keep business and individual customers' expectations of timely delivery at the core of their operations. If a firm is unable to deliver customer orders in a timely manner, it experiences an elevated level of customer dissatisfaction. Thus, businesses are pushed to set and advertise ambitious delivery deadlines in order to maintain their competitive edge and win customer orders. These delivery goals must be maintained even in the event of a disruption. While the traditional supply chain theory suggests minimizing the cost in developing and executing contingency plans in response to disruptive events, online retailers are aware that making deliveries on time is just as important, if not more, in order to keep their customers satisfied and/or avoid penalties for late deliveries. Conversations with the executives at our motivating firm have helped us identify a new risk measure that limits the probability of late deliveries.

Our paper conceptualizes this modern perspective by combining the traditional cost minimization approach with a novel "Service-at-Risk" (SaR) concept representing a firm's desire to avoid late deliveries. We capture this intention with a chance constraint where the probability of late deliveries exceeding a threshold is limited with a tolerable probability. It is important to mention that our SaR concept differs from the traditional value-at-risk (VaR) and conditional value-at-risk (CVaR) approaches. While VaR and CVaR are designed to limit monetary losses, our SaR constraint specifically focuses on the on-time delivery performance, making it an operational constraint. This is a novel representation of service operations performance under disruption risk that is critical for firms, including our motivating firm, that use on-time delivery as a winning criterion. The development of the SaR risk measure is an outcome of our interactions with the motivating firm's executives.

Inventory is commonly used as a buffer against disruptions. In manufacturing networks, for example, inventory decisions are primarily driven by disruption risks. Buffer inventory is kept around the nodes that are more susceptible to disruptive events to sustain continuity of manufacturing operations that feature predecessor-successor relationships between nodes. Delivery networks, on the other hand, do not possess a predecessor-successor structure between nodes, making the placement of excess inventory across the network less intuitive. In a delivery network, excess inventory kept at a node increases the firm's capability of serving as a backup to the other nodes in the event of a disruption. Our paper addresses two important decisions regarding inventories: Where and how much inventory should be kept in a delivery network so that the firm can keep its goal of delivering goods in a timely manner.

Our study combines proactive and reactive risk mitigation strategies to counter disruption risks in delivery operations and continue to serve the firm's contracted customers. Prior to the disruption, the firm considers building a sufficient level of inventory at each fulfillment center (FC) as a precautionary and proactive risk mitigation technique. The stocking quantity decisions resemble building capability to be able to fulfill orders at the time of a disruption at other facilities. If there is no additional inventory at a nearby FC, then diverting delivery decisions to that facility would not be beneficial. Thus, stocking quantity decisions at each FC constitute the *proactive risk mitigation* approach. The reactive perspective of our study determines the best way to deliver customer orders from the operational FCs and other vendors in the event of a disruption at an FC, while complying with its SaR constraint. Specifically, the firm solves a transportation problem under disruption length and demand uncertainty with the goal of minimizing the total transportation cost while complying with the SaR constraint. Thus, the firm's optimal shipment decisions represent a *reactive risk mitigation* approach in dealing with the disruption risk; shipments from operational FCs are only viable when there are excess inventories.

When the firm does not have sufficient inventory in nearby FCs, the orders can be filled by vendors. However, the shipments from vendors are almost always late, impairing the firm's delivery performance and winning criterion. Thus, vendor shipments decrease the chances of satisfying the SaR constraint.

We formulate the problem using a two-stage stochastic program. The firm determines the optimal levels of inventory in each FC in stage 1. After observing the disruption, corresponding to stage 2 of our model, the firm determines how to best satisfy the orders arriving at the disrupted FC. In stage 2, our model considers the length of disruption as random and that the firm complies with a SaR constraint that limits the probability of the number of late deliveries exceeding a threshold by a tolerable probability. Initially, our model ignores demand uncertainty in order to isolate the influence of disruption risks.[1] The structural properties developed under deterministic demand serves as a building block to the analysis with demand uncertainty. Demand uncertainty is incorporated into the model in Section 5.5 where we consider two different time epochs for its realization. We adjust the SaR constraint and the model accordingly and show that our main results continue to hold under demand uncertainty.

The analysis integrates a comprehensive set of potential disruptions including broad-impact events which can affect multiple facilities, and narrow-impact events affecting a single facility. The data for these disruptions are provided by the firm (a string of eight years' information) and collected from national sources.

This paper makes four main contributions. First, our study shows that the total amount of excess inventory committed as a result of disruption risk is not necessarily increasing with risk aversion; rather, the total excess stock can be decreasing with risk aversion. This is an unexpected result. One would intuit that the total inventory level should be nondecreasing, and even further monotonically increasing in risk aversion because excess inventory increases the capability of the firm in making on-time deliveries and the chances of satisfying the SaR constraint. However, our analysis describes several conditions under which the total inventory level can decrease in a delivery network at higher degrees of risk aversion. We show that this result primarily stems from the following two factors: (1) The SaR constraint causes a shift in the decision maker's attention from more likely narrow-impact disruptions to less likely broad-impact disruptions; and (2) adding excess inventory at one FC reduces the marginal benefit of excess inventory kept at other FCs in a delivery network. Therefore, this result is a direct consequence of the application of our novel SaR approach to a delivery network. Our study also demonstrates this phenomenon numerically in our motivating firm's network highlighting that the conditions needed for this result are sufficiently

---

[1] The setting with deterministic demand is consistent with the firm's operating environment as well as the focus of its executives. The average coefficient of variation for random disruption length (from data) is 2.23 whereas the average coefficient of variation for random demand (combined from all FCs) is 0.13.

general and practically applicable. Our numerical illustrations demonstrate that this finding leads to a minimum of 11% savings in inventory at the firm motivating our study.

Second, we introduce a new metric called Risk Dispersion Index (RDI) that evaluates the dispersion in risk exposure across facilities in the network. RDI complements the Risk Exposure Index (REI) introduced in Simchi-Levi et al. (2014). The REI metric is originally developed for a manufacturing setting with the aim of identifying the individual risk level at each node in the presence of predecessor relationships. Our new RDI metric, on the other hand, focuses on the system-wide risk in a delivery network. It is based on the mean absolute deviation in REI scores in a given network infrastructure:

RDI of a network = Σ |REI score at each FC – Average REI score| / [number of FCs].

Lower RDI scores imply that the network has a more balanced risk profile among its facilities and it is much less vulnerable to potential disruptions. Thus, RDI can be understood as extending the existing literature by introducing a dispersion perspective in supply chain network design under disruption risk. Our interactions with the firm executives reveal that companies would benefit by combining the REI scores for individual facilities with the RDI score for the entire system in designing resilient supply chains. Using RDI, we show that our proposed proactive and reactive risk mitigation strategies lead to more resilient supply chain operations with lower and balanced levels of risk exposure at the firm's FC network. Our numerical analysis in Section 6 demonstrates that our model leads to substantial improvements in comparison to the firm's current practice. Average risk exposure is decreased by 28.11% in the risk-neutral setting and by 23.44% in the risk-averse setting. Similarly, average dispersion in the risk exposure is decreased by 29.91% in the risk-neutral setting and by 25.52% in the risk-averse setting. The firm executives indicate that this is a substantial improvement in their risk mitigation efforts pertaining to disruptions in fulfillment center operations.

Third, the study shows that an FC can abandon its customers in order to serve the customers of a disrupted facility. We describe the proportion of late deliveries as a function of transportation distance. This surprising result occurs when the lateness proportion increases in distance in a convex manner. In this scenario, the closest facility serves the customers of a disrupted FC rather than its own customers, those customers that are not served by their region's operational facility are served by another nearby facility creating a chain of rerouting support. We coin this behavior as the abandonment policy and show that it becomes more prominent under risk aversion.

Fourth, we show that incorporating demand uncertainty diminishes the marginal benefit of inventory and reduces a risk-neutral retailer's initial inventory commitment. This finding is driven by the risks of overstocking and over-shipping stemming from the demand and supply mismatches. Despite the diminishing marginal benefit of inventory, our main contributions discussed above remain robust under demand uncertainty.

2. Research Context

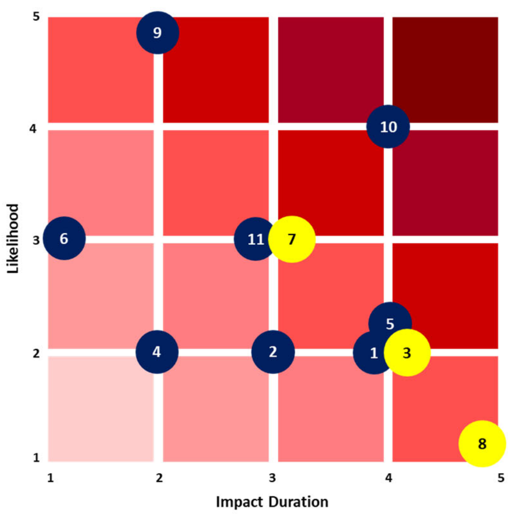
This section provides information regarding our motivating firm, a Fortune 150 online retailer whose business customers constitute the largest share in its revenues. As a result, the firm operates a dedicated supply chain to serve its contracted business customers. Figure 1 shows that the firm has 15 fulfillment centers (FC) located in the US, and these facilities are responsible for delivering orders within the next business day. Specifically, orders placed before 5:00pm in regional time are intended to be delivered the next day before 5:00pm. The firm utilizes its delivery performance (within one day) as its *winning criterion* in competition with other retailers. Specifically, the firm promises to its contracted customers that the likelihood of its late deliveries exceeding a certain nationwide threshold will not be greater than an advertised probability. We coin the firm's probability restriction as the "Service-at-Risk" (SaR) constraint.

Figure 1: Illustration of the firm's delivery supply chain.

The firm is already relying on deliveries that can be made from other operational FCs in the event of disruptions. Data from an eight-year time interval reveals that the firm experiences an average of 252 disruptions per year. The firm has a target for percentage of late deliveries stemming from disruptions; this value is set to 3.4%. However, the percentage of late deliveries increased to 5.06% in the most recent year from 3.92% a year before. As the firm was drifting away from the goal, executives felt a need to reverse the trend by being more proactive in order to mitigate the disruption risk.

The data shows that the disruption costs come in three areas: Additional transportation cost due to filling the demand (of non-operational FCs) through operational FCs is 9.01%; the penalty cost from late deliveries is 30.35%; and the vendor shipments (which are always late and incur the penalty cost) is 60.64% of the total cost from disruptions. These numbers exemplify that there was not sufficient extra inventory in the system to minimize the late deliveries and reduce the associated penalties. Thus, the firm needs help in determining the adequate level of extra inventories in each location.

Figure 2 provides the heat map for a comprehensive set of disruptions at the motivating firm (details of this data are provided in Section 6). Using data provided by the firm (a string of eight years' information) and data collected from national sources, we classify disruptions as narrow-impact and broad-impact events. Bomb threat, break-in, fire, flood, gas leak, tornado, power outage, and weather are classified as narrow-impact disruptions. Earthquake, hurricane, and chemical and nuclear plant failures are broad-impact disruptions that can potentially affect multiple facilities.



ID	Risk	I	L
1	Bomb Threat	4	2
2	Break In	3	2
3	Earthquake	4	2
4	Fire	2	2
5	Flood	4	2
6	Gas Leak	1	3
7	Hurricane	3	3
8	Chemical & Nuclear Plant	5	1
9	Power Outage	2	5
10	Tornado	4	4
11	Weather	3	3

I = Impact Duration L = Likelihood

- Dot color reflects impact range:

Narrow Impact Broad Impact

Figure 2: Heat map for disruptions on a scale of 1-5 where 1 represents the lowest and 5 represents the highest likelihood/impact duration. Node color denotes the impact range as narrow or broad impact.

3. Literature Review

There are three streams of literature related to our work. The first stream is the operational mitigation literature that examines proactive risk mitigation strategies a firm can take prior to a disruption. The second stream is the operational contingency literature that focuses on how a firm should react amid a disruption to minimize loss. The third stream is the literature pertaining to quantification of the disruption risk in a supply chain.

There is a growing literature that examines inventory decisions in FCs of online retailers (Chen and Graves 2014, Acimovic and Graves 2015). However, these publications ignore disruption risks. Our study employs a proactive inventory planning approach at fulfillment centers in the presence of disruption risk. In our study, proactive risk mitigation decisions are coupled with reactive contingency transportation planning decisions. We also develop a metric to measure the risk exposure across a supply chain. Because our work emphasizes disruption risk, we next present how our study contributes to the most closely-related three streams of literature.

Operational Mitigation Literature: This stream primarily examines disruption risks in the context of sourcing flexibility. These studies determine how many and which suppliers to utilize as a proactive measure, and how to split orders among backup suppliers when a supplier becomes unavailable due to a disruption as a reactive measure. Anunpindi and Akella (1993), Swaminathan and Shanthikumar (1999), Tomlin (2006), Chopra et al. (2007), Tomlin (2009), and Yang et al. (2012) investigate dual sourcing settings, while Berger et al. (2004), Berger and Zeng (2006), Ruiz-Torres and Mahmoodi (2007), Dada et al. (2007), Federgruen and Yang (2008, 2011), Hu and Kostamis (2014), and Zhao and Freeman (2019) analyze the multi-supplier models. Ang et al. (2017) and Bimpikis et al. (2019) extend these analyses to a multi-tier supply network. Tang et al. (2014) incorporate endogenous supplier reliability into their problem setting.

Our study contributes to this literature in three dimensions. First, we emphasize the inventory quantity decision at each facility as a proactive measure. This is different from incorporating external suppliers into a network of facilities from a modeling perspective. When the inventory commitments are determined within the firm's network facilities, our model accounts for a combination of distances between facilities, lateness in deliveries and shipments from vendors (external suppliers). Second, our study introduces a novel operational risk measure: Service-at-risk (SaR) constraint. This operational risk measure is critical to the success of the firm motivating our study with its goal of maintaining high service levels and timeliness in delivery operations. Thus, SaR is another distinct feature of our model that differentiates our study from earlier publications. Third, our setting incorporates the possibility of broad-impact disruptions that can affect multiple facilities together (as opposed earlier publications, which often feature only narrow-impact disruptions). Berger et al. (2004) and Ruiz-Torres and Mahmoodi (2007) also consider the possibility of multiple facilities getting disrupted. However, their study employs a decision-tree approach in order to choose the optimal number of suppliers. Our paper differs from these two publications because it emphasizes proactive inventory decisions and it introduces the novel SaR measure in developing risk mitigation strategies.

Operational Contingency Literature: While our paper investigates a delivery network, operational contingency has been widely examined in manufacturing settings. Meyer et al. (1979) is the first to consider using inventory to mitigate disruption risk in a single-stage production-storage system. Earlier publications focus on the optimal inventory replenishment policies and Snyder et al. (2016) provides a comprehensive review of this literature. Parlar and Berkin (1991), Berk and Arreola-Risa (1994), Ross et al. (2008) and Qi et al. (2009) examine the single-stage continuous-review policies amid disruption risks.

Parler et al. (1995), Song and Zipkin (1996), Li et al. (2004) and Schmitt et al. (2010) extend these findings to the single-stage periodic-review policies. DeCroix (2013) examines the period-review policy in a multi-period, multi-echelon setting to find the optimal base-stock policy. Yang et al. (2009) investigate the role of backup production.

Mitigating disruption risks with inventory in delivery operations differs from that in a manufacturing setting. In manufacturing, suppliers and manufacturers have predecessor relationships. Earlier literature considers this dependency on each other. Inventory is accumulated in a manufacturing facility in order to sustain its own production activity if a supplier is disrupted. In delivery operations, however, FCs are independent entities with their own customers to serve and are stocked with mostly identical goods. Therefore, in our setting, additional inventory is kept at surrounding facilities in order to sustain operations at a disrupted facility. Thus, nodes in a delivery network create a complementary backup capability rather than dependency. As a result, the analytical models developed for manufacturing and delivery operations are notably different.

Given the differences in manufacturing and delivery operations, our study contributes to the operational contingency literature with new results in proactive and reactive measures. While earlier publications advocate for a higher level of inventory in order to sustain continuity in successor operations, our study is the first to show that the total inventory commitment can be decreasing under risk aversion. Our study also contributes with new reactive measures: We show that it might be beneficial for an FC to abandon serving its own customers in order to serve the customers of the disrupted facility.

Literature that quantifies risk exposure: The last contribution of our paper to the literature is that we introduce a new metric called the Risk Dispersion Index (RDI) that measures the dispersion in risk exposure across a supply chain. RDI complements the Risk Exposure Index (REI) developed by Simchi-Levi et al. (2014), which identifies vulnerability at each node of an automobile assembly network. The REI metric starts from calculating the time to recovery (TTR) for each node, which then translates to estimated losses during the recovery period. Managers can easily identify the high-risk nodes using the REI metric. However, the REI metric itself does not reveal any information about the system-wide risk in the supply chain, which motivates the need to develop a new metric while preserving the REI framework. Simchi-Levi et al. (2015) further develop the analysis of Simchi-Levi et al. (2014) by introducing the concept of time to survive (TTS) which is the maximum amount of time a node can fully function in the midst of a disruption. Gao et al. (2019) extend this study with a stochastic TTR and incorporate scenarios of random nodes being disrupted. While these publications focus on manufacturing and assembly operations, our study examines delivery operations for contracted customers with tight deadlines for fulfilling customer demand. In assembly settings, earlier publications conclude that risk can be mitigated by increasing the supply of parts at the potentially disrupted facility. Our study, on the other hand, concludes that inventory should be kept in surrounding facilities (rather than the potentially disrupted facility) to enable delivery operations and continue to satisfy customer demand. In addition to differences in the proposed risk mitigation strategies, we introduce an operational risk measure called SaR which does not exist in earlier publications. Moreover, our study compares our operational risk measure SaR with the well-known financial risk measures VaR and CVaR. It shows how these measures are similar under certain conditions and differ in other conditions.

Hendricks and Singhal (2003, 2005a, 2005b) show that disruptions have significant financial implications. Financial mitigation literature primarily considers insurance as a way to mitigate the financial consequences. Dong and Tomlin (2012) and Dong et al. (2018) demonstrate that financial mitigation in the form of insurance is beneficial in minimizing the total cost against disruptions in manufacturing settings. However, insurance has no positive impact on the performance of deliveries. Because the motivating firm's winning criterion involves on-time delivery performance, our study focuses on operational mitigation and operational contingency, and it ignores financial mitigation in the form of insurance. Building more resilient supply chains by using metrics such as RDI and REI to identify and strengthen the vulnerable areas should reduce the cost of insurance.

4. The Model

We formulate the firm's delivery planning problem under disruption risk using a two-stage stochastic program. Table 1 summarizes the notation used in our model. In stage 1, the firm determines the amount of additional inventory at each FC in order to minimize the sum of two costs: The cost of stocking additional inventory and the expected cost from executing a contingency plan over the next time period (e.g., one year). The inventory decisions in stage 1 are made under disruption risk that would halt operations at FCs. Therefore, stage 1 decisions can be perceived as building the firm's capability to service its customers in the event of a disruption. If one or more FCs are disrupted, the firm implements a contingency plan in stage 2 with the objective of minimizing the execution cost of the contingency plan. Specifically, when a disruption occurs, the firm determines how to distribute the available inventory in stage 2 limited by the decisions made in stage 1.

Business customers of the firm are clustered in regions (in the form of a collection of zip codes) indexed as $j = 1, 2, \ldots, J$. The daily customer demand in region j is denoted D_j. The firm motivating our problem serves primarily business customers whose demand is stable over time and the firm is confident in its ability to forecast the demand accurately. Including demand uncertainty does not alter our main findings, but the deterministic demand setting enables us to isolate the influence of disruption risk in delivery operations. The structural properties developed under deterministic demand serve as a building block to the analysis pertaining to the problem setting with demand uncertainty. In Section 5.5, we show that our main findings continue to hold under demand uncertainty.

We use subscript $i = 1, \ldots, J$ to describe FC i. When there is no disruption in the network, all customers in region j are served by a single FC i located in the same region, i.e., $i = j$. If a disruption occurs at the facility serving region j, the demand D_j is diverted to other operational FCs (i.e., $i \neq j$) and vendors. We use index $n = 1, \ldots, N$ to describe the network of operating facilities where $N = 2^J - 1$, representing all subsets of $\{1, \ldots, J\}$ excluding the network structure when all facilities are disrupted. For example, in a three FC network with three regions to serve, there is a total of 7 ($= 2^3 - 1$) different operating network combinations where $n = 1$ represents only FC 1 is operational, $n = 2$ represents only FC 2 is operational, $n = 3$ represents only FC 3 is operational, $n = 4$ represents FCs 1 and 2 are operational, $n = 5$ represents FCs 1 and 3 are operational, $n = 6$ represents FCs 2 and 3 are operational and $n = 7$ represents all three FCs are operational. In a given network n, we describe set of disrupted facilities with $\Lambda(n)$. In the above three-facility network structure, $\Lambda(n = 4) = \{3\}$ indicates that FC 3 is disrupted.

D_j	Daily customer demand in region j ($j = 1, 2, \ldots, J$)
$\Lambda(n)$	Set of disrupted facilities in a given network n ($n = 1, \ldots, N$)
p_{mn}	Probability of a disruption event m causing a network n over the next period
$\tilde{\tau}^{mn}$	Random time length of disruption in days
K_i	Amount of excess inventory at each FC i ($i = 1, \ldots, J$)
$\Psi_1(\vec{\mathbf{K}})$	Objective function in stage 1 for vector $\vec{\mathbf{K}} = (K_1, \ldots, K_J)$
$\Psi^*_{2,mn}(\vec{\mathbf{K}})$	Expected cost from executing the optimal contingency plans in stage 2
c_K	Unit cost of holding additional inventory over the next period (e.g., one year)
c_T	Unit transportation cost
c_L	Unit late delivery cost
c_V	Unit vendor shipment cost
$\bar{\mathbf{x}}^{mn}$	Vector of daily shipment decisions x_{ij}^{mn} under disruption event m in network n
d_{ij}	Distance from FC i to region j
$l(d_{ij})$	Proportion of late deliveries for a given d_{ij}
α, β	Parameters used in the Service-at-Risk constraint – total lateness exceeding the total tolerable lateness of β cannot be greater than the tolerated probability of $1 - \alpha$.

Table 1: Summary of notation used in the model.

Each disruption event (e.g., hurricane) is represented by index $m = 1, \ldots, M$. The probability of a disruption event m causing a network n over the next period (e.g., one year) is described with p_{mn}. Our representation of disruption event index m and network structure index n yields a total of M x N unique pairs of (m, n); this allows for complete generalization of our model. However, in practice, many (m, n) pairs would have a probability of zero, e.g., an earthquake disrupting an FC located in Minnesota where major earthquakes are not seen. This reduces the computational complexity significantly.

The time length of disruption is random and is expressed as $\tilde{\tau}^{mn}$ (e.g., in days). We make no assumptions regarding the distribution of $\tilde{\tau}^{mn}$ other than having a non-negative support, thus, our conclusions remain valid for any type of probability distribution.

In stage 1, the firm determines the amount of excess inventory at each FC i, denoted K_i, where $i = 1, \ldots, J$. The vector $\overline{\mathbf{K}} = (K_1, \ldots, K_J)$ represents the additional inventory decisions in stage 1 that can be used to continue the delivery service in stage 2. Without loss of generality, we assume the present level of inventory at FC j to be equal to the region's demand D_j; the case when each facility owns excess inventory in the beginning does not bring new insights. We relax this assumption in the numerical illustrations in Section 6. Unit cost of holding additional inventory over the next period (e.g., one year) is described with c_K. The objective function in stage 1, denoted $\Psi_1\left(\overline{\mathbf{K}}\right)$, can be expressed as follows:

$$\min_{\overline{\mathbf{K}} \geq \overline{0}} \Psi_1\left(\overline{\mathbf{K}}\right) = \sum_i c_K K_i + \sum_{m=1}^{M} \sum_{n=1}^{N} p_{mn} \Psi_{2,mn}^{*}\left(\overline{\mathbf{K}}\right) \tag{1}$$

The first term in (1) describes the total cost of adding inventory in FCs. The second term is the expected cost from executing the optimal contingency plans in stage 2 where $\Psi_{2,mn}^{*}\left(\overline{\mathbf{K}}\right)$ is the minimal cost for servicing the customers under disruption m in network n for the given set of stocking decisions $\overline{\mathbf{K}}$ made in stage 1.

In stage 2, given a disruption $m \in \{1, \ldots, M\}$ in network $n \in \{1, \ldots, N\}$, the firm determines the optimal amount of daily shipments to be made from operational FC i to region j denoted with x_{ij}^{mn} where $j = 1, \ldots, J$ and $i \notin \Lambda(n)$. The vector $\overline{\mathbf{x}}^{mn}$ represents the vector of daily shipment decisions x_{ij}^{mn} under disruption event m in network n.

In stage 2, the firm experiences three types of additional costs due to the contingency plans:

Transportation: This cost includes the costs of transportation diverted to operational FCs. Unit transportation cost c_T is multiplied by the distance from FC i to region j described by d_{ij}. Thus, each unit shipped from FC i to region j costs the firm $c_T d_{ij}$. Our model examines the additional costs stemming from disruptions; therefore, we assume $d_{ii} = 0$ for $i = 1, \ldots, J$, and $d_{ij} > 0$ for $i \neq j$ can be perceived as the additional distance of travel when customers in region j are served from FC i.

Late deliveries: The firm experiences late deliveries for some proportion of shipments made from FCs that are located in other regions. The lateness proportion is described with $l(d_{ij})$ (≤ 1), which is a function of the distance between the serving FC i and the customers in region j denoted with d_{ij}. The FC located in the same region with its customers is capable of delivering on time, therefore, we set $l(d_{ii} = 0) = 0$ for $i = 1, \ldots, J$. The proportion of late deliveries increases in the distance from FC i to customers in region j, i.e.,

$\partial l(d_{ij})/\partial d_{ij} \geq 0$. Each unit of late delivery costs the firm c_L, and therefore, $c_L l(d_{ij})$ can be perceived as the realized unit cost of late deliveries for shipments from FC i to customers in region j.

Vendor shipments: All unsatisfied demand from FCs are fulfilled through vendors. The firm pays the vendor for the shipment made from the vendor, denoted c_V. All deliveries made from the vendor miss the firm's promised 24-hour delivery window. Thus, the firm incurs a total cost of $(c_V + c_L)$ for each unit of demand fulfilled from the vendors. Since the vendors are not expected to make these deliveries on time, without loss of generality, we assume that vendors have inventory to eventually fulfill these orders.

The second-stage objective function minimizes the total expected cost stemming from transportation, late deliveries, and vendor shipments under an uncertain disruption duration:

$$\Psi^*_{2,mn}\left(\overline{\mathbf{K}}\right) = \min_{\bar{\mathbf{x}}^{mn} \geq \bar{0}} \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right) = \begin{bmatrix} c_T \sum_j \sum_i d_{ij} x_{ij}^{mn} + c_L \sum_j \sum_i l\left(d_{ij}\right) x_{ij}^{mn} \\ +\left(c_V + c_L\right) \sum_j \left(D_j - \sum_i x_{ij}^{mn} \right) \end{bmatrix} E\left[\tilde{\tau}^{mn}\right] \tag{2}$$

s.t.

$$\sum_j x_{ij}^{mn} \leq K_i + D_i \;\; \forall i \tag{3}$$

$$\sum_i x_{ij}^{mn} \leq D_j \;\; \forall j \tag{4}$$

$$P\left[\left(\sum_j \sum_i l\left(d_{ij}\right) x_{ij}^{mn} + \sum_j \left(D_j - \sum_i x_{ij}^{mn} \right) \right) \tilde{\tau}^{mn} > \beta \right] \leq 1 - \alpha \tag{5}$$

$$x_{ij}^{mn} = 0 \;\; \forall i \in \Lambda(n) \tag{6}$$

Constraint (3) represents the supply constraint for each FC i. Constraint (4) represents the demand constraint at each region j. Constraint (6) avoids shipments from the disrupted facilities in network n.

Inequality (5) describes the Service-at-Risk (SaR) constraint. According to SaR, the firm complies with a chance constraint in which the probability of total lateness (during disruption m in network n) exceeding the total tolerable lateness of β cannot be greater than the tolerated probability of $1 - \alpha$. The total lateness in SaR is influenced by two factors in our model: First, a proportion of the shipments made from an operating FC i to region j, denoted $l(d_{ij})$, arrive late. Second, shipments made from the vendor miss the company's promised 24-hour delivery window.

It is important to distinguish our model with a SaR constraint from the traditional modeling approaches that employ Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) constraints. While VaR and CVaR constraints are designed to avoid monetary losses, our SaR constraint specifically focuses on the on-time delivery performance, and thus, it is an operational constraint. This unique focus of SaR becomes apparent when it is compared with the objective function (2) and other financial risk measures like Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) risk measures which represent monetary values. From the objective function (2), one can see that a shipment arriving late from the vendor costs (c_V + c_L) whereas a shipment arriving late from an operational FC i costs ($c_T d_{ij}$ + c_L). Therefore, the financial implication of a late delivery differs depending on whether it is fulfilled from a vendor or an operational FC. The SaR constraint, on the other hand, treats all late deliveries the same way regardless of deliveries being made by the vendors or operational FCs, i.e., a late delivery is a late delivery no matter where the shipment comes from. Our SaR constraint is a novel representation of delivery performance in the presence of disruptions. This is a critical operational measure for firms, including our motivating firm, that use on-time delivery as a winning criterion. Because SaR constraint is an operational requirement, it is incorporated into the second-stage formulation guaranteeing performance at every disruption in each network configuration. Inserting a similar constraint in the first-stage formulation is redundant because the second-stage model guarantees that a first-stage SaR constraint would be satisfied. Without the SaR constraint in the second-stage of the model, the first-stage formulation can lead to potentially infeasible solutions; our modeling approach eliminates this possibility.

5. Analysis

The model in (1)–(6) is complex and its dimensions increase tremendously with a larger network, making it difficult to obtain critical results that illuminate managerially insightful findings. We begin our analysis with the risk-neutral setting in Section 5.1 where we ignore the impact of SaR. Incorporating the SaR constraint (5) later in sections 5.2 and 5.3 enables us to highlight the impact of risk aversion. In Section 5.4, we compare our operational risk measure SaR with the traditional financial risk measures VaR and CVaR. In Section 5.5, we incorporate demand uncertainty into our analysis.

5.1. Risk Neutral Analysis

The risk-neutral version of the second-stage problem, with the objective function (2) subject to constraints (3), (4) and (6), is linear in its decision variables. Thus, combined with vendors having unlimited supply, there exists a set of optimal x_{ij}^{mn} decisions in stage 2 for each disruption event m in network n. We next examine the properties of these optimal contingency decisions.

The lateness function $l(d_{ij})$ influences the optimal second-stage decisions. When lateness increases linearly in distance, i.e., $\partial^2 l(d_{ij})/\partial d_{ij}^2 = 0$, each FC always prioritizes serving its own customers before sending supplies to the customers of disrupted regions.

Proposition 1 (*Loyal Policy*). *When* $\partial^2 l(d_{ij})/\partial d_{ij}^2 = 0$, *the optimal amount of shipment is* $x_{ii}^{mn*} = D_i$ *for each FC* $i \notin \Lambda(n)$.

We refer to the solution proposed in Proposition 1 as the Loyal Policy (Policy L) of distribution activities. If the excess inventory in FCs that are in the vicinity is insufficient to fulfill the demand in region j, then the vendor must be utilized. The next proposition establishes a threshold distance, denoted d_H, that identifies the potential backup FCs for each region.

Proposition 2 (*Backup shipment threshold*). *Suppose* $\partial^2 l(d_{ij})/\partial d_{ij}^2 = 0$ *and FC j is disrupted, i.e.,* $j \in \Lambda(n)$.

(a) $x_{ij}^{mn*} = 0$ *for each FC* $i \notin \Gamma(j)$ *where* $\Gamma(j)$ *includes every FC i such that* $d_{ij} \leq d_H$ *where there exists a unique* d_H *which solves* $c_T d_H + c_L l(d_H) = c_V + c_L$.

(b) *Region j receives vendor shipments if* $D_j > \sum_{i \notin \Lambda(n), i \in \Gamma(j)} K_i$.

The set $\Gamma(j)$ defined in Proposition 2(a) describes the set of backup FCs that can serve the customers of region *j*; the shipment amount from an FC outside of $\Gamma(j)$ is equal to zero. Proposition 2(b) states that, if the total excess inventory at the operational backup FCs for region *j* is not sufficient to cover the region's entire demand, then region *j* receives vendor shipments.

Proposition 2 establishes a backup shipment threshold d_H in the presence of a lateness function that increases linearly in distance. However, when lateness follows a convex increasing behavior in distance, then the optimal second-stage decisions feature a different behavior. In this case, an operational FC can abandon its own customers in order to serve the customers in the disrupted region. We next elaborate on this distribution behavior.

To clearly demonstrate the two different types of response mechanisms under a disruption, it is sufficient to create a triangular scenario ($J = 3$). Suppose FC 1 is disrupted ($1 \in \Lambda(n)$), FCs 2 and 3 are operational ($\{2, 3\} \notin \Lambda(n)$), and moreover, the distance between FC 3 and FC 1 is the longest among the three FCs, i.e., $\{d_{21}, d_{32}\} < d_{31}$. We also utilize the triangular inequality where the total distance of traveling through an intermediate FC is always longer than traveling directly. This implies that the sum of distances travelled from FC 3 to FC 2 and from FC 2 to FC 1 is greater than the distance travelled directly from FC 3 to FC 1. Thus,

$$d_{32} + d_{21} - d_{31} > 0. \quad (7)$$

Proposition 3 (*Loyal vs. Abandonment Policy*). *Suppose FC* 1 *with demand* D_1 ($> K_2$) *is disrupted, i.e.,* $1 \in \Lambda(n)$, *and FCs* 2 *and* 3 *are operational, i.e.,* $\{2, 3\} \notin \Lambda(n)$ *where* $\{d_{21}, d_{32}\} < d_{31} \leq d_H$. *We define the following condition*

$$c_L[l(d_{31}) - l(d_{32}) - l(d_{21})] \leq c_T[d_{32} + d_{21} - d_{31}]. \quad (8)$$

(a) (*Loyal Policy*): *If* (8) *holds*, *operational FCs* 2 *and* 3 *prioritize serving their own customers and then use their excess inventory to directly serve the customers of FC* 1, *i.e.,* $x_{22}^{mn*} = D_2$, $x_{21}^{mn*} = K_2$, $x_{33}^{mn*} = D_3$, $x_{31}^{mn*} = \min\{D_1 - K_2, K_3\}$, $x_{32}^{mn*} = 0$, $x_{23}^{mn*} = 0$. *We denote these decisions by* $\vec{x}_L^{mn}$.

(b) (*Abandonment Policy*): *If* (8) *does not hold, FC* 2 *prefers to abandon its customers in region* 2 *in order to serve the customers of FC* 1 *while FC* 3 *serves the customers that FC* 2 *abandons, i.e.,* $x_{22}^{mn*} = (D_2 - K_3)^+$, $x_{21}^{mn*} = \min\{D_1, K_2 + \min\{D_2, K_3\}\}$, $x_{33}^{mn*} = D_3$, $x_{31}^{mn*} = \min\{(D_1 - K_2 - \min\{D_2, K_3\})^+, (K_3 - D_2)^+\}$, $x_{32}^{mn*} = \min\{D_2, K_3\}$, $x_{23}^{mn*} = 0$. *We denote these decisions by* $\vec{x}_A^{mn}$.

Proposition 3 shows that there are two characteristically different ways of utilizing inventory in the event of a disruption; validity of (8) determines which policy is optimal. The right-hand side of (8), which is always positive, represents the additional transportation cost due to the longer total distance traveled in Abandonment Policy (from FC 3 to FC 2 and from FC 2 to FC 1) compared to Loyal Policy (from FC 3 to FC 1). Left-hand side of (8) can fetch a positive or a negative value depending on the lateness function $l(d_{ij})$. A positive (negative) value for the left-hand side of (8) indicates that Abandonment Policy reduces (increases) the proportion of late deliveries, and hence, the late delivery cost compared to Loyal Policy. A lateness function increasing in distance in a linear manner always yields a negative value for the left-hand side of (8). In that case, the optimal policy is the Loyal Policy as presented in Proposition 3(a). Conversely, a convex lateness function (i.e., $\partial^2 l(d_{ij})/\partial d_{ij}^2 > 0$) can lead to a positive value for the left-hand side of (8) depending on its degree of convexity. If the lateness function is sufficiently convex, then the left-hand side of (8) becomes greater than the right-hand side, i.e., (8) does not hold. In this case, the Abandonment Policy as shown in Proposition 3(b) is optimal where there exists an operational FC i with $x_{ii}^{mn*} < D_i$ that is $i = 2$. In Abandonment Policy, FC 2 abandons its own customers and prioritizes serving the customers of the disrupted facility FC 1. In turn, the customers of FC 2 are served by the other operational FC 3. This type of chain rerouting becomes practical when utilizing a long transportation link, which leads to significantly higher proportions of late deliveries as compared to utilizing two shorter links due to the convexity in lateness function.

Let us briefly elaborate on the first-stage decisions under risk neutrality. The first-order derivative of (1) with respect to K_i can be represented as

$$\partial \Psi_1\left(\overline{\mathbf{K}}\right)/\partial K_i = c_K + \sum_{m=1}^{M}\sum_{n=1}^{N} p_{mn}\theta_i^{mn}\left(\overline{\mathbf{K}}\right) \tag{9}$$

where $\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$ is the dual price of the supply constraint (3) for FC i evaluated at $\overline{\mathbf{K}}$ under a given (m, n). The value of $\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$ is a non-decreasing step function in $\overline{\mathbf{K}}$ and is non-positive. $\left|\theta_i^{mn}\left(\overline{\mathbf{K}}\right)\right|$ represents the marginal benefit of additional inventory for a given (m, n).

$\partial \Psi_1\left(\overline{\mathbf{K}}\right)/\partial K_i$ fetches its maximum value of c_K (> 0) at high values of K_i for each i since $\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$ is non-decreasing. This ensures that the minimization problem in stage 1 is bounded and there exists an optimal solution. Proposition 4 formalizes this result.

Proposition 4 (*Optimality in stage* 1). *The risk-neutral problem described with* (1)*,* (2)*,* (3)*,* (4) *and* (6) *is bounded and there always exists an optimal solution.*

Now that we established the results for a risk neutral firm, we next examine the role of risk aversion through SaR as described in expression (5).

5.2. Risk Aversion

In the event of a disruption, Proportion 3 establishes the conditions for the Loyal Policy (Policy L) and the Abandonment Policy (Policy A). How are these two policies impacted by the inclusion of risk aversion? Our analysis in this section incorporates the SaR constraint in (5) where the firm is required to satisfy the probability of late deliveries exceeding a threshold β to be less than the tolerated probability of $1 - \alpha$. Note that, for each α probability, there exists a range of β values that satisfy the SaR constraint (5). The next proposition shows that risk aversion increases the likelihood of following the Abandonment Policy in stage 2.

Proposition 5 (*Risk aversion in stage* 2). *For a given first-stage decision vector* $\overline{\mathbf{K}}$,

(a) *If* $l(d_{31}) \le (>)\ l(d_{32}) + l(d_{21})$, *then Policy L* (*Policy A*) *provides a superior compliance with the SaR constraint in* (5), *i.e.*, *the left-hand side of* (5), *when* $\vec{x}^{mn} = \vec{x}_L^{mn}$ *is smaller* (*greater*) *than that when* $\vec{x}^{mn} = \vec{x}_A^{mn}$;

(b) *If Policy A is optimal in the risk-neutral setting, then it is also optimal in the risk-averse setting;*

(c) *If Policy L is optimal in the risk-averse setting, then it is also optimal in the risk-neutral setting;*

(d) *The probability that Policy A is optimal in the risk-averse setting is greater than that in the risk-neutral setting, i.e.,*

P[*Policy A is optimal*] *in the risk-averse setting* > *P*[*Policy A is optimal*] *in the risk-neutral setting.*

The above proposition highlights the fact that risk aversion, as incorporated through the SaR constraint, encourages the deployment of Policy A more than it would in the risk-neutral setting. Recall that Proposition 3 established the fact that Policy A can only occur in a risk-neutral setting when (8) does not hold. Propositions 5(a) and 5(d), however, show that Policy A can be optimal under risk aversion, as long as $l(d_{31}) > l(d_{32}) + l(d_{21})$ holds, even when (8) holds (see online supplement for details). Thus, one way to mitigate the risk of late deliveries is altering the delivery patterns with Policy A for given inventory levels.

Now that we established preferences in the second stage under risk aversion, we can now turn our attention to the first-stage analysis.

5.3. Impact of Risk Aversion on Optimal Stocking Decisions

Earlier studies examining the inventory decisions showed that incorporating risk aversion (under exogenous prices) leads to a higher quantity investment (see Eeckhoudt et al. 1995). This section presents a surprising finding that the stocking investment in stage 1 may decrease with the introduction of risk aversion through our SaR constraint. This is an unexpected result because excess inventory can increase the capability of the firm in making on-time deliveries, and specifically, it increases the chances of satisfying the SaR constraint in (5). Therefore, one would intuit that the stocking investment should be nondecreasing (and even further monotonically increasing) in risk aversion. However, our analysis shows that, under several conditions, the excess stocking decisions can go down at higher degrees of risk aversion. We later demonstrate that this finding can occur in practice through our numerical analysis presented in Section 6.

A scenario of three FCs ($J = 3$) is sufficient to highlight the driving factors of this surprising result regarding how excess inventory can have a decreasing behavior in risk aversion. Suppose FC 1 and FC 2 are two nearby facilities and there is a third facility FC 3 which is located away from FCs 1 and 2, i.e., $d_{12} < \{d_{13}, d_{23}\}$. For simplicity, we consider two kinds of disruptions. A narrow-impact disruption with a low range ($m = L$) can influence each facility individually leading to three different operating network structures where $n = 1$ represents FC 1 is disrupted, $n = 2$ represents FC 2 is disrupted and $n = 3$ represents FC 3 is disrupted. A broad-impact disruption with a high range ($m = H$) can halt the operations of the two nearby facilities FC 1 and FC 2 together; this network is represented by $n = 4$. Because FC 3 is located distant from FC 1 and FC 2, the broad-impact event would not halt operations at FC 3.

Consider the case when the optimal risk neutral solution, denoted (K_1^N, K_2^N, K_3^N), is such that

(C1): $K_1^N > 0, K_2^N > 0, K_3^N = 0$.

From a risk neutral perspective, Condition (C1) entails that the firm is economically better off by carrying excess inventory only at the two nearby facilities, FC 1 and FC 2. This would allow these two facilities to serve as backup for each other. Furthermore, both FC 1 and FC 2 can serve FC 3 as a backup facility in the event that FC 3 is disrupted. Condition (C1) also entails that there is no excess inventory kept at FC 3, thus, FC 3 cannot make any shipment to FC 1 or FC 2 in case either one of them or both are disrupted.

SaR constraint in (5) implies that the probability of the total of late deliveries exceeding a certain threshold (indicated by β) cannot be greater than the tolerable probability of $1 - \alpha$. In the probability distribution of the random disruption length, this corresponds to the right tale of the random variable $\tilde{\tau}^{mn}$ at α–percentile, denoted τ_α^{mn}. Thus, the firm is concerned with late deliveries accumulating when the random disruption length exceeds the value of τ_α^{mn}. We define the following condition:

$$\text{(C2): } (D_1 + D_2)\tau_\alpha^{H4} = \beta + \varepsilon > \beta \geq \{D_1\tau_\alpha^{L1}, D_2\tau_\alpha^{L2}, D_3\tau_\alpha^{L3}\} \quad (10)$$

where ε is a small positive infinitesimal quantity representing the switch from risk-neutral setting to risk-averse setting. Condition (C2) makes two statements. First, it guarantees that the SaR constraint (5) is never violated for $(m, n) \in \{(L, 1), (L, 2), (L, 3)\}$, representing the narrow-impact disruptions. Second, it indicates that, when risk aversion is incorporated into the decision making process, the risk-neutral solution defined in Condition (C1) marginally violates the SaR constraint (5) for $(m, n) = (H, 4)$, representing the broad-impact disruption halting operations at both FC 1 and FC 2. In this case, the only viable action to satisfy (5) for $(m, n) = (H, 4)$ is to increase inventory at the distant facility FC 3.

How does then the total stocking quantity in stage 1 decrease with risk aversion? One might intuit that the inventory increase in FC 3 would lead to an increase in the total stock in the network. However, the increase in the inventory of FC 3 can trigger a decrease in the inventories of FC 1 and FC 2 when the following condition holds:

$$\text{(C3): } |p_{L2}\theta_1^{L2}(K_1^N, K_2^N, K_3^A) + p_{L3}\theta_1^{L3}(K_1^N, K_2^N, K_3^A)| \leq c_K,$$
$$|p_{L1}\theta_2^{L1}(K_1^N, K_2^N, K_3^A) + p_{L3}\theta_2^{L3}(K_1^N, K_2^N, K_3^A)| \leq c_K$$

where K_3^A denotes the stocking decision in FC 3 satisfying (5) for $(m, n) = (H, 4)$. Condition (C3) states that, when the inventory in FC 3 is increased to K_3^A, the first-order derivatives of stocking decisions at FC 1 and FC 2 evaluated at their risk-neutral values ($K_1 = K_1^N$ and $K_2 = K_2^N$) become positive. This represents that the marginal benefits of the last unit of inventory added at FC 1 and FC 2 do not justify the cost of carrying them. Therefore, the firm is better off by decreasing the inventories in FC 1 and in FC 2 by K_3^A each compared to the risk-neutral values given in Condition (C1). It is worth mentioning that, without Condition (C3), the firm would be able to comply with the SaR constraint by increasing inventory at FC 3 without decreasing the excess inventories at FC 1 and FC 2. The following proposition formalizes this finding that total inventory in the network can decrease when risk aversion is introduced. Note that this finding does not depend on the type of lateness function.

Proposition 6 (*Decreasing quantity in risk aversion*). *Suppose* $l(\min\{d_{31}, d_{32}\}) < 1$. *Let* (C1), (C2), (C3) *hold and* ε *be a positive infinitesimal quantity. The risk averse solution becomes* $(K_1^*, K_2^*, K_3^*) = (K_1^N - K_3^A, K_2^N - K_3^A, K_3^A)$, *where* $K_3^A = \varepsilon/[(1 - l(\min\{d_{31}, d_{32}\}))\tau_\alpha^{H4}]$, *representing the solution where the total risk averse inventory decision is smaller than the total risk neutral inventory decision*.

We next provide examples for each condition to better illustrate when this counterintuitive result in Proposition 6 can be observed in practice. Condition (C1) would hold if the likelihood of a broad-impact disruption impacting two nearby FCs is low enough (a small value of p_{H4}) such that the firm does not find it economic to carry excess inventory at a distant FC in the risk-neutral setting. This condition indicates that, in the absence of a serious risk of broad-impact disruptions (e.g., hurricane), the excess inventory decisions are made primarily to respond to narrow-impact disruptions (e.g., fire). Thus, nearby FCs can serve as each other's sole backup due to their proximity, which is a common practice observed at many firms including our motivating company.

Condition (C2) would hold if the (random) total demand at FC 1 and FC 2 represents a high volume because of a broad-impact disruption lasting quite long has a significant probability (i.e., a large value of τ_α^{H4}). This condition conceptualizes the shift in the firm's planning perspective when the SaR constraint is introduced. The SaR constraint enforces the firm to be more cautious for broad-impact disruptions that affect multiple facilities because they can lead to a greater number of late deliveries. As a result of the SaR constraint, the firm puts more emphasis on potentially catastrophic disruptions in determining the excess inventory quantities. This reaction is in line with risk-averse behavior in practice. For example, from our conversations with the executives at our motivating firm, we have learned that they are concerned about the possibility of hurricanes taking down multiple facilities located in the Southern US. As a result, they have indicated that risks pertaining to such broad-impact events should not be neglected in developing risk mitigation efforts even though the likelihood of those events are not high. Condition (C2) simply represents a case in which the risk-neutral inventory decisions prioritizing narrow-impact disruptions are not feasible when the SaR constraint is incorporated into the problem. This condition is adjusted slightly when we introduce demand uncertainty into the model in Section 5.5, but the interpretation remains the same.

Condition (C3) stems from the nature of delivery networks. When excess inventory is added to one FC, it causes a decrease in the marginal benefit of excess inventory kept at other FCs in the network. Excess inventory at one FC acts as a substitute for inventories at other FCs. In our case, inventory is added to FC 3 due to the SaR constraint in order to be used in the event of a broad-impact disruption. However, that inventory can also be used in the event of a narrow-impact disruption occurring at either FC 1 or FC 2. As a result, when inventory is added in FC 3, the marginal benefit of excess inventory at FC 1 and FC 2 decreases. Condition (C3) simply represents a case when the reduced marginal benefits of excess inventory at FC 1 and FC 2 do not justify the unit cost of adding inventory c_K. It is easy to see that this condition is satisfied when the unit cost of inventory c_K is sufficiently large.

In conclusion, Proposition 6 shows that risk aversion can result in a reduction in the systemwide inventory compared with the risk-neural setting. The conditions needed for this result are sufficiently general and practically applicable as discussed above. Our numerical analysis in Section 6 demonstrates this phenomenon in the network of our motivating firm, which has 15 facilities. Thus, this finding is not specific to a three-facility network and exists in larger network infrastructures with more than three facilities.

5.4. Service-at-Risk vs. Value-at-Risk and Conditional Value-at-Risk

In this section, we examine the relationships between the operational risk measure SaR with the traditional financial risk measures value-at-risk (VaR) and conditional value-at-risk (CVaR). VaR is a risk measure prescribed by Basel II and III accords and is widely employed by financial institutions. CVaR, on the other hand, is a coherent risk measure and is the preferred approach by academics because of its tractability. How is SaR related with these two financial risk measures? Are the delivery optimization solutions developed under a SaR perspective feasible and admissible under the financial risk measures VaR and CVaR? We begin our discussion by comparing SaR with VaR.

5.4.1. SaR vs. VaR

It is established in earlier derivations that when an optimal risk-neutral solution is not feasible in a risk-averse setting, increasing excess inventory enables the firm to comply with the SaR constraint in (5). In order to provide a fair comparison of the firm's operational risk measure SaR with the financial risk measure VaR, let us retain the same risk tolerance probability $(1 - \alpha)$. Defining β_V as the amount of loss tolerated under VaR, the corresponding VaR constraint can be written as:

$$P\left[\left(c_L \sum_j \sum_i l\left(d_{ij}\right) x_{ij}^{mn} + \left(c_L + c_V\right) \sum_j \left(D_j - \sum_i x_{ij}^{mn}\right)\right) \tilde{\tau}^{mn} > \beta_V\right] \le 1 - \alpha \,. \tag{11}$$

In (11), the probability expression has two cost terms. The first term $c_L \sum_j \sum_i l\left(d_{ij}\right) x_{ij}^{mn}$ is the cost stemming from internal late deliveries. The second term $\left(c_L + c_V\right) \sum_j \left(D_j - \sum_i x_{ij}^{mn}\right)$ is the cost stemming from vendor shipments (which are always late). These two cost terms are multiplied by the random disruption length $\tilde{\tau}^{mn}$ which leads to the total cost associated with a disruption event. The VaR constraint described in (11) ensures that the probability that the total cost from lateness and vendor deliveries exceeding the tolerable loss of β_V does not exceed the tolerated probability of $(1 - \alpha)$. For a given set of x_{ij}^{mn} values, let $\beta_V^{=}$ be the value that makes the VaR constraint in (11) satisfied at equality, i.e.,

$$P\left[\left(c_L \sum_j \sum_i l\left(d_{ij}\right) x_{ij}^{mn} + \left(c_L + c_V\right) \sum_j \left(D_j - \sum_i x_{ij}^{mn}\right)\right) \tilde{\tau}^{mn} > \beta_V^{=}\right] = 1 - \alpha \,. \tag{12}$$

The next proposition shows the conditions under which a delivery solution that satisfies the VaR constraint in (11) violates and satisfies the SaR constraint in (5).

Proposition 7. *For a given* $(1 - \alpha)$, *consider a set of* x_{ij}^{mn} *values that satisfy the* VaR *constraint in* (11).
(a) *If* $\beta \le (\beta_V^{=} /(c_L + c_V))$, *then the SaR constraint in* (5) *is not satisfied by the same set of* x_{ij}^{mn} *values; and,*
(b) *if* $\beta \ge (\beta_V^{=} /c_L)$, *then the SaR constraint in* (5) *is also satisfied by the same set of* x_{ij}^{mn} *values.*

Proposition 7 shows that a solution that satisfies the financial risk measure VaR does not always satisfy the operational risk measure SaR. It establishes the relationship between the firm's tolerance for lateness described by the value of β in the SaR constraint with the VaR constraint's tolerated monetary loss described with β_V. At the value of $\beta_V^{=}$ that makes the VaR constraint satisfied at equality, Proposition 7(a) shows that when the firm's tolerance for late deliveries is small, i.e., when $\beta < (\beta_V^{=} /(c_L + c_V)))$, the same set of delivery decisions described by x_{ij}^{mn} values that satisfy the VaR constraint in (11) lead to the violation of the SaR constraint in (5). However, when the firm's tolerance for late deliveries is large as described with $\beta \ge (\beta_V^{=}/c_L)$, then the same set of delivery decisions described by x_{ij}^{mn} values also satisfy the SaR constraint in (5). When tolerance for late deliveries is in intermediate values, i.e., $(\beta_V^{=} /(c_L + c_V)))$

$< \beta < (\beta_V^{=}/c_L)$, the SaR constraint in (5) may or may not be satisfied. In sum, the implication of Proposition 7(a) is the following: Even if the firm can limit the monetary loss by following the VaR risk measure, the high service level could be compromised and could lead to losses.

We next show that satisfying the firm's operational risk measure SaR does not always satisfy the financial risk measure VaR. For a given set of x_{ij}^{mn} values, let $\beta^{=}$ be the value that makes the SaR constraint in (5) in equality, i.e.,

$$P\left[\left(\sum_j \sum_i l\left(d_{ij}\right) x_{ij}^{mn} + \sum_j \left(D_j - \sum_i x_{ij}^{mn}\right)\right) \tilde{\tau}^{mn} > \beta^{=}\right] = 1 - \alpha \ . \tag{13}$$

The following corollary establishes the conditions in which a solution that satisfies the VaR constraint in (11) can violate the firm's tolerated lateness and the SaR constraint in (5).

Corollary 1. *For a given* $(1 - \alpha)$, *consider a set of* x_{ij}^{mn} *values that satisfy the SaR constraint in* (5) *at equality*.

(a) *If* $\beta_V \leq c_L \beta^{=}$, *then the VaR constraint in* (11) *is not satisfied by the same set of* x_{ij}^{mn} *values; and,*

(b) *if* $\beta_V \geq (c_L + c_V)\beta^{=}$, *then the VaR constraint in* (11) *is also satisfied by the same set of* x_{ij}^{mn} *values*.

Corollary 1(a) establishes that, when the firm's tolerated monetary loss described by β_V is small in comparison to the firm's tolerated lateness, then the delivery solution proposed by the SaR risk measure does not satisfy the VaR constraint. As expected, Corollary 1(b) indicates that both risk measures are satisfied when the firm's tolerated monetary loss is greater in value.

While Corollary 1(a) indicates that SaR may not always guarantee compliance from the VaR perspective, it is important to highlight that it warrants financial compliance in a CVaR risk measure perspective. This is described next in the comparison of SaR with CVaR.

5.4.2. SaR vs. CVaR

The objective function in (2) accounts for all potential losses incurred by costly operations stemming from late deliveries and deliveries made by vendors. This objective function is a special form of CVaR in which the tolerated loss (let us denote it with β_C) and the tolerated loss probability α are both set to zero. When $\beta_C = 0$, both our objective function and CVaR account for all losses, whether they constitute a large or a small monetary loss. When the value of β_C is a large positive value, CVaR counts only large losses. From this perspective, our objective function can be perceived as a more stringent operational risk measure as it does not want the firm to have any losses, small or large. More importantly, a solution proposed by SaR complies with CVaR – it can even be the optimal solution for some CVaR measures. This is established in the following proposition.

Proposition 8.

(a) *The objective function in* (2) *is equivalent to minimizing* CVaR *with* $\beta_C = 0$ *and* $\alpha = 0$. *Thus, any optimal delivery solution that complies with the SaR constraint in* (5) *is guaranteed to be optimal under this CVaR risk measure*;

(b) *The optimal solution developed under the SaR constraint in* (5) *continues to be optimal under a* CVaR *risk measure for a non-negative tolerated loss amount of* $\beta_C \geq 0$ *and with the tolerated loss probability of* $\alpha = 0$; *and*,

(c) *An optimal solution developed under the SaR constraint in* (5) *is always feasible under a* CVaR *risk measure with* $\beta_C \geq 0$ *and* $\alpha \geq 0$.

The implication of Proposition 8 is that any solution developed under a SaR risk measure complies with the financial risk measure CVaR. However, the opposite is not true. A solution developed under a CVaR perspective may not always comply with the operational risk measure SaR. Proposition 8(a) shows that the optimal delivery solution developed under the SaR constraint is also optimal under a CVaR risk measure with $\beta_C = 0$ *and* $\alpha = 0$ that accounts for all losses (big and small). As shown in Proposition 8(b), the same optimal delivery solution developed under the SaR risk measure is optimal under CVaR even if the firm's tolerated loss is positive, i.e., $\beta_C > 0$ when $\alpha = 0$. However, an optimal solution developed under SaR is not guaranteed to be the optimal solution under looser CVaR risk measures with $\beta_C > 0$ *and* $\alpha > 0$, despite being feasible solutions. As mentioned earlier, however, an optimal solution under a CVaR risk measure with $\beta_C > 0$ *and* $\alpha > 0$ is not guaranteed to be feasible under a SaR constraint. Thus, SaR guarantees operational feasibility with good financial solutions. These solutions can even be optimal financial solutions in risk-averse settings. Good financial solutions, however, do not guarantee feasible operational performance.

Our comparison of the operational risk measure SaR with the financial risk measures VaR and CVaR leads to the following conclusions. Even if the firm develops a solution using a financial perspective through a VaR or a CVaR perspective, it does not guarantee delivery performance described by the SaR risk measure. Solutions developed under a SaR risk measure, on the other hand, guarantee good financial performance under VaR and CVaR measures and are always feasible under a CVaR risk measure. And finally, solutions developed under a SaR constraint provide good financial performance even under a VaR perspective when the tolerated amount of late deliveries is sufficiently large as depicted in Proposition 7(b). These observations carry an important managerial implication: In the presence of disruption risks, a firm concerned about late deliveries might have to develop more geographically dispersed contingency strategies as compared to a firm concerned about its monetary losses. As a result, a firm using the SaR approach is likely to develop plans with greater reliance on the internal inventory within its own network.

Thus, we conclude that SaR can lead to lower operational costs stemming from late deliveries and deliveries made by vendors, leading to financially robust solutions.

We have completed analyzing a complex model featuring disruption length uncertainty and nonlinearity caused by the SaR constraint. Using deterministic demand, we have been able to develop the structural properties of the two-stage stochastic program where the first-stage inventory decisions serve as the proactive risk mitigation preparedness and the second-stage distribution decisions as the reactive measures when a disruption occurs. In the next section, we introduce demand uncertainty into this already complex problem. The structural properties developed under deterministic demand setting serve as a building block in characterizing the optimal solution under demand uncertainty.

5.5. Incorporating Demand Uncertainty

We incorporate demand uncertainty into our analysis with two different model settings. Both model settings examine the first-stage decisions that are made under demand uncertainty. In the first setting, random demand is revealed before the firm determines optimal delivery decisions. It is important to highlight that our paper investigates a problem where shipments are made directly from operational FCs to the customers of the non-operational FCs – this is known as the last-mile delivery operations. Because the stage 2 model solves a last-mile delivery problem, the realization of demand uncertainty before delivery decisions is required for the firm to be able to generate purchase order numbers, shipping invoices, and charge its customers. As a result, the first model setting is a common representation of demand uncertainty in managing the last-mile delivery operations for many online retailers including the firm motivating our study. In the second model setting, we consider demand as uncertain at the time of determining delivery decisions. The idea in this setting can be viewed as a futuristic extension of Amazon's anticipatory shipping (NPR, 2018), which normally pertains to the first-mile operations, to the last-mile deliveries. However, this setting requires the firm to possess the capabilities to create purchase order numbers, print labels, perform packaging, etc. in trucks while shipments are *en route* to the customer and before the firm sees the actual customer demand. We show that our main findings continue to hold under both the present and the futuristic settings of demand uncertainty.

5.5.1. Demand Uncertainty Revealed Before Delivery Decisions

We begin our discussion with the setting in which the firm makes shipment decisions to its customer after observing the customer orders. This corresponds to the setting where demand uncertainty is revealed before making delivery decisions. We show that (1) our main findings regarding risk mitigation strategies continue to hold under this demand uncertainty setting; and (2) demand uncertainty leads to a reduction in the risk-neutral firm's proactive inventory commitment. The latter finding is caused by a reduction in the marginal benefit from the inventory commitment in stage 1. The marginal benefit from inventory decreases because of the probability of overstocking.

The sequence of events is as follows: In stage 1, the firm determines the optimal amount of inventory commitment in the presence of disruption and demand uncertainty. When disruption occurs, random demand is revealed at each FC. The firm then determines the optimal distribution and transportation decisions under a SaR constraint.

Let $\tilde{D}_j$ describe the random demand variable with an expected value of $\bar{D}_j$ at a disrupted FC j defined on a support $[D_j^L, D_j^H]$ where $D_j^H > D_j^L > 0$. Let D_j^r represent the realization of the demand random variable.

It is important to observe that the value of demand random variable is revealed at the beginning of stage 2. Therefore, the model that solves the stage 2 problem is identical to the expressions in (2)–(6) except that the values of D_j are replaced with the realized values of demand D_j^r at each region j in (2), (4) and (5). In addition, the right-hand side of the supply constraint (3) is replaced with $K_i + \bar{D}_i$ for each FC i where K_i represents the excess inventory.

In the online supplement, Corollary A1 shows that our main finding where an FC can abandon its customers in order to serve the customers of a disrupted facility continues to hold under demand uncertainty. The result follows from the same condition expressed in (8).

We next present an important distinction between the two settings with demand uncertainty and deterministic demand. For a given (m, n), $\sum_{i \notin \Lambda(n)} K_i + \bar{D}_i$ represents the total amount of supply at the operational FCs. If the total realized demand $\sum_j D_j^r$ is greater than $\sum_{i \notin \Lambda(n)} K_i + \bar{D}_i$, the dual price of the supply constraint (3) remains to have the same structure as in the deterministic demand case, i.e., $\theta_i^{mn}(\overline{\mathbf{K}})$ is a non-decreasing, non-positive step function in $\overline{\mathbf{K}}$. However, if $\sum_j D_j^r$ is smaller than $\sum_{i \notin \Lambda(n)} K_i + \bar{D}_i$, then the dual price of the supply constraint (3) becomes zero for each FC i. Let $\Psi_1^{S1}(\overline{\mathbf{K}})$ represent the first-stage objective function under this stochastic demand setting; its first-order derivative with respect to K_i becomes

$$\partial \Psi_1^{S1}(\overline{\mathbf{K}}) / \partial K_i = c_K + \sum_{m=1}^{M} \sum_{n=1}^{N} \left(p_{mn} \theta_i^{mn}(\overline{\mathbf{K}}) \left(1 - G\left(\sum_{i \notin \Lambda(n)} K_i + \bar{D}_i \right) \right) \right) \tag{14}$$

where $G(.)$ is the cumulative distribution function of the random aggregate demand $\sum_j \tilde{D}_j$. We denote the dual price of the supply constraint (3) in this setting using $\lambda_i^{mn}(\overline{\mathbf{K}})$:

$$\lambda_i^{mn}\left(\overline{\mathbf{K}}\right)=\theta_i^{mn}\left(\overline{\mathbf{K}}\right)\left(1-G\left(\sum_{i\notin\Lambda(n)}K_i+\bar{D}_i\right)\right)\geq\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$$

due to the fact that $0 \leq 1 - G(\cdot) \leq 1$ and $\theta_i^{mn}\left(\overline{\mathbf{K}}\right) \leq 0$. Let $\sum_i K_i^{D*}$ describe the optimal amount of total inventory commitment in stage 1 under deterministic demand; similarly, let $\sum_i K_i^{S1*}$ represent the optimal amount of total inventory commitment in stage 1 under random demand where its value is realized before shipments to the customer. The following proposition shows two important results. First, it proves that the marginal benefit from additional inventory under demand uncertainty is smaller than the marginal benefit under deterministic demand for each FC *i*, i.e., $\left|\lambda_i^{mn}\left(\overline{\mathbf{K}}\right)\right| \leq \left|\theta_i^{mn}\left(\overline{\mathbf{K}}\right)\right|$. As a result, demand uncertainty (when revealed before shipment decisions are made to the customer) leads to a reduction in the optimal total excess inventory in comparison to the deterministic demand setting.

Proposition 9. (a) *The marginal benefit of additional inventory under demand uncertainty revealed before delivery decisions is smaller than the marginal benefit of additional inventory under deterministic demand, i.e.,* $\left|\lambda_i^{mn}\left(\overline{\mathbf{K}}\right)\right| \leq \left|\theta_i^{mn}\left(\overline{\mathbf{K}}\right)\right|$. (b) *Let* $D_j = \bar{D}_j$ *at each FC j for ceteris paribus. The optimal amount of total inventory commitment in stage* 1 *is smaller under the setting when demand uncertainty is revealed before delivery decisions than that under the deterministic demand setting, i.e.,* $\sum_i K_i^{S1*} \leq \sum_i K_i^{D*}$.

Proposition 9(a) shows that the value from inventory preparedness decreases when demand is uncertain at the time of stocking decisions. When demand is deterministic, the firm plans with 100% confidence that every unit of excess inventory would be utilized in the event of a disruption. The marginal benefit from proactive inventory investment is lower under demand uncertainty because the firm faces the risk of overstocking expressed with the cdf term $G(\cdot)$. As a result, Proposition 9(b) proves that the firm prefers to carry a relatively less inventory under demand uncertainty in comparison with the deterministic demand setting.

In the online supplement, Corollary A2 shows that our main result where the total inventory commitment can be decreasing under risk aversion continues to hold under demand uncertainty. Conditions (C2) and (C3) are replaced with the revised conditions (C2′) and (C3′) to accommodate stochastic demand. Comparing the optimality conditions presented in Proposition 6 and Corollary A2, our analysis shows that the firm needs a less restrictive condition under demand uncertainty. Thus, we conclude that incorporating demand uncertainty into the model makes our earlier findings more pronounced.

5.5.2. Demand Uncertainty Revealed After Delivery Decisions

We next examine the setting where the firm makes shipment decisions before observing the customer demand. As a result, demand continues to be random in stage 2, requiring several adjustments in the model and its ensuing analysis. It is important to highlight that the realized demand can be smaller than the amount shipped by the firm, leading to waste in shipments and inventories. Thus, the SaR constraint described in (5) and the objective function expressed in (2) require revisions.

We begin our discussion with the adjustments in the SaR constraint. Due to demand uncertainty, the total number of late shipments can be considered as random. To understand the total lateness in a disrupted facility FC j, we compare the random demand rate $\tilde{D}_j$ with the proportion of shipments that are made on time. Recall that the firm's shipments from all operational FCs to the region j is equal to $\sum_i x_{ij}^{mn}$. Out of these shipments, a portion of them will be delivered late, i.e., $\sum_i l\left(d_{ij}\right) x_{ij}^{mn}$. The remaining portion of deliveries, equaling $\sum_i \left(1-l\left(d_{ij}\right)\right) x_{ij}^{mn}$, will be made on time. Thus, the total number of late deliveries (per day) to region j is equal to $\left(\tilde{D}_j - \sum_i \left(1-l\left(d_{ij}\right)\right) x_{ij}^{mn}\right)^+$; this term includes all late deliveries stemming from the firm's internal shipments as well as the vendor shipments. Then, the total number of late deliveries in all regions throughout a disruption with random length $\tilde{\tau}^{mn}$ becomes $\sum_j \left(\tilde{D}_j \tilde{\tau}^{mn} - \sum_i \left(1-l\left(d_{ij}\right)\right) x_{ij}^{mn} \tilde{\tau}^{mn}\right)^+$. This results in the revised SaR constraint expressed below:

$$P\left[\sum_j \left(\tilde{D}_j \tilde{\tau}^{mn} - \sum_i \left(1-l\left(d_{ij}\right)\right) x_{ij}^{mn} \tilde{\tau}^{mn}\right)^+ > \beta\right] \leq 1-\alpha . \tag{15}$$

For notational simplicity, we define $\tilde{\zeta}_j^{mn}$ to represent the random cumulative demand $\tilde{D}_j \tilde{\tau}^{mn}$ at FC j during a disruption with random length. We rewrite (15) and arrive at the revised SaR constraint:

$$P\left[\sum_j \left(\tilde{\zeta}_j^{mn} - \sum_i \left(1-l\left(d_{ij}\right)\right) x_{ij}^{mn} \tilde{\tau}^{mn}\right)^+ > \beta\right] \leq 1-\alpha . \tag{16}$$

The adjustment in the second-stage objective function (2) is similar. The term $\sum_j \left(\tilde{\zeta}_j^{mn} - \sum_i \left(1-l\left(d_{ij}\right)\right) x_{ij}^{mn} \tilde{\tau}^{mn}\right)^+$ represents the total number of late shipments inclusive of the late shipments from the firm's own delivery operations and the shipments made by the vendor. The total lateness cost stemming from firm's internal shipments and vendor shipments is expressed as

$c_L \sum_j \left(\tilde{\zeta}_j^{mn} - \sum_i \left(1 - l\left(d_{ij}\right)\right) x_{ij}^{mn} \tilde{\tau}^{mn} \right)^+$. In addition, the firm incurs $c_V \sum_j \left(\tilde{\zeta}_j^{mn} - \sum_i x_{ij}^{mn} \tilde{\tau}^{mn} \right)^+$ for shipments originating from vendors. Thus, the second-stage objective function in (2) can then be written as follows:

$$\Psi_{2,mn}^*\left(\overline{\mathbf{K}}\right) = \min_{\bar{\mathbf{x}}^{mn} \geq \bar{0}} \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right) = E\left[\begin{array}{c} c_T \sum_j \sum_i d_{ij} x_{ij}^{mn} \tilde{\tau}^{mn} + c_L \sum_j \left(\tilde{\zeta}_j^{mn} - \sum_i \left(1 - l\left(d_{ij}\right)\right) x_{ij}^{mn} \tilde{\tau}^{mn} \right)^+ \\ + c_V \sum_j \left(\tilde{\zeta}_j^{mn} - \sum_i x_{ij}^{mn} \tilde{\tau}^{mn} \right)^+ \end{array} \right]. \quad (17)$$

Because demand is unknown even at the operational FCs at the time delivery decisions are determined, constraint (3) associated with the amount of supply that can get shipped also requires revision. Let us define the base stock in the operational FC i with BS_i. The value of BS_i isolates the firm's inventory preparedness towards demand uncertainty; the value of K_i becomes strictly associated with inventory preparedness towards disruption uncertainty. We can then rewrite (3) as follows:

$$\sum_j x_{ij}^{mn} \leq K_i + BS_i \;\; \forall i. \quad (18)$$

Constraint (4) in the stage 2 model indicates that the shipment from operational FCs should not exceed the demand of the destination FC. This constraint is also revised so that the total shipment from operational FCs would not exceed the base stock of the destination FC:

$$\sum_i x_{ij}^{mn} \leq BS_j \;\; \forall j. \quad (19)$$

The stage 2 model for this setting becomes minimizing (17) subject to (16), (18) and (19) along with non-negativity constraints in (6). We next show that the approach utilized to solve the problem under random disruption duration continues to be the prevailing approach to solve the problem under the combination of demand and disruption duration uncertainty. Thus, the main findings continue to hold when demand uncertainty revealed after delivery decisions is incorporated into the model. We supplement our earlier conclusions with two new results. First, the value of additional inventory commitment decreases when demand uncertainty is incorporated into the problem due to the risk of over-shipping. As a result, the firm commits to a smaller inventory under demand uncertainty. Second, our earlier finding where the introduction of SaR can lead to less inventory becomes more prevalent when demand uncertainty is incorporated into the model. This finding occurs with one less condition than the earlier model settings, making it easier to experience under demand uncertainty.

Corollary A3 in the online supplement shows that Condition (8) continues to help determine when to employ Policy L vs. Policy A under demand uncertainty. Together with Corollary A1, Condition (8) determines the policy choice regardless of whether demand uncertainty is realized before or after delivery decisions.

We next elaborate on the first-stage solution approach. Let $\Psi_1^{S2}\left(\overline{\mathbf{K}}\right)$ represent the first-stage objective function under this stochastic demand setting; its first-order derivative with respect to K_i can be represented as

$$\partial \Psi_1^{S2}\left(\overline{\mathbf{K}}\right)/\partial K_i = c_K + \sum_{m=1}^{M}\sum_{n=1}^{N} p_{mn}\gamma_i^{mn}\left(\overline{\mathbf{K}}\right) \tag{20}$$

where $\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)$ is the dual price of the supply constraint (18) for FC i evaluated at $\overline{\mathbf{K}}$ under (m, n). The value of $\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)$ is a continuous non-decreasing function in $\overline{\mathbf{K}}$ and is non-positive. $\left|\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)\right|$ represents the marginal benefit of additional inventory for a given (m, n).

$\partial \Psi_1^{S2}\left(\overline{\mathbf{K}}\right)/\partial K_i$ fetches its maximum value of c_K (> 0) at high values of K_i for each i since $\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)$ is non-decreasing. As a result, the minimization problem in stage 1 is bounded, leading to an optimal solution (see Corollary A4 in the online supplement).

We next present an important feature of demand uncertainty on the value of inventory preparedness. Let us first compare the first-order derivative of (2) in the deterministic demand setting with that of (17) under demand uncertainty. The first-order derivative of (2) with respect to x_{ij}^{mn} is

$$E\left[\tilde{\tau}^{mn}\right]\left[c_T d_{ij} - c_L\left(1 - l\left(d_{ij}\right)\right) - c_V\right]. \tag{21}$$

The first-order derivative of (17) with respect to x_{ij}^{mn} is

$$E\left[\tilde{\tau}^{mn}\right]\left[c_T d_{ij} - c_L\left(1 - l\left(d_{ij}\right)\right)\left(1 - F_j\left(\sum_i\left(1 - l\left(d_{ij}\right)\right)x_{ij}^{mn}\right)\right) - c_V\left(1 - F_j\left(\sum_i x_{ij}^{mn}\right)\right)\right] \tag{22}$$

where $F_j(\cdot)$ denotes the cumulative distribution function of random demand $\tilde{D}_j$. One can see that the value of (21) is less than or equal to the value of (22) for all i, j, and (m, n). This indicates that making a shipment under deterministic demand is less costly than making a shipment under demand uncertainty. This stems from the fact that in the deterministic demand setting, the firm is 100% confident that every shipment made by the firm is utilized toward satisfying the readily known demand. Conversely, when the firm does not know the demand at the time of delivery decisions, it faces the risk of sending more shipments than the actual demand of customers. Moreover, this possibility of shipping more than the realized demand increases as the firm increases its shipment amount. As a result, from an expected cost minimization perspective, making a shipment when demand is known is more economical than making a shipment when demand is uncertain in stage 2. This structure pertaining to stage 2 also has immediate implications for the firm's inventory decisions in stage 1. Following from the fact that the value of (21) is less than that of (22), we have $\theta_i^{mn}\left(\overline{\mathbf{K}}\right) \le \gamma_i^{mn}\left(\overline{\mathbf{K}}\right) \le 0$. As a result, we conclude the following:

$\left|\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)\right| \le \left|\theta_i^{mn}\left(\overline{\mathbf{K}}\right)\right|$. This implies that demand uncertainty diminishes the marginal benefit of additional inventory. This leads to a reduction in the firm's inventory commitment in stage 1 under demand uncertainty. Let $\sum_i K_i^{S2*}$ denote the optimal total inventory level under random demand where the second-stage shipments to customers are made under demand uncertainty. The following proposition formalizes the findings with reduced marginal benefit from the first-stage inventory commitment under demand uncertainty, leading to a smaller level of optimal inventory under demand uncertainty.

Proposition 10. (a) *The marginal benefit from additional inventory under demand uncertainty in the setting where random demand is revealed after delivery decisions is smaller than the marginal benefit of additional inventory under deterministic demand, i.e.,* $\left|\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)\right| \le \left|\theta_i^{mn}\left(\overline{\mathbf{K}}\right)\right|$. (b) *Let $D_j = BS_j$ at each FC j for ceteris paribus. The optimal amount of total inventory commitment in stage* 1 *is smaller under the setting when demand uncertainty is revealed after shipments are made to the customer than that under the deterministic demand setting, i.e.,* $\sum_i K_i^{S2*} \le \sum_i K_i^{D*}$.

Propositions 9 and 10 collectively suggest that demand uncertainty, regardless of when random demand is revealed, reduces the value of carrying additional inventory. When demand is revealed before the delivery decisions, the possibility of overstocking is the driver for this finding. When demand is revealed after the delivery decisions, the possibility of over-shipping becomes an additional driving force. As a result of these two propositions, we conclude that the optimal inventory commitment under demand uncertainty is smaller than that under deterministic demand.

We have shown in Section 5.3 that the introduction of SaR constraint can lead to a reduction in the firm's initial inventory commitment. It is important to highlight that it would not have been easy to demonstrate the reduction in inventory due to SaR under the settings with demand uncertainty because demand uncertainty also reduces the firm's initial inventory commitment. While SaR can cause a reduction in the initial inventory commitments, this is not a monotone behavior; the conditions (C1) through (C3) highlight the conditions under which the firm reduces its initial inventory levels. When these conditions are not met, SaR can lead to an increase in the firm's initial inventory commitment. The reduction in inventory levels under demand uncertainty, on the other hand, is uniform in the risk-neutral setting.

In the online supplement, we show that our main finding pertaining to the decrease in total inventory commitment with risk aversion continues to hold under demand uncertainty where random demand is revealed after shipment decisions are made. Conditions (C2) and (C3) are replaced with the revised conditions (C2″) and (C3″) to accommodate stochastic demand. Lemma A1 proves that Condition (C3″) is always satisfied. Corollary A5 shows that our main conclusion, originally presented in Proposition 6, continues to hold under demand uncertainty. Moreover, Corollary A5 requires less restrictive conditions than Proposition 6, making our earlier finding more pronounced.

Using two different settings in Section 5.5, we have presented a comprehensive examination of the role demand uncertainty plays in the firm's delivery operations (stage 2 decisions) and its preparedness for disruptions (proactive stocking decisions in stage 1). Our analysis has illuminated three new findings that are specifically driven by demand uncertainty. First, Proposition 9 has shown that demand uncertainty, when revealed before delivery decisions, reduces the value of inventory preparedness because of the risk of overstocking. As a result, we have shown that the firm's initial inventory commitment is reduced in comparison to the deterministic setting. Second, Proposition 10 has confirmed this finding when random demand is revealed after shipment decisions are made. In this case, the driving force becomes the risk of over-shipping. Collectively, we conclude that the firm's first-stage stage optimal inventory commitment is smaller under both settings of demand uncertainty than its deterministic demand equivalent. Third, using Lemma A1 and Corollary A5, we have identified that the finding pertaining to the total inventory decreasing in response to risk aversion through SaR becomes a more general result when demand is uncertain at the time of delivery decisions.

In sum, we have shown that our paper's insights go beyond the case of deterministic demand by investigating demand uncertainty comprehensively through two different settings. Our analysis identifies three new results driven solely by demand uncertainty. These new results lead to the conclusion that inventory preparedness is less valuable under demand uncertainty.

6. Numerical Analysis with Company and Public Data

Using our motivating firm's network with 15 FCs, this section presents numerical analyses confirming that our analytical findings apply in practical settings. Specifically, the analysis demonstrates how the optimal inventory investment can decrease when the firm switches from a risk-neutral to risk-averse setting. We also introduce a new metric that is useful in practice: Risk Dispersion Index (RDI). RDI measures the degree of dispersion in the amount of financial exposure across all facilities stemming from disruption risks. Through RDI, we show that our model leads to a reduction in dispersion of risk exposure across facilities, and thus, results in a more balanced supply chain architecture.

6.1. 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: 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, which shows that earthquakes that have a Richter magnitude less than 5.5 result in minor or no impact. Therefore, earthquakes with a Richter magnitude greater than or equal to 5.5 help determine the probability of an earthquake impacting an FC.

The data for hurricanes and tornadoes is obtained from the National Oceanic and Atmospheric Administration. Facilities are influenced by hurricanes geographically (primarily in the East Coast of the US) and seasonally (from June 1st through November 30th). We derive the probability of an 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 use the impact region for an average tornado (one mile in width with a 50-mile-long travel distance) 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 employ the impact radius of 100 miles for chemical and nuclear disruptions and use the proximity of the company's FCs to these nuclear power reactors 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. The mean and standard deviations regarding the length of disruption are provided in the same data set. We assume normal distribution for this set of disruptions. For the disruption events whose data is collected from national sources, our granular data regarding the frequency of disruptions leads to varying probabilities across FCs. 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 considers the impact of eleven disruptions. Table 2 provides the mean duration and its standard deviation of each disruption event, and their corresponding probabilities over a year at each FC.

Which FC can service a disrupted facility? The firm indicates that an FC located more than 600 miles apart is not capable of serving customers in a disrupted region within the 24-hour delivery window. The threshold of 600 miles is determined by incorporating order preparation time, packaging, loading and unloading in addition to the transportation time. Moreover, on-time delivery performance declines and the proportion of late deliveries increases with distance. The firm indicates that the proportion of late deliveries increase linearly in time, and it is described as $l(d_{ij}) = d_{ij}/600$ for $0 \leq d_{ij} \leq 600$ and $l(d_{ij}) = 100\%$

for $d_{ij} > 600$. It is important to note that our findings and insights do not rely on the linearly increasing late deliveries; they continue to hold under convex increasing functions.

		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
FC ID	198	2.44%	0.00%	18.29%	0.30%	0.61%	0.00%	3.05%	0.30%	0.63%	0.00%	0.08%
	268	2.01%	0.00%	15.95%	0.29%	0.43%	0.14%	2.23%	0.72%	0.63%	0.00%	0.08%
	269	1.00%	0.13%	7.55%	0.77%	0.00%	0.13%	0.26%	0.27%	0.00%	0.21%	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%
	297	0.72%	0.00%	17.31%	0.24%	0.48%	0.00%	5.23%	0.24%	9.38%	0.00%	0.02%
	368	1.55%	0.31%	3.31%	0.41%	1.14%	0.21%	1.06%	0.62%	3.75%	0.00%	0.06%
	397	0.42%	0.00%	12.92%	0.00%	0.00%	0.00%	0.15%	0.83%	0.00%	0.04%	0.00%
	398	2.53%	0.00%	20.73%	0.16%	0.47%	0.00%	4.02%	0.32%	3.75%	0.00%	0.06%
	469	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%
	598	2.50%	0.00%	12.50%	0.00%	3.75%	0.00%	3.86%	0.00%	0.00%	0.00%	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%
	868	0.00%	0.00%	8.33%	0.00%	0.00%	0.00%	2.39%	4.17%	0.00%	0.00%	0.02%
	948	1.30%	0.00%	13.54%	0.00%	5.47%	0.00%	2.87%	0.52%	0.00%	0.00%	0.04%
	983	0.72%	0.00%	17.31%	0.24%	0.48%	0.00%	5.23%	0.24%	9.38%	0.00%	0.02%

Table 2: The mean and standard deviation of duration (in hours) and the probability over a year of each disruption at fulfillment centers.

The cost data is provided by the company that motivated this study. In the event of a disruption at an FC, three types of additional costs are incurred: (1) Cost of transportation; (2) cost of late deliveries; and (3) cost of vendor shipments. If there is available inventory at the nearby operational FC, the delivery would take place using the firm's own transportation vehicles. On-time deliveries would incur the firm's standard transportation cost per mile, c_T, and is multiplied by the distance between the disrupted facility and the operational facility. 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 company motivating our study indicates that the cost of delivery using vendors, c_V, is 6.8 times the cost of late delivery, c_L. Finally, the unit cost of additional inventory, c_K, is estimated through the inventory holding cost which is inclusive of the additional space, risks such as obsolescence and the firm's internal rate of return.

The demand data is also provided by the firm. We are given current demand levels along with historical data. Using this historical data, we estimate the range of demand values at each FC providing the support for the random demand variable. In our analysis pertaining to risk aversion, our SaR constraint emphasizes demand realizations at extreme cases, corresponding to the right tail of the distribution. In the case of the firm motivating our study, this corresponds to demand realizations that are greater than or equal to the value at 97th percentile.

We denote the present excess stocking levels with K_i^0. We compare the performance of our model's proposed solution with that of the firm's current stocking decisions.

Using the disruption probabilities in Table 2 and the unit holding cost $c_K = 2$, we first calculate the "Risk Exposure Index" (REI) score established by Simchi-Levi et al. (2014) representing the financial impact of disruptions at each FC. Table 3 presents the REI scores (on a scale of 0 to 100) at each FC at the current stocking levels K_i^0. REI score of an FC in the table below represents the cost contribution of that FC in the first-stage objective function (1) where $K_i = K_i^0$ for all i. The FC with the highest cost contribution is given a score of 100 (representing the highest exposure) and all remaining FCs are assigned scores proportionally based on the ratio of their cost contribution to the FC with the highest cost contribution. Table 3 shows that FC 983 has the highest monetary exposure followed by FCs 268 and 398. FCs 497, 281, 198, 397, 694, 868 and 697 have less than 10% of the maximum monetary risk.

FC ID	983	268	398	368	297	598	948	269	497	198	281	397	694	868	697
REI	100.00	50.74	45.35	27.18	26.55	20.87	14.22	13.95	9.96	9.88	8.82	6.70	6.11	5.38	0.61

Table 3: REI scores of each FC where $K_i = K_i^0$.

6.2. Findings under the Risk-Neutral Setting

We examine the optimal first-stage stocking decisions using the model presented in Section 4 in the absence of the SaR constraint. Table 4 presents the results of optimal stocking decisions for the risk-neutral analysis under various values of the unit inventory cost c_K. We use the expected total cost at $c_K = 2$ as the benchmark case in Table 4 and show the increase in expected total cost at higher values of c_K.

	c_K				
FC ID	2	4	6	8	10
198	1,509	987	987	772	433
268	0	0	0	0	89
269	0	0	0	0	0
281	274	526	0	0	0
297	1,395	252	0	0	0
368	0	339	339	339	0
397	0	0	0	0	0
398	0	0	0	0	0
469	772	0	0	0	251
497	0	0	0	0	0
598	265	0	0	0	0
697	0	0	0	0	0
868	389	0	0	0	0
948	0	0	0	0	0
983	252	0	0	0	0
ΣK_i^*	4,856	2,104	1,326	1,111	773
Expected Total Cost	100%	134%	150%	162%	170%

Table 4: The impact of the unit inventory holding cost (c_K) on the optimal excess stocking decisions under the risk-neutral setting.

We make two observations from the results pertaining to the risk-neutral analysis. First, the aggregate excess stocking investment in the network, described as $\sum K_i^*$, decreases with higher levels of c_K. Second, the FC-level excess inventory levels do not necessarily decrease monotonously with respect to unit inventory cost c_K. For example, the optimal inventory at FC 281 first increases when the unit inventory cost increases from $c_K = 2$ to $c_K = 4$, however, it drops to zero when $c_K = 6$. This is an example of non-monotone inventory investment in an FC. It is important to note that FCs 983 and 297 are within 600-mile distance of FC 281, and thus, these three facilities can serve as a backup to each other when there is a disruption. With increasing c_K values, the firm initially prefers an extra unit of inventory at a relatively safer location, FC 281, and decreases the investment at the FCs at riskier locations, FCs 983 and 297. When the unit inventory cost increases further, i.e., when $c_K \geq 6$, it becomes too expensive to justify any amount of excess inventory at these three facilities; thus, the optimal values drop to zero.

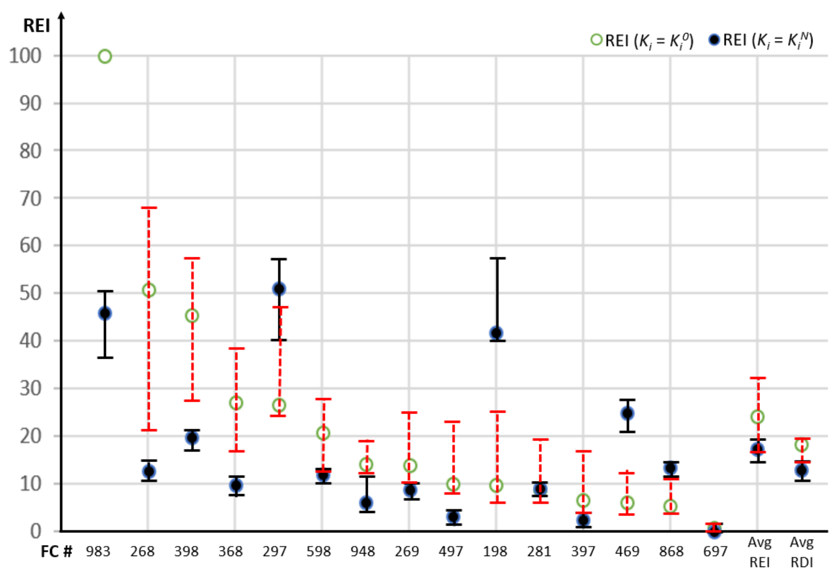
We next introduce a new metric that is beneficial in describing the relative risk levels at a multi-facility delivery network. The REI score approach described in Simchi-Levi et al. (2014) calculates the monetary risk and helps determine the most vulnerable facilities in a manufacturing setting. However, the REI approach does not provide how that risk is distributed across different delivery facilities. To fill in this void, we introduce another risk metric called the Risk Dispersion Index (RDI) of network that is based on the mean absolute deviation in REI scores in a given network infrastructure:

RDI in a network = Σ |REI score at each FC – Average REI score| / [number of FCs]. (23)

Lower RDI scores imply that the network has a more balanced risk profile among its facilities and is much less vulnerable to potential disruptions.

We next present how our proposed model leads to a more balanced risk profile for the firm in question. We first solve the optimal stocking decisions using the risk-neutral version of the model presented in Section 4, excluding (5), at the unit inventory cost $c_K = 2$. We denote these optimal inventory levels in the risk-neutral setting as $K_i = K_i^N$ for all i. Figure 3 presents the comparison of REI and RDI scores from our model with the firm's current inventory levels, denoted $K_i = K_i^0$ for all i. We assign a score of 100 to the FC with the highest cost contribution for $K_i \in \{K_i^0, K_i^N\}$ for all i in order to clearly demonstrate the change in risk exposure. Figure 3 establishes the range of REI and RDI scores for the given range of demand variations establishing the minimum and maximum risks. In Figure 3, REI and RDI scores at the current expected demand levels are expressed with dots.

Several observations can be made from the results presented in Figure 3. Recall that at the current inventory levels ($K_i = K_i^0$ for all i), FC 983 has the highest risk exposure, and therefore, is the most vulnerable facility in the network. In order to provide a fair benchmark for comparison, we anchor the highest risk exposure to FC in order to determine the maximum and minimum REI scores establishing the range for each FC. The following observations are based on the current expected demand case. First, we observe that when the optimal inventory levels are used, the maximum risk exposure at a facility is approximately half the risk exposure of FC 983. Under our proposed solution, FC 983 is no longer the facility with the highest risk exposure; rather FC 297 becomes the most vulnerable facility. The REI score of FC 297 is 51.0 and is substantially lower than the initial REI score at FC 983. This observation demonstrates the significant level of risk reduction obtained through our model at the most vulnerable facilities. Second, the average REI score decreases from 24.19 to 17.39, representing a 28.11% overall improvement in the average REI score. Given the definition of the REI score, this means that the firm may decrease its expected total cost stemming from disruption risk by 28.11% using our model. Third, the RDI score shows that the dispersion in the risk exposure is reduced, and therefore, the firm would have a more balanced supply chain, making it resilient to disruptions. The RDI score at the current network is 18.42, and the optimal solution of our model reduces the RDI score to 12.91. The reduction in RDI scores represents a 29.91% improvement and demonstrates the fact that the proposed network is prepared in a more balanced manner to mitigate potential disruptions than the firm's present supply chain. In addition to the above three observations, it can be observed that demand variations do not have much impact on REI scores of most FCs (can be seen from narrow ranges). In conclusion, our proposed solution leads to a more resilient supply chain network.



FC ID	983	268	398	368	297	598	948	269	497	198	281	397	469	868	697	Avg. REI	RDI
REI ($K_i = K_i^0$)	100.0	50.74	45.35	27.18	26.55	20.87	14.22	13.95	9.96	9.88	8.22	6.70	6.11	5.38	0.61	24.19	18.42
REI ($K_i = K_i^N$)	46.03	12.66	19.81	9.74	51.00	11.96	5.99	8.70	3.17	41.78	9.13	2.46	24.94	13.38	0.06	17.39	12.91

Figure 3: Minimum and maximum REI scores at each FC, average REI and RDI scores, comparing the optimal inventory investment decisions ($K_i = K_i^N$) with the firm's current inventory levels ($K_i = K_i^0$) under the risk-neutral setting when $c_K = 2$.

6.3. Findings under the Risk-Averse Setting

This section presents the impact of risk aversion through the SaR constraint in designing a supply chain to mitigate the negative consequences of disruptions. The firm promises its business customers next day delivery, i.e., orders placed before 5:00pm are delivered the next day by 5:00pm, with an on-time delivery performance of 97% corresponding to the value of α = 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 promise and tracks the performance throughout its network. Our analysis examines the optimal stocking decisions under various values of β and α = 97% using our SaR constraint.

Table 5 provides the results pertaining to the optimal inventory decisions under increasing risk aversion at c_K = 2. 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.

FC ID	Risk Neutral	β 1200	1150	1100	1050	1000	950	900
198	1,509	1,509	1,509	1,509	1,509	1,509	1,509	-
268	0	0	0	0	0	0	0	-
269	0	0	0	0	0	0	0	-
281	274	538	820	1,103	1,385	1,668	2,105	-
297	1,395	1,132	850	567	285	0	0	-
368	0	0	0	0	0	0	0	-
397	0	0	0	0	0	0	0	-
398	0	0	0	0	0	0	0	-
469	772	772	772	772	772	772	772	-
497	0	0	0	0	0	0	0	-
598	265	265	265	265	265	265	265	-
697	0	0	0	0	0	0	0	-
868	389	389	389	389	389	389	389	-
948	0	0	0	0	0	0	0	-
983	252	0	0	0	0	250	252	-
ΣK_i^*	4,856	4,605	4,605	4,605	4,605	4,853	5,292	-
Expected Total Cost	100%	100.2%	101.1%	102.0%	102.9%	104.4%	108.0%	-

Table 5: The impact of risk aversion (with decreasing values of β) on the optimal inventory decisions when c_K = 2.

The results in Table 5 show one of the most important findings in this paper: When the firm switches from a risk-neutral setting (i.e., ignoring the SaR constraint) to a risk-averse setting (i.e., incorporating the SaR constraint), it might end up reducing its initial stocking decision. One would intuit that, given a constant value of initial cost c_K, the firm would always increase its inventory investment under risk aversion. However, the total inventory level, designated with $\sum K_i^*$ in Table 5, decreases when the firm introduces risk aversion. While the total inventory is 4,856 units under the risk neutral setting, it is equal to 4,605 units for β = 1200 representing a risk-averse setting. This observation highlights the theoretical findings presented in Proposition 6, demonstrating that it is possible for this phenomenon to be observed in practice. Moreover, this data-driven demonstration shows that this finding can be present in broader network configurations with more than three FCs. In the network of our focal firm, FCs 281, 983 and 297 are the main drivers of this result. FCs 983 and 297 are located relatively close to each other and can be affected together by broad-impact disruptions similar to how FC 1 and FC 2 are disrupted in the analysis in Section 5.3. FC 281 is located farther from these facilities resembling FC 3 of the same analysis. If the inventory increase at FC 281 were not accompanied by a decrease in FCs 297 and 983 as our model suggests, the firm would have ended up with approximately 31% more inventory than the optimal risk-averse amount at these three FCs. In the entire network, this would correspond to approximately 11% more inventory than the optimal risk-averse amount of 4,605. Thus, our paper's important finding adds value to the firm by cutting a significant amount of unnecessary inventory by preventing the firm from blindly increasing inventory in response to risk aversion.

Second, as the value of β further decreases below 1050 (representing higher degrees of risk aversion) the firm starts investing more heavily in inventory in order to comply with the SaR constraint. As a result, the total stocking decision starts to increase. Figure 4 demonstrates this result by showing the total optimal inventory decisions under the risk-neutral and risk-averse settings (with various values of β).

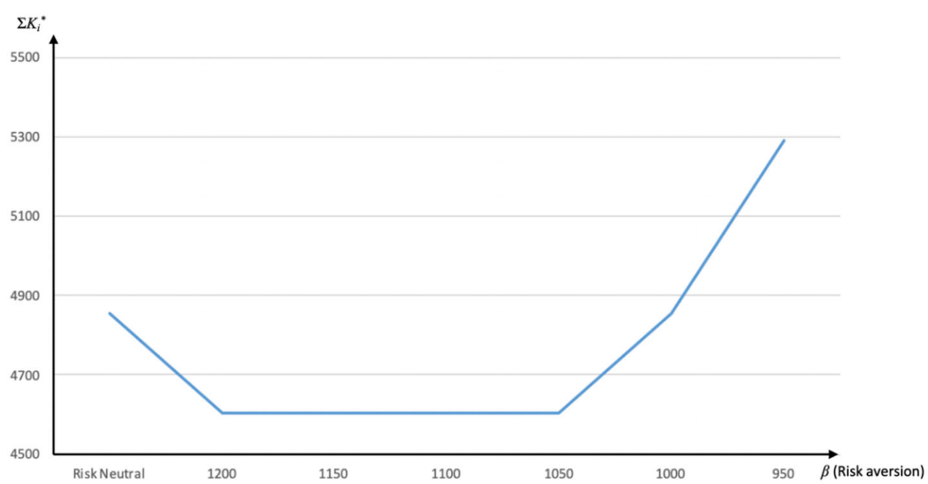
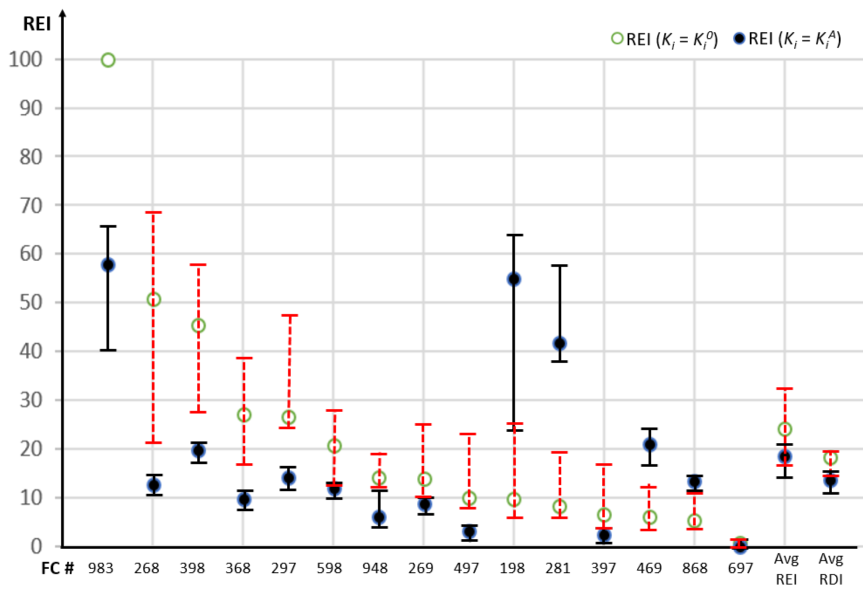


Figure 4: The impact of risk aversion on the sum of optimal inventory decisions.

Third, FCs with the highest initial REI scores at $K_i = K_i^0$ continue to get limited excess inventory, while FCs with the lower initial REI scores continue to receive excess inventory under risk aversion. For example, at β = 1,200, we observe that the facilities with the highest initial REI scores (FCs 983, 268, 398) receive no excess inventory. On the other hand, facilities that are tagged with lower initial REI scores carry excess inventory such as FCs 198, 469, and 868, all with initial REI scores lower than 10.

Thus, risk aversion leads the firm to add excess inventory at FCs that are initially perceived as less vulnerable in order to be able to make faster deliveries and provide better customer service.

Our proposed model continues to lead to balanced supply chain preparedness against disruptions. Figure 5 presents the REI and RDI scores under risk aversion when $\beta = 950$ and compares the results with the firm's current network. The analysis uses the demand fluctuations at each FC from earlier analysis pertaining to the risk-neutral setting. We denote the optimal inventory decisions under risk aversion with $K_i = K_i^A$.



FC ID	983	268	398	368	297	598	948	269	497	198	281	397	694	868	697	Avg. REI	RDI
REI ($K_i = K_i^0$)	100.0	50.74	45.35	27.18	26.55	20.87	14.22	13.95	9.96	9.88	8.82	6.70	6.11	5.38	0.61	24.19	18.42
REI ($K_i = K_i^A$)	57.89	12.66	19.81	9.74	14.22	11.96	5.99	8.70	3.17	55.05	41.78	2.46	20.97	13.38	0.06	18.52	13.72

Figure 5: Minimum and maximum REI scores at each FC, average REI and RDI scores, comparing the optimal inventory investment decisions ($K_i = K_i^A$) with the firm's current inventory levels ($K_i = K_i^0$) under risk-averse setting when $\beta = 950$ and $c_K = 2$.

Several observations can be made from the results presented in Figure 5. Recall that at the current inventory levels ($K_i = K_i^0$ for all i), FCs 983 and 268 had the highest REI scores with 100 and 50.74, respectively. First, our proposed model reduces the REI scores of these two facilities to 57.89 and 12.66, respectively. FC 983 continues to be the facility with the highest risk exposure, and thus, it still is the most vulnerable facility in the network. However, its monetary risk exposure is reduced from 100 to 57.89, representing a substantial improvement at the most vulnerable facility. Second, the average REI score decreases from 24.19 to 18.52, representing a 23.44% overall improvement. This indicates that the firm may decrease its expected total cost stemming from disruption risk by 23.44% under risk aversion. Third, the RDI scores show that the dispersion in the risk exposure is reduced, and therefore, the firm would have a more balanced supply chain that is resilient to disruptions. The RDI score is reduced from 18.42 to 13.72 in the risk-averse optimal solution. FCs 198 and 281 see an increase in their REI scores because of the increase in their initial inventory commitment due to risk aversion. However, the RDI scores are reduced by 25.52%, leading to a more balanced supply chain in terms of preparing for disruptions when compared with the current supply chain infrastructure. In conclusion, our proposed solution leads to a more resilient supply chain network with the ability to provide better customer service.

7. Conclusions and Managerial Insights

This paper examines disruption risks and develops risk mitigation strategies for a Fortune 150 firm that considers on-time delivery performance its winning criterion for business customers. It introduces an operational requirement coined as Service at Risk (SaR) constraint where the probability of late deliveries exceeding a threshold is limited with a tolerable probability. The SaR constraint differs from the traditional value-at-risk (VaR) and conditional value-at-risk (CVaR) constraints, and financial risk mitigation methods in the form of insurance, which does not positively impact firm's operational performance.

The proposed model combines proactive and reactive risk mitigation strategies to counter disruption risks in delivery operations. The analysis integrates a comprehensive set of disruptions provided by the firm and collected from national sources in a granular manner. Prior to the disruption, the firm determines the optimal amount of inventory to be kept in each FC as a proactive measure. In the event of a disruption at an FC, the firm solves a transportation problem that minimizes the total contingency cost while complying with the SaR constraint; this corresponds to the reactive measure.

This paper makes four main contributions. First, it shows that the total amount of excess inventory committed as a result of disruption risk can decrease with risk aversion. This finding departs from the results presented in earlier publications where inventory increases with higher disruption risks. The result stems from the fact that the firm prefers to locate one unit of additional inventory at a distant FC by enabling the reduction in the inventories of two nearby FCs. We show that this finding is robust as it holds under various settings that incorporate demand uncertainty. We also demonstrate numerically that this phenomenon occurs in the network setting of the firm motivating our study.

Second, the study introduces a new metric called Risk Dispersion Index (RDI) that evaluates the dispersion in risk exposure across facilities in the network. Through RDI, our proposed proactive and reactive risk mitigation strategies lead to more resilient supply chain operations with lower and balanced levels of risk exposure at the firm's FC network. Comprehensive numerical illustrations with varying demand values at each FC demonstrate that our model leads to substantial improvements in REI and RDI scores. In the risk-neutral setting, REI and RDI scores are reduced by 28.11% and 29.91% on average, respectively. In the risk-averse setting, our model reduces REI and RDI scores by 23.44% and 25.52% on average, respectively.

Third, the study shows that an FC can abandon its customers in order to serve the customers of a disrupted facility. This result occurs when late delivery percentages increase in distance in a convex manner. Rather than serving its own customers, the network can create a chain of rerouting support. This is referred to as the abandonment policy, which is found to occur more frequently under risk aversion.

Fourth, we show that incorporating demand uncertainty causes a decrease in the firm's initial proactive inventory commitment. This finding is driven by a reduced marginal benefit of inventory commitment under demand uncertainty. In our analysis, we examine the influence of demand uncertainty with two different settings. In the first variation, random demand is realized before delivery decisions are made. In the second setting, delivery decisions are made before observing the customer order; as such, they are made in the presence of demand uncertainty. Diminishing marginal benefit of inventory is caused by the risk of overstocking in the first variation and the risk of over-shipping in the second variation. Collectively, both settings lead us to the same conclusion: Demand uncertainty causes a reduction in the risk-neutral firm's initial inventory commitment. We also conclude that our main findings continue to hold under demand uncertainty (including both variations). Specifically, under demand uncertainty, our study shows that (1) the firm's total inventory commitment can be decreasing with risk aversion and (2) an operational FC can abandon its own customers to serve the customers of a disrupted facility.

Our study examines distribution operations in a single country, and it can be used to study disruption risks in global supply chain operations. One of the important features in our study is that two fulfillment centers can get disrupted at the same time; this leads to new insights: (1) Inventory preparedness does not have to be anchored to geographic proximity, i.e., backup inventories can be located in a distant facility rather than a nearby facility, and (2) inventory commitment can decrease with risk aversion. Recent developments associated with coronavirus shows the importance of analyzing disruptions at multiple facilities at the same time, as the pandemic occurred in many countries at the same time and in distant locations. Our study advocates for diversifying the location of inventories in preparing for disruptions while determining the necessary amount of stocking levels in the presence of disruption risks. During catastrophic events such as the spread of coronavirus, diversification of inventory locations is particularly beneficial both from the ability to use supplies from the regions that are not immediately impacted and from those geographies that recover from the pandemic earlier. By shedding light on the influence of disruption possibilities on multiple facilities occurring concurrently, our study will be beneficial for future research examining preparedness for disruptions at multiple geographies from a broader scope and with a global context.

References

Acimovic, J., S. 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. Iancu, R. Swinney. 2017. Disruption risk and optimal sourcing in multitier supply networks. *Management Science* **63**(8), 2397-2419.

Anunpindi, R., R. Akella. 1993. Diversification under supply uncertainty. *Management Science* **39**(8), 944-963.

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

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

Berk, E., A. Arreola-Risa. 1994. Note on: future supply uncertainty in EOQ models. *Naval Research Logistics* **41**(1), 129-132.

Bimpikis, K., O. Candogan, S. Ehsani. 2019. Supply disruptions and optimal network structures. *Management Science* **65**(12), 5504-5517.

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

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

Dada, M., N. Petruzzi, L. Schwarz. 2007. A newsvendor's procurement problem when suppliers are unreliable. *Manufacturing & Service Operations Management* **9**(1), 9-32.

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

Dong, L., S.Y. Tang, B. Tomlin. 2018. Production chain disruptions: Inventory, preparedness, and insurance. *Production and Operations Management* **27**(7), 1251–1270.

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

Eeckhoudt, L., C. Gollier, H. Schlesinger. 1995. The risk-averse (and prudent) newsboy. *Management Science* **41**(5), 786-794.

Federgruen, A., N. Yang. 2008. Selecting a portfolio of suppliers under demand and supply risks. *Operations Research* **56**(4), 916–936.

Federgruen, A., N. Yang. 2011. Procurement strategies with unreliable suppliers. *Operations Research* **59**(4), 1033–1039.

Hendricks, K.B., V.R. Singhal. 2003. The effect of supply chain glitches on shareholder wealth. *Journal of Operations Management* **21**(5), 501-522.

Hendricks, K.B., V.R. Singhal. 2005a. An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production and Operations Management* **14**(1), 35–52.

Hendricks, K.B., V.R. Singhal. 2005b. Association between supply chain glitches and operating performance. *Management Science* **51**(5), 695-711.

Hu, B., D. Kostamis. 2014. Managing supply disruptions via multiple sourcing when output determines market price. *Production and Operations Management* **24**(5), 808–820.

Gao, S.Y., D. Simchi-Levi, C. Teo, Z. Yan. 2019. Disruption risk mitigation in supply chains: The risk exposure index revisited. *Operations Research* **67**(3), 831-852.

Li, Z., S. Xu, J. Hayya. 2004. A periodic-review inventory system with supply interruptions. *Probability in the Engineering an Informational Sciences* **18**(1), 33–53.

Meyer, R., M. Rothkopf, S. Smith. 1979. Reliability and inventory in a production-storage system. *Management Science* **25**(8), 799-807.

NPR. 2018. Optimized Prime: How AI and anticipation power Amazon's 1-hour deliveries. https://www.npr.org/2018/11/21/660168325/optimized-prime-how-ai-and-anticipation-power-amazons-1-hour-deliveries

Parlar, M., D. Berkin. 1991. Future supply uncertainty in EOQ models. *Naval Research Logistics* **38**(1), 107-121.

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

Ross, A., Y. Rong, L. Snyder. 2008. Supply disruptions with time dependent parameters. *Computers and Operations Research* **35**(11), 3504-3529

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

Schmitt, A., L. Snyder, Z. Shen. 2010. Inventory systems with stochastic demand and supply: properties and approximations. *European Journal of Operational Research* **206**(2), 313-328.

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. 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.

Snyder, L., Z. Atan, P. Peng, Y. Rong, A. Schmitt, B. Sinsoysal. 2016. OR/MS models for supply chain disruptions: A review. *IIE Transactions* **48**(2), 89-109.

Song, J., P. Zipkin. 1996. Inventory control with information about supply conditions. *Management Science* **42**(10), 1409-1419.

Swaminathan, J., J. Shanthikumar. 1999. Supplier diversification: Effect of discrete demand. *Operations Research Letters* **24**(5), 213-221.

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., 2009. Impact of supply learning when suppliers are unreliable. *Manufacturing & Service Operations Management* **11**(2), 192-209.

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.

Zhao, M., N. Freeman. 2019. Robust sourcing from suppliers under ambiguously correlated major disruption risks. *Production and Operations Management* **28**(2), 441-456.

Online Supplement
for
Service at Risk in Delivery Operations

Proof of Proposition 1. Recall from Section 4 that $d_{ii} = 0$ and $l(d_{ii}) = 0$. Therefore, we have the following: $\partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{ii}^{mn} = -(c_V + c_L) \leq \partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{ij}^{mn} = c_T d_{ij} + c_L l(d_{ij}) - (c_V + c_L)$ for all (i, j) – note that $\Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)$ is linear in x_{ij}^{mn}. Combined with $\partial^2 l(d_{ij})/\partial d_{ij}^2 = 0$ and (4), it follows that $x_{ii}^{mn*} = D_i$ for each FC $i \notin \Lambda(n)$. ❑

Proof of Proposition 2. From (2), one can see that each unit of vendor shipment to a disrupted FC j (i.e., $j \in \Lambda(n)$) costs $c_V + c_L$ whereas each unit of shipment made from FC i costs $c_T d_{ij} + c_L l(d_{ij})$. Following from $\partial l(d_{ij})/\partial d_{ij} \geq 0$, there exists a unique threshold d_H that solves $c_V + c_L = c_T d_H + c_L l(d_H)$. We define $\Gamma(j)$ as the set that includes every FC i such that $d_{ij} \leq d_H$. This set $\Gamma(j)$ represents the FCs that can potentially back up FC j in case of a disruption where $c_T d_{ij} + c_L l(d_{ij}) \leq c_V + c_L$. The FCs that are outside of $\Gamma(j)$ will not be utilized as backup, i.e., $x_{ij}^{mn*} = 0$ for each FC $i \notin \Gamma(j)$, because vendor shipment becomes more economic, i.e., $c_T(d_H + \varepsilon) + c_L l(d_H + \varepsilon) > c_V + c_L$ for any $\varepsilon > 0$. Also, recall from Proposition 1 that operational FCs always first serve their own demand region when $\partial^2 l(d_{ij})/\partial d_{ij}^2 = 0$, i.e., $x_{ii}^{mn*} = D_i$ for each FC $i \notin \Lambda(n)$. Therefore, combined with (3), we can conclude that a disrupted region j receives vendor shipment if its demand D_j is greater than the total excess inventory of the operational FCs within d_H miles, denoted $\sum_{i \neq j, i \notin \Lambda(n), i \in \Gamma(j)} K_i$. ❑

Proof of Proposition 3. Following from $d_{21} < d_{31} \leq d_H$, we have

$$\partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{21}^{mn} < \partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{31}^{mn} \leq 0.$$

This implies that FC 2 is the primary backup option for the disrupted region 1. Therefore, the excess inventory at FC 2 (i.e., K_2) should certainly be utilized to partially recover the demand in region 1 (i.e., $D_1 > K_2$). For the remaining demand $D_1 - K_2$, there are two possible strategies that the firm can follow. In the first strategy, FC 2 serves its own region with the dedicated portion of its inventory (i.e., D_2) while the excess inventory at FC 3 (i.e., K_3) is utilized to make shipments directly to the disrupted region 1. The following expression

$$\partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{22}^{mn} + \partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{31}^{mn} \tag{24}$$

represents the incremental cost associated to this strategy which we refer to as the loyal policy. In the second strategy, FC 2 abandons its own region and uses its dedicated inventory to make shipments to the disrupted region 1 while FC 3 uses its excess inventory K_3 to serve the abandoned customers in region 2. The following expression

$$\partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{21}{}^{mn} + \partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{32}{}^{mn} \tag{25}$$

represents the incremental cost associated to this strategy which we refer to as the abandonment policy. Note that, following from $\{d_{21}, d_{32}, d_{31}\} \leq d_H$, $d_{22} = 0$ and $l(d_{22}) = 0$, both (24) and (25) are negative meaning that both policies can decrease the cost. The optimal strategy is the loyal policy if (24) ≤ (25) which is the case when (8) holds. This implies that $x_{21}{}^{mn*} = K_2$, $x_{22}{}^{mn*} = D_2$, $x_{32}{}^{mn*} = 0$, $x_{31}{}^{mn*} = \min\{D_1 - K_2, K_3\}$. The optimal strategy is the abandonment policy if (24) > (25) which is the case when (8) does not hold. This implies that $x_{21}{}^{mn*} = \min\{D_1, K_2 + \min\{D_2, K_3\}\}$, $x_{22}{}^{mn*} = (D_2 - K_3)^+$, $x_{32}{}^{mn*} = \min\{D_2, K_3\}$, $x_{31}{}^{mn*} = \min\{(D_1 - K_2 - \min\{D_2, K_3\})^+, (K_3 - D_2)^+\}$.

It is worth noting that $l(d_{31}) - l(d_{32}) - l(d_{21})$ in (8) increases with the degree of convexity of the lateness function $l(d_{ij})$ because $\{d_{21}, d_{32}\} < d_{31}$. Therefore, (8) does not hold if the lateness function $l(d_{ij})$ is sufficiently convex. Otherwise, (8) holds, i.e., when the lateness function $l(d_{ij})$ is linear, concave or weakly convex.

Note that, regardless of (8), FC 3 always serves its own region with the dedicated portion of its inventory (i.e., D_3) because

$$\partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{33}{}^{mn} + \partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{21}{}^{mn} < \partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{31}{}^{mn} + \partial \Psi_{2,mn}\left(\bar{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right)/\partial x_{23}{}^{mn}$$

due to $d_{21} < d_{31}$, $d_{33} = 0$ and $l(d_{33}) = 0$. This implies that $x_{33}{}^{mn*} = D_3$ and $x_{23}{}^{mn*} = 0$. ❑

Proof of Proposition 4. The risk-neutral second stage, described with (2), (3), (4) and (6), is a transportation problem where operational FCs along with the vendors have sufficient supply to meet the demand in all regions. Therefore, there always exists an optimal solution to the risk-neutral problem in stage 2. We define $\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$ as the dual price of the supply constraint (3) for FC i evaluated at $\overline{\mathbf{K}}$ under a given (m, n). Dual price $\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$ fetches a negative value if (3) is binding; otherwise, it is zero. The value of $\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$ is a non-decreasing step function in K_i.

The first-order derivative of the stage-1 objective function (1) with respect to K_i can be written as

$$\partial \Psi_1\left(\overline{\mathbf{K}}\right)/\partial K_i = c_K + \sum_{m=1}^{M}\sum_{n=1}^{N} p_{mn}\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$$

where the second term is the weighted average of the dual price values for FC i. At high values of K_i, the second term will drop to zero. Therefore, $\partial \Psi_1\left(\overline{\mathbf{K}}\right)/\partial K_i$ will fetch its maximum value of c_K which is positive. This guarantees that the minimization problem in stage 1 is bounded and there exists an optimal solution. ❑

Proof of Proposition 5. Part (a): Left-hand side of (8) can be interpreted as the change in lateness due to a switch from Policy L to Policy A. If $l(d_{31}) - l(d_{32}) - l(d_{21})$ is positive (negative), lateness decreases (increases) when switched from Policy L to Policy A for given $\overline{\mathbf{K}}$. Recall that the service-at-risk constraint (5) is defined over late deliveries. Therefore, if $l(d_{31}) - l(d_{32}) - l(d_{21}) > 0$, then Policy A leads to a less tight (5) than Policy L following from

$$\sum_j \sum_i l\left(d_{ij}\right) x_{ij}^A < \sum_j \sum_i l\left(d_{ij}\right) x_{ij}^L$$

where x_{ij}^A and x_{ij}^L denote the optimal Policy A and Policy L decisions, respectively. On the other hand, if $l(d_{31}) - l(d_{32}) - l(d_{21}) \leq 0$, then Policy L leads to a less tight (5) than Policy A following from

$$\sum_j \sum_i l\left(d_{ij}\right) x_{ij}^L \leq \sum_j \sum_i l\left(d_{ij}\right) x_{ij}^A .$$

Part (b): Policy A being optimal in the risk-neutral setting indicates that it is more economic than Policy L. Following from Proposition 3, this implies that $[l(d_{31}) - l(d_{32}) - l(d_{21})] > (c_T/c_L)[d_{32} + d_{21} - d_{31}] > 0$. Since we have $l(d_{31}) - l(d_{32}) - l(d_{21}) > 0$, Policy A leads to a less tight (5) than Policy L as shown above. Therefore, Policy A continues to be optimal in the risk-averse setting.

Part (c): Policy L being optimal in the risk-averse setting indicates that it leads to a less tight (5) than Policy A. This implies that $[l(d_{31}) - l(d_{32}) - l(d_{21})] \leq 0$ as shown above. Therefore, we have $[l(d_{31}) - l(d_{32}) - l(d_{21})] < (c_T/c_L)[d_{32} + d_{21} - d_{31}]$ which indicates that Policy L is more economic than Policy A as shown in Proposition 3. Therefore, Policy L continues to be optimal in the risk-neutral setting.

Part (d): In part (b), we established that Policy A is always optimal when $[l(d_{31}) - l(d_{32}) - l(d_{21})] > (c_T/c_L)[d_{32} + d_{21} - d_{31}] > 0$. In part (c), we established that Policy L is always optimal when $[l(d_{31}) - l(d_{32}) - l(d_{21})] \leq 0 < (c_T/c_L)[d_{32} + d_{21} - d_{31}]$. We next examine the case when $0 < [l(d_{31}) - l(d_{32}) - l(d_{21})] \leq (c_T/c_L)[d_{32} + d_{21} - d_{31}]$. In this case, Policy L is more economic, however, Policy A leads to a less tight (5). Let us characterize the risk neutral optimal distribution policies in the following four sets:

$\Omega_{LS} = \{[l(d_{31}) - l(d_{32}) - l(d_{21})] > (c_T/c_L)[d_{32} + d_{21} - d_{31}] > 0$ and (5) satisfied: Policy L is optimal in the risk-neutral and risk averse settings$\}$

$\Omega_{LN} = \{[l(d_{31}) - l(d_{32}) - l(d_{21})] > (c_T/c_L)[d_{32} + d_{21} - d_{31}] > 0$ and (5) is not satisfied: Policy L is optimal in the risk-neutral, but it violates SaR constraint in the risk-averse setting$\}$

$\Omega_{AS} = \{[l(d_{31}) - l(d_{32}) - l(d_{21})] \leq 0 < (c_T/c_L)[d_{32} + d_{21} - d_{31}]$ and (5) satisfied: Policy A is optimal in the risk-neutral and risk averse settings$\}$

Ω_{AN} = {$[l(d_{31}) - l(d_{32}) - l(d_{21})] \leq 0 < (c_T/c_L)[d_{32} + d_{21} - d_{31}]$ and (5) is not satisfied: Policy A is optimal in the risk-neutral, but it violates SaR constraint in the risk-averse setting}.

Thus,

P[Policy A is optimal] in the risk-neutral setting = $P[\Omega_{AS}] + P[\Omega_{AS}]$.

In the set Ω_{AN}, when the SaR constraint (5) is violated, the optimal solution in the risk-averse setting will not switch to the loyal allocation policy; rather, it will adjust quantity decisions in the same allocation scheme. On the other hand, Proposition 3 establishes that, when the SaR constraint (5) is violated in the set Ω_{LN}, the optimal policy switches from Policy L to Policy A. Thus, the probability set that Policy A is optimal in the risk-averse setting is:

P[Policy A is optimal] in the risk-neutral setting = $P[\Omega_{AS}] + P[\Omega_{AS}] + P[\Omega_{LN}]$

$>$ P[Policy A is optimal] in the risk-neutral setting = $P[\Omega_{AS}] + P[\Omega_{AS}]$.

The above proof implies that, even though Policy L can be optimal in the risk-neutral setting, SaR constraint in (5) can cause a switch to Policy A in the risk-averse setting, leading to a higher probability that Policy A is the optimal distribution allocation in the risk-averse setting. ❑

Proof of Proposition 6. Condition (C1) states that the risk-neutral solution is $(K_1^*, K_2^*, K_3^*) = (K_1^N, K_2^N, 0)$ such that $\{K_1^N, K_2^N\} > 0$. We have a total of four (m, n) pairs: $(L, 1)$, $(L, 2)$, $(L, 3)$, $(H, 4)$. In the risk-averse setting, the risk-neutral solution $(K_1^N, K_2^N, 0)$ does not comply with the risk constraint (5) for $(m, n) = (H, 4)$ following from (C2) where ε is a small positive infinitesimal quantity. Again from (C2), we see that the risk-constraint (5) for $(m, n) \in \{(L, 1), (L, 2), (L, 3)\}$ never becomes binding. In order to comply with (5) for $(m, n) = (H, 4)$, following from $l(\min\{d_{31}, d_{32}\}) < 1$, the only viable action is to increase the value of K_3 to K_3^A which solves $[l(\min\{d_{31}, d_{32}\})K_3^A + (D_1 + D_2 - K_3^A)]\tau_\alpha^{H4} = \beta$; this can be written as $K_3^A = \varepsilon/[(1 - l(\min\{d_{31}, d_{32}\}))\tau_\alpha^{H4}]$. This causes a deviation from the risk-neutral optimal decision. Condition (C3) states that, at this new solution (K_1^N, K_2^N, K_3^A), the first-order derivatives with respect to K_1 and K_2 are positive: $\{\partial \Psi_1(\overline{\mathbf{K}})/\partial K_1 \mid_{(K_1^N, K_2^N, K_3^A)}, \partial \Psi_1(\overline{\mathbf{K}})/\partial K_2 \mid_{(K_1^N, K_2^N, K_3^A)}\} \geq 0$. This means that the values of both decisions should decrease. Recall that the risk-neutral solution in (C1) yields $K_1 + K_3 = K_1^N$ and $K_2 + K_3 = K_2^N$. Following from the proof of Proposition 4, (1) is piecewise linear in K_i. Therefore, increasing K_3 from zero to K_3^A should be accompanied by the same amount of decrease in K_1 and K_2 each. As a result, the risk-averse optimal solution is $(K_1^*, K_2^*, K_3^*) = (K_1^N - K_3^A, K_2^N - K_3^A, K_3^A)$. ❑

Proof of Proposition 7.

(a) Using the same set of x_{ij}^{mn} values that satisfy the VaR constraint in (11) implies that

$$P\left[\left(\left(\frac{c_L}{c_L + c_V}\right)\sum_j \sum_i l\left(d_{ij}\right)x_{ij}^{mn} + \sum_j\left(D_j - \sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn} > \frac{\beta_V^{=}}{\left(c_L + c_V\right)}\right] = 1 - \alpha\ .$$

Note that $\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\left(\left(\frac{c_L}{c_L+c_V}\right)\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}$ because $c_L/(c_L+c_V)<1$. If $\beta\leq(\beta_V^{=}/(c_L+c_V))$, it can be seen that the SaR probability in (5) is violated, i.e.,

$$P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\beta\right]$$

$$>P\left[\left(\left(\frac{c_L}{c_L+c_V}\right)\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\frac{\beta_V^{=}}{\left(c_L+c_V\right)}\right]=(1-\alpha).$$

(b) Using the same set of x_{ij}^{mn} values that satisfy the VaR constraint in (11) implies that

$$P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\left(\frac{c_L+c_V}{c_L}\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\frac{\beta_V^{=}}{c_L}\right]=1-\alpha .$$

Note that $\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}<\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\left(\frac{c_L+c_V}{c_L}\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}$ because $(c_L+c_V)/c_L>1$. If $\beta\geq(\beta_V^{=}/c_L)$, it can be seen that the SaR probability in (5) is also satisfied, i.e.,

$$P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\beta\right]$$

$$<P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\left(\frac{c_L+c_V}{c_L}\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\frac{\beta_V^{=}}{c_L}\right]=(1-\alpha).\ \square$$

Proof of Corollary 1.

(a) Using the same set of x_{ij}^{mn} values that satisfy the SaR probability in (5) implies that

$$P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\beta^{=}\right]=1-\alpha .$$

Note that $\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}<\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\left(\frac{c_L+c_V}{c_L}\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}$ because $(c_L+c_V)/c_L>1$. If $\beta_V\leq c_L\beta^{=}$, it can be seen that the VaR probability in (11) is violated, i.e.,

$$P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\left(\frac{c_L+c_V}{c_L}\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\frac{\beta_V}{c_L}\right]>$$

$$P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\beta^{=}\right]=(1-\alpha).$$

(b) Using the same set of x_{ij}^{mn} values that satisfy the SaR constraint in (5) implies that

$$P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\beta^{=}\right]=1-\alpha .$$

Note that $\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}<\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\left(\frac{c_L+c_V}{c_L}\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}$ because $(c_L+c_V)/c_L>1$. If $\beta_V\geq(c_L+c_V)\beta^{=}$, it can be seen that the VaR probability in (11) is also satisfied, i.e.,

$$P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\left(\frac{c_L+c_V}{c_L}\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\frac{\beta_V}{c_L}\right]<$$

$$P\left[\left(\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}+\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\right)\tilde{\tau}^{mn}>\beta^{=}\right]=(1-\alpha).\ \square$$

Proof of Proposition 8. (a) The second-stage objective function is a special form of minimizing CVaR. Let us rewrite the objective function here.

$$\min_{\vec{\mathbf{x}}^{mn}\geq\vec{0}}\Psi_{2,mn}\left(\vec{\mathbf{x}}^{mn}\middle|\overline{\mathbf{K}}\right)=\begin{bmatrix}c_T\sum_j\sum_i d_{ij}x_{ij}^{mn}+c_L\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}\\+\left(c_V+c_L\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\end{bmatrix}E\left[\tilde{\tau}^{mn}\right].$$

The above second-stage objective function is equivalent to minimizing a CVaR objective function where the tolerable loss β_C and the tolerated loss probability α are set to zero, i.e.,

$$\min_{\vec{\mathbf{x}}^{mn}\geq\vec{0}}E\left[\begin{pmatrix}c_T\sum_j\sum_i d_{ij}x_{ij}^{mn}+c_L\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}\\+\left(c_V+c_L\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\end{pmatrix}\tilde{\tau}^{mn}\middle|\Psi_{2,mn}\left(\vec{\mathbf{x}}^{mn}\middle|\overline{\mathbf{K}}\right)\geq\beta_C=0\right]=\min_{\vec{\mathbf{x}}^{mn}\geq\vec{0}}\Psi_{2,mn}\left(\vec{\mathbf{x}}^{mn}\middle|\overline{\mathbf{K}}\right).$$

Given that $\beta_C = 0$ and $\alpha = 0$, the above CVaR minimization is the most stringent form of CVaR risk measure.

For a given set initial inventory commitment vector $\overline{\mathbf{K}}$, we define the optimal solution of a delivery configuration as $\vec{\mathbf{x}}^{mn*}$. Let us also define β_C^0 as follows:

$$\beta_C^0=E\left[\begin{pmatrix}c_T\sum_j\sum_i d_{ij}x_{ij}^{mn*}+c_L\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn*}\\+\left(c_V+c_L\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn*}\right)\end{pmatrix}\tilde{\tau}^{mn}\right].$$

The above optimal solution would minimize CVaR even in the most stringent CVaR risk measure with $\beta_C = 0$ and $\alpha = 0$, corresponding to all possible disruption length scenarios.

(b) We next compare the optimal solution developed under a SaR constraint with one developed under a CVaR risk measure with $\beta_C > 0$ and $\alpha = 0$.

$$\min_{\vec{\mathbf{x}}^{mn}\geq\vec{0}}E\left[\begin{pmatrix}c_T\sum_j\sum_i d_{ij}x_{ij}^{mn}+c_L\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}\\+\left(c_V+c_L\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\end{pmatrix}\tilde{\tau}^{mn}\middle|\Psi_{2,mn}\left(\vec{\mathbf{x}}^{mn}\middle|\overline{\mathbf{K}}\right)\geq\beta_C=0\right]$$

$$=\min_{\vec{\mathbf{x}}^{mn}\geq\vec{0}}E\left[\begin{pmatrix}c_T\sum_j\sum_i d_{ij}x_{ij}^{mn}+c_L\sum_j\sum_i l\left(d_{ij}\right)x_{ij}^{mn}\\+\left(c_V+c_L\right)\sum_j\left(D_j-\sum_i x_{ij}^{mn}\right)\end{pmatrix}\tilde{\tau}^{mn}-\beta_C\right]$$

$$= \min_{\vec{\mathbf{x}}^{mn} \geq \vec{0}} \Psi_{2,mn}\left(\vec{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right) - \beta_C.$$

Because β_C is a constant term, minimizing CVaR with $\beta_C > 0$ and $\alpha = 0$ leads to the same optimal delivery solution with the objective function in (2). Thus, minimizing the CVaR for a positive tolerated loss is equivalent to minimizing CVaR with a fixed term of $-\beta_C$ subtracted from the derived optimal objective function value. Note that this result holds true for all $\beta_C \leq \beta_C^0$; otherwise, no solution is feasible under both criteria.

(c) Consider CVaR minimization with $\beta_C = 0$ and $\alpha > 0$. Let $\tau_{(1-\alpha)}$ represent the point where the distribution of $\tilde{\tau}$ is at $(1 - \alpha)$ percentile, i.e., $P\left[\tilde{\tau} \middle| \tilde{\tau} \leq \tau_{(1-\alpha)}\right] = (1 - \alpha)$. It should be observed that the optimal solution of a delivery configuration as $\vec{\mathbf{x}}^{mn*}$ is a feasible solution for the CVaR minimization where

$$\min_{\vec{\mathbf{x}}^{mn} \geq \vec{0}} E\left[\left(\begin{array}{c} c_T \sum_j \sum_i d_{ij} x_{ij}^{mn} + c_L \sum_j \sum_i l\left(d_{ij}\right) x_{ij}^{mn} \\ +\left(c_V + c_L\right) \sum_j \left(D_j - \sum_i x_{ij}^{mn}\right) \end{array}\right) \tilde{\tau}^{mn} \middle| \Psi_{2,mn}\left(\vec{\mathbf{x}}^{mn} \middle| \overline{\mathbf{K}}\right) \geq \beta_C = 0 \;\&\; \tilde{\tau}^{mn} \leq \tau_{(1-\alpha)}\right]$$

because the same as $\vec{\mathbf{x}}^{mn*}$ solution for all $\tilde{\tau}$ realizations including those where $\tilde{\tau} > \tau_{(1-\alpha)}$. In conclusion, an optimal solution that complies with SaR constraint in stage 2 is also optimal for CVaR with all tolerated losses of β_C and is feasible for all tolerated loss probabilities of $(1 - \alpha)$, making the same solution to be compliant with CVaR objective. However, the solution developed under a SaR constraint cannot be guaranteed to be the optimal solution for a CVaR minimization with $\beta_C > 0$ and $\alpha > 0$. ❑

Corollary A1 (*to Proposition* 3). *Suppose FC* 1 *is disrupted, i.e.,* $1 \in \Lambda(n)$, *and FCs* 2 *and* 3 *are operational, i.e.,* $\{2, 3\} \notin \Lambda(n)$ *where* $\{d_{21}, d_{32}\} < d_{31} \leq d_H$. *Consider the scenario when the realized demand* D_1^r *in FC* 1 *is greater than the excess inventory in FC* 2.

(a) (*Loyal Policy*): *If* (8) *holds*, *operational FCs* 2 *and* 3 *prioritize serving their own customers.*

(b) (*Abandonment Policy*): *If* (8) *does not hold, FC* 2 *prefers to abandon its customers in region* 2 *in order to serve the customers of FC* 1 *while FC* 3 *serves the customers that FC* 2 *abandons.*

Proof of Corollary A1. The proof follows from the proof of Proposition 3 by replacing D_1, D_2 and D_3 with $D_1^{\,r}$, $D_2^{\,r}$ and D_3^r. ❑

Proof of Proposition 9. (a) For a given (m, n), $\sum_{i \notin \Lambda(n)} K_i + \overline{D}_i$ is the total amount of supply at the operational FCs. In stage 2, if $\sum_j D_j^r \geq \sum_{i \notin \Lambda(n)} K_i + \overline{D}_i$, the dual price of the supply constraint (3) remains to have the same structure as in the deterministic demand case, i.e., $\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$ is a non-decreasing, non-positive step function in $\overline{\mathbf{K}}$. If $\sum_j D_j^r < \sum_{i \notin \Lambda(n)} K_i + \bar{D}_i$, then the dual price of the supply constraint (3) becomes zero for each i. Thus, we have

$$\partial \Psi_1\left(\overline{\mathbf{K}}\right)/\partial K_i = c_K + \sum_{m=1}^{M}\sum_{n=1}^{N}\left(p_{mn}\theta_i^{mn}\left(\overline{\mathbf{K}}\right)\left(1 - G\left(\sum_{i \notin \Lambda(n)} K_i + \bar{D}_i\right)\right)\right).$$

Following from $G(.)$ being $0 \le G(\cdot) \le 1$ by definition and $\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$ being non-positive, we have $\left|\lambda_i^{mn}\left(\overline{\mathbf{K}}\right)\right| \le \left|\theta_i^{mn}\left(\overline{\mathbf{K}}\right)\right|$ for all (m, n).

(b) The proof follows from part (a). The value of (14) reaches zero faster than the value of (9) as K_i increases. This leads to $\sum_i K_i^{S1*} \le \sum_i K_i^{D*}$. ❑

Development of Corollary A2:

We next show that our main result where the total inventory commitment can be decreasing under risk aversion continues to hold under stochastic demand in the setting where demand uncertainty is revealed before shipment decisions. In a three FC setting, this finding occurs when there is a disruption probability that impacts two nearby facilities at the same time. It is important to observe that the SaR constraint (5) is now required to be satisfied in the worst-case scenario which corresponds to the case when (i) disrupted FCs 1 and 2 have realized demand values at their corresponding upper support, i.e., $D_1^r = D_1^H$ and $D_2^r = D_2^H$, and (ii) the operational FC 3 has no leftover inventory after satisfying its own customers' demand D_3^r. For simplicity in item (ii), let us consider the case when the demand in FC 3 has a point distribution, i.e., $D_3^r = \bar{D}_3$, with probability 1 such that, after satisfying its own customers there is no excess inventory left in FC3. This represents that Condition (C1) developed in Section 5.3 holds in this scenario. We next adjust Condition (C2) to reflect the realizations of demand uncertainty in FCs 1 and 2:

$$(\text{C2}'): (D_1^H + D_2^H)\tau_\alpha^{H4} = \beta + \varepsilon > \beta \ge \{D_1^H\tau_\alpha^{L1}, D_2^H\tau_\alpha^{L2}, \bar{D}_3\, \tau_\alpha^{L3}\}$$

where ε is a small positive infinitesimal quantity representing the switch from risk-neutral setting to risk-averse setting. The revised Condition (C2′) enforces the substitution effects described earlier in satisfying the SaR constraint to hold true in the worst-case scenario where $D_1^r = D_1^H$ and $D_2^r = D_2^H$. We also adjust Condition (C3) as follows:

$$(\text{C3}'): |p_{L2}\lambda_1^{L2}(K_1^N, K_2^N, K_3^A) + p_{L3}\lambda_1^{L3}(K_1^N, K_2^N, K_3^A)| \le c_K,$$
$$|p_{L1}\lambda_2^{L1}(K_1^N, K_2^N, K_3^A) + p_{L3}\lambda_2^{L3}(K_1^N, K_2^N, K_3^A)| \le c_K$$

The revised Condition (C3′) under demand uncertainty accounts for the risk of overstocking. It is worth mentioning that Condition (C3′) is more likely to hold than Condition (C3) following from Proposition 9(a).

Using the above-described worst-case scenario, Corollary A2 shows that our main finding where the total inventory commitment in stage 1 can be decreasing with risk aversion continues to exist under stochastic demand in the setting where demand uncertainty is revealed before shipment decisions.

Corollary A2 (*to Proposition* 6). *Suppose* $l(\min\{d_{31}, d_{32}\}) < 1$. *Let* (C1), (C2′), (C3′) *hold and let* $D_3^r = \bar{D}_3$ *with probability* 1. *The total inventory commitment under risk aversion is smaller than the total risk neutral inventory commitment.*

Proof of Corollary A2. The proof follows from the proof of Proposition 6 by replacing (i) conditions (C2) and (C3) with conditions (C2′) and (C3′); (ii) D_1, D_2 with random variables $\tilde{D}_1$ and $\tilde{D}_2$; and (iii) satisfying the SaR constraint in (5) at the worst-case scenario where $D_1^r = D_1^H$, $D_2^r = D_2^H$ and $D_3^r = \bar{D}_3$.

❑

Corollary A3 (*to Proposition* 3 *and Corollary* A1). *Suppose FC* 1 *is disrupted, i.e.,* $1 \in \Lambda(n)$, *and FCs* 2 *and* 3 *are operational, i.e.,* $\{2, 3\} \notin \Lambda(n)$ *where* $\{d_{21}, d_{32}\} < d_{31} \leq d_H$. *Consider the scenario when the base stock* BS_1 *in FC* 1 *is greater than the base stock* BS_2 *in FC* 2.

(a) (*Loyal Policy*): *If* (8) *holds*, *operational FCs* 2 *and* 3 *prioritize serving their own customers.*

(b) (*Abandonment Policy*): *If* (8) *does not hold, FC* 2 *prefers to abandon its customers in region* 2 *in order to serve the customers of FC* 1 *while FC* 3 *serves the customers that FC* 2 *abandons.*

Proof of Corollary A3. The proof follows from the proofs of Proposition 3 (and Corollary A1) by replacing D_1, D_2 and D_3 (and D_1^r, D_2^r and D_3^r in Corollary A1) with BS_1, BS_2 and BS_3. ❑

Corollary A4 (*to Proposition* 4). *The risk-neutral problem described with* (1), (17), (18), (19) *and* (6) *is bounded and there exists an optimal solution.*

Proof of Corollary A4. The proof is similar to that of Proposition 4. The risk-neutral second stage problem is described with the objective function in (17) subject to constraints in (18), (19) and (6). This second-stage formulation is a transportation problem where operational FCs along with the vendors have sufficient supply to meet the demand in all regions. Therefore, there always exists an optimal solution to the risk-neutral problem in stage 2. We define $\gamma_i^{mn}\left(\vec{\mathbf{K}}\right)$ as the dual price of the supply constraint (18) for FC i evaluated at $\vec{\mathbf{K}}$ under a given network (m, n). Dual price $\gamma_i^{mn}\left(\vec{\mathbf{K}}\right)$ is continuously non-decreasing in K_i because

$$\frac{\partial \Psi_{2,mn}\left(\vec{\mathbf{x}}^{mn} \middle| \vec{\mathbf{K}}\right)}{\partial x_{ij}^{mn}} = E\left[\tilde{\tau}^{mn}\right]\left[c_T d_{ij} - c_L\left(1 - l\left(d_{ij}\right)\right)\left(1 - F_j\left(\sum_i \left(1 - l\left(d_{ij}\right)\right) x_{ij}^{mn}\right)\right) - c_V\left(1 - F_j\left(\sum_i x_{ij}^{mn}\right)\right)\right]$$

is continuously non-decreasing in x_{ij}^{mn} where $F_j(\cdot)$ denotes the cumulative distribution function of random demand $\tilde{D}_j$. $\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)$ fetches a negative value if (18) is binding; otherwise, it is zero. The first-order derivative of the stage-1 objective function (1) with respect to K_i can be written as

$$\partial \Psi_1\left(\overline{\mathbf{K}}\right)/\partial K_i = c_K + \sum_{m=1}^{M}\sum_{n=1}^{N} p_{mn}\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)$$

where the second term is the weighted average of the dual price values for FC i. At high values of K_i, the second term will drop to zero. Therefore, $\partial \Psi_1\left(\overline{\mathbf{K}}\right)/\partial K_i$ will fetch its maximum value of c_K, which is a positive value. This guarantees that the minimization problem in stage 1 is bounded and there exists an optimal solution. ❑

Proof of Proposition 10. (a) We have (21) ≤ (22) because both $F_j\left(\sum_i\left(1-l\left(d_{ij}\right)\right)x_{ij}^{mn}\right)$ and $F_j\left(\sum_i x_{ij}^{mn}\right)$ have values between 0 and 1 where $F_j(\cdot)$ denotes the cumulative distribution function of random demand $\tilde{D}_j$. $\theta_i^{mn}\left(\overline{\mathbf{K}}\right) \le \gamma_i^{mn}\left(\overline{\mathbf{K}}\right)$ follows from (21) ≤ (22). Since both $\theta_i^{mn}\left(\overline{\mathbf{K}}\right)$ and $\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)$ are non-positive, we have $\left|\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)\right| \le \left|\theta_i^{mn}\left(\overline{\mathbf{K}}\right)\right|$.

(b) The proof follows from part (a). The value of (20) reaches zero faster than the value of (9) as K_i increases. This leads to $\sum_i K_i^{S2*} \le \sum_i K_i^{D*}$. ❑

<u>Development of Lemma A1 and Corollary A5:</u>

We next examine the impact of risk aversion under demand uncertainty in the setting where shipment decisions are made before observing the customer demand. In what follows, we show that our main finding pertaining to the decrease in total inventory commitment with risk aversion continues to hold under demand uncertainty where random demand is revealed after shipment decisions are made. Moreover, we show that one of the three optimality conditions that led to this finding (in Section 5.3) is always satisfied in this new setting.

In our earlier analysis, Proposition 6 has established that the firm's first-stage stocking decisions can be decreasing with increasing risk aversion. This proposition has relied on three conditions. We retain Condition (C1) in the same way, however, we adjust Condition (C2) in order to accommodate the revised SaR constraint in (16). Recall that the original form of Condition (C2) in (10) accounts for the cumulative demand during broad-impact and narrow-impact disruptions. We revise this condition using the random variable $\tilde{\zeta}_j^{mn}$ describing the cumulative random demand during the disruption with random length at the disrupted FC j. Let $\zeta_{j\alpha}^{mn}$ represent the α-percentile (corresponding to the right tale of the distribution) of random cumulative demand during disruption. The firm is concerned about late deliveries when the cumulative demand exceeds the α-percentile. We describe the revised condition with (C2″) and express it as follows:

$$\text{(C2}''\text{)}: \ \zeta_{1\alpha}^{H4} + \zeta_{2\alpha}^{H4} = \beta + \varepsilon > \beta \geq \left\{ \zeta_{1\alpha}^{L1}, \zeta_{2\alpha}^{L2}, \zeta_{3}^{L3} \right\}$$

where ε is a small positive infinitesimal quantity representing the switch from risk-neutral setting to risk-averse setting. Similar to the earlier analysis, Condition (C2″) makes two statements. First, it guarantees that the SaR constraint in (16) is never violated for narrow-impact disruptions where $(m, n) \in \{(L, 1), (L, 2), (L, 3)\}$. Second, it ensures that, when risk aversion is incorporated into the decision making process, the risk-neutral solution defined in Condition (C1) marginally violates the SaR constraint in (16) during broad-impact disruption $(m, n) = (H, 4)$, halting operations at both FC 1 and FC 2. In this case, the only viable action to satisfy (16) for $(m, n) = (H, 4)$ is to increase inventory at the distant facility FC 3.

We next revise Condition (C3) to account for this new setting where random demand is realized after delivery decisions are made. The revised Condition (C3″) is as follows:

$$\text{(C3}''\text{)}: |p_{L2}\gamma_1^{L2}(K_1^N, K_2^N, K_3^A) + p_{L3}\gamma_1^{L3}(K_1^N, K_2^N, K_3^A)| \leq c_K,$$
$$|p_{L1}\gamma_2^{L1}(K_1^N, K_2^N, K_3^A) + p_{L3}\gamma_2^{L3}(K_1^N, K_2^N, K_3^A)| \leq c_K$$

where K_3^A denotes the stocking decision in FC 3 satisfying (16) for $(m, n) = (H, 4)$. The following lemma shows that the revised Condition (C3″) always holds in this setting.

Lemma A1. *$|p_{L2}\gamma_1^{L2}(K_1^N, K_2^N, K_3^A) + p_{L3}\gamma_1^{L3}(K_1^N, K_2^N, K_3^A)| \leq c_K$, and $|p_{L1}\gamma_2^{L1}(K_1^N, K_2^N, K_3^A) + p_{L3}\gamma_2^{L3}(K_1^N, K_2^N, K_3^A)| \leq c_K$.*

Proof of Lemma A1. Due to the continuous functional form of $\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)$, the risk neutral solution provided in Condition (C1) satisfies the following:

$$|p_{L2}\gamma_1^{L2}(K_1^N, K_2^N, 0) + p_{L3}\gamma_1^{L3}(K_1^N, K_2^N, 0)| = c_K,$$
$$|p_{L1}\gamma_2^{L1}(K_1^N, K_2^N, 0) + p_{L3}\gamma_2^{L3}(K_1^N, K_2^N, 0)| = c_K.$$

Since $\gamma_i^{mn}\left(\overline{\mathbf{K}}\right)$ is non-decreasing in $\overline{\mathbf{K}}$ and non-positive, increasing the value of K_3 from 0 to K_3^A leads to the following:

$$|p_{L2}\gamma_1^{L2}(K_1^N, K_2^N, K_3^A) + p_{L3}\gamma_1^{L3}(K_1^N, K_2^N, K_3^A)| \leq c_K,$$
$$|p_{L1}\gamma_2^{L1}(K_1^N, K_2^N, K_3^A) + p_{L3}\gamma_2^{L3}(K_1^N, K_2^N, K_3^A)| \leq c_K$$

which satisfies Condition (C3″). ❑

The following corollary confirms that our earlier finding where the first-stage inventory commitment can be decreasing under risk aversion continues to hold under the setting with demand being uncertain at the time of delivery decisions. It is important to note that we do not need to specify Condition (C3″) in the following corollary because this condition always holds under this demand uncertainty setting. As a result, Corollary A5 requires fewer conditions than Proposition 6 and Corollary A2. This indicates that our earlier finding would become more pronounced if demand is uncertain at the time of delivery decisions.

Corollary A5 (*to Proposition* 6 *and Corollary* A2). *Suppose* $l(\min\{d_{31}, d_{32}\}) < 1$. *Let* (C1) *and* (C2″) *hold. The total inventory commitment under risk aversion is smaller than the total risk neutral inventory commitment*.

Proof of Corollary A5. The proof follows from the proof of Proposition 6 and Corollary A2 with some adjustments. Condition (C1) remains the same. Condition (C2″) replaces Condition (C2) in Proposition 6 and Condition (C2′) in Corollary A2. Lemma A1 states that Condition (C3″) (counterpart of Condition (C3) in Proposition 6 and Condition (C3′) in Corollary A2) always holds, thus, it is not needed as a requirement in Corollary A5. ❑