Operations Research

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

Compensating for Dynamic Supply Disruptions: Backup Flexibility Design

Soroush Saghafian, Mark P. Van Oyen

OPERATIONS RESEARCH
Vol. 64, No. 2, March–April 2016, pp. 390–405
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.2016.1478

Compensating for Dynamic Supply Disruptions: Backup Flexibility Design

Soroush Saghafian
Harvard Kennedy School, Harvard University, Cambridge, Massachusetts, Soroush_Saghafian@hks.harvard.edu

Mark P. Van Oyen
Department of Industrial & Operations. Engineering, University of Michigan, Ann Arbor, Michigan, vanoyen@umich.edu

To increase resilience in supply chains, we investigate the optimal design of flexibility in a backup system. We model the dynamics of disruptions as Markov chains, and consider a multiproduct, multisupplier supply chain under dynamic disruption risks. Using our model, we first show that a little flexibility in the backup system can go a long way in mitigating dynamic disruption risks. This raises an important and fundamental question in designing flexibility in the backup system: to achieve the benefits of full backup flexibility, which unreliable suppliers should be backed up? To answer this question, we connect the supply chain to various queueing and dam models by analyzing the dynamics of the inventory shortfall process. Using this connection, we show that backing up suppliers merely based on first moment considerations such as their average reliability or average product demand can be misleading. All else equal, it is better to back up suppliers with (1) longer but less frequent disruptions, and (2) lower demand uncertainty. In addition to such second moment effects, by employing the Renyi's Theorem, we demonstrate that when disruptions are relatively long (if they occur), backing up the suppliers for which the expected wasted backup capacity is minimum provides the best backup flexibility design. We also develop easy-to-compute and yet effective indices that (a) guide the supply chain designer in deciding which suppliers to backup, and (b) provide insights into the role of various factors such as inventory holding and shortage costs, purchasing costs, suppliers reliabilities, and product demand distributions in designing backup flexibility.

Keywords: backup capacity; flexibility; dynamic disruption risk; inventory shortfall process.
Subject classifications: inventory/production: multi-item/echelon/stage; dynamic programming: optimal control.
Area of review: Operations and Supply Chains.
History: Received June 2013; revisions received March 2014, February 2015, September 2015; accepted December 2015. Published online in *Articles in Advance* March 11, 2016.

1. Introduction

Some companies, like Ericsson, have learned the hard way that even minor incidents can cause disruptions of major economic consequence. For Ericsson, a very small fire in a small production cell was put out in ten minutes; however, the impact on a critical clean room resulted in a serious loss of production capacity. Ultimately, $200M in insurance compensation was paid out (see, e.g., Norrman and Jansson 2004). Although it is obvious that the frequency of supply disruptions varies greatly depending on the products involved, the business practices in place, the transportation modes, and stability of the political and infrastructural environment, disruptions are much more frequent than is commonly recognized. A recent large-scale international survey performed by Business Continuity Institute indicates that 85% of firms around the world experience at least one supply chain disruption each year, and more than 50% face between one and five (Business Continuity Institute 2011). We are particularly motivated to study effective backup flexibility designs for businesses with recurring disruptions.

Many firms, with electronics sector companies being among the leaders, are pursuing more rigorous business continuity and contingency plans. As identified in Norrman and Jansson (2004), Ericsson now frames contracts with specific attention to the identification of and plans for response at a backup site or resource. Ericsson is moving to incorporate both top tier suppliers and important subsuppliers into a risk management approach that considers both the length of disruptions (recovery time) and their financial cost (Norrman and Jansson 2004). Our study considers a microeconomic model of the disruption cost in terms of backorders (and indirectly, inventory holding costs when extra inventory is carried as a disruption mitigation strategy). We also introduce a Markov chain model of disruptions that permits heterogeneity in the rate of supplier disruptions (reliability) versus the mean length of the disruption duration (recovery time). This permits a business analytics approach that can expose the importance of business recovery time in addition to supplier reliability in predisruption planning. For example, we investigate in §4 the decision of which of two suppliers is more important to back up when one has higher availability than the other, but also longer disruption durations.

Investing in backup suppliers has been used by many leading companies. For example, Toyota used it to reduce its exposure to disruptions (see Kim 2011). Kouvelis and Li (2008, p. 184) provide an insightful description of a business environment that is particularly suited to this approach of creating an emergency backup supplier for a single high disruption risk product, stating that "Frequently, in just-in-time environments where the buyer (manufacturer) runs a continuous-flow system for high-volume low-uncertainty goods ("functional goods" . . .), the most frequent cause in creating supply-demand mismatches is not demand uncertainty but unreliable supply . . . " Although the majority of studies focusing on the benefits of creating backup capacity develop models that are essentially single-period, ours allows inventory to carry over, a mechanism that is typically used in practice to mitigate disruptions.

As Tang (2006) posits, one of the fundamental strategies for increasing the robustness of the supply chain is to increase the flexibility of the supply base. The area of operational flexibility provides a rich landscape of paradigms which could be used to increase the flexibility of the supply chain in the sense of better maintaining high service levels despite a disruption from a primary supplier. In contrast to the traditional approach of providing an inflexible backup supplier, we note that either full or partial flexibility in the backup supplier can be more economically attractive than the traditional use of "dedicated" backups. Considering this, firms are increasingly thinking about ways to create backup capacity in a flexible manner as an alternative to having a dedicated backup supplier for every product or carrying extra inventories, which are both expensive practices. It is often desirable (e.g., in the case of circuit board production or semiconductor manufacturing) to have a single pooled flexible backup supplier that is capable of ensuring supply continuity (at some capacity level) for multiple products in the event that one or more primary suppliers are disrupted. This type of approach has received some attention to date (see, e.g., Tomlin and Tang 2008, Saghafian and Van Oyen 2012).

It should be noted that process designs for a backup supplier's operation may be quite different than that for the primary supplier because of issues of volume and utilization. Particularly, when the volumes are large, primary suppliers are likely to have assembly line processes, which provide economy of scale at high volume but are usually expensive to make flexible. Unless the primary supplier has very frequent disruptions, production is required only intermittently from the backup supply infrastructure, and thus high volume assembly lines in the backup may not be economical. Rather, it is more likely that such backup suppliers have job shop, batch, or reconfigurable production processes, all of which are more flexible than automation-intensive assembly lines (and could easily process a wide range of products in areas such as circuit board assembly or semiconductor manufacturing). The higher the reliability of the primary supplier, the lower the average demand for the backup supplier and the greater the barrier to justifying the investment in the backup supplier. The above issues justify the "economy of scope" that can be harnessed by investing in *flexible backup* manufacturing infrastructure.

Although attractive, injecting full flexibility into the backup system is typically impossible due to various technological and economical burdens. The choice to pool capacity in a backup supplier when full flexibility is impossible, poses an interesting and fundamental predisruption risk management question: which suppliers should be backed up by the flexible backup capacity? In fact, the optimal design of backup flexibility is an important question in practice, which has not received enough attention in the academic literature. It is our goal in this paper to fill this gap: we attempt to generate insights into effective ways of designing flexibility in a supply chain backup system. Importantly, we first show that a little flexibility in the backup system can go a long way in mitigating dynamic disruptions, suggesting that partial flexibility can effectively achieve the benefits of full flexibility for mitigating dynamic disruption risks. However, to achieve such benefits, the partial flexibility should be designed intelligently. This, of course, is inextricably linked with the ability to carry inventory over time and update inventory safeguards for different products as an alternative way to mitigate disruptions.

To address the optimal design of flexibility in the backup system, we consider a supply chain with limited backup capacity, multiple products, and multiple unreliable suppliers. We allow for disruption risks to dynamically change over time, and model the dynamics of a supplier's disruption risk as a Discrete Time Markov Chain (DTMC)[1] with several threat levels indicating the "health level" of suppliers. As one example, the S&P credit risk rating system with states $\{1 = \text{AAA}, 2 = \text{AA}, 3 = \text{A}, 4 = \text{BBB}, 5 = \text{B/BB}, 6 = \text{CCC/CC/C}\} \cup \{0 = \textit{Default}\}$ is an analogous system for which Markov chain modeling is commonly used. We analyze the inventory shortfall process (the difference between desirable inventory safeguards and inventory levels, a potentially positive quantity because of a limited capacity when primary suppliers are disrupted) and connect it to single-server queueing systems as well as dam storage processes.

Using such a connection, we show that backing up suppliers merely based on first moment effects such as their average reliability or product demand can be misleading. We find that, all else equal, it is better to back up suppliers with (1) longer but less frequent disruptions, and (2) lower demand uncertainty. These shed light on the important role of disruption lengths and demand variabilities in design of backup flexibility. To generate further insights into the role of demand distributions (not just the first and second moments), we also consider situations where disruptions are relatively long when they occur, i.e., scenarios with long "time to recovery." By using the Renyi's Theorem (which provides an approximation for geometric random sum of i.i.d. random variables) for the inventory shortfall process under long disruptions, we show that backing up the suppliers for which the expected wasted backup capacity is minimum provides the best backup flexibility design. Computing the expected wasted backup capacity requires the supply chain designer to evaluate the cumulative distributions of the product demands at the available backup capacity. This sheds further light on the role of demand distributions (not just their averages) in the design of backup flexibility, especially in supply chains under long "time to recovery."

After generating insights into the important effects of length of disruptions, demand variability, and demand distributions, we factor out such effects, and focus on the role of other parameters such as inventory holding and shortage costs and purchasing costs. We do this by (a) considering the system under i.i.d. Bernoulli disruptions (a special case of our general Markov chain model), and (b) assuming exponential demand distributions. Under these assumptions, the dynamics of inventory shortfall process becomes equivalent to that of waiting time in a GI/M/1 queueing system. Taking advantage of this equivalence, we characterize both the optimal inventory base-stock levels and long-run average costs. In turn, this enables us to develop an easy-to-compute and yet effective index, which we term *Backup Effect Index* (*BEI*), that (a) guides the supply chain designer in deciding which suppliers to backup, and (b) provides interesting insights into the role of various factors such as inventory holding and shortage costs, purchasing costs, suppliers reliabilities, and product demands in designing backup flexibility. Our results suggest that, when demand distributions are close to exponential, following a *largest BEI policy* is optimal in deciding which unreliable suppliers to back up. We then extend this result by relaxing the exponential demand assumptions (i.e., by considering general distributions) and analyzing a GI/GI/1 queueing counterpart to the inventory shortfall process. This allows us to provide a generalized BEI (GBEI), and show that in general backing up suppliers based on a *largest GBEI first policy* provides an effective backup flexibility design, enabling supply chain designers to effectively compensate for disruption risks.

To generate further insights into the role of capacity pooling in the backup system, we also develop a numerical study and compare scenarios with dedicated backup suppliers versus a single pooled backup capacity. We find that the value of a flexible backup supplier is more than the summation of benefits that can be obtained separately for each of the products through dedicated backups: when one of the primary suppliers is in a high threat level and the other is in a low threat level, the pooled backup capacity can be shifted toward the unreliable supplier which is at a higher risk. Moreover, we observe that a firm will reserve at least as much capacity from a backup flexible supplier as the amount reserved in total from dedicated backup ones. Indeed, the flexibility of a backup supplier provides the firm with greater benefits, justifying reserving more backup capacity because of the economic advantage of shifting the orders whenever necessary. This observation sheds light on higher fees charged in practice by flexible suppliers for reserving their capacity compared to inflexible suppliers.

The remainder of the paper is organized as follows. We review the literature in the next section, then in §3 we describe our model. Section 4 generates insights into the important role of length of disruptions as well as demand variability in designing backup flexibility. Section 5 neutralizes the role of the factors studied in §4, and generates insights into the role of some other important factors. Section 6 develops a numerical study and generates insights into the capacity pooling advantage in the backup system. Section 7 summarizes the insights gained and concludes. All proofs are presented in the Online Appendix (available as supplemental material at http://www.dx.doi.org/10.1287/opre.2016.1478).

2. Literature Review and Contributions

The operational literature on supply chain disruption risks can be viewed from two different perspectives: (A) time relative to the disruption event, and (B) the way the disruption event is modeled.

From the perspective of time relative to the disruption event, Behdani et al. (2012) perform a literature review and describe how the literature can be classified in three categories: (A.1) predisruption studies, (A.2) postdisruption studies, and (A.3) integrated studies of both pre and post-disruptions. Most disruption management studies fall into the first two categories. This paper is among the very few studies to consider both pre and postdisruptions. We treat several common predisruption approaches to supply flow continuity: carrying additional inventory, investing in backup capacity, and taking into consideration the dynamics of a supplier's likelihood of disruption (i.e., dynamic supplier monitoring and assessment of its threat level). As a post-disruption mechanism, the modeling of dynamic inventory replenishment policies is especially meaningful in this paper. We consider not only inflexible backup suppliers (which also require a capacity investment decision), but also the proper dynamic use of a pooled flexible backup supplier that can serve multiple products out of its shared but limited capacity (in response to disruptions in unreliable suppliers).

Whether static (e.g., single-shot or repeated settings) or dynamic, the literature models disruptions in the following four categories: (B.1) random disruptions (i.e., all-or-nothing), (B.2) random yield, (B.3) random capacity, and (B.4) financial default. This paper is in the first category, dynamic random disruptions.

For studies that consider the case of random disruptions, we refer interested readers to Parlar and Perry (1996), Gürler and Parlar (1997), Moinzadeh and Aggarwal (1997), Arreola-Risa and DeCroix (1998), Tomlin (2006), Babich et al. (2007), Saghafian and Van Oyen (2012), and the references therein. Some studies including Wang et al. (2010) consider a combination of the above-mentioned types of disruptions. Moreover, although most studies have focused on static disruptions, a few consider dynamic of disruptions. Tomlin and Snyder (2006), for instance, develop multiperiod models with dynamic disruptions in which the firm has a single unreliable supplier, as well as models with a second, perfectly reliable supplier. Tomlin (2006) considers dynamic disruption risks in a single-product setting with two dedicated suppliers: one perfectly reliable and one unreliable. When the amount ordered from the reliable supplier is a fixed percentage of the demand in each period, Tomlin (2006) establishes the optimality of a state-dependent base-stock policy. Dong and Tomlin (2012) consider a setting that operationally resembles the disruption model of Tomlin (2006) to study the interplay between business interruption insurance and operational measures.

The inventory control literature with Markovian supply availability is also to some extent relevant to our study, although it typically studies single-sourcing models without any supply flexibility. Within this literature, Song and Zipkin (1996) present a fundamental study with periodic review inventory control where information about the evolution of the supply system is modeled as a Markov chain. Parlar et al. (1995) addresses a periodic-review setting with setup costs, where the probability that an order placed now is filled in full depends on whether supply was available in the previous period (see also Özekici and Parlar 1999). We contribute a new perspective by using the connections between the dynamics of inventory shortfall process (under dynamic disruptions) and various queueing and dam storage processes.

Another part of the literature includes multiperiod models with repeated (but not dynamic) disruptions. Tomlin (2009) uses a Bayesian approach for supply learning (i.e., reliability-forecast updating) with i.i.d. Bernoulli disruptions and characterizes the firm's optimal sourcing and inventory decisions. Anupindi and Akella (1993) study a finite-horizon, discrete-time continuous demand model with two zero lead-time random-yield suppliers.

In addition to considering a multiperiod dynamic disruption model with an arbitrary number of suppliers, another distinct feature of our modeling framework is that we allow for product mix flexibility in the backup system, and study how such flexibility can be used to selectively supplement the production of primary suppliers based on the firm's inventory levels and suppliers' threat levels. Operational mix flexibility has been studied in various papers including Jordan and Graves (1995), Van Mieghem (1998), Kouvelis and Vairaktarakis (1998), Graves and Tomlin (2003), Tomlin and Wang (2005), Iravani et al. (2005), Bassamboo et al. (2010), Saghafian and Van Oyen (2011, 2012), Simchi-Levi and Wei (2012), Simchi-Levi et al. (2013). Our study contributes to this literature by (1) considering the value of a flexible pooled backup supplier/resource to compensate for unreliability of dedicated suppliers, and (2) addressing the design of partial flexibility in the backup system.

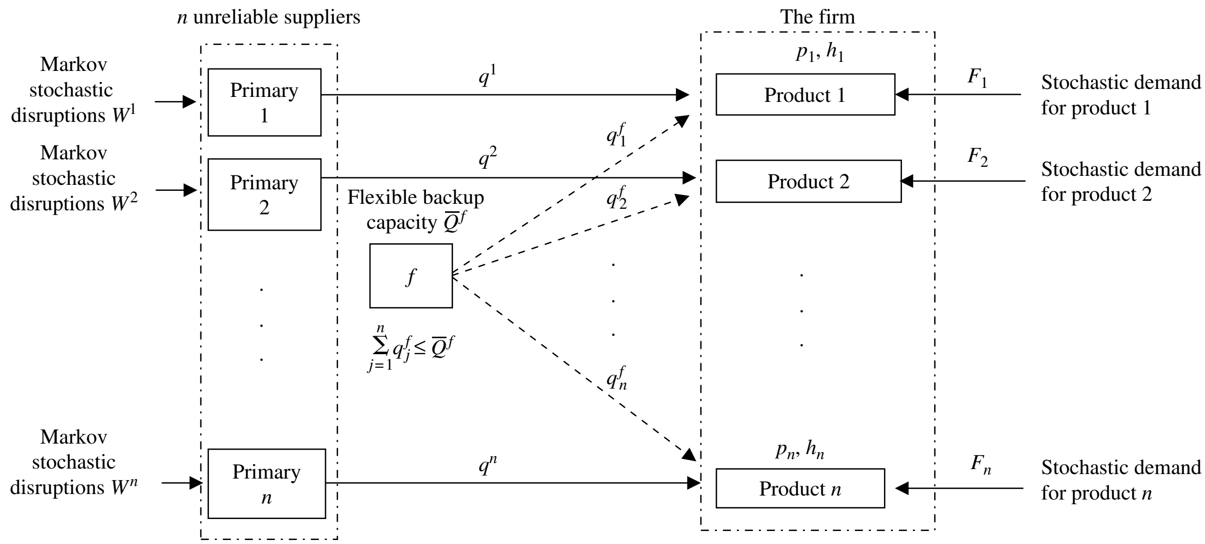
In closing this section, we note that several recent papers that study design of flexible queueing networks are also relevant to our work. For instance, Tsitsiklis and Xu (2015) show that two characteristics of fully flexible queueing structures, large stability regions, and asymptotically vanishing waiting times, can be persevered in partially flexible systems via expander-graph-based flexibility architectures. The graph expander approach is also studied in Chen et al. (2015) and Chou et al. (2011). Further insights on the design of partially flexible queueing structures can be found in Bassamboo et al. (2012), and the references therein. However, these papers are focused on queueing networks, and do not attempt to model supplier disruptions at the level of detail treated in this paper. Future work may consider carefully designed networks possessing structures of flexible backup suppliers.

3. The Model

The model is a multiperiod and multiproduct extension of the one in Saghafian and Van Oyen (2012) under a generalized capacity investment and flexibility setting. For readability, we employ the same notation where possible. Consider a firm that produces/sells $n = |\mathcal{N}|$ products, where $\mathcal{N} = \{1, 2, \ldots, n\}$ denotes the set of underlying products. The firm has a primary unreliable supplier, labeled supplier j, that supplies product $j \in \mathcal{N}$ (or perhaps one critical component for that product). To operationally insure the supply stream against future disruptions, the firm can also establish (or contract with) a flexible backup resource, namely f, at a limited capacity $\bar{Q}^f \in (0, \infty)$ that can produce on demand quantities of underlying products, the sum of which cannot exceed $\bar{Q}^f$. We let $g(u^f, \bar{Q}^f)$ denote the investment cost at the flexible backup capacity, which depends on the capacity level $\bar{Q}^f$ as well as a "per unit" investment cost, u^f. We allow for a general class of investment costs represented through the cost function $g(u^f, \bar{Q}^f)$. However, to represent a "well-behaved" investment cost function, we assume $g\colon \mathbb{R}_+^2 \to \mathbb{R}_+$ is continuous, increasing in u^f with $g(0, \cdot) = 0$, increasing convex in capacity $\bar{Q}^f$ with $g(\cdot, 0) = 0$, and supermodular (twice differentiable with positive cross partials). We note that a special case of this type of investment is that of reserving some backup capacity through a *capacity reservation* contract (also known as an option contract), where an up-front cost of $g(u^f, \bar{Q}^f) = u^f \bar{Q}^f$ reserves a backup capacity of $\bar{Q}^f$ units (see, e.g., Saghafian and Van Oyen 2012 and the references therein). We also permit a product-dependent per unit ordering cost c_j^f from the backup. For convenience, we use subscripts for products, superscripts for suppliers, and employ the following notation ($j \in \mathcal{N}$):

- h_j: Holding cost per unit of product j per period;
- p_j: Penalty cost per unit of unmet demand of product j;
- r_j: Revenue per unit of product j (set to zero when unmet demand is backlogged);
- c^j: Per unit purchasing cost of product j from dedicated/primary supplier j;
- c_j^f: Per unit purchasing cost of product j from the flexible backup supplier;
- u^f: Per unit capacity reservation cost of the flexible backup supplier;
- $\bar{Q}^f$: Reserved capacity from the flexible backup supplier;

Figure 1. The model with n products, n primary suppliers, and one flexible backup supplier.



Note. The up front investment in backup capacity $\bar{Q}^f$ and dynamic resupply orders are based on inventory levels and available information on supplier threat levels, which evolve as Markov chains.

$g(u^f, \bar{Q}^f)$: Investment cost function at the flexible backup capacity;

q^j: Order quantity from dedicated/primary supplier j;

q_j^f: Order quantity from the flexible backup supplier for product j.

Figure 1 depicts the two-echelon supply chain model under consideration.[2] The firm has an option to establish (or reserve) a desired amount of flexible backup capacity, $\bar{Q}^f$, at time 0 to insure its supply system against future disruptions. The firm then exercises a periodic review inventory control in every future period, during which it can procure either from the primary suppliers or from the limited reserved backup flexible capacity or both. Unmet demand is backordered and supply lead times and production cycles are negligible in comparison with the review period. We assume the following order of events within each review period: (1) The firm observes the state of the system (inventory levels and disruption threat levels). (2) The firm decides the order sizes and orders from all suppliers subject to the contracts. (3) Product demands are realized. (4) Holding costs or shortage costs accrue. (5) The state of the system is updated, including the inventory and disruption threat levels. The firm has to pay the purchasing cost c^j and c_j^f per order of product $j \in \mathcal{N}$ delivered by dedicated (and unreliable) supplier j and the flexible backup resource, respectively. The flexible backup resource has a shared and limited capacity $\bar{Q}^f$ (a decision variable). It can deliver any combination and quantity of products $(q_j^f\colon j \in \mathcal{N})$ as long as $\sum_{j\in\mathcal{N}} q_j^f \leqslant \bar{Q}^f$. We assume none of the products in set $\mathcal{N}$ can be procured for free; $c_j^f + u^f > 0$ and $c^j > 0$ for $j \in \mathcal{N}$.

To address the optimal design of flexibility in the backup system, we may also allow for only a subset of the product types to share the flexible backup. This is done by defining a *flexibility set* $\mathcal{F} \triangleq \{j \in \mathcal{N}\colon c_j^f < \infty\}$, and requiring it to be a strict subset of $\mathcal{N}$. When $\mathcal{F} = \mathcal{N}$, we say the backup supplier is *fully flexible*; otherwise, when $\mathcal{F} \subset \mathcal{N}$, we say the backup supplier is *partially flexible*.

For $j \in \mathcal{N}$, let $\mathcal{L}_j(x) = h_j[x]^+ + p_j[-x]^+$ and define the expected one-stage cost

$$\begin{aligned} G_j(x) &= E_{D_j}[\mathcal{L}_j(x - D_j)] \\ &= h_j \int_{-\infty}^{x} (x-\xi)\, dF_j(\xi) + p_j \int_{x}^{\infty} (\xi - x)\, dF_j(\xi), \quad (1) \end{aligned}$$

where $[x]^+ = \max\{0, x\}$, and $F_j(\cdot)$ is the cumulative distribution function (c.d.f) of the demand, D_j, for product j. We assume that demands for each product j across periods are i.i.d. random variables, and further, D_j and $D_{j'}$ are independent (for all $j, j' \in \mathcal{N}$ s.t. $j \neq j'$). We also assume unmet demand in a period is backlogged.

We model the disruption risk processes of the dedicated suppliers via a discrete time Markov process. Let s^j denote the threat level of dedicated supplier j (as an indicator of its health), where $s^j = 0$ means dedicated supplier j is in the down (default) state and $s^j = k > 0$ denotes that it is in threat level k. We assume that the dynamics of disruptions can be modeled as a Discrete Time Markov Chain (DTMC) with state space $\mathcal{S}^j = \{0, 1, \ldots, k^j\}$ for dedicated supplier j. Let $\mathbf{W}^j = [w_{lm}^j]$ denote the transition probability matrix (t.p.m.) of DTMC of supplier j, where w_{lm}^j is the probability that it will be in threat level m in the next period given that the current threat level is l. The set $\mathcal{W} = \{\mathbf{W}^j, j \in \mathcal{N}\}$ completely describes the dynamics of disruptions of all unreliable suppliers, where each supplier may have a different DTMC (regarding state space and/or transition probabilities). For every $j \in \mathcal{N}$, as a convention and without loss of generality, we assume $w_{k0}^j < w_{k'0}^j$ for every

$0 < k < k'$ in $\mathcal{S}^j$ (i.e., the higher the threat level, the higher the risk of disruption). We assume every element of $\mathcal{W}$ is *aperiodic* and *irreducible*. Thus the underlying DTMC's are all *ergodic* and have a steady-state distribution which for supplier j we denote by the vector $\pi^j = (\pi_0^j, \pi_1^j, \ldots, \pi_{k^j}^j)$. Hence, π_0^j is the long-run disruption probability of dedicated supplier j and $(1-\pi_0^j)$ is its reliability, the long-run fraction of time that it is not disrupted.

To perform our analysis, we let the vector $\mathbf{x}(t) = (x_j(t): j \in \mathcal{N})$ denote the inventory on hand at period t. Also, by $\mathbf{q}(t) = (q^j(t): j \in \mathcal{N})$ and $\mathbf{q}^f(t) = (q_j^f(t): j \in \mathcal{N})$ we denote the vectors of order sizes from the primary suppliers and the flexible supplier at period t, respectively. Additionally, we let $\beta \in (0,1)$ be the discount factor and $\mathbf{s}(t) = (s^j(t): j \in \mathcal{N}, s^j(t) \in \mathcal{S}^j)$ denote the state of disruption threat levels of the unreliable suppliers at period t.

Let $\tilde{J}(\mathbf{x}(0), \mathbf{s}(0))$ denote the optimal expected infinite-horizon discounted cost of the firm (including the investment cost at period $t=0$) if the initial disruption threat levels are $\mathbf{s}(0)$ and the firm starts with an inventory on-hand vector of $\mathbf{x}(0)$. This value can be computed by the following program:

$$\tilde{J}(\mathbf{x}(0), \mathbf{s}(0)) = \min_{\bar{Q}^f \in \mathbb{R}_+} \left\{ g(u^f, \bar{Q}^f) + J_{\bar{Q}^f}(\mathbf{x}(0), \mathbf{s}(0)) \right\}, \qquad (2)$$

where $J_{\bar{Q}^f}: \mathbb{R}^n \times (\prod_{j\in\mathcal{N}} \mathcal{S}^j) \to \mathbb{R}_+$ is the optimal infinite-horizon discounted cost of the firm given the established capacity $\bar{Q}^f \in \mathbb{R}_+$. In (2), $J_{\bar{Q}^f}(\cdot,\cdot)$ can be computed using the following Bellman equation for all $t \in \mathbb{Z}^+$:

$$\begin{aligned} &J_{\bar{Q}^f}(\mathbf{x}(t), \mathbf{s}(t)) \\ &= \min_{\mathbf{q}(t),\, \mathbf{q}^f(t) \geqslant 0:\, \sum_{j\in\mathcal{N}} q_j^f(t) \leqslant \bar{Q}^f} \Bigg\{ \sum_{j\in\mathcal{N}} [c_j^f q_j^f(t) + \mathbb{1}_{(s^j(t)>0)}(c^j q^j(t) \\ &\quad + G_j(x_j(t)+q^j(t)+q_j^f(t))) + \mathbb{1}_{(s^j(t)=0)} G_j(x_j(t)+q_j^f(t))] \\ &\quad + \beta \mathbb{E}_{\mathbf{D}(t)} \mathbb{E}_{\mathbf{s}(t+1)} [J_{\bar{Q}^f}(\mathbf{x}(t+1), \mathbf{s}(t+1)) \mid \mathbf{x}(t), \mathbf{s}(t)] \Bigg\}, \qquad (3) \end{aligned}$$

where the inventory transition rule from $\mathbf{x}(t)$ to $\mathbf{x}(t+1)$ is

$$\begin{aligned} \mathbf{x}(t+1) = \mathbf{x}(t) + \mathbf{q}^f(t) + (\mathbb{1}_{\{s^1(t)\geqslant 0\}}, \mathbb{1}_{\{s^2(t)\geqslant 0\}}, \ldots, \mathbb{1}_{\{s^n(t)\geqslant 0\}}) \\ \cdot \mathbf{q}(t) - \mathbf{d}(t), \qquad (4) \end{aligned}$$

and threat level transitions from $\mathbf{s}(t)$ to $\mathbf{s}(t+1)$ are defined through Markov processes governed by the set of t.p.m's $\mathcal{W}$.

Solving program (2)–(3) derives the firm's optimal expected infinite-horizon discounted cost as well as its optimal investment level. This in turn yields a measure for the value of the flexible backup resource:

$$\Delta^f(\mathbf{x}(0), \mathbf{s}(0)) = J_0(\mathbf{x}(0), \mathbf{s}(0)) - \tilde{J}(\mathbf{x}(0), \mathbf{s}(0)).$$

For instance, $\Delta^f(\mathbf{x}(0) = \mathbf{0}_{1\times n}, \mathbf{s}(0) = \mathbf{1}_{1\times n})$ provides a good measure for investigating the value of the flexible backup resource by setting all initial inventory levels to zero and placing all suppliers in their most reliable state.

In addition to the discounted cost, we may also use the firm's long-run average inventory related per-period cost (holding and backlogging only), since it is a more convenient measure for the purpose of connecting the inventory related cost of the system to the queueing or dam storage processes. The long-run average inventory related per-period cost under any given backup capacity $\bar{Q}^f$ can be obtained from (3), and is defined as

$$\liminf_{\beta\to 1^-} (1-\beta) J_{\bar{Q}^f}(\mathbf{x}, \mathbf{s}).$$

4. Backup Flexibility Design: The Role of Disruption Length and Demand Variability

In this section, we analyze the optimal design of flexibility in the backup system, and generate insights into the important role of disruption length and demand variabilities. To this end, we consider the case of partial backup flexibility in which the backup flexibility can cover only a subset of products $\mathcal{N}$ that belong to the flexibility set $\mathcal{F} \triangleq \{j \in \mathcal{N}: c_j^f < \infty\}$.

Before considering systems with partial flexibility, however, we shall establish a basic result under full flexibility: firms that procure relatively more products benefit more from establishing a fully flexible backup supplier, but the marginal benefit diminishes.

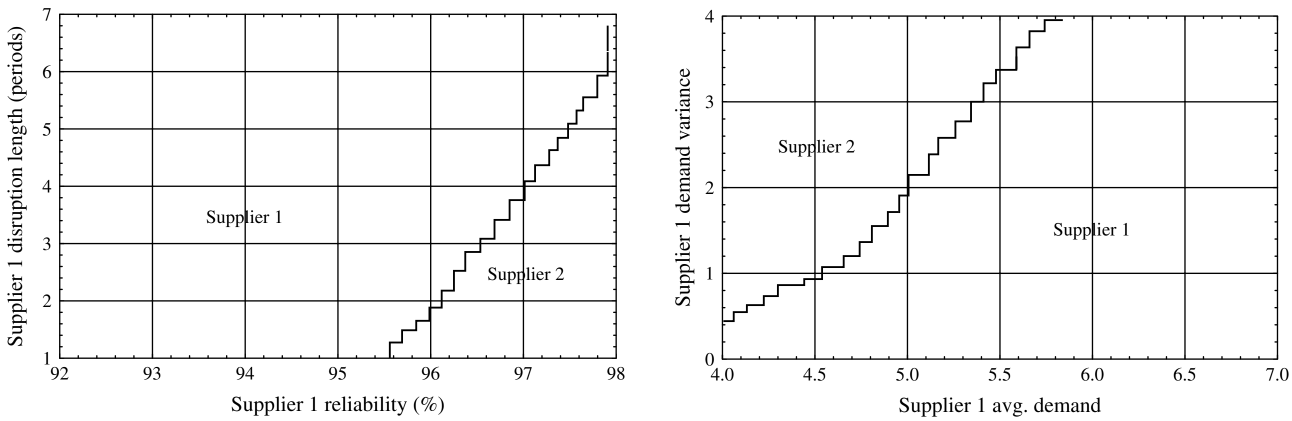
Proposition 1 (Diminishing Rate of Return). *Under full flexibility ($\mathcal{F} = \mathcal{N}$) and complete product symmetry (product or supplier independent parameters), the value of a fully flexible backup supplier has a diminishing rate of return (increasing concave) in $|\mathcal{N}|$.*

Comparing systems with full backup flexibility ($\mathcal{F} = \mathcal{N}$) and with partial backup flexibility ($\mathcal{F} \subset \mathcal{N}$) under full product symmetry, we next establish an important insight: for mitigating dynamic disruption risks, a *little backup flexibility can go a long way*.

Proposition 2 (Partial Backup Flexibility). *Consider a system with full product symmetry, and fix the product set $\mathcal{N}$. The difference between the value of the flexible backup supplier with flexibility set $\mathcal{N}$ (full backup flexibility) and with $\mathcal{F} \subset \mathcal{N}$ (partial backup flexibility) is decreasing and convex in the number of products in $\mathcal{F}$.*

Note that under each backup flexibility design, (e.g., for each level of $|\mathcal{F}|$) the available backup capacity $\bar{Q}^f$ is optimized accordingly, since the firm needs to pay a lump-sum investment cost that depends on the reserved backup capacity. Furthermore, the firm needs to pay a per-unit purchasing cost at each period when utilizing the backup capacity. Thus, even though the backup capacity depends on the flexibility level/design, Proposition 2 states that under product symmetry, partial backup flexibility provides a diminishing rate of return and may rapidly achieve the benefit of full flexibility as $\mathcal{F}$ grows to include more product types. However, it should be noted that Proposition 2 does

Figure 2. Which supplier to back up? Left: Effect of disruption length (supplier 2 reliability = 96%, supplier 2 average disruption length = 5/3); Right: Effect of demand uncertainty ($\mathbb{E}(D_2) = 5$, $\sigma^2_{D_2} = 2$).



not discuss practical/achieveable "levels of convexity." This investigation is performed in Online Appendix B, where we find that the level of convexity is indeed high. For instance, we typically observe that 60% to 85% of the benefit of full flexibility when $|\mathcal{N}| = 2$ can be obtained by only using a 1-flexibility design ($|\mathcal{F}| = 1$), noting that 50% corresponds to linearity.

It is also important to consider that in practice there may be a constraint that limits how much, if any, flexibility the backup supplier may have. Also, full flexibility may not be optimal. Under asymmetry, intuition suggests that backing up *some* of the suppliers will suffice in most cases, because either the cost of backup capacity will be very high or because disruptions are relatively rare for a subset of suppliers, or because of other model parameters rendering inventory carry-over a more cost-effective way to mitigate the disruption risks for some of the suppliers. These motivate the following question.

Question 1. *Under asymmetry, which unreliable supplier(s) should be backed up first?*

When suppliers only differ in disruption risks, one may expect that backing up the most unreliable supplier (lowest percentage uptime) would be the best strategy, because it allows using the backup more frequently, thereby more significantly offsetting the fixed investment cost in the backup supplier. Similarly, if they differ only with respect to their product demands, then backing up the supplier of the most "popular" product (highest average product demand) may be seen as the best strategy, because it prevents the accrual of high holding costs (if inventory is carried over to mitigate disruption risks) or high penalty costs (if inventory is not carried over, and hence, stockouts occur). However, we find that focusing on these first moment effects (average reliability and mean demand) can be misleading.

For instance, in Figure 2, we use our Markov decision process (MDP) framework for a representative example with $\mathcal{N} = 2$, and we model the disruption risk dynamics as two-state Markov chains to generate insights into Question 1. Specifically, in Figure 2 (left), we fix the supplier 2 reliability at 96% and its average disruption length at $5/3 = 1.67$ (review) periods (i.e., we fix $\mathbf{W}^2$). Next, we vary $\mathbf{W}^1$, while considering all other parameters to be equal for both suppliers or products ($c^j = 2$, $h_j = 1.5$, $r_j = 4.5$, $p_j = 3.5$, $u^f = 0.2$, $\beta = 0.9$, uniform demand distributions in [1,5], $c^f_j / c^j = 1.1$ for all $j \in \mathcal{F}$, and $g(u^f, \bar{Q}^f) = u^f \bar{Q}^f$, where $\bar{Q}^f$ is optimized for each scenario). As can be seen in Figure 2 (left), even when supplier 1 has a higher reliability than supplier 2 (i.e., 96%), it can be better to back up supplier 1 than 2 because it faces sufficiently longer disruptions. This shows the importance of considering the *dynamics* of disruptions rather than only average reliabilities. Similarly, in Figure 2 (right), for this representative example, we fix $\mathbb{E}(D_2) = 5$ and $\sigma^2_{D_2} = 2$, and vary the mean demand of product 1 (within the family of uniform distributions) while keeping all other parameters the same, to investigate the effect of demand variability. Interestingly, we observe that the intuition of backing up the supplier of the best-selling product is also not sufficiently subtle; it can be better to back up the supplier with the lower mean product demand, if its product demand is less *uncertain* than the other product. In summary, our numerical investigations reveal that, in designing backup flexibility, focusing only on first moment effects can be misleading.

Observation 1. All else equal (including average supplier reliability and mean product demand), it is better to back up the supplier with (1) longer but less frequent disruptions (than the supplier with shorter but more frequent disruptions), and/or (2) lower demand uncertainty (than the supplier of the product with less predictable demand).

The intuition behind part (1) is that longer disruptions (i.e., greater recovery times) reduce the advantage of carrying inventories over time as a means to mitigate disruptions. Even with frequent disruptions, the strategy of using inventory safeguards remains more effective if disruptions are sufficiently brief (and resulting in equal average reliability). Similarly, part (2) states that inventory safeguards are more beneficial when used for the product with higher demand uncertainty. This may seem counter-intuitive; however, the inventory safeguards can also be used to respond to demand volatility, i.e., a *safety stock functionality*. Hence, it is better to use the backup resource for the supplier of the product with more predictable demand, and benefit from inventory safeguards for the other product. In addition, backing up the supplier with the more predictable demand requires a lower up-front investment in the backup system (as the optimum investment level is increasing in demand uncertainty), making this choice even more attractive.

Observation 1 provides insights into Question 1 through a representative numerical experiment. But we can gain broader insights by investigating the related trade-offs in a more general setting. We do this by introducing and exploring the dynamics of *threat-dependent inventory shortfall*,[3] which is the difference between the threat-dependent base-stock levels and the inventory level caused by the limited supply capacity during disruptions as well as the random demand realizations. In particular, we will observe that the dynamics of threat-dependent inventory shortfalls resemble the waiting time of a customer in a single-server G/GI/1 type queue where the distribution of inter-arrival times is given by a two-state deterministic Markov modulated process. This fact along with an analogous simple dam/storage process enables us to answer Question 1 in more depth.

To this end, consider two suppliers indexed by $j = 1, 2$ that have the same average reliability, but let supplier 1 experience longer (in terms of first order stochastic dominance) but less frequent disruptions than supplier 2. For simplicity, we assume both suppliers have the same threat level state space denoted by $\boldsymbol{\mathcal{S}} \triangleq \boldsymbol{\mathcal{S}}^1 = \boldsymbol{\mathcal{S}}^2$. If the transition probability matrices $\mathbf{W}^j = [w^j_{lm}]_{l,\,m\in\boldsymbol{\mathcal{S}}}$ are such that $w^1_{00} \geqslant w^2_{00}$ but $\pi^1_0 = \pi^2_0$, then the required conditions hold. In particular, although both suppliers have the same average reliability $(1 - \pi^j_0)$, the geometric random variables representing the length of disruptions denoted by L^j satisfy $L^2 \leqslant_{st} L^1$, where $\leqslant_{st}$ denotes first order stochastic dominance. Moreover, since $\pi^1_0 = \pi^2_0$, we have that the frequency of disruptions $\sum_{i\in\boldsymbol{\mathcal{S}},\, i\neq 0} \pi^j_i w^j_{i0}$ is higher for supplier 2 than supplier 1. Using this setting, we first study the effect of disruption length in the absence of any backup capacity, and then extend the result to situations where they can be backed up. To do so, we note that the firm's product j inventory at the beginning of period $t+1$ under the optimal base-stock inventory control policy is recursively given by

$$X^j_{t+1} = \min\{X^j_t - D^j_t + K^j(s^j_t),\, y^j(s^j_t)\}, \tag{5}$$

where $y^j(s^j_t)$ is the base-stock level of product j when its supplier's threat level at period t is s^j_t, D^j_t denotes the demand of product j at period t, and $K^j(s^j_t)$ can be thought of as a state-dependent "random supply capacity" (of supplier j when its threat level at period t is s^j_t): $K^j(s^j_t) = \mathbb{1}_{\{s^j_t \neq 0\}} M$, for some sufficiently large number M that approximates the capacity of the backup supplier.[4] Denote by $\Psi^j_t(s^j_t) = y^j(s^j_t) - X^j_t$ the threat-dependent inventory shortfall of product j at period t. Then, $\Psi^j_t(s^j_t)$ can be recursively written as

$$\Psi^j_{t+1}(s^j_{t+1}) = \max\{\Psi^j_t(s^j_t) - K^j(s^j_t) + D^j_t,\, 0\}. \tag{6}$$

The dynamics of threat-dependent inventory shortfall presented above is similar to that of the *waiting time* (in queue) of the tth customer in a single server queueing system with inter-arrival and service times being represented by $K^j(s^j_t)$ and D^j_t, respectively. However, first it should be noted that this interpretation of the shortfall as a waiting time is not intuitively straightforward and differs from manufacturing queues where time is continuous and queues are discrete: the shortfall in period t in our framework is interpreted as the waiting time (in queue) of customer t, and hence, time is discrete whereas queues are continuous. Second, we note that if the "random supply capacity" was not threat-dependent (i.e., random variables $K^j(s^j_t)$ were i.i.d. across periods), then the Lindley type dynamics presented in (6) would be exactly the same as that of waiting time (in queue) of a customer in a $GI/GI/1$ queueing system. However, (6) shows that the dynamics of shortfall is threat-dependent and resemble the waiting time (in queue) of a customer in a $G/GI/1$ queueing system in which the inter-arrival times (random variable $K^j(s^j_t)$) are deterministically defined based on an exogenous two-state Markov-modulated process: $K^j(s^j_t) = \mathbb{1}_{\{s^j_t \neq 0\}} M$.

Since the Markov chain describing the dynamics of s^j_t is ergodic, both s^j_t and s^j_{t+1} converge in distribution (denoted by $\xrightarrow{d}$) to the same random variable s^j as $t \to \infty$, which has the distribution π^j (i.e., the steady-state distribution). Hence, there exists a random variable $K^j(s^j)$ such that $K^j_t(s^j_t) \xrightarrow{d} K^j(s^j)$ as $t \to \infty$. Thus, assuming that the underlying queueing system is stable, $\Psi^j_{t+1}(s^j_{t+1}) \xrightarrow{d} \Psi^j(s^j)$ where

$$\Psi^j(s^j) \stackrel{d}{=} \max\{\Psi^j(s^j) - K^j(s^j) + D^j,\, 0\}. \tag{7}$$

Using the discussion above, we first note that the firm's long-run average inventory cost associated with product j (under a threat-dependent base-stock $y^j(s^j)$) is

$$\begin{aligned}
&\mathbb{E}_{s^j,\Psi^j(s^j),D^j}[h_j(y^j(s^j) - \Psi^j(s^j) - D^j)^+ \\
&\qquad\qquad + p_j(D^j + \Psi^j(s^j) - y^j)^+] \\
&\quad = \mathbb{E}_{s^j,\Psi^j(s^j)}[G_j(y^j(s^j) - \Psi^j(s^j))] \\
&\quad = \sum_{i\in\boldsymbol{\mathcal{S}}\triangleq\boldsymbol{\mathcal{S}}^1=\boldsymbol{\mathcal{S}}^2} \pi^j_i\, \mathbb{E}_{\Psi^j(i)}[G_j(y^j(i) - \Psi^j(i))],
\end{aligned} \tag{8}$$

where $G_j(\cdot)$ is defined in (1). Moreover, to use (8), it can be easily seen that:

$$\begin{aligned}
&\mathbb{E}_{\Psi^j(i)}[G_j(y^j(i) - \Psi^j(i))] \\
&\quad = h_j \int_{-\infty}^{y^j(i)} (y^j(i) - \xi)\, dF_{D^j+\Psi^j(i)}(\xi) \\
&\qquad + p_j \int_{y^j(i)}^{\infty} (\xi - y^j(i))\, dF_{D^j+\Psi^j(i)}(\xi),
\end{aligned} \tag{9}$$

where $F_{D^j+\Psi^j(i)}$ is the *convolution* of distributions of demand, D^j, and shortfall, $\Psi^j(i)$ (by independence). Since the above function is convex (due to convexity of $G_j(\cdot)$), it follows from the first order condition that the optimal base-stock level satisfies

$$\Pr(D^j \leqslant y^j(i) - \Psi^j(i)) = \mathbb{E}_{\Psi^j(i)}[F_D^j(y^j(i) - \Psi^j(i))] = \frac{p_j}{p_j + h_j}, \tag{10}$$

or equivalently

$$y^j(i) = F^{-1}_{D^j+\Psi^j(i)}\left(\frac{p_j}{p_j + h_j}\right), \tag{11}$$

where for any random variable Ξ with a c.d.f. $F_\Xi(\xi)$,

$$F_\Xi^{-1}(y) = \inf\{\xi\colon F_\Xi(\xi) \geqslant y\}. \tag{12}$$

Replacing the optimal base-stock level characterized by (11) in (8) and (9) will enable us to compare the optimal cost of product/supplier 1 to that of product/supplier 2. Using this, we first establish that a stochastically larger "waiting time" (in queue) in the corresponding queueing system means both a higher optimal base-stock level and cost in the original inventory system. This will later enable us to show the important insight that, keeping average reliability the same, longer but less frequent disruptions cause higher average costs.

Lemma 1 (Queueing Comparison). *In the absence of any backup capacity, if $\Psi^2(i) \leqslant_{st} \Psi^1(i)$ (for all $i \in \mathcal{S} \triangleq \mathcal{S}^1 = \mathcal{S}^2$), then (i) the state-dependent optimal base-stock levels satisfy $y^2(i) \leqslant y^1(i)$, and (ii) having only product 2 is preferred to having only product 1: $\mathcal{N} = \{1\} \preceq \mathcal{N} = \{2\}$, where $\preceq$ represents preference in terms of associated expected long-run average cost.*

We next show that longer but less frequent disruptions (all else equal including average reliabilities) indeed implies a stochastically larger "waiting time" in the corresponding queueing system (i.e., larger shortfalls in the inventory system), which is analogous to a well-known fact in queueing systems: a stochastically higher interarrival time variability typically implies a longer average waiting time. To show this result, we consider the following dam model or storage process[5] with an input D_t at the beginning of period t. In each period, a decision regarding whether to open the dam is made: the dam is kept completely closed if $s_t = 0$, and is completely opened (i.e., instantaneous release) otherwise. This dam model allows us to characterize the threat-dependent inventory shortfall as a compound random variable (i.e., a random sum of i.i.d. random variables) and state the following.

Lemma 2 (Role of Disruption Length). *In the absence of any backup capacity, the inventory shortfall of product j given risk state i (for all $i \in \mathcal{S} \triangleq \mathcal{S}^1 = \mathcal{S}^2$) satisfies*

$$\Psi^j(i) \stackrel{d}{=} \mathbb{1}_{\{i=0\}} \sum_{t=1}^{L^j} D_t, \tag{13}$$

where random variables D_t are i.i.d., and D_t is the demand in period t. Hence, if supplier 1 has longer but less frequent disruptions than supplier 2 but all else is equal, then in the absence of any backup capacity $\Psi^2(i) \leqslant_{st} \Psi^1(i)$ (for all $i \in \mathcal{S} \triangleq \mathcal{S}^1 = \mathcal{S}^2$).

It is noteworthy that although the base-stock levels are different for different threat-levels, the shortfall process characterized in the above lemma only depends on whether the system is up or down, providing a much simpler process to analyze.

Combining Lemmas 1 and 2 we observe that, in the absence of any backup capacity, if supplier 1 has longer but less frequent disruptions than supplier 2, then (a) the firm tends to keep stochastically higher levels of on-hand inventory of product 1 than 2, and (b) product 1 is associated with a higher optimal average cost. The following lemma extends this result to the case where a given secondary capacity can be assigned to back up unreliable suppliers.

Lemma 3 (Disruption Length and Backup Capacity). *When a given capacity $\bar{Q}^f$ is assigned to back up supplier $j \in \{1, 2\}$, for all $i \in \mathcal{S} \triangleq \mathcal{S}^1 = \mathcal{S}^2$*

$$\Psi^j(i) \stackrel{d}{=} \mathbb{1}_{\{i=0\}} \sum_{t=1}^{L^j} (D_t - \bar{Q}^f)^+. \tag{14}$$

Hence, if supplier 1 has longer but less frequent disruptions than supplier 2 but all else is equal, then (i) $\Psi^2(i) \leqslant_{st} \Psi^1(i)$ (for all $i \in \mathcal{S} \triangleq \mathcal{S}^1 = \mathcal{S}^2$), and (ii) $\mathcal{N} = \{1\} \preceq \mathcal{N} = \{2\}$.

In addition to characterizing the effect of backup capacity on inventory shortfall, the above lemma shows that disruption length and backup capacity are *substitutes*: the effect of a longer disruption length can be offset by a higher backup capacity level. Using the above lemma, we next establish the interesting insight that when a backup capacity exists, it is better to back up the supplier with longer but less frequent disruptions (all else equal). Recalling the problem of determining the optimal flexibility set $\mathcal{F}$ (which indicates the set of suppliers to backup), this insight is presented in the following result.

Proposition 3 (The Effect of Disruption Length on Backup Flexibility Design). *Let $\mathcal{N} = \{1, 2\}$, and while keeping all else equal (including average reliabilities), suppose supplier 1 has longer but less frequent disruptions than supplier 2.* (i) *For any available backup capacity $\bar{Q}^f$, $\mathcal{F} = \{2\} \preceq \mathcal{F} = \{1\} \preceq \mathcal{F} = \{1, 2\}$.* (ii) *The result of part* (i) *holds even when the backup capacity is optimized separately for each of the three flexibility designs.*

We note that the above result is congruent with the numerical insights captured in Observation 1. Whether the backup capacity is fixed across the alternative designs or optimized separately for each, it proves the following important insight in designing backup flexibility: the length of disruptions plays a critical role in deciding which supplier(s) to back up. Therefore, paying attention only to suppliers' average reliabilities can be misleading, because it might be better to backup a supplier with higher average reliability, if its disruptions are lengthier. Moreover, Proposition 3 can be easily extended to a setting with an arbitrary number of products as follows. Let $\mathcal{N} = \{1, 2, \ldots, n\}$ for some n, and suppose that suppliers are only different in their length of disruption in the sense of stochastic dominance while all else is equal. If only a partial flexibility with $|\mathcal{F}| = k < n$ is possible, then the optimal backup flexibility design is the one in which $\mathcal{F}$ is chosen to be the set of k suppliers with the lengthiest disruptions: a *longest disruption length first* policy. This agrees with the focus on "time to recover" recently emphasized in Simchi-Levi et al. (2014).

We now turn our attention from the supply side to the demand side. Specifically, we establish that, similar to the supply side, merely paying attention to the average demand can be misleading. To this end, we can again benefit from the dynamics of inventory shortfall. Since a higher demand variability in the original system implies a higher service time variance in the queueing counterpart, and the performance of queueing systems typically is roughly inversely proportional to service time variance, we can explore the answer to Question 1 in more depth. This is done in the following proposition, where variability is captured through *convex stochastic ordering*.

Proposition 4 (The Effect of Demand Variability on Backup Flexibility Design). *Let $\mathcal{N} = \{1, 2\}$, and while keeping all else equal (including average demand for products* 1 *and* 2), *suppose $D^1 \leqslant_{cx} D^2$ (where $\leqslant_{cx}$ denotes convex stochastic ordering*[6]), *implying that demand for product* 2 *has a higher variability than product* 1. (i) *For any given backup capacity $\bar{Q}^f$, $\mathcal{F} = \{2\} \preceq \mathcal{F} = \{1\} \preceq \mathcal{F} = \{1, 2\}$.* (ii) *The result of part* (i) *holds even when the backup capacity is optimized separately for each of the three flexibility designs.*

To gain further insights into the effect of demand distributions on backup flexibility design, we next consider scenarios where disruptions are long (i.e., in long time-to-recovery situations). This allows us to factor out the effect of disruption lengths and focus on the effect of demand distributions. The reader should note that, because of the Markov process governing disruptions, disruption durations are geometrically distributed in our framework.

Relatively Long Geometric Disruptions. When disruptions are long, to characterize the inventory shortfall distribution and thereby generating more insights, we can use the *Renyi's Theorem* for geometric sums of random variables. Renyi's Theorem can be stated as follows: Let $S_L = \sum_{i=1}^{L} X_i$ be the sum of L i.i.d. random variables having a common positive but finite mean $\mathbb{E}[X]$. If L is geometric with mean l, then

$$\Pr\left\{\frac{S_L}{l} > \xi\right\} = e^{-\xi/(\mathbb{E}[X])} + o(1), \tag{15}$$

as $l \to \infty$. That is, S_L is approximately exponential with rate $(l\,\mathbb{E}[X])^{-1}$ (see Blanchet and Glynn 2007 for more discussions and a connection to the *Cramer-Lundberg* approximation when $\xi l \to \infty$). It should be noted that the approximation remains exact in some cases. For instance, when X_i's are i.i.d. exponentials, S_L has an exponential distribution with rate $(l\,\mathbb{E}[X])^{-1}$ for any finite $l > 1$.

Let $d^j = \mathbb{E}[D^j]$ and $l^j = \mathbb{E}[L^j]$. Since L^j in our Markov chain model of disruptions has a geometric distribution, using Lemma 3 along with the Renyi's Theorem, we observe that, when disruptions are relatively long (i.e., as $l \to \infty$), the inventory shortfall in steady-state satisfies

$$\Psi^j(i) \stackrel{d}{=} \mathbb{1}_{\{i=0\}} \Xi^j, \tag{16}$$

where Ξ^j has an approximately exponential distribution with mean $\mathbb{E}[(D^j - \bar{Q}^f)^+]\mathbb{E}[L^j] = (d^j - \bar{Q}^f)l^j(1 - F_D^j(\bar{Q}^f))$. Furthermore, as $l \to \infty$, it can be shown that $D^j + \Psi^j(i)$ can also be approximated as an exponential random variable with mean $(d^j - \bar{Q}^f)l^j(1 - F_D^j(\bar{Q}^f))$ when $i = 0$. Also, from (16), when $i \neq 0$, $D^j + \Psi^j(i) \stackrel{d}{=} D^j$. These, together with Equations (8) and (11), allow us to establish the following result.

Proposition 5 (Long Disruptions: The Effect of Demand Distributions). *Suppose disruptions are relatively long in duration* (*i.e., let $l \to \infty$*), *and a capacity of $\bar{Q}^f$ is assigned to back up supplier j. Then*:

(i) *The optimal base-stock level for product j is*

$$y^j(i) \simeq \begin{cases} (d^j - \bar{Q}^f)l^j(1 - F_D^j(\bar{Q}^f)) \ln\left(1 + \dfrac{p_j}{h_j}\right): & i = 0, \\[2ex] F_{D^j}^{-1}\left(\dfrac{p}{p+h}\right): & i \neq 0. \end{cases}$$

(ii) *Under the optimal base-stock level of part* (i), $\mathbb{E}_{\Psi^j(0)}[G_j(y^j(0) - \Psi^j(0))] \simeq h_j y^j(0)$. *That is, the long-run average inventory cost related to default periods is linear in the base-stock level $y^j(0)$ with a slope equal to h_j.*

(iii) *Provided $\mathcal{N} = \{1, 2\}$ and suppliers differ only in their product demand distribution, then, assuming $(d^2 - \bar{Q}^f) F_{D^2}(\bar{Q}^f) \leqslant (d^1 - \bar{Q}^f)F_{D^1}(\bar{Q}^f)$ without loss of generality, we have $\mathcal{F} = \{2\} \preceq \mathcal{F} = \{1\} \preceq \mathcal{F} = \{1, 2\}$.*

The above result provides an important insight about which supplier to back up: when disruptions are relatively long if they occur, backing up the suppliers with the highest value of $(d^j - \bar{Q}^f)F_{D^j}(\bar{Q}^f)$ provides the best backup flexibility design. This strategy is equivalent to backing up the suppliers with the lowest $\mathbb{E}[(\bar{Q}^f - D^j)^+]$, suggesting that the goal should be backing up the suppliers for which the expected excess/wasted backup capacity will be minimum. Again, we observe that even when all else is equal, the simplistic belief that it is best to back up the supplier with the most "popular" (i.e., highest average demand) product provides only limited insight into a more complex reality.

5. Backup Flexibility Design: The Role of Other Factors

In the previous section, we gained insights into the role of two important factors in designing backup flexibility: disruption length and demand variability/distribution. In this section, we generate insights into the role of other factors such as inventory costs, purchasing costs, average demand and supplier reliabilities, etc. To this end, we consider a special case of our model in which disruptions are i.i.d. Bernoulli, allowing us to factor out the effect of disruption lengths. We start by considering the case where the demand random variables are exponentially distributed, which enable us to also neutralize the role of demand variability, and focus on our goal of understanding the effect of other factors such as inventory costs, purchasing costs, supplier reliabilities, etc. We then expand our results to other demand distributions.

5.1. Exponential Demand Distribution

Consider an i.i.d. random supply capacity counterpart to our model, and assume the supply capacity in each period is (independent of anything else) a large number, M (i.e., the primary supplier is up) with probability $1-\pi_0^j$, and is $\bar{Q}^f$ (i.e., the primary supplier is down) with probability π_0^j. Let K^j denote a random variable which gets value M with probability $1-\pi_0^j$ and $\bar{Q}^f$ otherwise. The Lindley dynamics of shortfall, similar to (7), results in the steady-state equation:

$$\Psi^j \stackrel{d}{=} \max\{\Psi^j - K^j + D^j, 0\}, \tag{17}$$

which is that of waiting time in a GI/GI/1 queueing system. In particular, if we assume demand is exponential, we obtain the dynamics of waiting time in a GI/M/1 queueing system. Suppose $\mathbb{E}[D^j] = d^j$, and let γ^j be such that the Laplace transform of K^j evaluated at $(d^j\gamma^j)^{-1}$ is $1-1/\gamma^j$:

$$\begin{aligned}\mathbb{E}[e^{-K^j/(d^j\gamma^j)}] &= \pi_0^j e^{-\bar{Q}^f/(d^j\gamma^j)} + (1-\pi_0^j)e^{-M/(d^j\gamma^j)}\\ &= 1-\frac{1}{\gamma^j}.\end{aligned} \tag{18}$$

From standard results for GI/M/1 queueing systems (see, e.g., Kleinrock 1975), we note that there exists a unique $\gamma^j > 1$ satisfying the above equation when $\mathbb{E}[K^j] > d^j$. Furthermore, from such results, we obtain that the shortfall (representing the waiting time in queue) in steady-state has a mixture distribution: it is zero with probability $1/\gamma^j$ and is exponential with mean $d^j\gamma^j$ with probability $1-1/\gamma^j$. These allow us to characterize the optimal base-stock level and cost.

Theorem 1 (GI/M/1). *Suppose the demand related to supplier j is exponential with mean $\mathbb{E}[D^j] = d^j$, and a capacity of $\bar{Q}^f$ is assigned to back up supplier j. Then:*

(i) $F_{D^j+\Psi^j}(\xi) = 1-e^{-\xi/(d^j\gamma^j)}$.

(ii) *The optimal base-stock level for product j is $y^j = d^j\gamma^j \ln(1+p_j/h_j)$.*

(iii) *The long-run average inventory related cost of product j is $h_j y^j = h_j d^j\gamma^j \ln(1+p_j/h_j)$.*

From the above result, we observe that the optimal base-stock levels are linear in the average demand but logarithmic in p_j/h_j. Furthermore, the long-run average inventory related cost of each product is linear in its optimal base-stock level with a slope equal to the per unit holding cost. Both the optimal base-stock levels and inventory costs are also linear in parameter γ^j (i.e., inversely proportional to the steady-state probability of observing a zero shortfall), which depends on various system parameters including the backup capacity and the supplier's reliability among others. Or, one could note that they are also linear in $d^j\gamma^j$, which is the conditional mean shortfall given that a shortfall has occurred. Below, we generate further insights by characterizing the value of γ^j. To this end, we note that Theorem 1 presents results for a general backup capacity $\bar{Q}^f$, but considering some special cases can be fruitful.

In particular, it should be noted that if the average demand can be fully satisfied via the backup supplier, then an enormous investment in the backup supplier's infrastructure is required. Indeed, it might not be accurate to think of it as a backup, but rather a dual supplier more on the level of a primary supplier. In practice, however, backup capacity is typically expensive, and therefore limited compared to average demands. This makes the question of which supplier to back up more challenging for managers, and is reflected in our model via the up-front capacity investment/reservation cost function. We are able to gain further relevant insights and closed-form results by exploiting this observation and considering the special case of Theorem 1 where $\bar{Q}^f$ is small relative to average demand d^j.

Corollary 1 (Limited Backup Capacity). *Suppose a capacity of $\bar{Q}^f$ is assigned to back up supplier j, the demand related to supplier j is exponential with mean $\mathbb{E}[D^j] = d^j$, and $\bar{Q}^f$ is small relative to average demand d^j. Then:*

(i) $\gamma^j \simeq (1-\pi_0^j(\bar{Q}^f/d^j))/(1-\pi_0^j)$.

(ii) *The optimal base-stock level for product j is $y^j \simeq ((d^j - \pi_0^j\bar{Q}^f)/(1-\pi_0^j))\ln(1+p_j/h_j)$.*

(iii) *The long-run average inventory related cost of product j is $h^j y^j \simeq h_j((d^j-\pi_0^j\bar{Q}^f)/(1-\pi_0^j))\ln(1+p_j/h_j)$.*

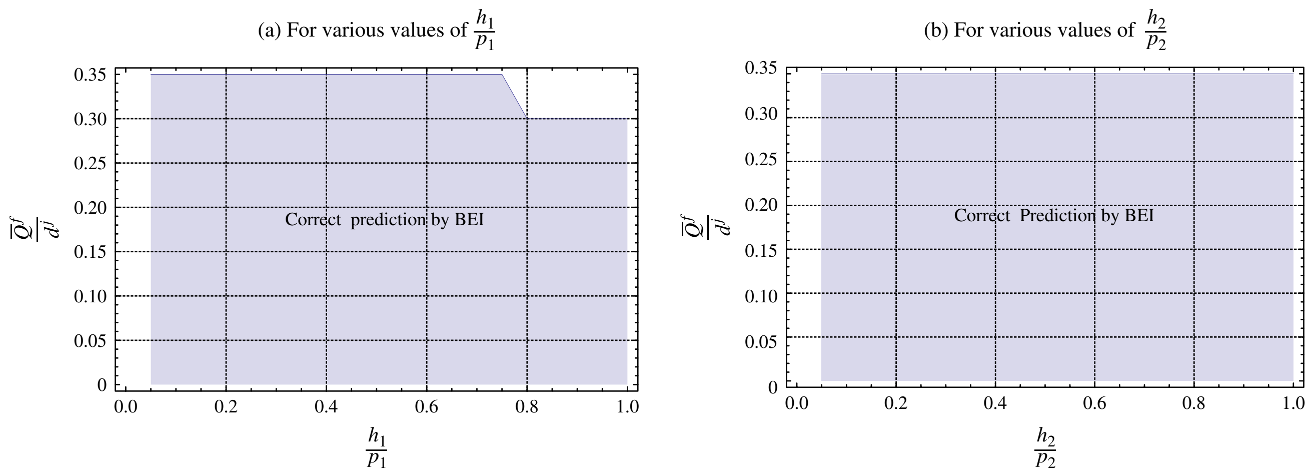
Using Corollary 1, we can also characterize the optimal level of backup capacity investment.

Proposition 6 (Backup Capacity Investment). *Under the conditions of Corollary 1 and with $g(\bar{Q}^f)$ being a convex function denoting the average per-period cost of investing in $\bar{Q}^f$ units as a backup, the optimal back up capacity level $\bar{Q}^{f*}$ is*

$$\bar{Q}^{f*} \simeq \left[g'^{-1}\left(\frac{\pi_0^j}{1-\pi_0^j}h_j \ln\left(1+\frac{p_j}{h_j}\right)\right)\right]^+, \tag{19}$$

where $g'^{-1}(\cdot)$ is the inverse function of the derivative of $g(\cdot)$.

Figure 3. (Color online) Sensitivity of the predictions made by the Backup Effect Index (BEI) to the value of $\bar{Q}^f/d^j$.



Note. Shaded area represents correct predictions [test cases used: $\pi_0^1 = 0.1$, $\pi_0^2 = 0.01$, $c^1 = c^2 = 0.6$, $c_1^f = 0.8$, $c_2^f = 0.6$, $p_1 = p_2 = 1$, $d^1 = d^2 = 1$, exponential demand distributions, i.i.d. disruptions, $h_2 = 0.3$ for part (a), and $h_1 = 0.3$ for part (b)].

In Online Appendix C, we investigate the value of a backup supplier as the amount of capacity reserved is varied. We find that even a small backup capacity can greatly reduce the costs.

In addition to characterizing the optimal backup investment level, Corollary 1 enables us to achieve our ultimate goal of developing a simple-to-use tool for supply chain designers in deciding which unreliable suppliers to back up.

Theorem 2 (Backup Effect Index (BEI)). *Let $\boldsymbol{\mathcal{N}} = \{1, 2\}$. Define the Backup Effect Index (BEI)*

$$\mathcal{I}^j = \frac{\pi_0^j}{1-\pi_0^j} h_j \bar{Q}^f \ln\left(1 + \frac{p_j}{h_j}\right) - \pi_0^j d^j (c_j^f - c_j), \tag{20}$$

and without loss of generality assume $\mathcal{I}^2 \leqslant \mathcal{I}^1$. Under the conditions of Corollary 1, $\boldsymbol{\mathcal{F}} = \{2\} \tilde{\preceq} \boldsymbol{\mathcal{F}} = \{1\} \tilde{\preceq} \boldsymbol{\mathcal{F}} = \{1, 2\}$, where $\tilde{\preceq}$ denotes preference with respect to the total long-run average approximate cost.

Theorem 2 provides an important index for deciding what supplier to back up first (in response to Question 1 we raised earlier): it suggests to follow a *largest BEI first* policy, where BEI is easily calculable and is given by (20). It also clearly shows why backing up the supplier of the most "popular" product or the supplier with lowest reliability is too simplistic. For instance, even if the per unit purchasing cost from the backup and from the primary supplier is equal for all products, the manager still needs to consider supplier reliabilities along with inventory holding and backlogging costs. When suppliers are equal in terms of (a) their average product demand, (b) reliabilities, and (c) inventory holding and backlogging costs, what matters according to the BEI is the per unit purchasing cost difference between the backup and the unreliable supplier. Setting aside the logarithmic effect of p_j/h_j, an appealing and interesting fact about BEI is that it is linear in all the other system parameters.

When there are n unreliable suppliers and only $k < n$ of them can be backed up, Theorem 2 suggests a design method in which one sorts the suppliers based on a decreasing BEI order, and back up the first k ones. This is a powerful result derived for cases in which (a) demand distributions are close to exponential, and (b) the back up capacity is small compared to average demands. We relax the first assumption in the next section by extending our results to cases with general demand distributions. To investigate the dependency of our results to the second assumption, we perform sensitivity analyses on the value of $\bar{Q}^f/d^j$ to observe whether the predictions made by BEI remain useful even when $\bar{Q}^f/d^j$ is not so small. To this end, in Figure 3, we use the recommendations/predictions made by the BEI, and compare it with the true results (using Theorem 1). As can be seen in Figure 3, the BEI provides robustly correct results even when $\bar{Q}^f/d^j$ is not so small; we find that the recomendations/predictions made by BEI are not sensitive and typically work well even when $\bar{Q}^f/d^j$ is as large as 0.35 (i.e., when 35% of the average demand can be supplied from the secondary, backup capacity).

5.2. General Demand Distribution

Consider the Lindley equation (17), and let η^j be the solution to $\mathcal{M}_{D^j-K^j}(\eta^j) = 1$, where $\mathcal{M}_{D^j-K^j}(\cdot)$ is the moment generating function of $D^j - K^j$. Observe that η^j is such that the Laplace Transform of the supply random variable K^j evaluated at η^j is equal to the $(\mathcal{M}_{D^j}(\eta^j))^{-1}$ (assuming the moment generating function of demand exists):

$$\mathbb{E}[e^{-\eta^j K^j}] = \pi_0^j e^{-\eta^j \bar{Q}^f} + (1-\pi_0^j) e^{-\eta^j M} = (\mathcal{M}_{D^j}(\eta^j))^{-1}, \tag{21}$$

which is a generalization of (18). Since M is large, when $\eta^j \bar{Q}^f$ is small (which is typically the case when $\bar{Q}^f$ is smaller than average demand), it can be seen using the

Taylor expansion of $e^{-\eta^j \bar{Q}^f}$ that η^j is approximately the solution to

$$\mathcal{M}_{D^j}(\eta^j) = \frac{1}{\pi_0^j}(1+\eta^j \bar{Q}^f). \tag{22}$$

Recalling that Ψ^j in our model is equivalent to the steady-state waiting time in a GI/GI/1 system, we can use the approximation results for the waiting time distribution of a GI/GI/1 system to characterize the optimal base-stock level and inventory costs in our system. This will in turn enable us to develop a generalized version of BEI, and provide a comprehensive answer to Question 1.

Analyzing waiting times in a GI/GI/1, Kingman (1964 and 1970) showed that when η^j is the solution to $\mathcal{M}_{D^j - K^j}(\eta^j) = 1$:

$$ce^{-\eta^j \xi} \leqslant \Pr(\Psi^j > \xi) \leqslant e^{-\eta^j \xi}, \tag{23}$$

for some constant c, Ross (1974) improved such bounds, and Abate et al. (1995) argued that

$$\Pr(\Psi^j > \xi) \simeq \alpha^j e^{-\eta^j \xi}, \tag{24}$$

where

$$\alpha^j \simeq \eta^j \mathbb{E}[\Psi^j]. \tag{25}$$

Using (24), we observe that:

$$\begin{aligned}\Pr(\Psi^j + D^j > \xi) &= \mathbb{E}_{D^j}[\Pr(\Psi^j + D^j > \xi \mid D^j)] \\ &\simeq \mathbb{E}_{D^j}[\alpha^j e^{-\eta^j(\xi - D^j)}] = \tilde{\alpha}^j e^{-\eta^j \xi},\end{aligned} \tag{26}$$

where $\tilde{\alpha}^j = \alpha^j \mathcal{M}_{D^j}(\eta^j)$. That is, the convolution of Ψ^j and D^j has an approximately exponentially decaying tail. This fact allows us to approximate the optimal base-stock level and inventory cost.

Theorem 3 (GI/GI/1). *Suppose a capacity of $\bar{Q}^f$ is assigned to back up supplier j. Then*:

(i) $F_{D^j+\Psi^j}(\xi) \simeq 1 - \tilde{\alpha}^j e^{-\eta^j \xi}$.

(ii) *The optimal base-stock level for product j is* $y^j \simeq (1/\eta^j)[\ln(1+p_j/h_j) + \ln\tilde{\alpha}]$.

(iii) *The long-run average inventory related cost of product j is approximately* $(h_j \tilde{\alpha}^j/\eta^j)[\ln(1+p_j/h_j) + \ln\tilde{\alpha}^j + 1/\tilde{\alpha}^j - 1]$.

When demand is exponentially distributed, $\tilde{\alpha}^j = 1$ and $\eta^j = (d^j \gamma^j)^{-1}$. Hence, Theorem 1 is recovered from Theorem 3, and the approximations are exact.

Theorem 3 allows us to develop a generalized version of the Backup Effect Index (BEI) introduced in Theorem 2 for designing backup flexibility in supply chains. Since our goal is to develop a simple-to-calculate index, we use the inventory cost from part (iii) of Theorem 3, and further approximate it by assuming that $\ln\tilde{\alpha}^j + 1/\tilde{\alpha}^j - 1 \simeq 0$ (which holds as an exact equality when demand is exponential, and as a useful and accurate approximation for a variety of other distributions). Doing so, we define the cost function

$$\begin{aligned} f^j(\bar{Q}^f) &= \frac{h_j \tilde{\alpha}^j}{\eta^j}\ln\left(1+\frac{p_j}{h_j}\right) \\ &\simeq \frac{1}{\pi_0^j} h_j \mathbb{E}[\Psi^j](1+\eta^j \bar{Q}^f)\ln\left(1+\frac{p_j}{h_j}\right) \qquad (27) \\ &\simeq \frac{1}{\pi_0^j} h_j d^j \left(\frac{\rho^j}{1-\rho^j}\right)\left(\frac{c_{k^j}^2 + c_{d^j}^2}{2}\right) \\ &\quad \cdot \left(1+\eta^j \bar{Q}^f\right)\ln\left(1+\frac{p_j}{h_j}\right), \qquad (28)\end{aligned}$$

where the first part of (27) follows from part (iii) of Theorem 3, and the second part of it follows from (22) and (25). Also, with $\rho^j = \mathbb{E}[K^j]/d^j$ denoting the traffic intensity, and $c_{k^j}^2$ and $c_{d^j}^2$ denoting the squared coefficients of variation of supply and demand, respectively, (28) follows from Kingman's approximation for average waiting time in a GI/GI/1 system.

Next, we note that the inventory cost reduction for supplier/product j because of assigning a small backup capacity $\bar{Q}^f$ is

$$f^j(0) - f^j(\bar{Q}^f) \simeq -\bar{Q}^f f^{j'}(0), \tag{29}$$

where $f^{j'}(0)$ is the derivative of f^j evaluated at zero. This allows us to develop a simple index (for the general demand distributions case) as follows.

Theorem 4 (Generalized Backup Effect Index (GBEI)). *Let $\mathcal{N} = \{1, 2\}$, and define the Generalized Backup Effect Index* (*GBEI*)

$$\mathcal{GI}^j = -\bar{Q}^f f^{j'}(0) - \pi_0^j d^j (c_j^f - c_j). \tag{30}$$

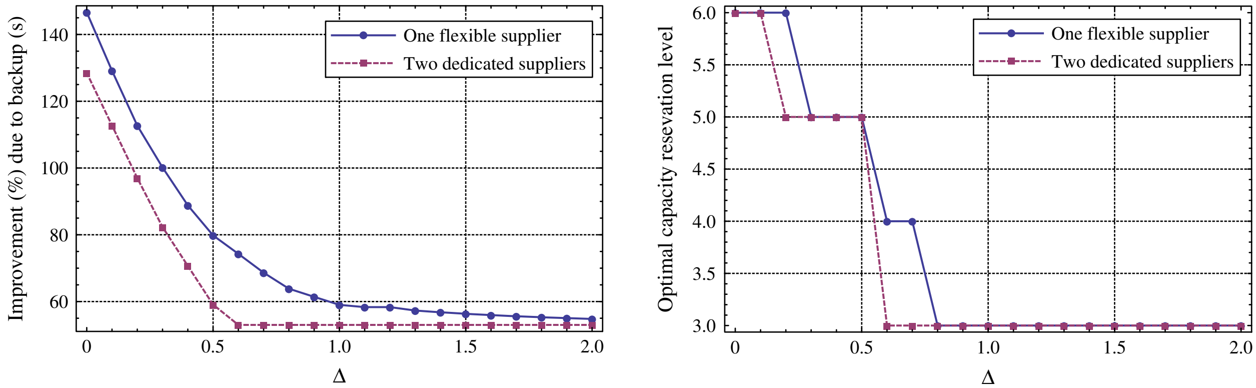
Suppose a relatively small backup capacity is available, and without loss of generality assume $\mathcal{GI}^2 \leqslant \mathcal{GI}^1$. Then, $\mathcal{F} = \{2\} \precsim \mathcal{F} = \{1\} \precsim \mathcal{F} = \{1, 2\}$.

Theorem 2 provides an important but simple-to-compute index for deciding what supplier to back up: it suggests to follow a *largest GBEI first* policy, where GBEI is given by (30). Similar to BEI established by Theorem 2, the first and second terms in GBEI represent the effect of choosing supplier j to back up on inventory and purchasing costs, respectively. It is worth noting that, for BEI (which was developed under an exponential demand assumption), $f^{j'}(0) = -(\pi_0^j/(1-\pi_0^j))h_j \ln(1+p_j/h_j)$, but in general $f^{j'}(0)$ should be obtained from (28).

6. Backup Capacity Pooling Advantage: A Numerical Study

In this section, we shed further light on the capacity pooling advantage in the backup system. We do so by numerically comparing the supply chain under consideration under

Figure 4. (Color online) An analysis of two backups vs. one flexible backup.



Note. Left: Cost improvement (%) compared to a no backup case; Right: Optimal capacity investment level (with the sum of the dedicated backups depicted).

dedicated backups with that under a single, pooled (i.e., fully flexible) backup capacity. To this end, consider a firm procuring two products from unreliable suppliers and let $u^f = 10$, $c_1^f = 0.3$, $c_2^f = \Delta c^2$ (where Δ scales the cost of dedicated supplier 2), $\beta = 0.9$, $c^1 = c^2 = 2$, $r_1 = r_2 = 3.5$, $p_1 = p_2 = 2.5$, $h_1 = 1$, $h_2 = 1.2$, and $g(u^f, \bar{Q}^f) = u^f \bar{Q}^f$, where $\bar{Q}^f$ is optimized for each scenario. Assume the demand distributions are uniform in $[1, 5]$, and the dynamics of disruption risks for both unreliable suppliers are defined by the t.p.m.

$$\mathbf{W}^1 = \mathbf{W}^2 = \begin{bmatrix} 0.10 & 0.40 & 0.20 & 0.10 & 0.10 & 0.10 \\ 0.10 & 0.10 & 0.40 & 0.20 & 0.10 & 0.10 \\ 0.15 & 0.05 & 0.10 & 0.30 & 0.20 & 0.20 \\ 0.20 & 0.10 & 0.10 & 0.10 & 0.30 & 0.20 \\ 0.25 & 0.10 & 0.10 & 0.10 & 0.20 & 0.25 \\ 0.30 & 0.10 & 0.10 & 0.10 & 0.20 & 0.20 \end{bmatrix}.$$

Figure 4 (left) reveals the insight that the value of the fully flexible backup supplier is more than the summation of benefits that can be obtained separately for each of the products through dedicated backups (assuming that the dedicated backup suppliers are priced similar to the flexible backup one). This is mainly due to the *capacity pooling advantage* of the flexible backup supplier; when one of the primary suppliers is in a high risk threat level and the other is in a low threat level, the reserved pooled capacity can be used as needed. However, using the difference between the two curves depicted in Figure 4 (left), we can make the following observation.

Observation 2. The pooling advantage is not monotone in Δ and has its maximum effect when Δ is in a middle range. However, as Δ increases, the pooling advantage vanishes: the backup flexible supplier can only be used for a single product, performing as a dedicated backup supplier.

Figure 4 (right) depicts the corresponding optimal investment levels in the backup suppliers. From this figure we observe the following.

Observation 3. The sum of optimal capacities required for product 1 and 2 in the case of two dedicated backup suppliers is never larger than the optimal capacity for the flexible backup supplier.

The observation above highlights the justification for higher levels of investments (or capacity reservation) fees charged in practice by flexible suppliers.

7. Summary of Findings and Concluding Remarks

We investigated the optimal design of flexibility in the backup system as a potent supply risk mitigation mechanisms. Analogous to some models for credit risk rating systems (e.g., S&P), we modeled the dynamics of disruptions as discrete time Markov chains, and considered a multiproduct, multisupplier supply chain under dynamically evolving disruption risks.

We analytically showed the important insight that in mitigating dynamic disruptions *a little backup flexibility can go a long way*. When suppliers are asymmetric and full flexibility is not possible (or is too expensive) this raises a fundamental question: which unreliable supplier(s) should be backed up? We addressed this question and observed that focusing merely on the first moment effects such as average demand (product popularity) or suppliers' reliability (percentage uptime) is insufficient to decide which supplier to back up. Rather, managers should also consider the effect of disruption dynamics (e.g., duration in addition to frequency) as well as demand uncertainty (e.g., variance in addition to average). In particular, characterizing the dynamics of the threat-dependent inventory shortfall process as a queueing system, we analytically established an important managerial insight: all else equal, it is better to back up the supplier with (1) longer but less frequent disruptions (than the supplier with shorter but more frequent disruptions), and (2) lower demand variability (than the supplier with less predictable demand).[7]

We then considered supply chains with long "time to recovery" and generated further insights into the role of demand distributions. When disruptions are relatively long, using the Renyi's Theorem, we characterized both the optimal base-stock levels and inventory costs. Consequently, we found that a strategy in which unreliable suppliers are backed up according to a minimum excess/wasted backup capacity is optimal. Such a strategy requires evaluation of demand distributions at the available back up capacity, and sheds more light on the importance of considering the whole demand distributions (and not just their first or second moments).

We next focused on the goal of understanding the effect of other factors such as inventory holding and shortage costs as well as purchasing costs in effectively designing flexibility in the backup system. We first considered scenarios in which demand distributions are close to exponential. We established a connection between waiting time in a GI/M/1 queueing system and the inventory shortfall process in our system, which enabled us to characterize both the inventory base-stock levels and costs. In turn, this allowed us to develop an easy-to-compute and yet effective index, which we termed Backup Effect Index (BEI). We showed that when demand distributions are close to exponential, following a *largest BEI policy* is optimal in deciding which unreliable suppliers to back up. We then extended this result by relaxing the exponential demand assumptions. For general demand distributions, we established and analyzed a GI/GI/1 queueing counterpart to the inventory shortfall process in our system. This allowed us to provide a generalized BEI (GBEI), and show that in general backing up suppliers based on a *largest GBEI first policy* provides an effective backup flexibility design. Our indices (BEI and GBEI) provide supply chain designers with easy-to-compute tools to decide which unreliable suppliers to backup, enabling them to effectively compensate for dynamic disruption risks.

Finally, we explored the backup capacity pooling advantage by comparing dedicated backups versus a pooled (i.e., fully flexible) backup capacity in a numerical study. We found that the value of a flexible backup supplier is more than the summation of benefits that can be obtained separately for each of the products through dedicated backups. Furthermore, a firm will reserve at least as much capacity from a backup flexible supplier as the amount reserved in total from dedicated backup ones. Indeed, the flexibility of a supplier provides the firm with greater benefits, justifying reserving more backup capacity because of the economic advantage of shifting the orders whenever necessary (capacity pooling). This observation also justifies flexible suppliers charging higher fees for reserving their capacity compared to inflexible suppliers.

The analyses, modeling framework, and insights presented in this paper can guide new practices to effectively increase the resilience of supply chains. Increasing the resilience of supply chains can in turn enable firms to deliver products with better availability and better prices to end customers, yielding social benefits. Although this study focused on the cost to a firm, a fruitful path for future research is to examine the possibility of creating such broader social advantages. Moreover, future research may expand this study to consider issues such as risk aversion or potential correlations between the dynamic disruption risks of different suppliers. Relaxing the assumption of known demand distributions in the vein of recent data-driven studies (see, e.g., Saghafian and Tomlin 2016 and the references therein) is also another fruitful avenue for future research.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/opre.2016.1478.

Acknowledgments

The authors sincerely thank Brian Tomlin, Wallace Hopp, Xiuli Chao, area editor, Prof. Chung Piaw Teo, and the anonymous associate editor and referees for their invaluable comments and insightful discussions which significantly improved the paper. This work was supported in part by the National Science Foundation [Grant CMMI 1068638] and Office of Naval Research [Grant N00014-08-1-0579]. The authors are also grateful to David Singer for his support through the above ONR grant.

Endnotes

1. Our Markov model is a step forward from the prevalent assumption of i.i.d. Bernoulli disruptions in the literature which does not model the effect of the length of a disruption.
2. We note that Figure 1 (which is our focus) may only depict part of a supply chain: we only treat the n products that have particularly unreliable suppliers and so the flexible backup need only cover these n suppliers modeled. This model also treats the extension to subsets of products forming a partition, each with a fully flexible backup suppliers for one subset of products in the partition. As we will see, our model also allows for partial flexibility.
3. For analysis of regular inventory shortfall in traditional inventory models with a single, fully reliable supplier (under limited capacity) we refer to Tayur (1993) and Glasserman (1997).
4. We consider M to be a sufficiently large number throughout this paper, but our results rigorously follow by taking the limit as M goes to infinity.
5. See, e.g., Prabhu (1965) and Tayur (1993) for some results on inventory models with single, fully reliable, but capacitated suppliers.
6. Note that $D^1 \leqslant_{cx} D^2$ if, and only if, $\mathbb{E}[f(D^1)] \leqslant \mathbb{E}[f(D^2)]$ for all convex functions f defined on the support of D^j $(j \in \{1, 2\})$, which results in $\mathbb{E}(D^1) = \mathbb{E}(D^2)$ but $\mathrm{Var}(D^1) \leqslant \mathrm{Var}(D^2)$; see, e.g., Stochastic Orders by Shaked and Shanthikumar (2007), Chapter 3.
7. In addition to the backup flexibility design, these observations can be also used for supplier selection purposes.

References

Abate J, Choudhury GL, Whitt W (1995) Exponential approximations for tail probabilities in queues, I: Waiting times. *Oper. Res.* 43(5):885–901.
Anupindi R, Akella R (1993) Diversification under supply uncertainty. *Management Sci.* 39(8):944–963.
Arreola-Risa A, DeCroix GA (1998) Inventory management under random disruptions and partial backorders. *Naval Res. Logist.* 45(7):687–703.

Babich V, Burnetas AN, Ritchken PH (2007) Competition and diversification effects in supply chains with supplier default risk. *Manufacturing Service Oper. Management* 9(2):123–146.

Bassamboo A, Randhawa RS, Van Mieghem JA (2010) Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. *Management Sci.* 56(8):1285–1303.

Bassamboo A, Randhawa RS, Van Mieghem JA (2012) A little flexibility is all you need: On the asymptotic value of flexible capacity in parallel queuing systems. *Oper. Res.* 60(6):1423–1435.

Behdani B, Adhitya A, Lukszo Z, Srinivasan R (2012) How to handle disruptions in supply chains—An integrated framework and a review of literature. Working paper, Delft University of Technology.

Blanchet J, Glynn P (2007) Unified renewal theory with applications to expansions of random geometric sums. *Adv. Appl. Prob.* 39(4):1070–1097.

Business Continuity Institute (2011) Supply Chain Resilience, 3rd Annual Survey.

Chen X, Zhang J, Zhou Y (2015) Optimal sparse designs for process flexibility via probabilistic expanders. *Oper. Res.* 63(5):1159–1176.

Chou MC, Teo C-P, Zheng H (2011) Process flexibility revisited: The graph expander and its applications. *Oper. Res.* 59(5):1090–1105.

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

Glasserman P (1997) Bounds and asymptotics for planning critical safety stocks. *Oper. Res.* 45(2):244–257.

Graves SC, Tomlin BT (2003) Process flexibility in supply chains. *Management Sci.* 49(7):907–919.

Gürler Ü, Parlar M (1997) An inventory problem with two randomly available suppliers. *Oper. Res.* 45(6):904–918.

Iravani SM, Van Oyen MP, Sims KT (2005) Structural flexibility: A new perspective on the design of manufacturing and service operations. *Management Sci.* 51(2):151–166.

Jordan WC, Graves SC (1995) Principles on the benefits of manufacturing flexibility. *Management Sci.* 41(4):577–594.

Kim CR (2011) Toyota aims for quake-proof supply chain. *Reuters* (September).

Kingman JFC (1964) A martingale inequality in the theory of queues. *Math. Proc. Camb. Phil. Soc.* 60(2):359–361.

Kingman JFC (1970) Inequalities in the theory of queues. *J. Roy. Statist. Soc. B* 32(1):102–110.

Kleinrock L (1975) *Queueing Systems: Theory*, Vol. 1 (John Wiley & Sons, New York).

Kouvelis P, Li J (2008) Flexible backup supply and the management of lead-time uncertainty. *Production Oper. Management* 17(2):184–199.

Kouvelis P, Vairaktarakis G (1998) Flowshops with processing flexibility across production stages. *IIE Trans.* 30(8):735–746.

Moinzadeh K, Aggarwal P (1997) Analysis of a production/inventory system subject to random disruptions. *Management Sci.* 43(11):1577–1588.

Norrman A, Jansson U (2004) Ericsson's proactive supply chain risk management approach after a serious sub-supplier accident. *Internat. J. Phys. Dist. Log. Management* 34(5):434–456.

Özekici S, Parlar M (1999) Inventory models with unreliable suppliers in random environment. *Ann. Oper. Res.* 91(1999):123–136.

Parlar M, Perry D (1996) Inventory models of future supply uncertainty with single and multiple suppliers. *Naval Res. Logist.* 43(2):191–210.

Parlar M, Wang Y, Gerchak Y (1995) A periodic review inventory model with Markovian supply availability. *Internat. J. Prod. Econom.* 42(2):131–136.

Prabhu NU (1965) *Queues and Inventories* (Wiley, New York).

Ross SM (1974) Bounds on the delay distribution in GI/G/1 queues, *J. Appl Prob.* 11(2):417–421.

Saghafian S, Tomlin BT (2016) The newsvendor under demand ambiguity: Combining data with moment and tail information, *Oper. Res.* 64(1):167–185.

Saghafian S, Van Oyen MP (2011) The "W" network and the dynamic control of unreliable flexible servers. *IIE Trans.* 43(12):893–907.

Saghafian S, Van Oyen MP (2012) The value of flexible suppliers and disruption risk information: Newsvendor analysis with recourse. *IIE Trans.* 44(10):834–867.

Shaked M, Shanthikumar JG (2007) *Stochastic Orders* (Springer, New York).

Simchi-Levi D, Wei Y (2012) Understanding the performance of the long chain and sparse designs in process flexibility. *Oper. Res.* 60(5):1125–1141.

Simchi-Levi D, Schmidt W, Wei Y (2014) From superstorms to factory fires: Managing unpredictable supply-chain disruptions. *Harvard Bus. Rev.* (January–February):96–101.

Simchi-Levi D, Wang H, Wei Y (2013) Increasing supply chain robustness through process flexibility and strategic inventory. Working paper, MIT, Cambridge, MA.

Song J-S, Zipkin PH (1996) Inventory control with information about supply conditions. *Management Sci.* 42(10):1409–1419.

Tang CS (2006) Robust strategies for mitigating supply chain disruptions. *Inter. J. of Log. Res. Appl.* 9(1):33–45.

Tayur SR (1993) Computing the optimal policy for capacitated inventory models. *Comun. Statist.* 9(4):585–598.

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

Tomlin B (2009) The impact of supply-learning on a firms sourcing strategy and inventory investment when suppliers are unreliable. *Manufacturing Service Oper. Management* 11(2):192–209.

Tomlin B, Tang CS (2008) The power of flexibility for mitigating supply chain risk. *Inter. J. Production Econom.* 116(1):12–17.

Tomlin B, Snyder L (2006) On the value of a threat advisory system for managing supply chain disruptions. Working paper, University of North Carolina at Chapel Hill, Chapel Hill, NC.

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

Tsitsiklis JN, Xu K (2015) Flexible queueing architectures. Working paper, Stanford University, Stanford, CA.

Van Mieghem JA (1998) Investment strategies for flexible resources. *Management Sci.* 44(8):1071–1078.

Wang Y, Gilland W, Tomlin B (2010) Mitigating supply risk: Dual sourcing or process improvement? *Manufacturing Service Oper. Management* 12(3):489–510.

Soroush Saghafian is an assistant professor at Harvard Kennedy School, Harvard University. His research focuses on the application and development of operations research methods in studying stochastic systems with specific applications in healthcare and operations management.

Mark Van Oyen is a professor of industrial and operations engeering (IOE) at the University of Michigan. His research emphasizes the analysis, design, control, flexibility, and management of operational systems and supply chains. He has a longstanding interest in the stochastic control of queueing network models. He also enjoys healthcare operations and medical decision making. He was the IOE Dept. Faculty of the Year for 2008–9, an ALCOA Manufacturing Systems Faculty Fellow, and the 2003 Researcher of the Year at Loyola University Chicago's School of Business.