On the Value of a Threat Advisory System for Managing Supply Chain Disruptions

Brian T. Tomlin • Lawrence V. Snyder

Kenan-Flagler Business School, University of North Carolina–Chapel Hill, Chapel Hill, NC, USA

Department of Industrial and Systems Engineering Lehigh University, Bethlehem, PA, USA

brian_tomlin@unc.edu • *larry.snyder@lehigh.edu*

Current version: April 26, 2007

Original version: February 27, 2006

Most firms face some risk of disruption to their supply due to labor strikes, supplier bankruptcies, manufacturing quality failures, natural disasters, terrorism, or other events. In recognition of the fact that the risk of disruption may change over time, some firms have begun to continually monitor their disruption risk and to adjust their safeguards, e.g., inventory, in accordance with the current risk level. Such a system allows the firm to increase its protection level when the disruption risk is high and to operate "lean" otherwise. In this paper, we propose and investigate a threat advisory system in which the firm, operating a periodic-review inventory system, dynamically adjusts its inventory in response to changes in the threat level. We consider a model in which the firm has a single unreliable supplier, for which inventory is the only disruption-management tactic, as well as a model in which a second, perfectly reliable supplier is available, and, thus, sourcing becomes an additional disruption-management tactic. We consider both infinite- and finite-horizon settings. We characterize the optimal threat-dependent inventory levels and show that a threat advisory system can result in substantial cost savings. We establish that supplier capacity and the structure of the disruption risk process (the relative disruption risk in different threat levels and the nature of transitions between threat levels) significantly influence the value of a threat advisory system. We find that the presence of a threat advisory system and the structure of the disruption risk process can have a significant impact on the optimal disruption-management strategy, i.e., the choice of tactics (inventory and/or sourcing) used to manage the firm's disruption risk.

1. Introduction

Supply disruptions present a significant risk for companies, as evidenced by the experiences of both Harley-Davidson and Ford in February 2007. In the case of Harley-Davidson, a three-week strike at its largest plant resulted in the company announcing that motorcycle shipments in the first quarter would be 20% less than planned and that its 2007 earnings-per-share growth would fall in the 4-6% range instead of the projected 11-17% range (Wall Street Journal 2007). Ford was forced to temporarily halt production at their Louisville, Ky., truck plant because their engine supplier, Navistar, cut off supply as a result of a contract dispute. Ford reported that March sales of its profitable F-series pickup truck would be reduced because of the supply stoppage (Kosdrosky 2007).

Strikes and contract disputes are only two of many potential sources of disruptions. Natural disasters, terrorism, supplier bankruptcies, manufacturing quality failures and fires have all contributed to major disruptions in recent years. Firms are increasingly recognizing the significant risk presented by supply disruptions and are taking actions to better manage disruption risks. One strategy that firms are beginning to adopt is to continuously monitor their supply chain for warnings of potential supply-continuity problems. Open Ratings, a Dun and Bradstreet company, has deployed supply-chain monitoring software for large corporations such as Eaton and UTC.

> Eaton is working with Open Ratings to create a company-wide monitoring system that will give advance notice of potential supplier instability in time to put safeguards in place. "We are pleased to expand the use of Open Ratings' solution as part of a comprehensive supply chain strategy," said Richard B. Jacobs, vice president of supply chain management at Cleveland-based Eaton. "Open Ratings will help us gain critical visibility and actionable insight into any potential issues before a disruption occurs." (From Supply & Demand Chain Executive, February 2006.)

> The software toolset uses pattern recognition technology to constantly monitor supplier data to determine if any of UTC's 18,000 suppliers are heading for trouble. In August 2004 the system generated a financial alert based on a recognized pattern of events for a key castings supplier. That partner was immediately identified as being important to a number of product lines, and a system-generated e-mail was sent to the OTL staff warning of a potential bankruptcy ... [and] UTC increased its inventory buffer as an added layer of protection. (From Global Logistic & Supply Chain Strategies, December 2005.)

The decision by UTC to increase its inventory in response to a heightened risk of disruption captures an essential benefit of risk-monitoring systems: rather than continuously maintaining disruption safeguards, such as inventory, firms can increase/decrease their safeguards in response to the current risk level. Sophisticated supply-base monitoring software is not the only approach to tracking the risk of disruption and adjusting safeguards accordingly. Firms typically have some knowledge as to whether the risk of a labor strike is higher than usual. The increased threat of a strike during labor negotiations is a motive for temporarily increasing inventories that is often ascribed to automotive companies. Before the UAW negotiation in July of 2003, there was a significant inventory buildup in the late spring. While the Big-Three automakers would not "acknowledge that the coming UAW talks, which kick-off July 16-18, have influenced the buildup of inventory ... Wall Street analysts think it only natural for automakers to stock up on inventory just in case [of a strike]," (McCracken 2003).

The enhanced focus on disruption risks and the growing ability of firms to monitor the risk of disruption make it likely that firms will increasingly recognize the dynamic nature of disruption risk management. That is, rather than assessing risks at infrequent intervals, e.g., annually, and deploying a static safeguard based on that risk assessment, firms will monitor the risk of disruption on a continuous basis and adjust their safeguards in reaction to changes in the level of risk. Eaton and UTC are examples of firms already using a dynamic threat advisory system to monitor and manage disruption risks. The disruption literature to date has not explored the notion of a threat advisory system. Our paper addresses this gap by proposing and investigating a threat advisory system in which the firm (operating a periodic-review inventory system) can adjust its inventory in response to changes in the current threat level.

We first explore a single-supplier model in which inventory is the only disruption-management tactic available to the firm. We characterize the optimal threat-dependent inventory levels (for both the infinite-horizon and finite-horizon settings) and explore how the optimal levels are influenced by the structure of the disruption risk process, the nature of disruptions, and the remaining horizon length. We investigate the value of using a threat advisory system and establish that the value is significantly influenced by the structure of the disruption risk process (the relative disruption risk in different threat levels and the nature of transitions between threat levels) and the supplier capacity. We then explore a two-supplier model in which the firm can use sourcing tactics and inventory to manage disruption risks. We find that a threat advisory system may change the form of the optimal disruption-management strategy. For example, it may change the optimal strategy from one of acceptance (doing nothing to protect against the disruption risk) to one that adds protection during times of elevated risk. We show that the optimal strategy is heavily influenced by both the relative disruption risk (among threat levels) and by the product lifecycle.

The remainder of this paper is organized as follows. We review the literature on supply disruptions in §2. We describe the model in §3. We consider the single-supplier setting in §4 and the two-supplier setting in §5. Conclusions and directions for future work are presented in §6. All proofs are contained in Appendix A.

2. Literature Survey

The issue of supply uncertainty has received much attention of late in the literature. Hendricks and Singhal (2003, 2005a,b) empirically establish that supply chain disruptions (also called glitches) have a significant negative impact on both the operating performance and the stock price of firms.

Kleindorfer and Saad (2005) present a conceptual framework for the management of disruption risks and, based on empirical data on chemical-industry accidents, provide guidance for the creation of appropriate management systems. Dada et al. (2007) and Federgruen and Yang (2007) analytically explore the benefit of supply diversification in random-supply newsvendor settings. Swinney and Netessine (2007) investigate the role of contracts in mitigating the risk of supplier defaults in environments with uncertain costs.

A number of interpretations of the term "supply uncertainty" have appeared in the analytical literature, namely, disruptions, random yield, random capacity, and stochastic lead times. In *disruption* models, a supplier alternates between up-states and down-states. In up-states orders are filled on time and in full. In down-states, the supplier is unavailable. In *random-yield* models, the quantity received may differ from the quantity ordered by a random amount. We refer the reader to Yano and Lee (1995) for a review of the random-yield literature. Disruptions are binary events, while yield uncertainty often gives rise to smaller but more frequent variability in the delivered quantity. Chopra et al. (2007) consider the differences between these two types of supply uncertainty, and the error inherent in failing to distinguish between the two, in a single-period model, and Schmitt and Snyder (2007) extend their analysis to an infinite-horizon setting. In *random-capacity* models, there is an stochastic upper bound on the production quantity, e.g., Ciarallo et al. (1994). Finally, in *stochastic lead-time* models, orders placed with a supplier are filled in-full but subject to a random delivery time (e.g., Hadley and Whitin (1963), Kaplan (1970), Bagchi et al. (1986), Ramasesh et al. (1991), and Bradley and Robinson (2005)). In this paper, we focus on the disruption type of supply uncertainty.

The first treatment of supply disruptions in the literature appears to be that of Meyer et al. (1979), who consider a production facility subject to stochastic disruptions and repairs. Items produced by the facility are stored in a capacitated buffer that sees constant, deterministic demand. Their model is descriptive rather than prescriptive, characterizing the stockout percentage for a given inventory policy rather than finding the optimal policy.

Parlar and Berkin (1991) present an EOQ-like model with deterministic demand but stochastic disruptions and repairs with the aim of finding the optimal order quantity. Berk and Arreola-Risa (1994) point out two errors in Parlar and Berkin's original model and offer a corrected model. Snyder (2006) proposes an approximation for Berk and Arreola-Risa's cost function that, unlike the exact model, can be solved in closed form and permits a number of analytical results.

A number of papers extend Parlar and Berkin's EOQ-based model. Parlar and Perry (1995) relax the zero-inventory ordering (ZIO) assumption and consider random as well as deterministic yields. Both the reorder point and the waiting time between unsuccessful orders are decision variables, leading to what the authors call a (Q, r, T) inventory policy. Moinzadeh and Aggarwal (1997) consider a model based on the economic production quantity (EPQ) model. They propose a continuous-review (s, S) policy, rather than a (Q, r) policy, since the inventory level may fall strictly below the reorder point during a disruption. Parlar and Perry (1996) consider an EOQ-based model with multiple identical, unreliable suppliers, and Gürler and Parlar (1997) generalize their two-supplier model by allowing more general failure and repair processes.

The few papers to consider stochastic demand in addition to stochastic supply include those of Gupta (1996), who formulates a (Q, r)-based model, Parlar (1997), who studies a more general model but formulates only an approximate cost function, and Arreola-Risa and DeCroix (1998), who propose an (s, S) policy for a periodic-review system with supply disruptions and partial backorders. Chao (1987) and Chao et al. (1989) propose a Markov decision process model for an electricity market with random demand and supply. Snyder and Shen (2006) demonstrate using simulation that different, often opposite, strategies are required for coping with demand uncertainty and supply uncertainty (in the form of disruptions) in a variety of multi-echelon supply chains.

It is worth pointing out that nearly all of the papers on supply failures consider either single-period models or infinite-horizon models. Exceptions include the Bernoulli-yield models of Anupindi and Akella (1993), Parlar et al. (1995), Swaminathan and Shanthikumar (1999), Tomlin (2005a). In addition, with a few exceptions (Chopra et al. 2007, Snyder 2006, Schmitt and Snyder 2007), the models discussed thus far require numerical solution methods. In contrast, this paper considers both finite- and infinite-horizon settings, as well as several special cases that can be solved in closed form.

The papers that are most closely related to ours are those of Song and Zipkin (1996), Li et al. (2004), and Tomlin (2006). Song and Zipkin (1996) consider a very general supply process that can model, for example, supply disruptions or stochastic lead times. Their work established the optimality of a state-dependent base-stock policy (assuming no fixed cost of ordering). Li et al. (2004) consider a model which, like ours, allows the risk of disruption to change over time. However, their model considers *age-dependent* changes to the disruption risk—that is, the probability of a disruption depends only on the number of periods since the last disruption. In contrast, we present a *threat-dependent* model, in which the disruption risk depends on external factors that may be independent of the number of consecutive "up" periods. While an age-dependent model is appropriate for some types of disruptions (e.g., machine breakdowns), a threat-dependent model is more appropriate when (a) the disruption probability does not depend directly on the length of time since the last disruption and (b) the firm can periodically revise its estimate of the disruption risk. Disruptions due to labor disputes, weather, and terrorism, for example, fall into this category, since the disruption probability is dependent on external factors rather than on the time of the last disruption, and since likelihood indicators may be available. Song and Zipkin (1996) and Li et al. (2004) consider only a single, infinite-capacity supplier, whereas our models relax both assumptions. Finally, Tomlin (2006) considers a infinite-horizon, dual-sourcing model in which the firm may order from a cheap but unreliable supplier and/or an expensive but reliable supplier. Tomlin examines the conditions under which the firm's optimal strategy is to manage disruptions by holding extra inventory, by dual sourcing, by emergency sourcing, or by taking no action and simply accepting the disruption risk. Our model extends Tomlin's by allowing the disruption risk to vary stochastically over time (i.e., multiple versus single threat levels) and by considering finite-horizon (as well as infinite-horizon) models. We note also that Tomlin (2005b) considers a single-period, two-product supply-failure model in which the firm can select its disruption-management tactics from dual-sourcing, emergency sourcing or demand management.

3. The Model

We consider a firm that operates a periodic-review inventory system in which it procures product from a single supplier. (A description of the two-supplier generalization is deferred until §5.) There is no fixed cost of ordering and the firm pays the supplier a cost of c per unit. Production at the supplier is instantaneous, but orders incur a transit lead time of L periods. The supplier is subject to random disruptions, that is, in each period the supplier may either be in an up-state or a down-state. When in an up-state the supplier can produce product subject to a (possibly infinite) capacity limit. When in a down-state, the supplier is completely inoperative and can produce no product for the firm. The firm knows the supplier state (up or down) at all times. We will fully characterize the supply process after first describing the remaining features of the model.

We consider both infinite-horizon and finite-horizon settings. In the finite-horizon case, there are $T \geq 1$ periods and the time periods are indexed by $t = 1, \ldots, T$, with T being the final period in the horizon. Demand in each period is drawn from a stationary discrete distribution with strictly positive support and unmet demand is backordered. Excess inventory at the end of a period incurs a holding cost of h per unit, while backorders incur a cost of p per unit. In the finite-horizon case, excess inventory at the end of period T incurs a terminal holding cost of h_T per unit and unmet demand at the end of period T incurs a terminal stockout penalty of p_T per unit, that is, demand not filled by the end of the horizon is lost at cost of p_T per unit. A positive (negative) terminal holding cost h_T indicates ending inventory is disposed (salvaged). If ending inventory can be salvaged, the salvage value is less than or equal to the procurement cost ($-h_T \leq c$). To ensure the firm attempts to fill demand (rather than simply incur the terminal cost of unmet demand) we assume that $c < \beta^T p_T$ where $\beta \in (0, 1]$ is the one-period discount factor. We make the standard assumption that the terminal costs are incurred in addition to the normal holding and stockout costs at the end of period T. The firm minimizes the expected total discounted cost, although, at times, we will consider the long-run average cost criterion for the infinite-horizon setting.

The events in each period t occur in the following sequence:

1. Supplier's state-space transition occurs and the new state is observed.
2. Demand is observed.
3. Order (if any) is placed.
4. Units ordered in period $t - L$ arrive.
5. Demand is satisfied to the extent possible, and excess demands are backordered (or lost in the final period of the finite-horizon case.)
6. All costs are incurred.

The sequence of events given above is the same as that used by Song and Zipkin (1996) with the exception that demands are observed before orders are placed rather than after. Both conventions are common in the literature and are equivalent except that in our sequence of events, lead times are effectively shorter by one period. (See, e.g., Graves (1988), or most of the literature on the beer game, for examples of models using the same sequence as ours.)

We now characterize the supply process. Recall that the supplier is either up (i.e., operative) or down (i.e., not operative). The up-states are divided into N states, called *threat levels*, which represent the likelihood of being disrupted in the *next* period. The transition probability from up-state n to up-state m is given by λ_{nm}. The transition probability from up-state n to a down-state (i.e., the disruption probability in up-state n) is given by λ_{nd}. Neither of these transition probabilities depend on the number of periods that the supplier has been in up-state n. Without loss of generality, the threat levels are indexed in increasing order of their disruption probability; that is, $\lambda_{1d} \leq \lambda_{2d} \leq \ldots \leq \lambda_{Nd}$. When in up-state $n = 1, \ldots, N$, the supplier has a capacity $v_n \leq \infty$, i.e., the supplier's capacity may depend on the threat level. The capacity applies to the order quantity: the firm can order at most v_n units per period when the supplier is up (and 0 units when down).

Define the threat-transition matrix $\mathbf{M}$ as the $N \times N$ matrix containing the transition probabilities among the up-states (i.e., the probabilities of transitioning among threat levels.) Element (n, m) of $\mathbf{M}$ is therefore the transition probability λ_{nm}. We will see later that the threat-transition matrix is an important characteristic of the threat advisory system. We note that the threat-transition matrix completely specifies the disruption probabilities in each threat level. To see this, recall that $\lambda_{nd} = 1 - \sum_{m=1}^{N} \lambda_{nm}$ and λ_{nm} is given by $[\mathbf{M}]_{nm}$. Therefore, $\lambda_{nd} = 1 - \sum_{m=1}^{N} [\mathbf{M}]_{nm}$.

We now define another important characteristic of the threat advisory system, namely, the *threat-ratio* vector $\mathbf{r} = (r_1, \ldots, r_{N-1})$ where r_n is the ratio of the disruption probabilities in threat-levels $n+1$ and n, i.e., $r_n = \lambda_{n+1,d}/\lambda_{nd}$. We refer to r_n as the threat ratio for up-state n, $n = 1, \ldots, N-1$. Because the probability of disruption is non-decreasing in n, we have $r_n \geq 1$. If r_n is large, then threat level $n+1$ is a significantly higher-risk state (in terms of the disruption probability) than threat level n, whereas if $r_n \approx 1$, then the risks are comparable in both threat levels.

Let us now turn our attention to the down states. Let $i = 1, \ldots, \infty$ denote the number of periods (including the current one) for which the supplier has been down. The probability of a disruption ending by transitioning to some up-state m is assumed to depend only on the length of the disruption i and the up-state n in which the disruption originated. This probability is denoted by μ_{nim}. Down-states are denoted by (n, i). When in down-state (n, i), the overall probability of a disruption ending (the "repair" probability) is given by $\mu_{ni} = \sum_{m=1}^{N} \mu_{nim}$ and the probability of the disruption continuing is given by $1 - \mu_{ni}$.

Given the above characterization, the firm's supply process can be modeled as a discrete-time Markov process with up-states indexed by $n = 1, \ldots, N$, representing the threat levels, and down-states indexed by (n, i). The steady-state probabilities of being in up-state n and down-state (n, i) will be denoted by $\pi_u(n)$ and $\pi_d(n, i)$, respectively. The overall steady-state probabilities of being up and down, respectively, are $\pi_u = \sum_{n=1}^{N} \pi_u(n)$ and $\pi_d = \sum_{n=1}^{N} \sum_{i=1}^{\infty} \pi_d(n, i)$.

In closing this section, we note that many of our results can be directly extended to the case of Markov-modulated demand in which the demand distribution depends on the state of the supply process. However, for ease of exposition we focus our attention on the case where the demand distribution is state-independent. Demand in each period can take on values $d_1, d_2, \ldots, d_K$, with d_k a positive integer for each $k = 1, \ldots, K$ and K possibly equal to positive infinity.

4. Single-Supplier Analysis and Results

In the single-supplier setting, the firm's only available disruption-management tactic (Tomlin 2006) is to hold inventory in anticipation of potential disruptions. The existing disruption literature has not considered the possibility of a threat advisory system and, therefore, one key area that we explore in this section is whether there is much value in having threat-level information and what influences this value. We first consider the case of an infinite-capacity supplier, i.e., $v_n = \infty$ for $n = 1, \ldots, N$, and then investigate the effect of finite supplier capacity.

4.1 Infinite Supplier Capacity

As ours is a Markovian inventory-control problem, one can apply the results of Song and Zipkin (1993, 1996) to establish that a state-dependent base-stock policy is optimal. For the general model, it is not possible to analytically specify the optimal base-stock levels. Therefore, we will make some simplifying assumptions in this subsection to allow the development of analytical results. All of the following assumptions will be relaxed when we consider the finite-capacity case. We will assume that (i) demand is deterministic and equal to d units in each period, (ii) there is no discounting, i.e., $\beta = 1$, and (iii) the firm can return excess inventory for a full reimbursement of the purchase cost[1].

4.1.1 Optimal Base Stock Levels

We will first consider the infinite-horizon setting and then turn our attention to the finite-horizon setting. Because there is no discounting, i.e., $\beta = 1$, we use the long-run average cost criterion for the infinite-horizon setting. Let y_n denote the target inventory position at the end of a period in up-state n, after accounting for the (deterministic) demand. We refer to y_n as the *base-stock level* even though, unlike the common definition of that term, it represents the target inventory position after, rather than before, demand is subtracted. There is no benefit to choosing a base-stock level that is less than the lead-time demand Ld, otherwise a shortage cost is unnecessarily incurred in some periods. In addition, there is no benefit to ordering fractional demands. Therefore, we have the following results (for which formal proofs are omitted) for all n: (i) $y_n^* \geq Ld$, and (ii) y_n^* is an integer multiple of d. Define $\hat{y}_n = y_n - Ld$. Clearly, the optimal $\hat{y}_n$ is also an integer multiple of d, so we can write $\hat{y}_n = j_n d$, where j_n is a nonnegative integer. The quantity $\hat{y}_n$ is analogous to safety stock (protecting against supply uncertainty in this case), and j_n represents the number of down-periods covered by up-state n's base-stock level. We therefore refer to j_n as the *coverage* in up-state n. The coverage vector $(j_1, \ldots, j_N)$ completely characterizes a state-dependent base-stock policy. Because unmet demand is backlogged, the procurement cost c is not relevant to determining the optimal infinite-horizon coverages. Let $C(j_1, \ldots, j_N)$ denote the long-run average inventory-shortage cost of a base-stock policy $(j_1, \ldots, j_N)$.

[1]When using a state-dependent base-stock policy, the firm may find itself with more inventory than it wishes to have in a given period. This will happen when transitioning from an up state with a high base-stock level to one with a low base-stock level. If the firm cannot, or does not want to, return excess inventory to the supplier, then its inventory position will exceed the desired base-stock level until this excess inventory has been drained down by demand (assuming another state transition has not occurred in the meantime). In contrast, the firm can and will attain its desired inventory position in every up-state if the supplier allows it to return excess inventory for a full reimbursement. Please see Appendix B for (i) an analytical characterization of the optimal base stock level for a special case of the supply model when returns are not allowed, and (ii) for an analysis of the effect of the returns policy on the firm's cost.

Define $G_n[j] = \pi_u(n) + \sum_{i=1}^{j} \pi_d(n,i)$ and $\overline{G}_n[j] = \sum_{i=j+1}^{\infty} \pi_d(n,i)$. $G_n[j]$ is the steady state probability of being in up-state n or being in a disruption (originating from up-state n) that has lasted j periods or less. For ease of exposition, we refer to $G_n[j]$ as the steady state probability of being in a disruption (originating from up-state n) that has lasted j periods or less, with the understanding that a disruption lasting 0 periods means the supplier is up. Let us define $F_n[j] = G_n[j]/G_n[\infty]$, where $G_n[\infty] = \lim_{j\to\infty} G_n[j]$. $F_n[j]$ is the steady-state probability of being in a disruption (originating from up-state n) that has lasted j periods or less, conditioned on being in up-state n or in a down-state originating from up-state n.

Theorem 1 *The optimal coverage in up-state (threat level) n is $j_n^* = F_n^{-1}\left[\frac{p}{p+h}\right]$.*

This theorem generalizes a result in Tomlin (2006) to systems with multiple threat levels. The optimal coverage is characterized by a newsvendor ratio and by an appropriately defined distribution $F_n[j]$ for the supply process. Since disruptions are more likely in higher threat levels, $F_n[j]$ is increasing in n if $\mu_{ni} = \mu_i$ for $n = 1, \ldots, N$, i.e., the probability of a disruption ending after i periods is independent of the originating up-state n. This leads to the following result.

Corollary 2 *If $\mu_{ni} = \mu_i$ for $n = 1, \ldots, N$, then the optimal coverage in up-state (threat level) n, j_n^*, is nondecreasing in n; that is, the coverage increases as the probability of disruption increases.*

This confirms what one would intuitively expect, the higher the threat level the more inventory is held. This result would not necessarily hold if different threat levels had different disruption characteristics. For example, if disruptions originating from a high threat level tend to be shorter than those originating from a lower threat level, then less inventory may be held in the high threat level.

We now turn our attention to characterizing the optimal base-stock levels for the finite-horizon setting in which the firm minimizes the expected total horizon cost. Let y_{nt} denote the target inventory position (base-stock level[2]) in up-state n at the end of period t, after accounting for the (deterministic) demand. As in the infinite-horizon setting, there is no benefit to choosing a base-stock level that is less than the demand over the lead time, otherwise a shortage cost is unnecessarily incurred in some periods. Note that if $T - t < L$, then demand over the lead time is the demand over the remaining horizon, i.e., $(T - t)d$. Also, there is no benefit to ordering fractional demands. Therefore, we have the following results (again, formal proofs are omitted) for all n and t (i) $y_{nt}^* \geq \min\{L, T - t\}d$, and (ii) y_{nt}^* is an integer multiple of d. Analogous to the

[2]As in the infinite-horizon setting, we refer to y_{nt} as the *base-stock level* even though, unlike the common definition of that term, it represents the target inventory position after, rather than before, demand is subtracted.

infinite-horizon setting, define $\hat{y}_{nt} = y_n - \min\{L, T-t\}d = j_{nt}d$ where j_{nt} is a nonnegative integer that represents the number of down-periods covered by up-state n's base-stock level in period t. The coverage j_{nt} fully specifies the base stock level for up-state n in period t. We note the firm will incur no backorders or terminal stockouts if its coverage is at least as large as the maximum demand that can occur during a disruption. Demand over the length of a disruption must be less than or equal to demand over the remaining horizon, i.e., $(T-t)\,d$ and, therefore, $j^*_{nt} \leq (T-t)\,d$.

Theorem 3 *The optimal coverage in up-state n in period t is $j^*_{nt} = \min\{\hat{j}_{nt}, T-t\}$ where $\hat{j}_{nt}$ is the minimum j such that:*

$$hG_n[j] - p\left(G_n[T-t] - G_n[j]\right) - \left(p_T - c\right)\pi_d\left(n, T-t\right) \geq 0. \tag{1}$$

When the end-of-horizon is far away, the likelihood that a disruption (occurring next period) will last until the end-of-horizon is very low and, therefore, any customers backordered during a disruption will have their demands filled before the end of the horizon with a very high probability. Therefore, the primary purpose of any inventory coverage when far from the end-of-horizon is to protect against backorders. This is the same role played the coverage in an infinite-horizon setting, and one would, therefore, expect the finite-horizon coverage to approach the infinite-horizon coverage as the remaining horizon tends to infinity. This is formalized in the following corollary.

Corollary 4 $j^*_{nt} \rightarrow j^*_n$ *as $T-t \rightarrow \infty$.*

Intuitively, one might expect that the firm would reduce its coverage as the end-of-horizon approaches. In fact, this is not necessarily the case. As the end-of-horizon approaches, there is a higher probability that a disruption (occurring next period) will last until the end-of-horizon. If the supplier is still down at the end of the horizon, then the firm incurs terminal stockout costs, in addition to the accumulated backorder costs, for any demands not filled from the inventory coverage. Therefore, the purpose of the inventory coverage as the end-of-horizon approaches is to protect both against backorders and terminal stockouts. Backorder costs begin to be incurred once the coverage runs out and they accumulate until either the disruption ends or the end-of-horizon is reached. From the perspective of backorders then, disruptions are of less concern as the end-of-horizon approaches because the expected disruption length effectively decreases because the relevant length is limited by the remaining horizon. However, from the perspective of terminal stockouts, disruptions are of more concern as the end-of-horizon approaches because the probability that a disruption lasts until the end-of-horizon increases. Because the backorder concern is dampened whereas the terminal stockout concern is amplified as the end-of-horizon approaches, the terminal stock-out cost p_T and the backorder cost p play a crucial role in determining whether the optimal coverages decrease or increase in the time t, i.e., as the end-of-horizon approaches.

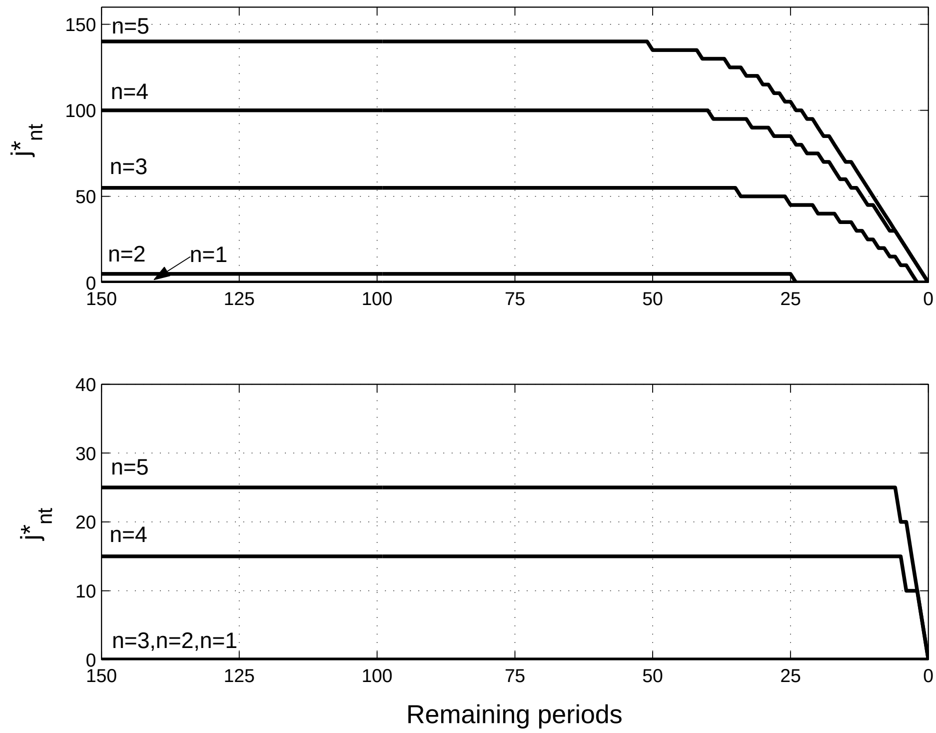
Theorem 5 *(i)* $\hat{\jmath}_{n,t+1} \le \hat{\jmath}_{nt}$ *if* $p_T \le c + p\left(\frac{1}{\mu_{n,T-t-1}} - 1\right)$ *and* $\hat{\jmath}_{n,t+1} \ge \hat{\jmath}_{nt}$ *if* $p_T > c + p\left(\frac{1}{\mu_{n,T-t-1}} - 1\right)$.

Corollary 6 *For the case of geometrically distributed disruption lengths, i.e.,* $\mu_{ni} = \mu_n$ *for* $i = 1, \ldots, \infty$, *(i) the optimal coverage* j^*_{nt} *is non-increasing in* t *if* $p_T \le c + p\left(\frac{1}{\mu_n} - 1\right)$, *(ii) the optimal coverage* j^*_{nt} *is at first non-decreasing and then non-increasing in* t *if* $p_T > c + p\left(\frac{1}{\mu_n} - 1\right)$.

To interpret the inequalities in Corollary 6, first note that, for geometrically distributed disruption lengths, the expected remaining disruption length is $\frac{1}{\mu_n}$, regardless of how long the disruption has lasted. Therefore, a backorder occurring far from the end-of-horizon incurs an expected cost of $c + \frac{p}{\mu_n}$ (p for each period until the end of the disruption, plus c to purchase the unit after the disruption). In contrast, a stockout that is outstanding when the horizon ends incurs a cost of $p_T + p$. Therefore, terminal stockouts are more [less] consequential than normal backorders if $p_T > c + p\left(\frac{1}{\mu_n} - 1\right)$ $\left[p_T \le c + p\left(\frac{1}{\mu_n} - 1\right)\right]$. If the terminal stockout cost is not too high relative to the backorder cost, i.e., $p_T \le c + p\left(\frac{1}{\mu_n} - 1\right)$, then the concern about normal backorders dominates that of terminal stockouts, so as the former concern decreases (because the end-of-horizon is approaching), the optimal coverages decrease, as well. If, instead, p_T is high, then the optimal coverage can increase in t. Even in this case, however, the coverage will eventually start to decrease because (as discussed earlier) the optimal coverage is always less than or equal to the demand over the remaining horizon, i.e., $j^*_{nt} \le (T - t)\, d$.

When far from the end-of-horizon, the optimal coverages are the same as the infinite-horizon coverages. How close must the end-of-horizon be for it to influence the optimal coverages? As we will see, the disruption frequency/duration profile is a key factor in determining when the end-of-horizon begins to influence the coverages. (Frequency refers to how often a disruption occurs and duration refers to the average disruption length.) We illustrate this by considering a problem with a horizon length of $T = 150$ periods, each representing one week. Demand equals 5 units in each period. We use holding and stockout costs $h = 0.005$ and $p = 0.248$, which results in a newsvendor ratio of $\frac{p}{p+h} = 0.98$. We set the terminal holding and stockout costs to $h_T = 0.05$ and $p_T = 1.248$. There are five threat levels and direct transitions between non-adjacent threat levels cannot occur, i.e., $\lambda_{nm} = 0$ if $|m - n| > 1$. The threat-ratio vector is $\mathbf{r} = (3, 3, 3, 3)$. We consider two different disruption profiles: one with rare/long disruptions and one with frequent/short disruptions. Disruptions are geometrically distributed with an expected length of ten weeks for the rare/long profile and an expected length of three weeks for the frequent/short profile. The threat-transition matrix elements for adjacent threat levels are set to achieve an overall (steady-state) probability of being up of $\pi_u = 0.97$ in both cases. Because the disruption

Figure 1: Optimal base-stock level vs. remaining horizon. Upper (lower) graph shows coverage for rare/long (frequent/short) profile.



probabilities are equal for both profiles, the rare/long profile spends much more time in the lower threat-level states than does the frequent/short profile. In Figure 1, we illustrate the optimal base-stock levels as a function of the remaining periods in the horizon. (The optimal coverages are zero in threat levels 1, 2 and 3 for the frequent/short profile, whereas the optimal coverage is zero only in threat level 1 for the rare/long profile.) We see that the end-of-horizon effect starts much sooner for the rare/long profile than it does for the frequent/short profile. The reason is as follows: For disruptions with short expected durations, the probability of a disruption (if it occurs) lasting until the end of the horizon is negligible unless the end-of-horizon is very close, whereas this probability can be significant for disruptions with long expected durations even when far from the end of the horizon.

We note that firms may unwittingly encounter end-of-horizon type effects if managers make inventory decisions based on their remaining tenure in the job rather than on the remaining product life. One can imagine the disruption effect leading managers to draw down inventory that protects against disruptions if the managers are measured solely on the performance during their tenure. Firms are therefore advised to put some procedural safeguards in place to prevent unwanted (from the firm's perspective) inventory draw down. This is especially true for firms facing rare/long disruptions.

4.1.2 Value of a Threat Advisory System

A threat advisory system is valuable because it allows the firm to adapt its inventory coverage to the prevailing conditions as reflected by the current threat level. How valuable is an adaptive (i.e., threat-dependent) policy and what influences the value? We address these two questions by comparing the cost of the optimal threat-dependent policy with the cost of the optimal constant, i.e. threat-independent, base-stock policy. To do so, we first need to characterize the optimal coverage for this threat-independent base-stock policy. We focus on the infinite-horizon setting in what follows. Define $F_c[j] = \sum_{n=1}^{N} \left(\pi_u(n) + \sum_{i=1}^{j} \pi_d(n,i) \right)$. $F_c[j]$ is the steady-state probability of being up (in any threat level) or in a disruption (originating from any threat level) that has lasted j periods or less. This is analogous to $F_n[j]$, defined above, but $F_n[j]$ was conditioned on being in up-state n or in some down-state originating from up-state n.

Theorem 7 *The optimal constant coverage is* $j_c^* = F_c^{-1}\left[\frac{p}{p+h}\right]$.

Before investigating the value of using a threat-dependent policy, it is helpful to contrast the optimal threat-dependent and constant coverages. We note that for a system with a single threat level, i.e., $N = 1$, the threat-dependent and constant policies are identical, that is, $j_c^* = j_1^*$. Explicit solutions (i.e., solutions not using the notation $F^{-1}[\cdot]$) for the optimal threat-dependent coverages, j_n^*, $n = 1, \ldots, N$, and the optimal constant coverage j_c^* can be obtained for certain special cases of our supply system.

Theorem 8 *For the case in which disruption lengths are geometrically distributed (i.e., a constant repair probability) and the repair probability is independent of threat level from which the disruption originated, let the repair probability be denoted by* μ. *Then*

$$j_n^* = \begin{cases} 0, & \frac{p}{p+h} \le \frac{\mu}{\mu+\lambda_{nd}} \\ \left\lceil \frac{\ln\left(\frac{1-\frac{p}{p+h}}{1-\frac{\mu}{\mu+\lambda_{nd}}}\right)}{\ln(1-\mu)} \right\rceil, & \frac{p}{p+h} > \frac{\mu}{\mu+\lambda_{nd}} \end{cases} \tag{2}$$

$$j_c^* = \begin{cases} 0, & \frac{p}{p+h} \le \pi_u \\ \left\lceil \frac{\ln\left(\frac{1-\frac{p}{p+h}}{1-\pi_u}\right)}{\ln(1-\mu)} \right\rceil, & \frac{p}{p+h} > \pi_u \end{cases} \tag{3}$$

where $\lceil x \rceil$ *is the smallest integer greater than or equal to* x.

Observe that j_n^* is identical to j_c^* except that π_u (the steady-state probability of being up) is replaced by $\frac{\mu}{\mu+\lambda_{nd}}$. We note that $\frac{\mu}{\mu+\lambda_{nd}}$ is the steady-state probability of being up conditioned on being in up-state n or some down-state that originated from up-state n. Therefore, there is a direct analogy between $\frac{\mu}{\mu+\lambda_{nd}}$ in the threat-dependent coverage and π_u in the constant coverage. Moreover, as established in the following corollary, the threat-dependent coverage (in up-state n) is higher [lower] than the constant coverage if the conditional steady-state probability of being up is lower [higher] than the overall steady-state probability of being up.

Corollary 9 $j_n^* = j_c^*$ *if* $\frac{\mu}{\mu+\lambda_{nd}} = \pi_u$, $j_n^* \geq j_c^*$ *if* $\frac{\mu}{\mu+\lambda_{nd}} < \pi_u$, *and* $j_n^* \leq j_c^*$ *if* $\frac{\mu}{\mu+\lambda_{nd}} > \pi_u$

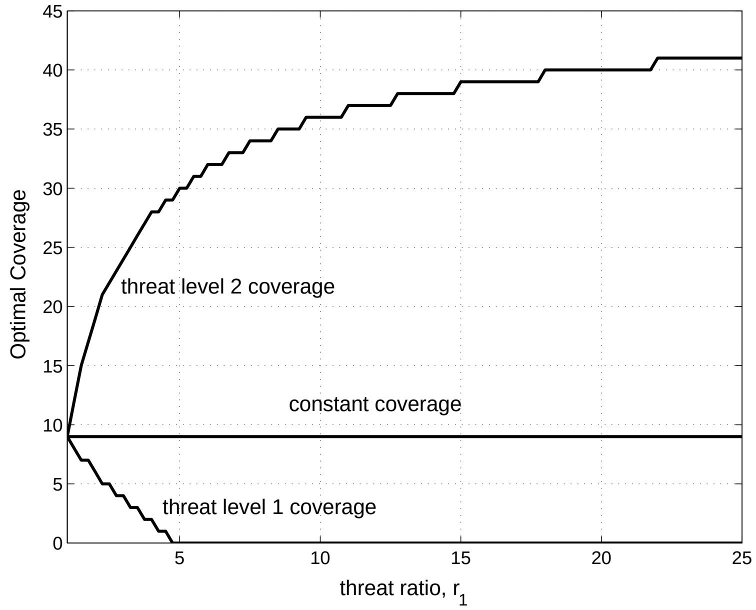
Recall that we previously defined the threat ratio r_n as the disruption probability in threat level $n+1$ relative to that in threat level n. Systems with high r_n values are ones in which the threat levels are highly differentiated. For a system with two threat levels, i.e., $N = 2$, the threat-ratio vector $\mathbf{r} = (r_1, \ldots, r_{N-1}) = r_1$. In Figure 2, we plot the state-dependent coverages j_1^* and j_2^* and the constant coverage j_c^* as a function of the threat ratio r_1 for a system with two threat levels in which $\frac{p}{p+h} = 0.98$, $\pi_u = 0.97$, $\lambda_{12} = 0.005$, $\lambda_{21} = 0.025$, $\mu_1 = 0.045$, $\mu_2 = 0.005$. Disruptions last an average of twenty periods and, therefore, periods are therefore best thought of as days for this example. The disruption probabilities λ_{1d} and λ_{2d} are completely specified by the threat ratio r_1 and the other probability parameters. Observe in Figure 2 that the constant coverage j_c^* is independent of the threat ratio. This is because j_c^* depends on π_u and μ but does not depend on the relative disruption probabilities. In contrast, the threat-dependent coverages are highly sensitive to the threat ratio. As the threat ratio increases, the system becomes more differentiated in terms of the disruption probabilities and so the optimal coverages become increasingly differentiated. In particular, j_2^* increases in the threat ratio while j_1^* decreases. This is because λ_{2d} [λ_{1d}, resp.], the probability of disruption in threat level 2 [1], is increasing [decreasing] in the threat ratio for a constant π_u and a constant repair probability μ. As the threat ratio increases, λ_{1d} eventually becomes so small that the coverage in the low-threat state reaches zero, whereas the coverage in the high-threat state continues to increase.

Let us now turn our attention to exploring the value of using a threat-dependent policy. Recall that $C(j_1, \ldots, j_N)$ denotes the long-run average inventory-shortage cost of a base-stock policy $(j_1, \ldots, j_N)$. Let Δ_{TD} denote the relative savings in inventory-shortage cost obtained by implementing the threat-dependent policy rather than the constant policy. Then,

$$\Delta_{TD} = \frac{C(j_c^*, \ldots, j_c^*) - C(j_1^*, \ldots, j_N^*)}{C(j_c^*, \ldots, j_c^*)}. \tag{4}$$

As established in the proof of Theorem 1, $C(j_1, \ldots, j_N) = \sum_{n=1}^{N} \left(hK_n[j_n] + p\overline{K}_n[j_n]\right)$ where $K_n[j] = jG_n[j] - E_n[j]$, $\overline{K}_n[j] = \overline{E}_n[j] - j\overline{G}_n[j]$, $E_n[j] = \sum_{i=1}^{j} i\pi_d(n,i)$, and $\overline{E}_n[j] = \sum_{i=j+1}^{\infty} i\pi_d(n,i)$.

Figure 2: Optimal coverages (j_1^*, j_2^*, and j_c^*) vs. r_1.

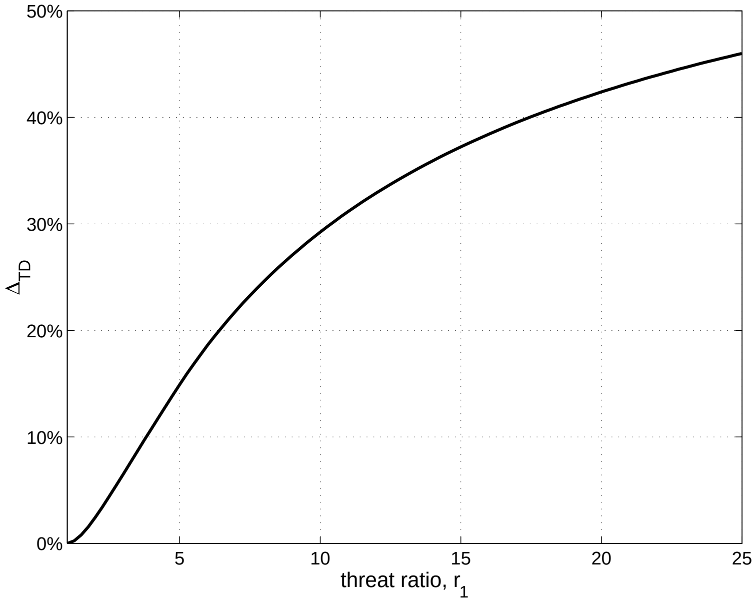


Substituting this into (4) and rearranging terms, we obtain

$$\Delta_{TD} = 1 - \frac{\sum_{n=1}^{N} \left(K_n \left[j_n^* \right] + \frac{p}{p+h} \left(\overline{K}_n \left[j_n^* \right] - K_n \left[j_n^* \right] \right) \right)}{\sum_{n=1}^{N} \left(K_n \left[j_c^* \right] + \frac{p}{p+h} \left(\overline{K}_n \left[j_c^* \right] - K_n \left[j_c^* \right] \right) \right)}. \tag{5}$$

The optimal coverages j_n^* , $n = 1, \ldots, N$, and j_c^* depend on the inventory and shortage costs, h and p respectively, only through the newsvendor ratio $\frac{p}{p+h}$. Therefore Δ_{TD} also depends on these costs only through the newsvendor ratio. An explicit expression for Δ_{TD} can be obtained (but is omitted here) for the two-threat-level case in which disruption lengths are geometrically distributed. Using the same parameters as for Figure 2 above, we plot Δ_{TD} as a function of the threat ratio in Figure 3. We see that the value of the threat-dependent policy, i.e., the relative savings in inventory-shortage cost obtained by using threat-dependent policy, is increasing in the threat ratio and that the relative savings is very high at high threat ratios. The reasons for this are as follows. In the constant-coverage policy, the coverage and resulting expected inventory-shortage cost are independent of the threat ratio. In the threat-dependent policy, the low-threat coverage decreases (eventually to zero) and the high-threat coverage increases as the threat ratio increases but the steady-state probability of being in the low-threat [high-threat] state increases [decreases] in the threat ratio, and so the expected cost decreases.

Figure 3: Value of threat-dependent policy (Δ_{TD}) vs. r_1.



4.2 Finite Supplier Capacity

We now explore the influence of finite capacity on the performance of a threat advisory system. As described in §3, we assume that the supplier has a (possibly threat-dependent) capacity $v_n \leq \infty$, and, therefore, the firm can order at most v_n units per period when the supplier is up (in threat level n) and 0 units when down. Because a state-dependent base-stock policy is optimal in the infinite-capacity case, we assume a modified base-stock policy here: the firm orders either the quantity required to reach the desired (threat-dependent) base-stock level or the supplier's (threat-dependent) capacity, whichever is lower.

In this section, we relax the assumptions that demand is deterministic and that there is no discounting. Furthermore, we no longer allow the firm to return unused inventory. Please see Appendix B for a treatment of the finite-capacity case in which returns are allowed (but possibly not for full reimbursement). In what follows, we first describe the dynamic programming formulation and algorithm and then proceed to explore the effect of capacity both on the performance of a threat-dependent system and on the end-of-horizon behavior of the optimal coverages.

4.2.1 Dynamic Programming Formulation and Algorithm

We focus on the finite-horizon setting in which the firm minimizes its expected total discounted cost. We restrict our attention to the case in which the lead time $L = 0$ in order to avoid a state-space explosion in our dynamic programming algorithm. We also assume that the repair probabilities are independent of both the originating up-state and the length of the disruption.

Let x_t be the inventory position (IP) in period t after demand is observed and subtracted from the IP. For example, if the firm begins the period with 10 units on-hand and the demand equals 20, then $x_t = -10$. Let y_t be the IP at the end of period t, i.e., after an order is placed and received. For example, if $x_t = -10$ and the firm wishes to end the period with 5 units on-hand, then $y_t = 5$ and an order of size 15 must be placed. For ease of exposition, we will at times suppress the subscript t in y_t and x_t. The dynamic program (DP) has two cost-to-go functions, one for up-states and one for down-states. In particular, let

- $f_t(n, x)$ be the optimal expected cost in periods $t, \ldots, T$ if the firm is in up-state n in period t and the IP after demand is observed is x.
- $g_t(x)$ be the optimal expected cost in periods $t, \ldots, T$ if the firm is in a down-state in period t and the IP after demand is observed is x.

An up-period ends with IP equal to y (and a down-period ends with IP equal to x), and so holding and stockout costs are computed based on this value. Therefore, the DP recursions are as follows:

$$f_t(n,x) = \min_{x \le y \le x+v_n} \Bigg\{ c(y-x) + hy^+ + p(-y)^+ + \beta \left[\sum_{m=1}^{N} \lambda_{nm} \sum_{k=1}^{K} q_k f_{t+1}(m, y-d_k) + \lambda_{nd} \sum_{k=1}^{K} q_k g_{t+1}(y-d_k) \right] \Bigg\}, \qquad (6)$$

$$g_t(x) = hx^+ + p(-x)^+ + \beta \left[\sum_{n=1}^{N} \mu_n \sum_{k=1}^{K} q_k f_{t+1}(n, x-d_k) + (1-\mu) \sum_{k=1}^{K} q_k g_{t+1}(x-d_k) \right], \qquad (7)$$

where $a^+ \equiv \max\{a, 0\}$. Recall that μ_n is the probability that a disruption ends by returning to up-state n, $1-\mu$ is the probability of a disruption continuing, and q_k is the probability that demand equals d_k. We note that, for ease of exposition, we have presented the DP for state-independent demand as we do not focus on Markov-modulated demand in this paper. However, the algorithm has been coded to allow for Markov-modulated demand in which the demand distribution can depend on whether the supplier is up or down, and on which down-state the firm is in. The following description of the implementation (including run time) also applies to a Markov-modulated demand DP formulation.

The number of reachable up-states (n, x) and down-states x may be large and even potentially infinite. To avoid computing $f_t(n, x)$ and $g_t(x)$ for every possible state, we define a range of reasonable IPs such that the probability of reaching other IPs is sufficiently small, then compute the cost-to-go functions only for those states. In particular, we assume that the IP in threat level n is always in the range $\underline{x}(n), \ldots, \bar{x}(n)$, and the IP in down-states is always in the range $\underline{x}_d, \ldots, \bar{x}_d$. If equations (6) and (7) require the evaluation of $f_t(n, x)$ or $g_t(x)$ for x values not in their permissible ranges, we set

$$
\begin{aligned}
f_t(n,x) &= f_t(n,\underline{x}(n)) \quad \text{if } x < \underline{x}(n) \\
f_t(n,x) &= f_t(n,\bar{x},(n)) \quad \text{if } x > \bar{x}(n) \\
g_t(x) &= g_t(\underline{x}_d) \quad \text{if } x < \underline{x}_d \\
g_t(x) &= g_t(\bar{x}_d) \quad \text{if } x > \bar{x}_d
\end{aligned} \tag{8}
$$

The period-by-period state probabilities may be obtained as a by-product of our DP algorithm, and these may be examined after the algorithm terminates to ensure that the ranges $\underline{x}(n),\ldots,\bar{x}(n)$ and $\underline{x}_d,\ldots,\bar{x}_d$ are sufficiently large. Using this state-space truncation, the DP can be solved as follows:

Initialization.

Set $f_{T+1}(n,x) = h_T x^+ + p_T(-x)^+$ for all $n = 1,\ldots,N$, $x = \underline{x}(n),\ldots,\bar{x}(n)$.

Set $g_{T+1}(x) = h_T x^+ + p_T(-x)^+$ for all $x = \underline{x}_d,\ldots,\bar{x}_d$.

Set $t = T$.

Recursion.

Set $f_t(n,x)$ using (6) and (8) for all $n = 1,\ldots,N$, $x = \underline{x}(n),\ldots,\bar{x}(n)$.

Set $g_t(x)$ using (7) and (8) for all $x = \underline{x}_d,\ldots,\bar{x}_d$.

Determine $y_t(n)$ for all $n = 1,\ldots,N$ as described below.

Termination.

If $t > 1$, set $t = t-1$ and go to 2. Otherwise ($t = 1$), STOP; the optimal cost is given by $f_1(n_1,x_1)$, where n_1 and x_1 are the (exogenously specified) initial state and inventory position. (We assume for convenience that the system starts in an up-state, but this is an easy assumption to relax).

If the firm is in up-state n and has IP x_t in period t, the optimal base-stock level, denoted $y_t^*(n,x)$, is given by replacing "min" with "argmin" in the right-hand side of (6). However, we are interested in $y_t^*(n)$, not $y_t^*(n,x)$; that is, the base-stock level should be independent of the IP x. To find $y_t^*(n)$, we can simply set $y_t^*(n)$ equal to the optimal base-stock level for the smallest possible IP x *assuming infinite capacity in period* t. This base-stock level represents the desired target IP, though this IP may not be attainable from every starting IP x because of the capacity constraints. (To be clear, the resulting base-stock levels are optimal for the finite-capacity case. The infinite-capacity assumption is necessary simply to avoid the artificially small order-up-to levels that occur if the starting inventory is small but the capacity is finite. We perform this calculation in a separate step, after the DP has completed, accounting fully for the finite capacity in future periods.)

The CPU time required for the algorithm to execute increases with the size of the problem (number of time periods and threat levels), as well as with the size of the ranges $\underline{x}(n),\ldots,\bar{x}(n)$

and $\underline{x}_d, \ldots, \bar{x}_d$. The CPU time also increases with the supplier capacity, since smaller capacities reduce the range of feasible actions in each state. For a reasonably large problem instance, e.g., 150 periods, 5 threat levels, with $\underline{x}(n) = \underline{x}_d = -300$ and $\bar{x}(n) = \bar{x}_d = 300$, the algorithm (coded in VBA) took between 30 seconds and several minutes on a desktop computer, with the run time primarily dependent on the capacity. Since we are primarily interested in analyzing the firm's strategies, rather than the computational performance of the algorithm, we omit a formal analysis of CPU times.

4.2.2 The Effect of Capacity on a Threat Advisory System

Capacity influences the optimal base stock levels (and the resulting performance) of a threat advisory system in two ways:

- Recovery Effect: If the supplier capacity is infinite, then the firm can recover fully from a disruption, i.e., attain an inventory position equal to the desired base-stock level, in the first period after a disruption ends. If the supplier capacity is finite, however, then it may take the firm a number of periods to attain the desired base-stock level. As the supplier capacity decreases, the time to recover from a disruption (i.e., attain the desired base-stock level) increases. A disruption that occurs during this recovery phase is more significant than one that occurs after recovery because the inventory position has yet to reach the optimal base-stock level. To protect against this "recovery effect", the optimal base-stock level increases as the supplier capacity decreases. This recovery effect was discussed by Tomlin (2006) for a single-threat system but is still relevant for multiple-threat systems.

- Threat-Transition Effect: In a multiple-threat system, the firm will want to increase its inventory position after transitioning between certain threat levels. For example, as proven earlier for the infinite-capacity setting, if disruption repair probabilities are threat-independent, then the optimal base-stock levels are increasing in the threat level n, and, so, the firm will want to increase its inventory position upon transitioning to a higher threat level. If supplier capacity is finite, then there will be a transient phase during which the firm has yet to attain the desired base-stock level. The lower the supplier capacity is, the longer this post-transition transient phase will be. The consequence of a disruption during this transient phase is more significant than that of a disruption after the desired base-stock level has been attained. To protect against this, the optimal base-stock level in lower threat levels will increase because this ensures that the firm will start from a higher inventory position when transitioning to a higher threat level. We call this the "threat-transition effect". This threat-transition effect increases base-stock levels above and beyond any increase caused by the recovery effect discussed above.

In what follows, we explore how the threat-transition matrix $\mathbf{M}$ influences the magnitude of these capacity-related effects. We focus on two key characteristics of a threat advisory system: (i) the structure of the threat transition matrix $\mathbf{M}$, and (ii) the threat-ratio vector $\mathbf{r} = (r_1, \ldots, r_N)$, defined earlier. By the structure of $\mathbf{M}$, we mean the set of possible threat transitions. For example, a "completely connected" structure, denoted $\mathbf{M}_C$, is one in which any threat level can transition directly to any other threat level, i.e., $\lambda_{nm} > 0$ for any threat-level combination m and n, whereas a "sequentially connected" structure, denoted $\mathbf{M}_S$, is one in which a threat level can transition directly only to an adjacent threat level, i.e., $\lambda_{nm} > 0$ if and only if $|m - n| \leq 1$.

We now introduce some metrics that will be helpful in what follows. Let v be the capacity of the supplier (assumed for convenience to be the same in all up-states), let $E[D]$ be the expected demand, and let $\rho = E[D]/v$. The quantity ρ is analogous to utilization, although not identical to it because expected production does not necessarily equal expected demand due to base-stock-level transitions. We will, however, refer to ρ as utilization. Let $J^*(\rho)$ denote the optimal (discounted) horizon cost for a given ρ. Let $\Psi(\rho) = \frac{J^*(\rho)}{J^*(0)}$, i.e., $\Psi(\rho)$ is the horizon cost with finite capacity relative to that with infinite capacity. We note that $\Psi(\rho) \geq 1$ and, for simplicity, we refer to $\Psi(\rho)$ to as the relative cost.

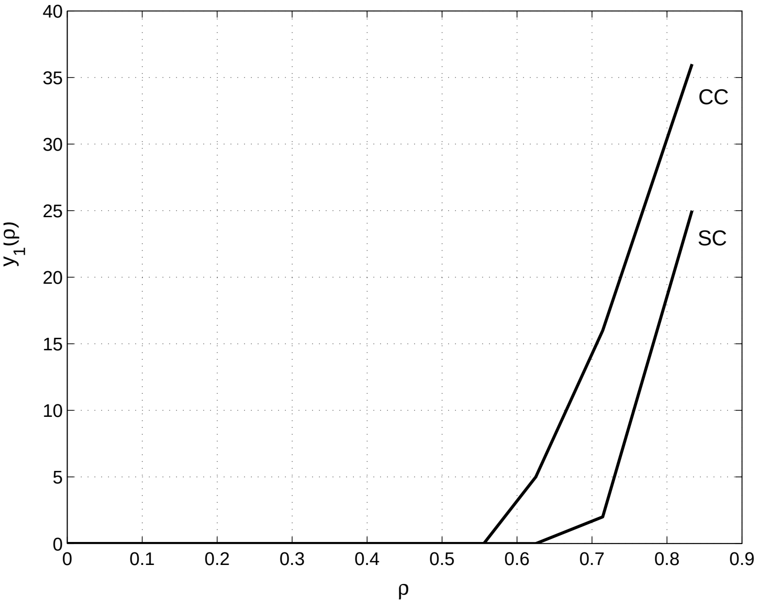
For illustration, we consider a five-threat-level system with a horizon length of $T = 150$ periods, each representing one week. We use holding and stockout costs $h = 0.005$ and $p = 0.248$, which results in a newsvendor ratio of $\frac{p}{p+h} = 0.98$. We set the terminal costs to $h_T = 0.05$ and $p_T = 1.248$. Demand is stationary but stochastic with a discrete distribution: the possible values are 3, 4, 5, 6 and 7, with probabilities 0.1, 0.2, 0.4, 0.2, and 0.1 respectively, which results in an expected demand of 5 units per period. We assume the disruption length is geometrically distributed, with a repair probability of $\mu = 0.211$, i.e., disruptions last an average of approximately five weeks. We set the disruption probabilities as follows: $\lambda_{1d} = 0.000725$ and the threat-ratio vector $\mathbf{r} = (r_1, \ldots, r_4) = (r, r, r, r)$ where $r = 3$. (The impact of changing the threat ratio is explored later.) We consider both a completely connected and a sequentially connected threat-transition matrix structure. We set the transition probabilities such that the long-run probability of the supplier being up is $\pi_u = 0.97$ for both threat-transition structures, although the system may never attain this probability due to the finite horizon. The elements of the threat transition matrices are available upon request.

Figure 4(a) shows the optimal base-stock level in threat level 1 (y_1^*) as a function of utilization ρ for both threat-transition matrices. In the infinite-capacity case, i.e., $\rho = 0$, the optimal base-stock level is 0 for both threat-transition matrices as the probability of a disruption is very low. As ρ increases, i.e., capacity decreases, the optimal base-stock levels initially remain at 0 because there is still sufficient capacity to immediately attain a higher threat level's base-stock level. As ρ increases further, the optimal base-stock levels start to increase. The probability of a disruption in threat level 1 does not change, and so the increase in the base-stock level is due solely to the recovery and threat-transition effects discussed above. We note that the base-stock level increases more rapidly for the completely connected structure. This is because the threat-transition effect is more pronounced for this structure since threat level 1 can transition directly to much higher threat levels, e.g., threat level 5, whereas multiple transitions have to occur in the sequentially connected structure to move from threat level 1 to threat level 5. Therefore, the sequentially connected structure has more time to increase its inventory position if the threat level evolves from 1 to 5. The threat-transition effect (on the base-stock level in threat level 1) is, therefore, more significant for the completely connected structure.

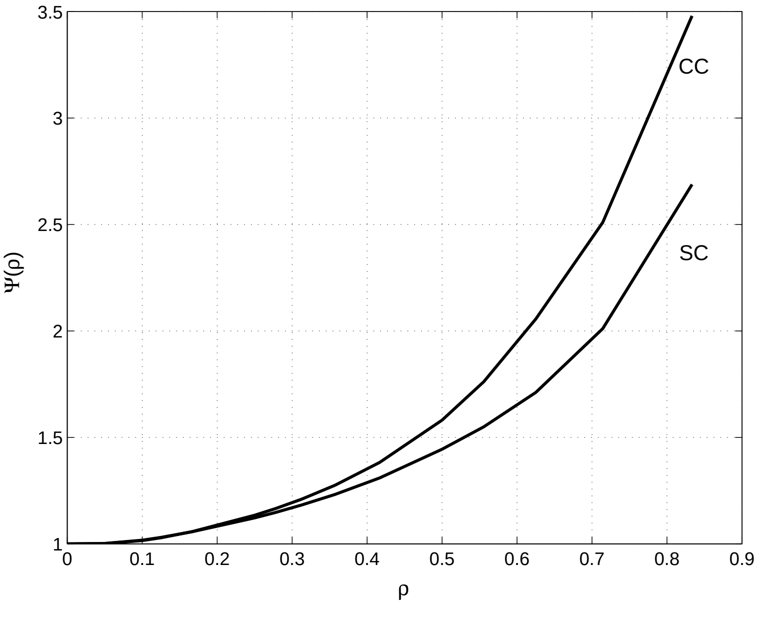
In Figure 4(b), we plot the relative cost $\Psi(\rho)$ as a function of ρ. We see that the relative cost increases significantly as ρ increases, i.e., as capacity decreases and approaches the expected demand. In particular, the capacity effect is greater for the completely connected structure. Again, this is because the threat-transition effect is more significant for this structure. In general, the threat-transition effect is more significant for transitions between non-adjacent threat levels than for transitions between adjacent threat levels because the difference in base-stock levels is larger. Capacity effects are thus greater for threat-transition structures in which non-adjacent transitions can occur.

We next investigate how the relative cost $\Psi(\rho)$ is affected by the threat-ratio vector $\mathbf{r} = (r_1, \ldots, r_4)$. We focus on the completely connected threat-transition structure and assume $\mathbf{r} = (r, r, r, r)$, i.e. the same threat ratio for all threat levels. We use the same problem parameters as before, except that the threat-transition probabilities are adjusted as r increases so as to keep $\pi_u = 0.97$. The specific threat-transition matrices are available from the authors upon request. In Figure 5, we plot $\Psi(\rho)$ as a function of ρ for $r = 1, 3, 5$. We see that $\Psi(\rho)$ increases more rapidly as the threat ratio increases. A threat ratio of 1 indicates a system in which all threat levels have the same probability of disruption. Because the repair probabilities are independent of the originating up state in our example, the $r = 1$ case is effectively a single-threat-level system and so the optimal base-stock levels are identical in all threat levels. As such, the cost increase observed for the $r = 1$ case in Figure 5 is caused purely by the recovery effect of finite capacity. The cost increase observed for the $r = 3$ and $r = 5$ cases results from both the recovery effect and the threat-transition effect of finite capacity. This combined effect, therefore, causes the relative cost increase to be larger when $r > 1$. We see that $\Psi(\rho)$ increases more rapidly as the threat ratio increases. This is be-

Figure 4: (a) Optimal base-stock level in threat-level 1 (y_1^*) and (b) percentage cost increase ($\Psi(\rho)$) vs. ρ for completely connected (CC) and sequentially connected (SC) transition structures.

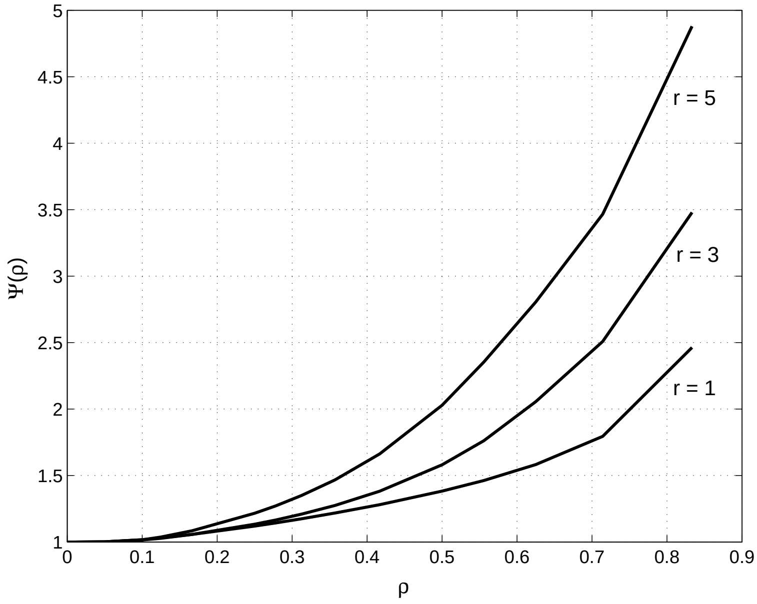


(a)



(b)

Figure 5: Percentage cost increase ($\Psi(\rho)$) vs. ρ for various threat ratios r.



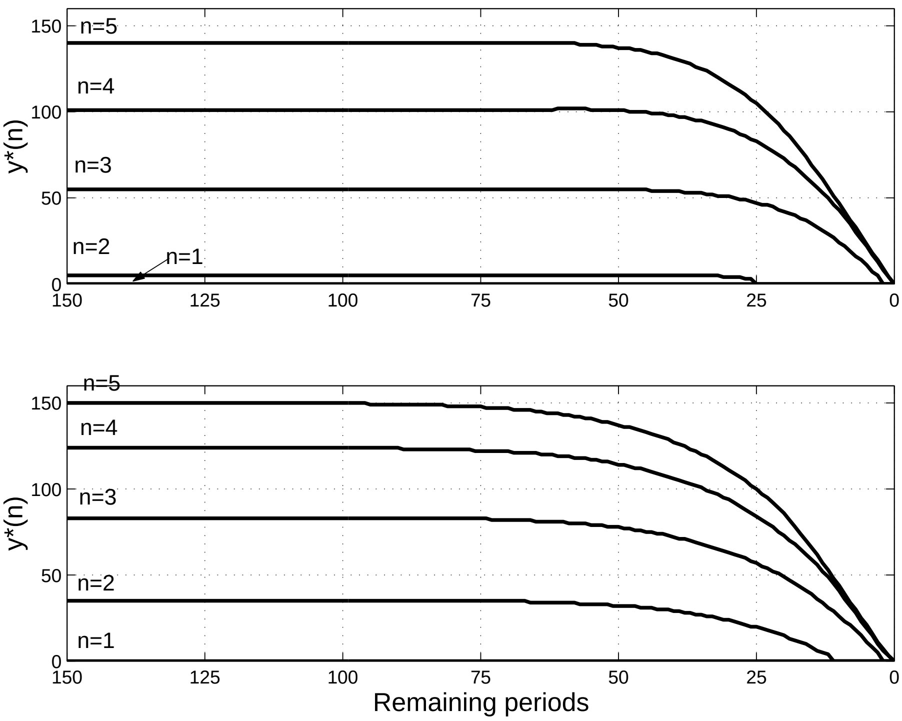
cause for a given π_u, threat levels become increasingly differentiated as the threat ratio increases. Increasingly differentiated threat levels result in increasingly differentiated base-stock levels in the infinite-capacity case, and therefore the threat-transition effect (in the finite-capacity case) is more significant as the threat ratio increases.

In summary, the lower the supplier capacity, the less beneficial is a threat advisory system because the threat-transition effect (whereby low-threat coverages are increased because high-threat coverages cannot be instantaneously attained if the capacity is low) reduces the firm's ability to have highly differentiated coverages, and this differentiated coverage is central to the value of a threat advisory system. This effect is more pronounced if the threat-transition matrix is highly connected or if the threat ratios are high.

4.2.3 The Influence of Capacity on the End-of-Horizon Coverage Effect

In exploring the optimal coverages in the infinite-capacity case, we explored the end-of-horizon effect, whereby the optimal coverages change as the end-of-horizon approaches. In closing this section, we briefly discuss the influence of capacity on the end-of-horizon effect. Figure 1 in §4.1.1 illustrated the optimal coverages (for both rare/long and short/frequent disruption profiles) as a function of the remaining horizon length in the case of infinite capacity. Using the same parameters (except that demand is stochastic with a mean of 5 rather than deterministically equal to 5), Figure 6 illustrates the optimal base-stock levels for the rare/long disruption profile for both infinite and finite capacity (equal to 10). As can be seen, the optimal coverages begin to change sooner in

Figure 6: Optimal base-stock level vs. remaining horizon. Upper (lower) graph shows coverage for infinite-capacity (finite capacity) case.



the finite-capacity case. In the infinite-capacity case, the firm can instantaneously recover from a disruption that finishes before the end of the horizon and, therefore, the coverage is only to protect against a disruption that may occur in the next period. Thus, the remaining horizon only affects the coverage through the possibility that a disruption (occurring next period) will last until the end of the horizon. In the case of finite capacity, the firm cannot recover immediately when a disruption ends or attain a higher inventory position immediately following a transition to a higher threat state. Therefore, the coverage in the finite-capacity case protects against both the possibility of a disruption in the next period and the possibility of future disruptions. The probability of future disruptions decreases as the remaining horizon decreases and, so, the end-of-horizon influences the optimal coverages sooner in the finite-capacity case. Moreover, because the probability of future disruptions decreases as the end-of-horizon approaches, this end-of-horizon effect acts to reduce the coverages. As noted before, the other end-of-horizon effect (the higher probability of a disruption lasting for the remaining horizon) can act to either decrease or increase the coverage depending on the terminal stockout cost relative to the backorder cost.

5. Two-Supplier Extension

We now extend our model to the situation in which the firm can source from two suppliers. This enables the firm to use sourcing in addition to (or in place of) inventory to manage disruptions. We first present the two-supplier model and then proceed to explore how the firm's disruption-management strategy is influenced by (i) the presence of a threat-advisory system and (ii) the product lifecycle. Throughout this section, we assume demand to be stochastic as described in §3.

5.1 Two-Supplier Model

There are two suppliers available to the firm: an unreliable supplier (U) that is subject to random disruptions and a reliable supplier (R) that is not subject to disruptions but is more costly than supplier U. Units procured from supplier U (R) cost c_u (c_r) per unit and $c_u \leq c_r$. Both suppliers have an instantaneous lead time. The firm has a threat advisory system for supplier U and this is modeled in exactly the same manner as described in §3. All other features of the single-supplier model continue to apply here. In the two-supplier system, there are two additional disruption-management tactics, beyond holding inventory, available to the firm: sourcing mitigation and contingent rerouting.

Sourcing mitigation refers to a sourcing policy in which the firm routinely, i.e., every period, sources a positive fraction $0 < w \leq 1$ of its demand from supplier R (regardless of whether supplier U is up or down). Sourcing mitigation dampens the effect of disruptions because a disruption only affects a portion $1 - w$ of the firm's supply. If the firm chooses $w = 0$, then it single sources from supplier U and the single-supplier model is recovered.

Contingent rerouting refers to a sourcing policy whereby the firm temporarily increases the quantity sourced from supplier R during a disruption to supplier U, that is, the firm procures additional units beyond the routine fraction w. These additional units have a cost of $c_f \geq c_r$ per unit. Rerouting is only allowed during a disruption to supplier U. As noted by Tomlin (2006), contingent rerouting may be limited by supplier R's volume-flexibility profile and the firm's response time to a disruption. Since Tomlin (2006) has explored the impact of volume-flexibility profiles and firm response time (in the context of a single-threat-level system), we choose to focus on what Tomlin (2006) calls the zero-flexibility and the instantaneous/infinite-flexibility (II-flexibility) cases. In the zero-flexibility case, the firm cannot reroute any quantity to supplier R, i.e., contingent rerouting is not possible. In the II-flexibility case, the firm can reroute any quantity and can begin this rerouting in the first period of a disruption. We note that zero-flexibility can be obtained as a special case of II-flexibility by setting $c_f = \infty$.

For a given $0 \le w \le 1$, the sequence of events in each period t mirror those in the single-supplier model except for the ordering event (#3), which is modified to reflect the routine fraction ordered from supplier R and the fact units can be ordered from supplier R (albeit at a higher cost c_f) during a disruption to supplier U under the contingent rerouting strategy

1. Supplier U's state-space transition occurs and the new state is observed.
2. Demand d_t is observed.

3a. An order is placed to supplier R for wd_t units.

3b. An order is placed with supplier U (during up-states) or supplier R (during down-states) for any remaining items desired.

4. Units ordered arrive.
5. Demand is satisfied to the extent possible, and excess demands are backordered (or lost in the final period of the finite-horizon case.)
6. All costs are incurred.

In choosing a disruption-management strategy, the firm can choose a any combination of inventory mitigation, sourcing mitigation and contingent rerouting (assuming $c_f < \infty$.) Sourcing mitigation can either be partial, i.e., $0 < w < 1$, or complete, i.e., $w = 1$. In what follows, the term "sourcing mitigation" refers to the complete case unless it is preceded by the word "partial". The set of possible optimal strategies is the same in the multiple-threat-level case as in a single-threat, infinite-horizon model (Tomlin 2006) but what we wish to explore is how the optimal strategy is influenced by a threat-advisory system and the product lifecycle.

5.2 Influence of a Threat Advisory System on the Disruption-Management Strategy

In this subsection, we first consider the infinite-capacity case and then briefly discuss the impact of finite supplier capacity. We focus on the infinite-horizon setting in what follows.

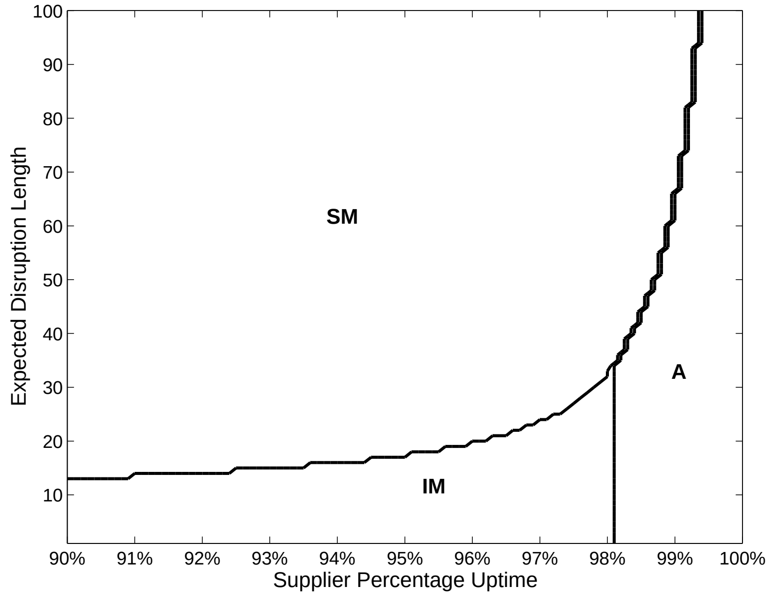
Theorem 10 *If supplier U has infinite capacity, then (i) For a given supplier-R allocation w, a state-dependent base-stock policy is optimal. (ii) Single-sourcing is optimal, that is, $w^* \in \{0, 1\}$.*

Partial sourcing mitigation, therefore, is not optimal in the case of infinite supplier capacity. In the zero-flexibility case, then, only the following three disruption-management strategies can be optimal: acceptance (A), i.e., source exclusively from supplier U and carry no inventory to mitigate disruptions; inventory mitigation (IM), i.e., source exclusively from supplier U but carry some inventory to mitigate disruptions; and sourcing mitigation (SM), i.e., source exclusively from supplier R, in which case there are no disruptions. In the II-flexibility case, there are two additional strategies that can be optimal. One is contingent rerouting (CR), in which the firm sources exclusively from supplier U (when U is up) and carries no inventory to mitigate disruptions but instead reroutes some lost supply to supplier R during a disruption. The other possible optimal strategy is inventory mitigation combined with contingent rerouting (IMCR). This strategy is identical to CR except that the firm carries some inventory to mitigate disruptions, as this allows it to reduce the amount of lost supply rerouted to supplier R.

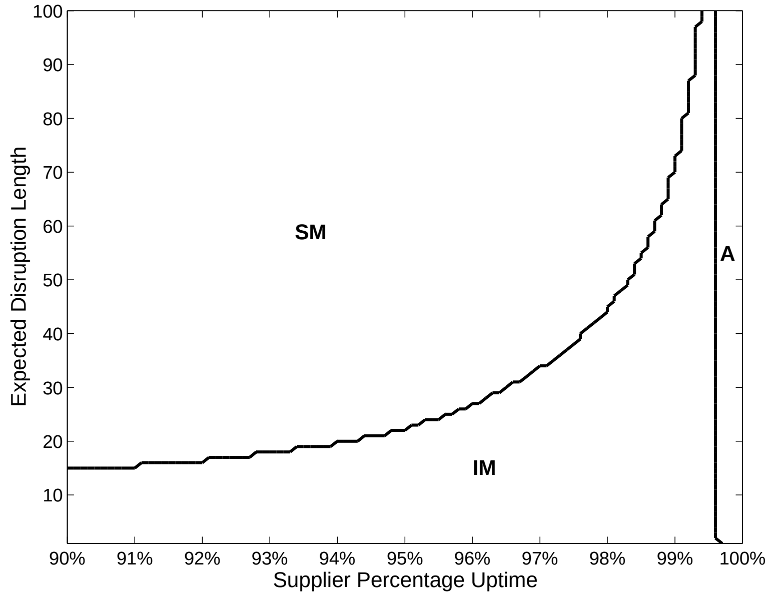
For the zero-flexibility case, Figure 7 illustrates the firm's optimal disruption-management strategy as a function of supplier U's percentage uptime π_u and the expected disruption length. Figure 7(a) illustrates the firm's optimal strategy if it does not have a threat advisory system (and so has to use a constant inventory policy) and Figure 7(b) illustrates the optimal strategy if the firm has a threat advisory system (and so can use an adaptive inventory policy). These examples assume a two-threat system with a threat ratio of $r = 10$. Demand is deterministic (assumed equal to 1 without loss of generality) and excess inventory can be returned for a full reimbursement. The other relevant problem parameters are: $d = 1$ $c_u = 1$, $c_r = 1.05$, $p = 0.075$, $h = 0.0015$, $\lambda_{12} = 0.005$, $\lambda_{21} = 0.025$, and $\frac{\mu_1}{\mu_2} = 9$.

The optimal strategy is significantly altered by the presence of a threat advisory system. In comparing Figures 7(a) and (b), we observe that acceptance (A) is replaced by inventory mitigation (IM) over quite a large area in Figure 7(b). We also observe that the boundary between inventory mitigation and sourcing mitigation (SM) is slightly higher in Figure 7(b), i.e., inventory mitigation displaces sourcing mitigation. A threat advisory system allows the firm to increase its inventory when the threat is high and reduce it when the threat is low. By enabling the firm to increase/decrease its inventory in response to the current threat level, a threat advisory system makes inventory mitigation a more attractive disruption-management strategy. Because a threat advisory system reduces the cost of mitigating disruptions with inventory, the firm can cost-effectively manage its disruption risk (even at high supplier percentage uptimes) rather than simply accepting it, i.e., IM displaces A in comparing Figure 7(b) to Figure 7(a). As discussed by Tomlin (2006) for a single-threat-level system, inventory mitigation is too expensive if disruptions are rare but long (i.e., as the expected disruption length increases for a fixed percentage uptime) because the firm has to carry a large amount of inventory over long periods without a disruption. While this observation remains true in the presence of a threat advisory system, the boundary at which inventory mitigation is no longer optimal is a little higher in the presence of a threat advisory system (i.e., comparing Figure 7(b) to Figure 7(a)) because a threat advisory system enables the firm to carry a large quantity of inventory only if the threat level is high rather than carrying it continuously.

Figure 7: Optimal disruption-management strategy (a) without and (b) with a threat advisory system.



(a)



(b)

As discussed in §4.2.2, capacity diminishes the attractiveness of a threat advisory system as it limits the firm's ability to take advantage of threat-dependent coverages. Therefore, as capacity becomes tighter, the threat advisory system will have less of an effect on the optimal disruption-management strategy. It is important to note that the single-sourcing result of Theorem 10 does not hold if supplier U has finite capacity. It may be optimal to routinely dual source, i.e., $0 < w^* < 1$ in this case. See (Tomlin 2006) for a discussion of this effect in a single-threat system.

5.3 Influence of Product Lifecycle on the Disruption-Management Strategy

The disruption literature to date has focused either on infinite-horizon models (appropriate for long-lifecycle products) or on single-period models (appropriate for short-lifecycle products with long lead times). Our finite-horizon model allows us to investigate products whose lifecycles lie between these two ends of the spectrum.

The firm's problem is to choose the fraction $0 \leq w \leq 1$ of demand to routinely source from supplier R and, given this w, the base-stock level for each period and supply state. For a given w, we adapt our single-supplier DP formulation to reflect the presence of supplier R. In what follows, we describe the necessary modifications to the DP. As before, we suppress the time-dependence of the inventory position x_t and the base-stock levels y_t for ease of exposition.

We define x to be the inventory position (IP) after event #3a (rather than #2 as was done in the single-supplier case); this IP accounts for the routine order placed on supplier R. Therefore, if the firm is in an up-state, begins the period with IP equal to x', and experiences a demand of d, then $x = x' - d + wd$. If the firm wishes to bring its IP to y, it incurs a total ordering cost of $c_r wd + c_u(y - x)$ if in an up-state and $c_r wd + c_f(y - x)$ if in a down state, reflecting the fact that the firm can place emergency orders (at a higher cost) with supplier R if supplier U is down. As noted previously, the case in which emergency orders are not allowed is captured by setting $c_f = \infty$. As in the single-supplier DP, $f_t(n, x)$ is the minimum possible expected cost in periods $t, \ldots, T$ if the firm is in up-state n in period t and the IP is x, and $g_t(x)$ is the minimum possible expected cost in periods $t, \ldots, T$ if the firm is in a down-state in period t and the IP is x. The DP recursions are as follows:

$$f_t(n,x) = \min_{x \le y \le x+v_n} \Bigg\{ c_u(y-x) + hy^+ + p(-y)^+ + \beta \Bigg[\sum_{m=1}^{N} \lambda_{nm} \sum_{k=1}^{K} q_k[f_{t+1}(m, y-d_k+wd_k) + c_r wd_k] + \lambda_{nd} \sum_{k=1}^{K} q_k[g_{t+1}(y-d_k+wd_k) + c_r wd_k] \Bigg] \Bigg\} \tag{9}$$

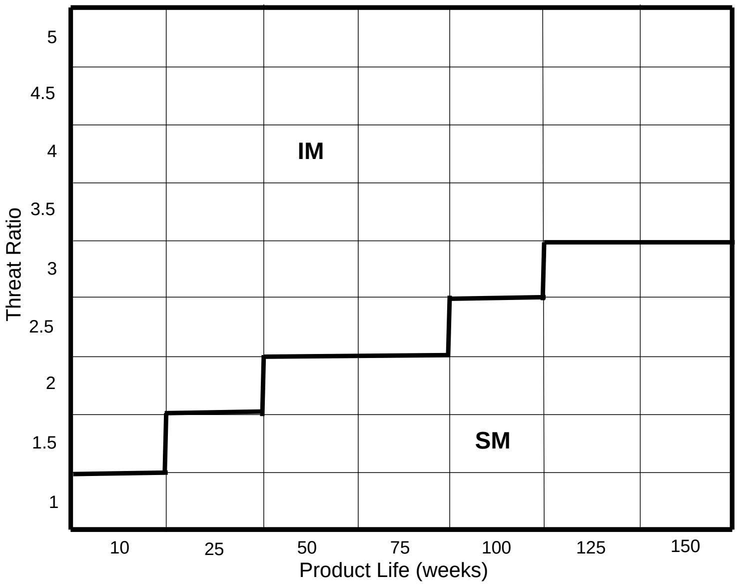
$$g_t(x) = \min_{x \le y} \Bigg\{ c_f(y-x) + hy^+ + p(-y)^+ + \beta \Bigg[\sum_{n=1}^{N} \mu_n \sum_{k=1}^{K} q_k[f_{t+1}(n, y-d_k+wd_k) + c_r wd_{nk}] + (1-\mu) \sum_{k=1}^{K} q_k[g_{t+1}(y-d_k+wd_k) + c_r wd_k] \Bigg] \Bigg\} \tag{10}$$

We note that (i) there is no capacity limit on the emergency orders in this formulation but such a limit can be readily accommodated, and (ii) because the DP algorithm considers only integer order quantities, we round wd to the nearest integer, and wd should be interpreted to mean "wd rounded to the nearest integer". Unlike in the single-supplier case, (10) now reflects the fact that emergency orders can be placed with supplier R if supplier U is down. Equations (9)–(10) mirror the singe-supplier equations (6)–(7) except for the inclusion of the $c_r wd$ terms in the cost-to-go functions, both in the IP in period $t+1$ (e.g., $f_{t+1}(m, y-d_{mk}+wd_{mk})$) and in the cost to order those units ($c_r wd_{mk}$). It is necessary to account for period $t+1$'s required supplier-R units in period t's cost functions because, when period $t+1$'s cost functions are computed, we know x, which already accounts for the demand in period $t+1$, but we don't know the actual demand in period $t+1$. The cost-to-go functions already include expectations over the demands in period $t+1$, so it is natural to include this cost there. However, this accounting method requires a slight modification to the recursions in periods 1 and T. First, since no state transition or demand occurs after period T, the d and $c_r wd$ terms must be omitted from the cost-to-go functions in (9)–(10) for period T's cost functions, $f_T(\cdot)$ and $g_T(\cdot)$. Second, to account for the required supplier-R units in period 1, we must add $c_r wd_1$ to the formulas in (9)–(10) for period 1's cost functions[3], $f_1(\cdot)$ and $g_1(\cdot)$.

For a given supplier-R allocation w, the algorithm for finding the optimal expected horizon cost and optimal base-stock levels is essentially the same as that described in §4.2.1 for the single-supplier

[3] In addition to the initial conditions introduced for the single-supplier DP in §4.2.1, we must also define the initial demand d_1. This is because the initial IP x_1 already accounts for the first-period demand, so it is impossible to determine this demand if we only know x_1. In the two-supplier model, we need to know d_1 to determine the cost of the items that are required to be procured from supplier R in period 1.

Figure 8: Optimal disruption-management strategy vs. product life and threat ratio.



model. The optimal w can then be found by searching over $0 \leq w \leq 1$.

As previous literature on disruption-management strategies has not considered either finite horizons or threat advisory systems, we focus our attention here on the impact of these two aspects on the firm's optimal disruption-management strategy. For a five-threat-level system (with a completely connected threat transition matrix), we present the optimal strategy as a function of the product life and the threat ratio for the case in which supplier R has zero flexibility (and so contingent rerouting is not possible); see Figure 8. We consider seven product lives (T = 10, 25, 50, 75, 100, 125 and 150 weeks) and nine threat ratios ($r = 1$ through 5, in increments of 0.5) for a total of 63 combinations. The procurement costs are $c_u = 1$, $c_r = 1.015$, and $c_f = \infty$ (since supplier R has zero flexibility and cannot accommodate emergency orders). The inventory costs and demand parameters[4] are the same as in the example in §4.2.2. Supplier U's capacity is 50 and its steady-state uptime percentage is $\pi_u = 0.97$. The probability of disruption in threat level 3 is 0.006525. Repair times are geometrically distributed with the same repair probability, $\mu = 0.211$, for all down-states.

We first note that single sourcing from either U or R is optimal in all 63 instances, even though Theorem 10 does not apply if supplier U has finite capacity. (Dual sourcing was found to be optimal if the capacity was reduced.) Either inventory mitigation (IM), i.e., source from U but carry inventory, or sourcing mitigation (SM), i.e., source from R, is optimal in all instances. Acceptance is

[4]We assume that demand is stationary over the life of the product. In reality, demand would be non-stationary and, in particular, would decrease as the end-of-life approaches.

not optimal in any instance. The product life and the threat ratio both a have a significant impact on the optimal strategy. Observe that inventory mitigation is optimal for higher threat ratios and/or shorter product lives. This result reflects our earlier findings in the single-supplier system regarding the impact of the threat ratio and the horizon length. For a given supplier percentage uptime, the cost of mitigating disruptions with inventory decreases as the threat ratio increases because the probability of being in threat levels that require high inventory levels decreases. As such, inventory mitigation becomes an increasingly attractive strategy as the threat ratio increases. As discussed earlier, if the terminal stockout cost is not too high, the amount of inventory required to mitigate disruptions decreases as the remaining horizon decreases. This makes inventory mitigation a more attractive strategy as the product life decreases. We note that one can create instances (using very high terminal stockout costs) in which sourcing mitigation replaces inventory mitigation as the lifecycle is decreased; however, for reasonable problem instances, inventory mitigation is preferred as the product lifecycles decrease.

The crucial point is that the optimal strategy is influenced by the product lifecycle, and, therefore, firms should consider lifecycles when determining their disruption-management strategy. In closing this section, we note that our model does not reflect newsvendor situations in which all inventory must be ordered in advance of the selling season. In such cases, inventory is not a useful mitigation tactic as the inventory arrives only if the supplier is not disrupted.

6. Conclusions

While the concept of a threat advisory system is well established in the national security domain, the essence of such a system (warning levels with different threat probabilities) has not, to the best of our knowledge, been considered in the literature on supply chain disruptions. This is in spite of the fact that firms, such as UTC, have, in effect, been using the threat advisory notion to adapt their inventory levels according to the threat of a disruption. In this paper, we formally investigate how the concept of a threat advisory system can influence a firm's disruption-management strategy. Our investigation uncovers a number of important findings.

A threat advisory system enables the firm to implement an adaptive inventory policy in which the firm can increase or decrease its inventory depending on the current threat level. This adaptive policy can lead to substantial savings in comparison to the constant inventory policy used in the absence of a threat advisory system. The cost savings increase with the threat ratio, i.e., with the relative disruption probability between adjacent threat levels. The value of the adaptive policy is diminished if the supplier capacity is tight. This capacity effect is amplified if the threat ratio is large or if direct transitions from low to high threat levels can occur.

The presence of a threat advisory system can significantly influence the firm's optimal disruption-management strategy. In the absence of such a system, it can be optimal simply to accept the risk of potential disruptions if the supplier reliability is high. With such a system, however, it is cost effective to mitigate these potential disruptions with inventory. At lower supplier reliabilities, the range of disruption frequencies at which inventory mitigation is the preferred strategy is higher in the presence of a threat advisory system. For a given supplier reliability, inventory mitigation is increasingly preferred as the threat ratio increases.

The existing disruption literature has focused on single-period or infinite-horizon models. Our finite-horizon model has enabled us to explore the effect of horizon length on the inventory held to protect against disruptions and, thus, the influence of product lifecycles on the firm's optimal disruption-management strategy. We find that (unless the terminal stockout cost is very high), the optimal inventory coverages decrease as the remaining horizon decreases and, as a result, inventory mitigation is more attractive as the product lifecycle decreases (assuming the lead time is small relative to the lifecycle.)

This work has laid the groundwork for exploring the use of threat advisory systems to manage the risk of supply disruptions. Future work should extend this research in a number of directions. For example, the firm's threat updates may lag behind changes in the actual risk levels and, so, the firm may have imperfect threat information. Also, as discussed in the introduction, firms are using threat advisory systems to simultaneously monitor multiple suppliers across different products, and, therefore, it would be valuable to explore the role of threat-advisory systems in more complex supply networks.

References

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

Arreola-Risa, Antonio, Gregory A. DeCroix. 1998. Inventory management under random supply disruptions and partial backorders. *Naval Research Logistics* **45** 687–703.

Bagchi, U., J. C. Hayya, C.-H. Chu. 1986. The effect of lead-time variability: The case of independent demand. *Journal of Operations Management* **6**(2) 159–177.

Berk, Emre, Antonio Arreola-Risa. 1994. Note on "Future supply uncertainty in EOQ models". *Naval Research Logistics* **41** 129–132.

Bradley, J.R., L.W. Robinson. 2005. Improved base-stock approximations for independent stochastic lead times with order crossover. *Manufacturing and Service Operations Management* **7**(4) 319–329.

Chao, Hung-Po. 1987. Inventory policy in the presence of market disruptions. *Operations Research* **35**(2) 274–281.

Chao, Hung-Po, Stephen W. Chapel, Jr. Clark, Charles E, Peter A. Morris, M. James Sandling, Richard C. Grimes. 1989. EPRI reduces fuel inventory costs in the electric utility industry. *Interfaces* **19**(1) 48–67.

Chopra, Sunil, Gilles Reinhardt, Usha Mohan. 2007. The importance of decoupling recurrent and disruption risks in a supply chain. Forthcoming in *Naval Research Logistics* .

Ciarallo, Frank W., Ramakrishna Akella, Thomas E. Morton. 1994. A periodic review, production planning model with uncertain capacity and uncertain demand: Optimality of extended myopic policies. *Management Science* **40**(3) 320–332.

Dada, Maqbool, Nicholas C. Petruzzi, Leroy B. Schwarz. 2007. A newsvendor's procurement problem with suppliers are unreliable. *Manufacturing and Service Operations Management* **9**(1) 9–32.

Federgruen, A., N. Yang. 2007. Selecting a portfolio of suppliers under demand and supply risks. Forthcoming in *Operations Research* .

Graves, Stephen C. 1988. Safety stocks in manufacturing systems. *Journal of Manufacturing and Operations Management* **1** 67–101.

Gupta, Diwakar. 1996. The (Q, r) inventory system with an unreliable supplier. *INFOR* **34**(2) 59–76.

Gürler, Ülkü, Mahmut Parlar. 1997. An inventory problem with two randomly available suppliers. *Operations Research* **45**(6) 904–918.

Hadley, G., T. M. Whitin. 1963. *Analysis of Inventory Systems*. Prentice-Hall, Englewood Cliffs, NJ.

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

Hendricks, Kevin B., Vinod R. Singhal. 2005a. Association between supply chain glitches and operating performance. *Management Science* **51**(5) 695–711.

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

Kaplan, R. 1970. A dynamic inventory model with stochastic lead times. *Management Science* **16** 491–507.

Kleindorfer, P. R., G. H. Saad. 2005. Managing disruption risks in supply chains. *Production and Operations Management* **14**(1) 53–68.

Kosdrosky, Terry. 2007. Ford sales mgr: Navistar spat could hurt march sales. *Dow Jones News Service* .

Li, Zhaolin, Susan H. Xu, Jack Hayya. 2004. A periodic-review inventory system with supply interruptions. *Probability in the Engineering and Informational Sciences* **18** 33–53.

McCracken, Jeffrey. 2003. Vehicle backlog hits new highs. *Detroit Free Press* May 14, 2003.

Meyer, Robert R., Michael H. Rothkopf, Stephen A. Smith. 1979. Reliability and inventory in a production-storage system. *Management Science* **25**(8) 799–807.

Moinzadeh, Kamran, Prabhu Aggarwal. 1997. Analysis of a production/inventory system subject to random disruptions. *Management Science* **43**(11) 1577–1588.

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

Parlar, Mahmut. 1997. Continuous-review inventory problem with random supply interruptions. *European Journal of Operational Research* **99** 366–385.

Parlar, Mahmut, David Perry. 1995. Analysis of a (Q, r, T) inventory policy with deterministic and random yeilds when future supply is uncertain. *European Journal of Operational Research* **84** 431–443.

Parlar, Mahmut, David Perry. 1996. Inventory models of future supply uncertainty with single and multiple suppliers. *Naval Research Logistics* **43** 191–210.

Parlar, Mahmut, Yunzeng Wang, Yigal Gerchak. 1995. A periodic review inventory model with Markovian supply availability. *International Journal of Production Economics* **42** 131–136.

Ramasesh, R. V., J. K. Ord, J. C. Hayya, A. Pan. 1991. Sole versus dual sourcing in stochastic lead time (s, Q) inventory models. *Management Science* **37** 428–443.

Schmitt, Amanda, Lawrence V. Snyder. 2007. Infinite-horizon models for inventory control under yield uncertainty and disruptions. Submitted for publication, Lehigh University, Bethlehem, PA.

Snyder, Lawrence V. 2006. A tight approximation for a continuous-review inventory model with supplier disruptions. Working paper, Lehigh University.

Snyder, Lawrence V., Z. Max Shen. 2006. Supply and demand uncertainty in multi-echelon supply chains. Submitted for publication, Lehigh University.

Song, Jing-Sheng, Paul H. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Operations Research* **41** 351–370.

Song, Jing-Sheng, Paul H. Zipkin. 1996. Inventory control with information about supply conditions. *Management Science* **42**(10) 1409–1419.

Swaminathan, J.M., J.G. Shanthikumar. 1999. Supplier diversification: Effect of discrete demand. *Operations Research Letters* **24** 213–221.

Swinney, R., S. Netessine. 2007. Long-term contracts under the threat of supplier default. Working paper, The Wharton School, University of Pennsylvania.

Tomlin, Brian T. 2005a. The impact of supply-learning on a firms sourcing strategy and inventory investment when suppliers are unreliable. Working paper OTIM-2005-05, Kenan-Flagler Business School, UNC.

Tomlin, Brian T. 2005b. Selecting a disruption-management strategy for short life-cycle products: Diversification, contingent sourcing, and demand management. Working paper OTIM-2005-09, Kenan-Flagler Business School, UNC.

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

Wall Street Journal. 2007. Earnings digest brief: Harley Davidson inc.: Shipments hit by strike lead to lower outlook. *Wall Street Journal* .

Yano, Candace A., Hau A. Lee. 1995. Lot sizing with random yield: A review. *Operations Research* **43**(2) 311–334.

Zipkin, Paul H. 2000. *Foundations of Inventory Management*. Irwin/McGraw-Hill.

Appendix

A. Proofs

Proof of Theorem 1

$C(j_1, \dots, j_N)$ is the long-run average inventory-shortage cost of a base-stock policy $(j_1, \dots, j_N)$. Under the fully reimbursed returns assumption, the firm's inventory position will always equal y_n at the end of a period in up-state n. Therefore, y_n only influences the IP in state n and in the down-states that follow from it. Because of this, the coverage in state n does not affect the cost in up-states other than n or the cost in down-states that originate from some up-state other than n. $C(j_1, \dots, j_N)$ can now be expressed as the sum of N single-variable functions, i.e.,

$$C(j_1, \dots, j_N) = \sum_{n=1}^{N} C_n(j_n), \tag{11}$$

where

$$C_n(j_n) = h\left(dj_n\pi_u(n) + \sum_{i=1}^{j_n} d(j_n - i)\,\pi_d(n,i)\right) + p\left(\sum_{i=j_n+1}^{\infty} d(i - j_n)\,\pi_d(n,i)\right). \tag{12}$$

The N-variable coverage-optimization problem therefore separates into N different single-variable optimization problems. For any nonnegative integer j, define $G_n[j] = \pi_u(n) + \sum_{i=1}^{j} \pi_d(n,i)$,

$\overline{G}_n[j] = \sum_{i=j+1}^{\infty} \pi_d(n,i)$, $E_n[j] = \sum_{i=1}^{j} i\pi_d(n,i)$ and $\overline{E}_n[j] = \sum_{i=j+1}^{\infty} i\pi_d(n,i)$. We can then write

$$\begin{aligned} C_n(j_n) &= hd\left(j_n G_n[j_n] - E_n[j_n]\right) + pd\left(\overline{E}_n[j_n] - j_n \overline{G}_n[j_n]\right) \\ &= hdK_n[j_n] + pd\overline{K}_n[j_n], \end{aligned} \tag{13}$$

where $K_n[j] = jG_n[j] - E_n[j]$ and $\overline{K}_n[j] = \overline{E}_n[j] - j\overline{G}_n[j]$. Let $\Delta_{C_n}(j_n) = C_n(j_n+1) - C_n(j_n)$ and $\Delta^2_{C_n}(j_n) = \Delta_{C_n}(j_n+1) - \Delta_{C_n}(j_n)$ be the first- and second-order differences, respectively. Then,

$$\begin{aligned} \Delta_{C_n}(j_n) &= hd\left(K_n[j_n+1] - K_n[j_n]\right) + pd\left(\overline{K}_n[j_n+1] - \overline{K}_n[j_n]\right) \\ &= hdG_n[j_n] - pd\overline{G}_n[j_n] \\ &= hdG_n[j_n] - pd\left(G_n[\infty] - G_n[j_n]\right) \\ &= d(h+p)\,G_n[j_n] - pdG_n[\infty] \end{aligned}$$

and

$$\Delta^2_{C_n}(j_n) = d(h+p)\,\pi(n, j_n+1) \geq 0.$$

$C_n(j_n)$ is therefore convex in j_n, and j_n^* is the minimum $j_n \geq 0$ such that $\Delta_{C_n}(j_n) \geq 0$. Defining $F_n[j] = G_n[j]/G_n[\infty]$, we then have

$$\begin{aligned} \frac{\Delta_{C_n}(j_n)}{d \cdot G_n[\infty]} &= (h+p)\frac{G_n[j_n]}{G_n[\infty]} - p \\ &= (h+p)\,F_n[j_n] - p, \end{aligned}$$

and so $j_n^* = F_n^{-1}\left[\frac{p}{p+h}\right]$.

Proof of Corollary 2

If $\mu_{ni} = \mu_i$ for $n = 1, \ldots, N$, then $F_n[j]$ is increasing in n. The proof then follows from Theorem 1, since $F_n[j]$ is also increasing in j.

Proof of Theorem 3

Recall that (i) demand is deterministic and equal to d units in each period, (ii) there is no discounting, i.e., $\beta = 1$, and (iii) the firm can return excess inventory for a full reimbursement of the purchase cost. Let j'_{nt} be the coverage in period t after the (deterministic) demand is observed and subtracted but before the order is placed. Let j_{nt} be the coverage after the order is placed. The order size is then $(j_{nt} - j'_{nt})d$. For a given (j'_{nt}, j_{nt}), define $C_{nt}(j'_{nt}, j_{nt})$ as the expected remaining-horizon optimal cost if in up-state n in period t.

$$\begin{aligned}C_{nt}\left(j'_{nt},j_{nt}\right) =& c(j_{nt}-j'_{nt})d+hd(j_{nt})^{+}+pd(-j_{nt})^{+}+\sum_{m=1}^{N}\lambda_{nm}C_{m,t+1}\left(j_{nt}-1,j^{*}_{m,t+1}\right)\\&+\sum_{i=1}^{T-t-1}v(n,i)\left(hd(j_{nt}-i)^{+}+pd(i-j_{nt})^{+}+\sum_{m=1}^{N}\mu_{nim}C_{m,t+i+1}\left(j_{nt}-i-1,j^{*}_{m,t+i+1}\right)\right)\\&+v(n,T-t)\left(\left(h+h_T\right)d(j_{nt}-(T-t))^{+}+\left(p+p_T\right)d(T-t-j_{nt})^{+}\right),\end{aligned}$$

where $a^{+}\equiv\max\{a,0\}$ and $v(n,i)=\lambda_{nd}\Pi_{k=1}^{i-1}\left(1-\mu_{nk}\right)$ is the probability that a disruption lasting at least i periods occurs, starting in the next period, given that we are in up-state n in this period. If $t\geq T-1$, we take the summation over $i=1,\ldots,T-t-1$ to equal 0.

We note that $\sum_{m=1}^{N}\lambda_{nm}+\sum_{i=1}^{T-t-1}v(n,i)\sum_{m=1}^{N}\mu_{nim}+v(n,T-t)=1$, and that $C_{m,t+1}\left(j_{nt}-1,j^{*}_{m,t+1}\right)+c(j_{nt}-j'_{nt})d=C_{m,t+1}\left(j'_{nt}-1,j^{*}_{m,t+1}\right)$. Therefore,

$$\begin{aligned}C_{nt}\left(j'_{nt},j_{nt}\right) =& hd(j_{nt})^{+}+pd(-j_{nt})^{+}+\sum_{m=1}^{N}\lambda_{nm}C_{m,t+1}\left(j'_{nt}-1,j^{*}_{m,t+1}\right)\\&+\sum_{i=1}^{T-t-1}v(n,i)\left(hd(j_{nt}-i)^{+}+pd(i-j_{nt})^{+}+\sum_{m=1}^{N}\mu_{nim}C_{m,t+i+1}\left(j'_{nt}-i-1,j^{*}_{m,t+i+1}\right)\right)\\&+v(n,T-t)\left(c(j_{nt}-j'_{nt})d+\left(h+h_T\right)d(j_{nt}-(T-t))^{+}+\left(p+p_T\right)d(T-t-j_{nt})^{+}\right).\end{aligned}$$

Noting that

$$\begin{aligned}\left(h+h_T\right)\left(j_{nt}-(T-t)\right)^{+}+p_T(T-t-j_{nt})^{+} =& \left(\left(h+h_T\right)+c\right)\left(j_{nt}-(T-t)\right)^{+}\\&+\left(\left(p+p_T\right)-c\right)\left(T-t-j_{nt}\right)^{+}-c(j_{nt}-(T-t)),\end{aligned}$$

we can then write

$$\begin{aligned}C_{nt}\left(j'_{nt},j_{nt}\right) =& \sum_{m=1}^{N}\lambda_{nm}C_{m,t+1}\left(j'_{nt}-1,j^{*}_{m,t+1}\right)+\sum_{i=1}^{T-t-1}v(n,i)\sum_{m=1}^{N}\mu_{nim}C_{m,t+i+1}\left(j'_{nt}-i-1,j^{*}_{m,t+i+1}\right)\\&+hd(j_{nt})^{+}+pd(-j_{nt})^{+}+\sum_{i=1}^{T-t}v(n,i)\left(hd(j_{nt}-i)^{+}+pd(i-j_{nt})^{+}\right)\\&+v(n,T-t)\left(c(T-t-j'_{nt})d+\left(h_T+c\right)d(j_{nt}-(T-t))^{+}+\left(p_T-c\right)d(T-t-j_{nt})^{+}\right).\end{aligned}$$

Therefore, j^{*}_{nt} minimizes $\hat{C}_{nt}\left(j_{nt}\right)$ where

$$\begin{aligned}\hat{C}_{nt}\left(j_{nt}\right) =& hd(j_{nt})^{+}+pd(-j_{nt})^{+}+\sum_{i=1}^{T-t}v(n,i)\left(hd(j_{nt}-i)^{+}+pd(i-j_{nt})^{+}\right)\\&+v(n,T-t)\left(\left(h_T+c\right)d(j_{nt}-(T-t))^{+}+\left(p_T-c\right)d(T-t-j_{nt})^{+}\right),\end{aligned}$$

since the terms that are in $C_{nt}(j'_{nt}, j_{nt})$ but not $\hat{C}_{nt}(j_{nt})$ are constants with respect to j_{nt}. Equivalently, j^*_{nt} minimizes $\frac{\pi_u(n)}{d}\hat{C}_{nt}(j_{nt})$. Noting that $\pi_d(n,i) = \pi_u(n)v(n,i)$, we have

$$\begin{aligned}\frac{\pi_u(n)}{d}\hat{C}_{nt}(j_{nt}) =& \pi_u(n)h(j_{nt})^+ + \pi_u(n)p(-j_{nt})^+ + \sum_{i=1}^{T-t}\pi_d(n,i)\left(h(j_{nt}-i)^+ + p(i-j_{nt})^+\right) \\ &+ \pi_d(n,T-t)\left((h_T+c)(j_{nt}-(T-t))^+ + (p_T-c)(T-t-j_{nt})^+\right).\end{aligned}$$

It is straightforward to show that $\frac{\pi_u(n)}{d}\hat{C}_{nt}(j_{nt})$ is decreasing in j_{nt} if $j_{nt} < 0$ and increasing in j_{nt} if $j_{nt} > T-t$. Therefore, we can restrict attention to $0 \le j_{nt} \le T-t$, in which case

$$\begin{aligned}\frac{\pi_u(n)}{d}\hat{C}_{nt}(j_{nt}) =& \pi_u(n)hj_{nt} + \sum_{i=1}^{T-t}\pi_d(n,i)\left(h(j_{nt}-i)^+ + p(i-j_{nt})^+\right) \\ &+ \pi_d(n,T-t)(p_T-c)(T-t-j_{nt})^+.\end{aligned}$$

Letting the first- and second-order differences be denoted by $\Delta_{\hat{C}_{nt}}(j_{nt}) = \frac{\pi_u(n)}{d}\hat{C}_{nt}(j_{nt}+1) - \frac{\pi_u(n)}{d}\hat{C}_{nt}(j_{nt})$ and $\Delta^2_{\hat{C}_{nt}}(j_{nt}) = \Delta_{\hat{C}_{nt}}(j_{nt}+1) - \Delta_{\hat{C}_{nt}}(j_{nt})$, one can show that

$$\Delta_{\hat{C}_{nt}}(j_{nt}) = hG_n[j_{nt}] - p(G_n[T-t] - G_n[j_{nt}]) - (p_T-c)\pi_d(n,T-t),$$

and

$$\Delta^2_{\hat{C}_{nt}}(j_{nt}) = (p+h)\pi_d(n, j_{nt}+1) \ge 0,$$

where recall that $G_n[j] = \pi_u(n) + \sum_{i=1}^{j}\pi_d(n,i)$. Therefore, $\frac{\pi_u(n)}{d}\hat{C}_{nt}(j_{nt})$ is convex in j_{nt} and $j^*_{nt} = \min\{\hat{j}_{nt}, T-t\}$ where $\hat{j}_{nt}$ is the minimum j such that:

$$hG_n[j] - p(G_n[T-t] - G_n[j]) - (p_T-c)\pi_d(n,T-t) \ge 0. \tag{14}$$

Proof of Corollary 4

From Theorem 3, $j^*_{nt} = \min\{\hat{j}_{nt}, T-t\}$, where $\hat{j}_{nt}$ is the minimum j such that $hG_n[j] - p(G_n[T-t] - G_n[j]) - (p_T-c)\pi_d(n,T-t) \ge 0$. Now, $\pi_d(n,T-t) \to 0$ as $T-t \to \infty$. Therefore, as $T-t \to \infty$, j^*_{nt} approaches the minimum j such that $hG_n[j] - p(G_n[\infty] - G_n[j]) \ge 0$. Recalling that $F_n[j] = G_n[j]/G_n[\infty]$, then, as $T-t \to \infty$, j^*_{nt} approaches the minimum j such that $hF_n[j] - p(1 - F_n[j]) \ge 0$. Thus, as $T-t \to \infty$, j^*_{nt} approaches $F_n^{-1}\left[\frac{p}{p+h}\right]$. From Theorem 1, $j^*_n = F_n^{-1}\left[\frac{p}{p+h}\right]$. Therefore, $j^*_{nt} \to j^*_n$ as $T-t \to \infty$.

Proof of Theorem 5

From Theorem 3, $\hat{j}_{nt}$ is the minimum j such that $\Delta_{\hat{C}_{nt}}(j_{nt}) \ge 0$, where

$$\Delta_{\hat{C}_{nt}}(j_{nt}) = hG_n[j] - p(G_n[T-t] - G_n[j]) - (p_T-c)\pi_d(n,T-t). \tag{15}$$

Therefore,

$$\Delta_{\hat{C}_{nt}}(j_{nt}) - \Delta_{\hat{C}_{n,t+1}}(j_{nt}) = -p\pi_d\left(n, T-t\right) + \left(p_T - c\right)\left(\pi_d\left(n, T-t-1\right) - \pi_d\left(n, T-t\right)\right), \quad (16)$$

so

$$\begin{aligned}\frac{\Delta_{\hat{C}_{nt}}(j_{nt}) - \Delta_{\hat{C}_{n,t+1}}(j_{nt})}{\pi_d\left(n, T-t\right)} &= -p + \left(p_T - c\right)\left(\frac{\pi_d\left(n, T-t-1\right)}{\pi_d\left(n, T-t\right)} - 1\right) \\ &= -p + \left(p_T - c\right)\left(\frac{1}{1-\mu_{n,T-t-1}} - 1\right) \\ &= -p + \left(p_T - c\right)\left(\frac{\mu_{n,T-t-1}}{1-\mu_{n,T-t-1}}\right).\end{aligned}$$

Therefore, if $p_T \leq c + p\left(\frac{1}{\mu_{n,T-t-1}} - 1\right)$, then $\Delta_{\hat{C}_{nt}}(j_{nt}) - \Delta_{\hat{C}_{n,t+1}}(j_{nt}) \geq 0$ and so $\hat{j}_{n,t+1} \leq \hat{j}_{nt}$. Also, if $p_T > c + p\left(\frac{1}{\mu_{n,T-t-1}} - 1\right)$, then $\Delta_{\hat{C}_{nt}}(j_{nt}) - \Delta_{\hat{C}_{n,t+1}}(j_{nt}) \leq 0$ and so $\hat{j}_{n,t+1} \geq \hat{j}_{nt}$.

Proof of Corollary 6

We first note that if $\mu_{ni} = \mu_n$ for $i = 1, \ldots, \infty$, then using Theorem 5, $\hat{j}_{n,t+1} \leq \hat{j}_{nt}$ if $p_T \leq c + p\left(\frac{1}{\mu_n} - 1\right)$ and $\hat{j}_{n,t+1} \geq \hat{j}_{nt}$ if $p_T > c + p\left(\frac{1}{\mu_n} - 1\right)$. (i) If $p_T \leq c + p\left(\frac{1}{\mu_n} - 1\right)$, then $\hat{j}_{nt}$ is non-increasing in t. Furthermore, $j^*_{nt} = \min\{\hat{j}_{nt}, T-t\}$ and $T-t$ is decreasing in t. Therefore, j^*_{nt} is non-increasing in t. (ii) If $p_T > c + p\left(\frac{1}{\mu_n} - 1\right)$, then $\hat{j}_{nt}$ is non-decreasing in t. Therefore, if $\hat{j}_{nt} > T-t$, then $\hat{j}_{nt'} > T-t'$ for all $t' > t$. Thus, j^*_{nt} is non-decreasing in t until some $\hat{j}_{nt} > T-t$, after which, j^*_{nt} is non-increasing in t.

Proof of Theorem 7

In the constant coverage case, $j_1 = j_2 \ldots = j_N = j_c$, and so we write the long-run average cost as $C\left(j_c, \ldots, j_c\right) = C\left(j_c\right)$ for ease of exposition. $C\left(j_c\right) = \sum_{n=1}^{N} C_n\left(j_c\right)$. Let $\Delta\left(j_c\right) = C\left(j_c+1\right) - C\left(j_c\right)$ and $\Delta^2\left(j_c\right) = \Delta\left(j_c+1\right) - \Delta\left(j_c\right)$ be the first- and second-order differences, respectively. Because $C\left(j_c\right) = \sum_{n=1}^{N} C_n\left(j_c\right)$, we have $\Delta\left(j_c\right) = \sum_{n=1}^{N} \Delta_{C_n}\left(j_c\right)$ and $\Delta^2\left(j_c\right) = \sum_{n=1}^{N} \Delta^2_{C_n}\left(j_c\right)$ where $\Delta_{C_n}\left(j_c\right)$ and $\Delta^2_{C_n}\left(j_c\right)$ are as defined in the proof of Theorem 1. Therefore,

$$\begin{aligned}\Delta\left(j_c\right) &= d\left(h+p\right)\sum_{n=1}^{N} G_n\left[j_c\right] - pd\sum_{n=1}^{N} G_n\left[\infty\right], \\ \Delta^2\left(j_c\right) &= d\left(h+p\right)\sum_{n=1}^{N} \pi\left(n, j_c+1\right) \geq 0,\end{aligned}$$

and therefore $C\left(j_c\right)$ is convex in j_c. Therefore j^*_c is the minimum $j_c \geq 0$ such that $\Delta\left(j_c\right) \geq 0$. By definition, $G_n\left[j\right] = \pi_u\left(n\right) + \sum_{i=1}^{j} \pi_d\left(n, i\right)$. Therefore $\sum_{n=1}^{N} G_n\left[\infty\right] = 1$ and the first-order difference can be written as

$$\begin{aligned}\Delta(j_c) &= d\left((h+p)\sum_{n=1}^{N}\left(\pi_u(n)+\sum_{i=1}^{j_c}\pi_d(n,i)\right)-p\right)\\ &= d\left((h+p)F_c[j_c]-p\right),\end{aligned}$$

and so $j_c^* = F_c^{-1}\left[\frac{p}{p+h}\right]$.

Proof of Theorem 8

Proof follows from application of Theorems 1 and 7 tailored to the special case of the supply system described in the theorem statement. Algebraic details available upon request.

Proof of Corollary 9

Proof follows from comparison of j_c^* and j_n^* from Theorem 8.

Proof of Theorem 10

We prove this theorem for the long-run average cost criterion and note that it can be proved in an analogous manner for the discounted-cost criterion. We also note that Tomlin (2006) established the result for the single-threat system but that the proof did not rely on the single-threat level assumption. We present the relevant parts of the proof here for completeness.

Consider the following inventory system. Demand in each period is stochastic but stationary. The demand random variable, denoted by D, has a strictly positive support. Supply is completely reliable with a guaranteed lead time of $L \geq 0$. Ordering costs are linear but state-dependent; the marginal ordering cost is c_u in all "up" states and is c_f in all "down" states. The state space and state transitions are identical to those described for the unreliable supplier in §3. The following results hold for this inventory system. (We prove them for the long-run average cost criterion but they also hold for the discounted-cost criterion.) **(a)** A state-dependent, base-stock policy is optimal. *Proof.* This supply system is a special case of Song and Zipkin (1996) [§9] who prove that a state-dependent base stock policy is optimal. **(b)** If demand is $D_t' = kD_t$ where $k \geq 0$, then the optimal base-stock levels and the optimal cost are $ky^*(i)$ and kV^* respectively, where V^* is the optimal cost when $k = 1$. *Proof.* This proof follows a similar logic to §7 of Song and Zipkin (1996) but we tailor $G^+(i,y)$ for the case where lead-times are constant but order costs are state dependent. The optimality condition can be expressed as $g + W(i,x) = \min\{H(i,y) : y \geq x\}$, where $H(i,y) = G^+(i,y) + E[W(i_+, y-D)]$, $G^+(i,y) = (c(i) - E[c(i_+)])\,y + C^L(y) + E[c(i_+)]\,E[D]$, $C^L(y) = E\left[\widehat{C}\left(y - D^{(L)}\right)\right]$, and i_+ is shorthand for the state reached after state i. The optimal base stock $y^*(i)$ minimizes $H(i,y)$ over y, that is $y^*(i) = \min\{y : \Delta_y H(i,y) \geq 0\}$, and $\Delta_x W(i,x) = 0$ for

$x < y^*(i)$ and $\Delta_x W(i,x) = \Delta_x H(i,x) \geq 0$ for $x \geq y^*(i)$, where Δ is the difference operator. Now, $\widehat{C}\left(ky - kD^{(L)}\right) = k\widehat{C}\left(y - D^{(L)}\right)$ and so $C^L_{D'}(ky) = kC^L_D(y)$. Therefore $G^+_{D'}(i, ky) = kG^+_D(i, y)$, and proof follows.

For a given supplier-R fraction w, the inventory-production system is then a special case of that described above with a demand of $(1-w)\,D_t$. We note that there is an additional expected cost of $c_r wE[D]$ in every period because of the Supplier R allocation. Proof that a base-stock policy is optimal follows directly from (a) above. Using (b) above and because there is an additional expected cost of $c_r wE[D]$ in every period, the optimal long-run average cost is $c_r wE[D] + (1-w)\,V^*$, which is linear in w, and so $w^* \in \{0, 1\}$.

B. Returns

As noted in the main body, under a state-dependent inventory policy, the firm may find itself with more stock than it wishes to have in a given period. If the firm cannot, or does not want to, return inventory to the supplier, then its inventory position (IP) will exceed the desired IP until the excess inventory has been drained down by demand (assuming another state transition has not occurred in the meantime). In contrast, the firm can and will attain its desired IP in every period if the supplier allows it to return excess inventory at no cost. The returns policy therefore plays a central role in the inventory dynamics.

For the infinite-capacity analysis in §4.1.1 we restricted attention to the case of fully reimbursed returns to facilitate the characterization of the optimal base-stock levels. In the first part of this appendix, we consider a special case of the general threat advisory model for which the optimal base-stock levels can be characterized when returns are not allowed. We then explore how the returns policy influences the firm's optimal cost. In the finite-capacity (finite-horizon) section (§4.2), we restricted attention to the case in which returns were not allowed. In the second part of this appendix, we generalize the dynamic programming formulation to allow for returns, under both full and partial reimbursement.

B.1 Influence of Returns Policy on Optimal Cost in Infinite-Horizon Case

In §4.1.1 we characterized the optimal base-stock levels in the infinite-capacity, deterministic-demand case under the assumption that returns were fully reimbursed. Here we consider the case in which returns are not allowed. We focus on the infinite-horizon case. The no-returns case results in excess inventory events whereby the IP can exceed the target base-stock level in certain periods. As a consequence, $C(j_1, \ldots, j_N)$, the long-run average cost of a policy $(j_1, \ldots, j_N)$, cannot be expressed as the sum of N single-variable cost functions as was possible in the free-returns case, so solving for the optimal coverage vector $(j_1^*, \ldots, j_N^*)$ is significantly more complex in the no-returns case. In contrast to the free-returns case, closed-form expressions for the optimal coverages do not exist in general.

So as to investigate the effect of the returns policy on the long-run average cost, we consider a particular two-threat-level system for which one can obtain closed-form expressions for the optimal coverages j_1^* and j_2^*. We assume that $\lambda_{1d} = 0$, that is, disruptions can only occur in threat level 2. Furthermore, we assume that disruptions always end by returning to threat level 2. Clearly the optimal coverage in up-state 1 is 0 under either returns policy as a disruption cannot occur in up-state 1. Let $C_{FR}(j_2)$ denote the long-run average cost in the free-returns case and $C_{NR}(j_2)$ denote the long-run average cost in the no-returns case. One can show that

$$C_{FR}(j_2) = hK_2[j_2] + p\overline{K}_2[j_2], \tag{17}$$

$$C_{NR}(j_2) = hK_2[j_2] + p\overline{K}_2[j_2] + h\left(\sum_{i=1}^{j_2}(j_2 - i)\,\pi_u(1,i)\right), \tag{18}$$

where $K_2[j]$ and $\overline{K}_2[j]$ are as defined in §4.1.2, and $\pi_u(1,i)$ is the steady-state probability of having been in up-state 1 for i periods. The last term in (18) reflects the additional cost incurred due to the excess inventory that occurs in the no-returns case after a transition from up-state 2 to up-state 1.

Let $j_{2,FR}^*$ and $j_{2,NR}^*$ denote the optimal threat-2 coverages under the free-returns and no-returns policies respectively. If the disruption lengths are geometrically distributed with a repair probability of μ, then the optimal coverages are as follows. If $\frac{p}{p+h} \leq \frac{\mu}{\mu+\lambda_{2d}}$, then $j_{2,FR}^* = j_{2,NR}^* = 0$. If $\frac{p}{p+h} > \frac{\mu}{\mu+\lambda_{2d}}$, then $j_{2,FR}^*$ is the minimum j such that

$$\frac{1+\lambda_{2d}\left(\frac{1-(1-\mu_2)^j}{\mu_2}\right)}{1+\frac{\lambda_{2d}}{\mu_2}} \geq \frac{p}{p+h}, \tag{19}$$

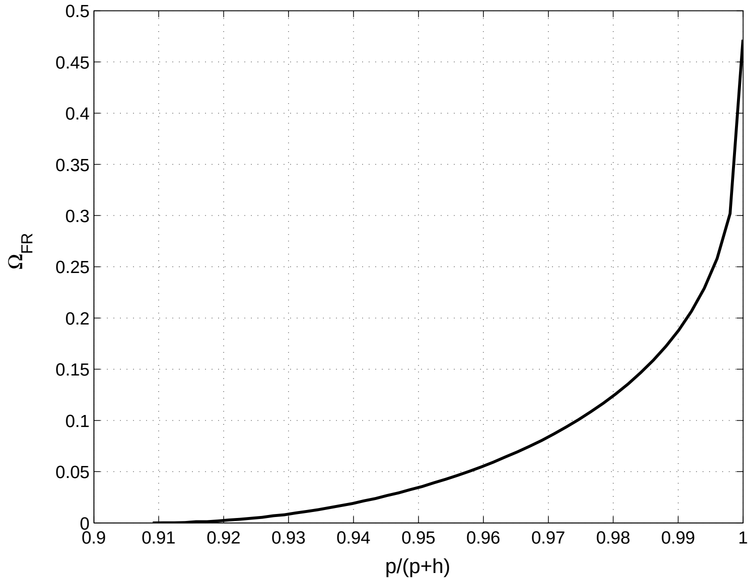
and $j_{2,NR}^*$ is the minimum j such that

$$\frac{1+\lambda_{2d}\left(\frac{1-(1-\mu_2)^j}{\mu_2}\right)}{1+\frac{\lambda_{2d}}{\mu_2}} + \left(\frac{h}{p+h}\right)\left(\frac{\lambda_{21}}{\lambda_{12}}\right)\left(\frac{1}{1+\frac{\lambda_{2d}}{\mu_2}}\right)\left(1-(1-\lambda_{12})^j\right) \geq \frac{p}{p+h}. \tag{20}$$

It then follows that $j_{2,FR}^* \geq j_{2,NR}^*$. The optimal threat-2 coverage is lower in the no-returns case because the no-returns case results in excess inventory and this excess inventory is increasing in the threat-2 coverage j_2.

Let Ω_{FR} denote the relative reduction in the firm's inventory-shortage costs if the free-returns

Figure 9: Relative cost reduction due to returns (Ω_{FR}) vs. newsboy ratio.



case prevails as compared to the no-returns case, i.e.,

$$\Omega_{FR} = \frac{C_{NR}\left(j^*_{2,NR}\right) - C_{FR}\left(j^*_{2,FR}\right)}{C_{NR}\left(j^*_{2,NR}\right)} \tag{21}$$

One can show that Ω_{FR} depends on p and h only through the newsvendor ratio $\frac{p}{p+h}$. In Figure 9, we plot the relative cost reduction as a function of $\frac{p}{p+h}$ for $\lambda_{2d} = 0.005$, $\mu = 0.05$, $\lambda_{12} = 0.011$, and $\lambda_{21} = 0.025$. The steady-state probability of the supplier being up is $\pi_u = 0.97$. For low values of $\frac{p}{p+h}$, the optimal coverage is zero under both returns policies and so $\Omega_{FR} = 0$. As $\frac{p}{p+h}$ increases the relative cost reduction increases, and so the returns policy is much more significant at higher $\frac{p}{p+h}$ values. While the relative cost reduction can be very substantial, we remind the reader that this is a reduction in the inventory-shortage related costs, and not the overall costs, which include procurement costs.

B.2 Generalization of Dynamic Program to Allow for Returns

In the finite-capacity (finite-horizon) section (§4.2), we restricted attention to the case in which returns were not allowed. We now generalize our dynamic program (DP) to allow for the case in which the firm can return inventory and earn a per-unit revenue b. Note that $b = 0$ corresponds to the free-returns policy and $b = -\infty$ corresponds to the no-returns policy. If $-\infty \leq b < 0$, then the firm incurs a cost for each return, whereas if $b > 0$, the firm obtains positive revenue for each return; we assume $\beta b \leq c$ to avoid the trivial case in which the firm earns a profit for ordering an item in one period and returning it the next. If returns are allowed, they are only allowed during up-periods.

This cost structure gives rise to a control-band policy (Zipkin 2000, p. 429) that has two parameters, y and z, with $y \leq z$. We refer to y as the "order-up-to level" and to z as the "return-down-to level." If the current IP is less than y, we order enough items to raise the IP to y; if the current IP is greater than z, we return enough items to lower the IP to z; and if the current IP is between y and z, we do nothing. One can show that if $b = -\infty$ (no returns), then $z^* = \infty$; if $b = c$ (free returns), then $y^* = z^*$; and if $-\infty \lneqq b \lneqq c$, then $y^* \lneqq z^* \lneqq \infty$.

The DP recursions are now:

$$f_t(n,x) = \min_{0 \leq y \leq x+v_n} \Bigg\{ c(y-x)^+ - b(x-y)^+ + hy^+ + p(-y)^+ + \beta \left[\sum_{m=1}^{N} \lambda_{nm} \sum_{k=1}^{K} q_k f_{t+1}(m, y-d_k) + \lambda_{nd} \sum_{k=1}^{K} q_k g_{t+1}(y-d_k) \right] \Bigg\}, \tag{22}$$

$$g_t(x) = hx^+ + p(-x)^+ + \beta \left[\sum_{n=1}^{N} \mu_n \sum_{k=1}^{K} q_k f_{t+1}(n, x-d_k) + (1-\mu) \sum_{k=1}^{K} q_k g_{t+1}(x-d_k) \right], \tag{23}$$

where $a^+ \equiv \max\{a, 0\}$. The up-state formula (22) differs from the no-returns case (see (6)) only in the lower bound of the minimization and in the procurement/returns terms $c(y-x)^+ - b(x-y)^+$. The down-state formula, i.e., (23), is unchanged from the no-returns case (see (7)) since returns are not allowed during down-states.

The DP algorithm is the same as described in §4.2.1 but now we have to determine $z_t(n)$ in addition to $y_t(n)$. As before, to find $y_t^*(n)$, we can simply set $y_t^*(n)$ equal to the optimal base-stock level for the smallest possible IP x assuming infinite capacity in period t. This base-stock level represents the desired target IP, though this IP may not be attainable from every starting IP x because of the capacity constraints. Similarly, in the costly-returns case we set $z_t^*(n)$ equal to the optimal base-stock level for the largest possible IP $\bar{x}(n)$. (In the fully-reimbursed-returns case, $z_t^*(n) = y_t^*(n)$, whereas in the no-returns case, $z_t^*(n) = \infty$.) Allowing for returns increases the execution time somewhat but as before, the execution time for reasonably sized problems, e.g. 150 periods and 5 threat levels, is on the order of minutes using a desktop computer.

Finally, we note that we also have generalized the two-supplier DP algorithm to allow for returns. Details of this formulation are available upon request.