Manufacturing & Service Operations Management

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

Coordinating Supply Chains by Controlling Upstream Variability Propagation

Anantaram Balakrishnan, Joseph Geunes, Michael S. Pangburn,

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 6, No. 2, Spring 2004, pp. 163–183
ISSN 1523-4614 | EISSN 1526-5498 | 04 | 0602 | 0163
DOI 10.1287/msom.1030.0031

Coordinating Supply Chains by Controlling Upstream Variability Propagation

Anantaram Balakrishnan
McCombs School of Business, University of Texas at Austin, Austin, Texas 78712, anantb@mail.utexas.edu
Joseph Geunes
Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida 32611, geunes@ise.ufl.edu
Michael S. Pangburn
Lundquist College of Business, University of Oregon, Eugene, Oregon 97403, pangburn@uoregon.edu

Effective distribution using collaborative fulfillment networks requires coordination among the multiple participating firms at different stages of the supply chain. Acting independently, supply chain partners fail to weigh the cost burden they impose on upstream suppliers when their replenishment order quantities vary from period to period. This paper explores a new approach to coordinate multiple stages in the supply chain by controlling, through appropriate downstream inventory management, the demand variability that is propagated to upstream stages. We propose and analyze a coordinated inventory replenishment policy that uses "order smoothing" to reduce order-size variability and thus reduce overall system costs, including both inventory and transportation costs. We characterize the optimal parameter values for smoothing alternatives (such as exponential smoothing and moving weighted average policies), assess their economic benefits, and develop insights regarding supply chain contexts that might benefit most significantly from reducing the variability of orders to upstream stages. Using the distribution network for specialty brand appliances as an illustrative example, we demonstrate the potential cost savings that order-smoothing strategies can yield compared to the uncoordinated case when individual firms separately minimize their costs. The magnitude of savings depends on several factors, including the variability in consumer demand, level of product variety, and degree of inventory aggregation in the distribution system. Based on our analytical results, we develop a framework to assess cost reduction opportunities through variability control for different supply chain scenarios.

Key words: supply chain coordination; inventory control; variability reduction; distribution systems; collaborative fulfillment; order smoothing
History: Received: July 24, 2002; accepted: November 3, 2003. This paper was with the authors 9 months for 4 revisions.

1. Introduction

Manufacturers, distributors, and retail firms rely increasingly on collaborative supply chain partnerships to cost effectively fulfill market needs. Different players in the supply chain must coordinate their operating policies and actions to ensure a lean and responsive supply chain. Acting independently, downstream supply chain members propagate demand variability upstream, thereby imposing high capacity and inventory costs. We propose a new approach for coordinating the supply chain by controlling the variability transmitted to upstream stages, study a collaborative replenishment policy to dampen this variability, and demonstrate its benefits. The motivation for this approach stems from the distribution challenges facing a leading building-products manufacturer. The manufacturer uses a nationwide network of distribution centers to supply national hardware retail chains. These "big-box" retailers are quite demanding, requiring frequent deliveries with short lead times and high service levels (e.g., retailers may place orders of any size, small or large, until the day before their scheduled delivery). Analysis of supply chain transactional data indicated that retailers' replenishment order quantities are highly variable, forcing distribution centers to incur substantial costs for maintaining high inventories and transportation resources. We refer to the transmission of demand variability to upstream stages as upstream variability propagation. The manufacturer wished to assess the distribution cost impact of the current high level of order variability, and develop options for reducing this variability and hence supply chain costs. Motivated by this context, we consider the multistage interactions in a generic distribution system, and focus on reducing total system costs by controlling upstream variability propagation.

The supply chain coordination challenge facing the building-products manufacturer can arise in other distribution contexts where the downstream stage faces exogenous (consumer) demand, and periodically orders replenishments from the upstream stage. We will refer to the downstream stage as the *retailer* and the upstream stage as the *supplier*. Retailers are primarily concerned with ensuring continuous in-stock availability of products while simultaneously controlling inventory costs. Suppliers must maintain sufficient inventory and delivery capacity to meet retailer orders. Observe that suppliers must manage two types of assets—those with short-term flexibility (e.g., inventory) and those that require long-term investments (e.g., transportation resources). Inventory levels can be varied dynamically in response to changing requirements and advance demand information. In contrast, for assets such as delivery resources (e.g., trucks, drivers), internal capacity levels are fixed in the short run, and so coping with order variability requires either maintaining excess safety capacity or using expensive outside capacity sources for large orders. Analogous situations involving both short-term and long-term fulfillment capacity decisions also occur in other supply chain settings. For example, make-to-order manufacturers (such as contract assembly facilities for electronic circuit boards), in their role as suppliers, must manage their component inventories and also decide long-run manufacturing and assembly capacities to respond dynamically to downstream orders. We refer to the resources needed to fulfill (i.e., make-to-order and/or deliver) retailer orders as *service* capacity. Order variability affects both inventory and service capacity costs, and therefore can have large upstream cost repercussions. To reduce these costs, suppliers prefer minimal variability in the replenishment orders from retailers. On the other hand, retailers, driven by the goal of reducing holding costs, prefer to use replenishment policies that propagate rather than dampen consumer demand variability. These conflicting concerns lead to a tension between the preferred order variability of retailers and suppliers. A coordinated system might entail increasing retailers' safety stocks to reduce the variance of their replenishment orders and decrease the supplier's inventory and capacity costs.

Recent literature documents the negative impacts of variability *amplification*, also known as the "bullwhip effect," on supply chain performance (e.g., Sterman 1989; Lee et al. 1997; Chen et al. 2000; Anderson et al. 2000; Dejonckheere et al. 2002, 2003). These papers motivate the importance of reducing variability amplification in supply chains by addressing the operational practices (e.g., forecasting, promotions, batching, forward buying, and capacity rationing) that cause the bullwhip effect. In contrast, we emphasize opportunities to reduce supply chain costs by *dampening* variability, even in systems where variability is not amplified. We introduce a variability-focused view of supply chain coordination and seek the optimal variability-dampening target to reduce overall logistics costs. This approach raises three important questions:

(1) How can a stage in a supply chain control the variability that it propagates to an upstream supplier?

(2) What is the "optimal" level of variability at each stage of the supply chain?

(3) Under what circumstances (e.g., in terms of product or supply chain characteristics) is variability control an effective way to coordinate supply chains?

We focus on using downstream inventory replenishment policies as a means to control the variability transmitted upstream. To our knowledge, researchers have not previously studied how *order smoothing* by retailers influences variability propagation and hence total system costs. We consider a generic two-stage supply chain model incorporating supplier warehousing and expediting costs, service capacity and outsourcing costs, and retailer holding and backorder costs. We show that order smoothing not only reduces the variation of upstream orders, but also serves to provide advance order information to the supplier.

After developing some properties of optimal smoothing parameters for minimizing total system cost, we consider two classes of policies: exponential smoothing and moving weighted average. We characterize (in terms of cost and variability parameters) supply chain contexts under which each of these policies outperforms the other. Our analysis establishes the surprising result that some degree of order smoothing is desirable even when the supplier's variability-induced costs are low compared to the retailer's costs. By extending the scope of our model to multiproduct, multiretailer systems and using representative data from a specialty brand appliance products distribution setting, we assess the practical effectiveness of order smoothing.

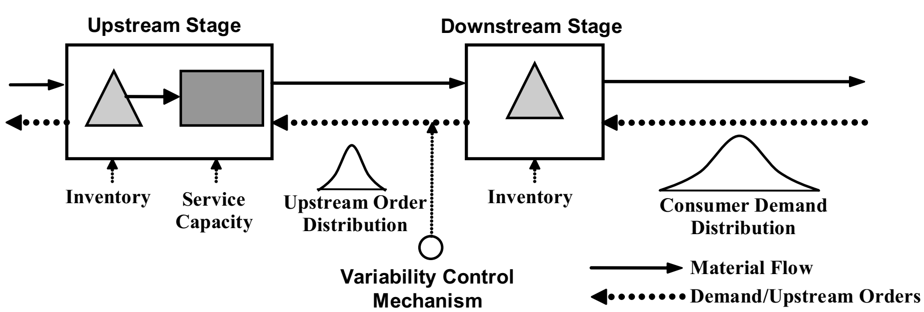
The remainder of this paper is organized as follows. Section 2 discusses approaches for tailoring inventory policies to control upstream variability propagation and defines a family of order-smoothing policies. In §3, we introduce a generic supply chain operations model, define how variability factors drive supply chain costs, and show how order smoothing influences these variability drivers. Section 4 analyzes the exponential smoothing policy and a moving weighted average policy, and discusses scenarios under which these two policies are optimal. We also provide a parsimonious characterization of supply chain cost structures, leading to important insights on the benefits of order smoothing. Section 5 generalizes our results to multiproduct, multiretailer systems. We then employ representative data from a specialty brand appliance distribution setting to assess the economic benefits of order smoothing in practice. Section 6 summarizes the learnings and implications of our analysis.

2. Variability Control in Supply Chains

In this section, we explore ways to control the variability that downstream stages propagate upstream, and discuss the properties of applying order-smoothing policies for variability control. As the setting for this discussion, we consider a two-stage partnered distribution chain consisting of retailers served by a single supplier. Figure 1 provides a supply chain schematic, emphasizing the role of variability control and the different types of assets (e.g., service capacity and inventory) at the upstream and downstream stages. The downstream stage (retailer) faces stochastic consumer demand; its inventory replenishment policy determines the variability of orders facing the upstream stage (supplier). In turn, this variability drives upstream inventory and service capacity decisions and costs. Our goal is to identify variability control mechanisms to moderate the variability of upstream orders and minimize the expected total system cost.

Recent literature has addressed coordination issues in multistage supply chain settings, encompassing scenarios with centralized or distributed control. Under centralized control, researchers address how integrated multiechelon inventory policies can reduce supply chain inventory costs (e.g., Cohen and Lee 1988, Graves and Willems 2000). Without centralized control, Baganha and Cohen (1998) show that introducing a central stocking point between a supplier and its retailers can reduce inventory cost, due to "risk pooling" benefits. Alternatively, appropriate buyer-supplier contracts (e.g., quantity discounts, buyback contracts, or other incentives) can serve to influence retailers' pricing and replenishment

Figure 1 **Supply Chain Structure and Variability Control**



decisions (e.g., Cachon 1999, Lee and Whang 1999, Chen 1999, Moinzadeh and Nahmias 2000). Tsay et al. (1999) provide a comprehensive review of the literature on supply chain contracts; economic and game theoretic models provide tools for optimizing such contracts (e.g., Cachon and Zipkin 1999, Corbett and DeCroix 2001). Typically, these models do not simultaneously address the inventory and service capacity decisions at the upstream stages that we consider.

Our work departs from conventional contracting literature by considering collaborative settings in which supply chain players enter into replenishment *policy* agreements to reduce total system costs. Within this policy framework, each stage minimizes its local costs. Despite this local autonomy, we will demonstrate that order smoothing effectively reduces supply chain costs, although we do not address how those savings are shared across the different stages of the supply chain (e.g., one option for gain sharing is to provide an appropriate discount to the retailer).

2.1. Inventory Policies to Control Upstream Variability

We next briefly describe three general approaches for controlling the variability of orders to upstream stages, namely, constrained ordering, variability penalties, and order smoothing.

Constrained ordering imposes minimum and maximum limits on retailer order quantities (e.g., Porteus 1990, Bassok and Anupindi 1997). Balakrishnan et al. (2002) study a constrained delivery policy under which the supplier can defer delivery of order quantities that exceed a prespecified upper limit. Rather than constrain the order quantity directly, Cachon (1999) considers constraints on order timing in a multiretailer system.

Variability penalties provide economic incentives for retailers to minimize order quantity deviations. For instance, if the penalty is proportional to the absolute deviation of the order quantity from the mean, retailers optimally follow a finite generalized base-stock policy with two base-stock levels (Porteus 1990). Henig et al. (1997) consider cost penalties for ordering above a contracted maximum value.

Order smoothing—the emphasis of this research—entails adopting an explicit replenishment policy that is tailored to reduce the variability of orders to upstream stages.

Smoothing, by considering a convex combination of a sequence of random variables, is a natural method to reduce the variability of an output stream. For instance, Graves et al. (1998) study production-smoothing policies to resolve trade-offs between production capacity and inventory requirements in manufacturing. Cruickshanks et al. (1984) use increased planning lead times to smooth production in a make-to-order job shop environment. Dejonckheere et al. (2003) use transfer function analysis to study how order variability changes by varying the parameters of an exponential smoothing model for demand forecasting and adjusting safety-stock levels. Our work departs from previous smoothing models by exploring the benefits (both in terms of variability dampening and system cost savings) of applying smoothing principles to replenishment orders directly. Next, we formally define the order-smoothing policy and discuss its variability reduction characteristics.

2.2. Order-Smoothing Principles

Let X_t denote the random *consumer demand* (at the retailer) for an item in period t, with *probability density function* (PDF) $\phi(x)$, *cumulative distribution function* (CDF) $\Phi(x)$, *mean* μ, and *standard deviation* σ. We assume that the demand distribution is stationary and demands in different periods are independent. If x_{t-1} is the actual demand that the retailer observes in period $t-1$, then the standard base-stock policy sets the replenishment order quantity q_t in period t equal to x_{t-1}. Hence, the variance of replenishment orders is σ^2. To reduce this variance, we consider a generalization of the base-stock policy by setting the order quantity equal to a convex combination of previous demand realizations. Given a sequence of nonnegative parameters $\mathbf{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_k, \ldots\}$ satisfying $\sum_{k=1}^{\infty} \alpha_k = 1$, suppose the retailer applies the following general (linear) *order-smoothing policy* to decide the order quantity q_t:

$$q_t = \sum_{k=1}^{\infty} \alpha_k x_{t-k}. \tag{2.1}$$

For convenience, we define the demand x_t, and hence order quantity q_t, to be zero for $t \leq 0$. We refer to the parameters α_k as *smoothing coefficients*; by varying these parameters, we can control the variability of upstream order quantities. When $\alpha_1 = 1$, the order-smoothing policy reduces to the "unsmoothed" base-stock policy of ordering $q_t = x_{t-1}$ in each period. We refer to the probability distribution of the (random) replenishment quantity Q_t as the *total-order* distribution. For the general order-smoothing policy (2.1), this distribution has mean $\mu_Q = \sum_{k=1}^{\infty} \alpha_k E(X_k) = \mu$ and variance

$$v_Q \stackrel{\Delta}{=} \sigma_Q^2 = \left(\sum_{k=1}^{\infty} \alpha_k^2\right)\sigma^2. \tag{2.2}$$

Because $\alpha_k \geq 0$ for all k, and $\sum_{k=1}^{\infty} \alpha_k = 1$, if two or more smoothing coefficients are positive, Equation (2.2) implies that $v_Q < \sigma^2$, i.e., the order-smoothing policy yields less upstream order variance than a base-stock policy. This variability reduction reflects a *temporal* risk pooling effect because the policy pools demand variation over multiple periods. As we shall see later, this pooling effect reduces both inventory and transportation costs, and complements other risk pooling effects in supply chain systems.

Within the broad family of general smoothing policies (2.1), we next consider two interesting subfamilies: *Moving Weighted Average* (MWA) and *Exponential Smoothing* (ES) policies. The *MWA policy* uses a *smoothing window size* W and sets the order quantity equal to a convex combination of the observed consumer demands during the last W periods. Thus, given the window size W and the smoothing coefficient vector $\mathbf{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_W\}$, the MWA policy sets:

$$q_t = \sum_{k=1}^{W} \alpha_k x_{t-k}. \tag{2.3}$$

Proposition 1: Variance-Minimizing MWA Policy. *Among all MWA order-smoothing policies with a given window size W, the Equal-Weight Moving Average (EMA) policy, with $\alpha_1 = \alpha_2 = \cdots = \alpha_W = 1/W$, minimizes the variance of the total-order distribution; for this policy, the total-order variance is σ^2/W.*

Proof. The optimal solution to the variance-minimizing convex optimization problem

$$\min\left\{v_Q = \left(\sum_{k=1}^{W} \alpha_k^2\right)\sigma^2 \colon 0 \leq \alpha_k \leq 1, \sum_{k=1}^{W} \alpha_k = 1\right\}$$

is $\alpha_k = 1/W$, for $k = 1, \ldots, W$. The corresponding minimum order variance is $\sum_{k=1}^{W} (1/W^2)\sigma^2 = \sigma^2/W$. □

The *ES policy* sets the quantity q_t in period t equal to a weighted combination using the *smoothing parameter* $\alpha \in (0, 1]$ of the latest consumer demand x_{t-1} and the previous order quantity q_{t-1}; that is,

$$q_t = \alpha_1 x_{t-1} + (1 - \alpha_1) q_{t-1}. \tag{2.4}$$

Recursively expressing the order quantity on the right-hand side of Equation (2.4) in terms of the consumer demand, we get

$$\begin{aligned} q_t &= \alpha x_{t-1} + (1-\alpha)(\alpha x_{t-2} + (1-\alpha) q_{t-2}) \\ &= \cdots = \sum_{k=1}^{\infty} \alpha (1-\alpha)^{k-1} x_{t-k}. \end{aligned} \tag{2.5}$$

Expression (2.5) shows that the ES policy is a special case of the general order-smoothing model (2.1), with $\alpha_k = \alpha(1-\alpha)^{k-1}$ for $k \geq 1$. Hence, the mean order quantity for the ES policy, denoted as μ_Q^E, is μ. Substituting $\alpha_k = \alpha(1-\alpha)^{k-1}$ in Expression (2.2), the variance of the total-order distribution simplifies to $v_Q^E = (\alpha/(2-\alpha))\sigma^2$, which shows that, for the ES policy, reducing α decreases total-order variance.

2.3. Information and Lead Time Implications of Order Smoothing

In addition to reducing upstream order variability, order smoothing also provides advance order information to the supplier. From Equation (2.1) we note that the retailer's replenishment orders in successive periods are *not* statistically independent because, for example, both Q_{t-1} and Q_t depend on x_{t-k} for all $k \geq 2$. The dependence between successive orders creates an opportunity for the supplier to use information embedded in past retailer orders. Let us first consider the ES policy (2.4). At time $(t-1)$, the supplier receives the replenishment order q_{t-1} to be filled in that period. However, because the retailer has agreed to use the ES policy, the supplier knows that in the next period t the retailer will order *at least* $(1-\alpha) q_{t-1}$ units. Therefore, if the manufacturer's delivery lead time to the supplier is one period, the supplier can order $(1-\alpha) q_{t-1}$ units from the manufacturer in period $(t-1)$, so as to receive and fill this portion of the retailer's order "just in time" in period t (the supplier's order would also include additional quantities needed to replenish the supplier's inventory). The supplier needs to maintain inventory only to service the stochastic portion of retailers' replenishment orders.

Thus, from the perspective of managing the supplier's inventory, the retailer's order quantity $q_t = \alpha X_{t-1} + (1-\alpha)q_{t-1}$ in period t consists of a deterministic portion $d_t = (1-\alpha)q_{t-1}$ and a stochastic portion $S_t = \alpha X_{t-1}$, which we refer to as the *net order* quantity. A similar argument applies to the general order-smoothing policy (2.1), if we assume that the supplier has visibility of past consumer demand realizations x_{t-k}.[1] In this case, the deterministic portion of the retailer's replenishment order in period t is $d_t = \sum_{k=2}^{\infty} \alpha_k x_{t-k}$, and the net order quantity is again $S_t = \alpha_1 X_{t-1}$. This net order quantity, which determines the supplier's (stochastic) inventory policy and holding costs, has mean $\mu_S = \alpha_1 \mu$ and variance $v_S = \alpha_1^2 \sigma^2$. Note that, for any $\alpha_1 < 1$, this variance is less than the total-order variance $\nu_Q = (\sum_{k=1}^{\infty} \alpha_k^2)\sigma^2$, which in turn is less than the order variance σ^2 without order smoothing.

As the preceding discussion illustrates, order smoothing can regulate the advance information (embedded in successive order quantities) available to the supplier. In particular, as the *first-period smoothing* parameter α_1 used by the retailer decreases from one, the supplier gains more advance order information. Several researchers have studied information sharing in supply chains (e.g., Bourland et al. 1996, Gavirneni et al. 1999, Lee et al. 2000, Cachon and Fisher 2000), and considered how collaborative planning and forecasting efforts can better match supply with demand. The cost to share information on future consumer demand must be weighed against the potential for using this information to lower inventory costs (e.g., Hariharan and Zipkin 1995, Gallego and Özer 2001, Chen 2001). The advance *order* information resulting from our order-smoothing approach complements opportunities for sharing advance *demand* information by further reducing the supplier's inventory costs beyond the reductions due to better forecasts of future demand.

We can also interpret the order-smoothing policy as a means to regulate the average time the supplier takes to replace stocks that the retailer sold in a particular period. When the retailer uses the general order-smoothing policy $q_t = \sum_{k=1}^{\infty} \alpha_k x_{t-k}$, we see that in period t the supplier replaces only the fraction α_k of period $(t-k)$'s consumer demand x_{t-k}. Therefore, even though the supplier replenishes the retailer's full order in each period, the effective time for the supplier to "replace" the retail stock depleted by consumer demand in a particular period is equal to $L = \sum_{k=1}^{\infty} k\alpha_k$. With no smoothing ($\alpha_1 = 1$), the stock replacement time L is one period. As we increase order smoothing, the effective stock replacement time increases, reducing the supplier's inventory cost while increasing the retailer's safety-stock requirements.

[1] Balakrishnan et al. (2003) show that even when suppliers do not have visibility of end-consumer demand, by retaining past order information they can reduce the replenishment order uncertainty under order smoothing.

3. Variability Drivers of Supply Chain Costs

Section 2 introduced order smoothing as a variability control mechanism and examined some of its structural properties. To obtain a deeper understanding of the potential benefits from order smoothing, we next develop a general two-stage supply chain model and identify the sources of uncertainty that drive inventory and capacity decisions. Identifying these pertinent cost drivers permits us to develop general properties of the optimal order-smoothing policy. Later, in §§4 and 5, we apply a specific cost structure and demand distribution, and use supply chain data in a practical setting to assess the benefits of order smoothing.

3.1. Supply Chain Model

We consider a supply chain consisting of retailers served by a single supplier. Consumer demand at each retailer follows an independent and stationary probability distribution. We assume that retailers and the supplier use a periodic-review inventory system with a common period length (e.g., one week). If consumer demand during a period exceeds available retail stock, then excess demand is backordered. At the end of every period, each retailer places replenishment orders with the supplier for delivery in full by the start of the next period. If the supplier has insufficient on-hand inventory to meet all orders, it requests expedited shipments (at additional cost) from the manufacturer to cover this shortage. If the total order size exceeds the supplier's in-house service (delivery) capacity, the supplier uses outside sources to handle any overage. After shipping to the retailers, the supplier places a replenishment order with the manufacturer, which is received before retailer orders arrive the next period. We assume, for convenience, that the supplier owns and manages the service capacity; however, our analysis applies even when the supplier is responsible only for managing inventory, and another supply chain partner handles deliveries. We will initially assume that the supplier provides a single product to a single retailer, and later (in §5) relax this assumption.

3.2. Cost Categories and Variability Drivers

The costs in this supply chain model naturally fall into three categories:

(1) *Supplier inventory* costs for holding inventory and expediting supplies, if necessary, from the manufacturer;

(2) *Service capacity* costs for in-house service capacity (e.g., vehicles and drivers) and any additional outsourcing costs when orders exceed in-house capacity; and

(3) *Retail inventory and consumer demand fulfillment* costs for holding retail stocks and for backordered consumer demand.

We will refer to the total cost for each of the three categories as *s-inventory*, *service capacity*, and *r-inventory* costs, respectively. We next identify the applicable variability (uncertainty) that drives each cost category. We measure variability in terms of the standard deviations of the relevant random variables; under the order-smoothing policy, these standard deviations depend on the smoothing parameters $\mathbf{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_k, \ldots\}$ and the standard deviation σ of consumer demand.

Supplier inventory management: As we noted in §2, the retailer's replenishment order in each period consists of a known quantity $d_t = \sum_{k=2}^{\infty} \alpha_k x_{t-k}$, which the supplier orders in advance from the manufacturer, and the net order quantity $S_t = \alpha_1 X_{t-1}$. Therefore, the variability driver of s-inventory costs (including holding and expediting costs) is simply the variability of the net order S_t in each period, measured by its standard deviation $\sigma_S = \alpha_1 \sigma$.

Service capacity management: In each period t, the supplier must process the retailer's total-order quantity Q_t, using a combination of in-house and contingency (outsourced) service capacity. Unlike inventory, service capacity cannot be adjusted dynamically on a period-by-period basis. Hence, the variability

$$\sigma_Q = \left(\sqrt{\sum_{k=1}^{\infty} \alpha_k^2}\right)\sigma$$

of the retailer's total order Q_t will drive the supplier's capacity cost.

Retailer inventory management: With order smoothing, the retailer orders $q_t = \sum_{k=1}^{\infty} \alpha_k x_{t-k}$ units in each period t, rather than maintain a fixed base-stock level. The variability of this inflow quantity together with uncertainty in outflow (consumer demand) constitutes the exogenous variability that drives retail inventory costs, i.e., the costs of retail stocks plus backorders. To formally identify the relevant random variable, suppose the retailer starts selling the product at time $t = 0$ with a starting inventory level of I_0. Recall that we define the demand X_t (and hence order quantity Q_t) to be zero for $t \leq 0$. We can express the ending inventory I_t in any period $t > 0$ as:

$$I_t = I_0 - \sum_{k=0}^{\infty} X_{t-k} + \sum_{k=1}^{\infty} Q_{t-k}. \tag{3.1}$$

We define the *retailer's effective demand* at time t, denoted as the random variable R_t, as the cumulative demand minus the cumulative replenishments until time t, i.e., $R_t = \sum_{k=0}^{\infty} X_{t-k} - \sum_{k=0}^{\infty} Q_{t-k}$, which implies that $I_t = I_0 - R_t$. Therefore, the variability of ending inventory is the same as the variability of the retailer's effective demand. Because $Q_t = \sum_{k=1}^{\infty} \alpha_k X_{t-k}$ for the order-smoothing policy, we can express effective demand in terms of the consumer demand X_t as

$$\begin{aligned} R_t &= \sum_{k=0}^{\infty} X_{t-k} - \sum_{k=1}^{\infty} Q_{t-k} \\ &= X_t + \sum_{k=1}^{\infty}\left(1 - \sum_{l=1}^{k} \alpha_l\right) X_{t-k}. \end{aligned} \tag{3.2}$$

Let us define $u_k = 1 - \sum_{l=1}^{k} \alpha_l$ as the *proportion* of period $(t-k)$'s demand that remains *unfilled* until (and including) period t, permitting us to express the effective demand as

$$R_t = X_t + \sum_{k=1}^{\infty} u_k X_{t-k}. \tag{3.3}$$

Because the consumer demand distribution is stationary and independent from period to period, the variability of ending inventory is

$$\sigma_R = \left(\sqrt{1+\sum_{k=1}^{\infty} u_k^2}\right)\sigma. \tag{3.4}$$

This variability in effective demand drives the variability of the retailer's ending inventory in each period, and thus determines the r-inventory costs.

In summary, we have identified three random variables—the total-order quantity Q_t, the retailer's effective demand R_t, and the supplier's net orders S_t—whose respective standard deviations σ_Q, σ_R, and σ_S drive the costs of service capacity, retailer inventory, and supplier inventory. Different order-smoothing policies produce different levels of variability for these three quantities. For instance, the unsmoothed base-stock policy with $\alpha_1 = 1$ minimizes σ_R (because $u_k = 0$ for all $k \geq 1$) and hence minimizes r-inventory costs, but can impose high supplier (s-inventory and capacity) costs. Similarly, the equal-weight moving average policy minimizes σ_Q and hence service capacity costs (Proposition 1), but does not minimize r-inventory and s-inventory costs. We wish to study smoothing policies that minimize the sum of these three costs, which we refer to as the *total system cost*.

3.3. Properties of Optimal Order-Smoothing Policies

We expect the s-inventory, capacity, and r-inventory costs to increase (or, more precisely, not decrease) with their respective variabilities σ_S, σ_Q, and σ_R. Under this condition, we can establish the following properties of the optimal smoothing policy.

Proposition 2: Monotonicity of Optimal Smoothing Coefficients. *For the general order-smoothing policy, $Q_t = \sum_{k=1}^{\infty} \alpha_k X_{t-k}$, the optimal smoothing coefficients α_k must be nonincreasing with k for all $k \geq 2$.*

Proof. See Appendix A. □

Proposition 2 implies that we should focus on order-smoothing policies with nonincreasing coefficients for $k \geq 2$. To identify specific cost-minimizing smoothing policies, we next consider a special case of the general supply chain cost model in which the variability-dependent component of service capacity cost is zero. This assumption might hold, for instance, if the supplier acquires no in-house service capacity, choosing instead to fully outsource this capacity to a firm that charges a per unit rate for actual capacity usage.[2] In this case, if g is the per unit shipping cost, the transportation cost is gQ_t in each period t, and the expected transportation cost is $gE(Q_t) = g\mu$, which is independent of the associated variability-driver σ_Q. We refer to this special case as the *third-party service provider* (3PS) scenario. In the next proposition, we characterize the optimal smoothing policy structure for the 3PS scenario.

Proposition 3: Optimal Policy Structure when Capacity Costs Are Variability Independent. *When capacity costs are variability independent, setting $\alpha_k = 0$ for all $k \geq 3$ is optimal, i.e., a policy that satisfies this property achieves the lowest total system cost among all order-smoothing policies.*

Proof. Let $\{\bar{\alpha}_1, \ldots, \bar{\alpha}_k, \ldots\}$ be the optimal set of smoothing parameters when capacity costs are variability independent. If $\bar{\alpha}_k > 0$ for at least one $k > 2$, then $\bar{\alpha}_2 < 1 - \bar{\alpha}_1$, and because $u_1 = 1 - \bar{\alpha}_1$, the variance of the retailer's effective demand is $\sigma_R^2 = \sigma^2(1 + (1-\bar{\alpha}_1)^2 + \sum_{k=2}^{\infty} u_k^2)$. By using a smoothing policy with coefficients $\alpha_2 = 1 - \bar{\alpha}_1$ and $\alpha_k = 0$ for all $k > 2$, we can reduce this variability to $\sigma^2(1 + (1-\bar{\alpha}_1)^2)$. Because $\sigma_S = \bar{\alpha}_1 \sigma$ remains unaffected, the total system cost with the new smoothing coefficients is lower, and so the given policy cannot be optimal. □

Because the optimal smoothing policy suggested by Proposition 3 has all smoothing coefficients except the first two (i.e., α_1 and α_2) equal to zero, we refer to this policy as a *two-period moving average* policy.

3.4. Variability Reduction by Exponential Smoothing and Moving Average Policies

Henceforth, we focus on two subclasses of the general smoothing model (2.1): The ES policy (from §2.2) and a *Balanced Moving Average* (BMA) policy that generalizes the EMA policy from Proposition 1. The BMA policy has two parameters: The smoothing window size W and the first-period smoothing coefficient α.

[2] Third-party logistics providers, for example, commonly offer per-unit fee structures for their services. Because they serve multiple customers, their capacity costs are not sensitive to the variability of individual customer requirements due to the pooling of demand uncertainty across customers.

Table 1 Variability Drivers for Special Smoothing Policies

	Variability multiplier $\theta_Y^P = \sigma_Y^P/\sigma$ for		
Smoothing policy P	s-Inventory	Capacity	r-Inventory
Exponential smoothing (ES): $P = E$	α	$\sqrt{\dfrac{\alpha}{2-\alpha}}$	$\dfrac{1}{\sqrt{\alpha(2-\alpha)}}$
Balanced moving average (BMA): $P = B(W)$	α	$\sqrt{\alpha^2 + \dfrac{(1-\alpha)^2}{W-1}}$	$\sqrt{1 + \dfrac{(1-\alpha)^2 W(2W-1)}{6(W-1)}}$
Two-period moving average (TMA): $P = T$	α	$\sqrt{\alpha^2 + (1-\alpha)^2}$	$\sqrt{1 + (1-\alpha)^2}$
Baseline $P = 0$	1	1	1

Given these two parameters, the policy sets $\alpha_1 = \alpha$ and $\alpha_k = (1-\alpha)/(W-1)$ for $k = 2, \ldots, W$ (and $\alpha_k = 0$ for all $k > W$). The *Two-Period Moving Average* (TMA) policy is a special case of the BMA policy with $W = 2$. The following observations motivate the BMA policy: (i) The first-period smoothing parameter α_1 determines the s-inventory cost, and so varying α_1 independent of the other coefficients is desirable; and (ii) equalizing the smoothing coefficients reduces the service capacity cost. Following the same line of argument as in Proposition 1, we can show that for a given α_1 and window size W the BMA policy minimizes the capacity cost among all MWA policies with window size W. Note that both the ES and BMA policies satisfy the coefficient-monotonicity condition of Proposition 2.

We adopt the following notation to distinguish between the results for different policy classes. We assign an index P to each policy class and employ this index as a superscript to denote the parameters and variables associated with this policy class, such as the smoothing vector, variability drivers, and costs. We omit the policy superscript when referring to the generic smoothing policy with unrestricted coefficients $\mathbf{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_k, \ldots\}$. We use the indices $P = E$ for the ES policy, $P = B(W)$ for the BMA policy with window size W, and $P = T$ for the TMA policy. With this notation, each policy that we consider has a single-policy parameter, namely, the first-period smoothing coefficient α. As an illustration of this notation, for the ES policy, we denote its smoothing vector as $\mathbf{A}^E = \{\alpha(1-\alpha)^{k-1}\}$ and its total-order quantity as Q_t^E. From §2, we know that this order quantity has mean $\mu_Q^E = \mu$ and $\sigma_Q^E = \sigma\sqrt{\alpha/(2-\alpha)}$. For the W-window BMA policy, $\mathbf{A}^{B(W)} = \{\alpha_1 = \alpha, \alpha_k = (1-\alpha)/(W-1)$, for $2 \le k \le W\}$, while the TMA policy has the smoothing vector $\mathbf{A}^T = \{\alpha_1 = \alpha, \alpha_2 = (1-\alpha)\}$. As a benchmark for comparing these policy classes, we also consider the unsmoothed (i.e., $\alpha_1 = 1$) *baseline* policy, denoted as $P = 0$. Table 1 provides expressions for the variability drivers σ_S, σ_Q, and σ_R, corresponding to the ES, BMA, TMA, and baseline policies. These expressions follow from substituting the smoothing coefficients for each policy into the general expressions for the standard deviations σ_S, σ_Q, and σ_R that we derived in §3.2. Because these standard deviations are all linear in the standard deviation σ of consumer demand, for each random variable Y (with $Y = S_t$, Q_t, or R_t), we show in Table 1 the expression for the ratio $\theta_Y^P \stackrel{\Delta}{=} \sigma_Y^P/\sigma$ for every policy class P. We refer to these ratios as *variability multipliers* and later use them to develop expressions for the total system cost.

From the expressions in Table 1, we see that for the ES policy the variability σ_R^E of the retailer's effective demand is convex and decreasing in α. However, the variability σ_Q^E of total orders and the variability σ_S^E of the supplier's net order quantity are increasing in α (σ_Q^E is also convex in α for $\alpha > 1/2$). Therefore, when s-inventory costs or capacity costs are highly sensitive to their respective variability drivers, the optimal value of α for the ES policy will be low. For the BMA policy with window size W, all three standard deviations are convex in α, and (for a given α) $\sigma_Q^{B(W)}$ decreases, but $\sigma_R^{B(W)}$ increases as W increases. Thus, for supply chain contexts with highly sensitive capacity costs, we expect the cost-minimizing BMA policy to have a large window size W and a low first-period coefficient α.

Henceforth, as we compare the performance of different classes of policies, we will use the term "best" policy to refer to the alternative among the ES, BMA, and TMA policies that provides the lowest total system cost, in contrast with the "optimal" policy that can have unrestricted smoothing coefficients. In Proposition 3, we showed that when capacity costs are variability independent, the TMA policy is the *optimal* policy among all (linear) order-smoothing policies. Under what circumstances will the ES policy or the general BMA policy, with $W > 2$, yield a lower total system cost than the TMA policy? The following two propositions address this question, again assuming only that the three cost components increase monotonically with their corresponding standard deviations. To identify scenarios under which the ES policy is effective, we consider the special case, analogous to the previous variability independent capacity cost case, in which *s*-inventory cost does not depend on the variability of the net orders. This situation applies if the manufacturer provides the supplier with a near-zero lead time, thus permitting the supplier to operate as a cross-docking facility that does not hold inventory (Stalk et al. 1992). We refer to this supply scenario as the *cross-docking* system.[3]

Proposition 4: Effectiveness of ES Policy when *s*-inventory Costs Are Variability Independent. *For systems with variability independent s-inventory costs, as in cross-docking systems, the ES policy achieves lower system cost than the BMA policy (including the TMA special case).*

Proof. See Appendix B. □

Next, we identify situations under which the general BMA policy achieves lower capacity cost.

Proposition 5: Effectiveness of BMA Policy. *Given any target supplier inventory cost and corresponding first-period smoothing coefficient α, the BMA policy achieves lower capacity cost than the ES policy if and only if $W > 2/\alpha$.*

Proof. Given α, σ_Q^2 is $\alpha/(2-\alpha)$ for the ES policy and $\alpha^2 + (1-\alpha)^2/(W-1)$ for the BMA policy. Therefore, the BMA policy has lower capacity cost if $\alpha^2 + (1-\alpha)^2/(W-1) < \alpha/(2-\alpha)$, i.e., if $W > 2/\alpha$. □

[3] While cross-docking reduces supplier inventory costs, it can increase handling costs. We can incorporate both fixed and variable handling costs in the model, and Proposition 4 continues to hold for this extension.

Proposition 5 has two implications. First, because $\alpha \le 1$, the ES policy always achieves lower capacity costs than the TMA policy (with $W = 2 < 2/\alpha$). On the other hand, for a sufficiently large window size W, the BMA policy yields a lower capacity cost than the ES policy. Thus, we see that *no one policy is best under all scenarios*. Depending on the relative sensitivity of the *s*-inventory, capacity, and *r*-inventory costs to their respective variabilities, any one of the three smoothing policies may dominate.

4. Optimizing Supply Chain Costs by Order Smoothing

Sections 3.3 and 3.4 have identified special situations under which the ES, BMA, or TMA policy is superior. To provide insights regarding which policy to use in more general settings and assess the resulting system cost savings relative to the no-smoothing baseline case, we next analyze the supply chain decision problems for a single-product, single-retailer system with a specific cost structure and consumer demand distribution.

4.1. Supply Chain Operating Costs

We assume that the retailer and supplier respectively incur holding costs at the rate of h_R and h_S per unit remaining in inventory at the end of each period. The retailer incurs a backorder cost of b per period for each unit that is backordered. To expedite shipments from the manufacturer when the retailer's order quantity exceeds the available supplier inventory, the supplier incurs an expediting cost of e per unit; this simplifying cost assumption facilitates our analysis and development of insights. The fixed cost to acquire and maintain one unit of in-house service capacity (e.g., delivery trucks, drivers) is f per period. The service capacity choice is made at the start of the planning horizon, and the fixed capacity costs apply in each period, irrespective of whether the capacity is utilized. In addition, the supplier incurs a variable cost of v for every unit actually processed using this service capacity. Although we assume that capacity is available in continuous increments, our analysis extends to cases where capacity is available only in a discrete size. External service capacity (e.g., deliveries using third-party carriers), needed when the order quantity in a period exceeds the internal capacity, costs g per unit. We assume, without loss of generality, that this *overflow cost* g exceeds the per-unit fixed plus variable cost $(f+v)$ of in-house capacity.

Given an order-smoothing policy with its associated set of smoothing parameters, each supply chain stage makes local decisions to minimize its own cost. Our goal is to determine the set of smoothing parameters that leads to the lowest overall system cost. Next, we discuss each stage's local decision problem.

4.2. Local Decisions

Recall that $\phi(\cdot)$ and $\Phi(\cdot)$ denote the PDF and CDF of *consumer* demand X_t in each period, with mean μ and standard deviation σ. The demand distribution and smoothing parameters determine the distribution of the three *derived* random variables: The net orders S_t, total orders Q_t, and effective demands R_t. If Y denotes one of these derived random variables, we define $\phi_Y(\cdot)$ and $\Phi_Y(\cdot)$ as the PDF and CDF of Y. The total-order quantity $Q_t = \sum_{k=1}^{\infty} \alpha_k X_{t-k}$ and the retailer's effective demand $R_t = X_t + \sum_{k=1}^{\infty} u_k X_{t-k}$ are both convolutions of consumer demands over multiple periods, while the net order quantity $S_t = \alpha_1 X_{t-1}$ is a scaled version of the consumer demand distribution, i.e., $\phi_S(x) = (1/\alpha)\phi(x/\alpha)$. Therefore, if consumer demand follows a normal distribution, then the distributions corresponding to these three derived variables are also normal. Using these distributions, we can now address the decision problem for the three cost categories (i.e., s-inventory, service capacity, and r-inventory costs).

Supplier's inventory decision: The supplier must maintain inventory or expedite supplies from the manufacturer to meet the stochastic portion $S_t = \alpha_1 X_{t-1}$ of retailer orders. Given this i.i.d. sequence of net orders, a base-stock policy minimizes the supplier's expected holding and backorder costs. To achieve the optimal trade-off between inventory and expediting costs, the supplier must set the inventory service level equal to the critical fractile value $\rho_S = e/(e+h_S)$ (Nahmias 2001). Therefore, the optimal inventory-stocking decision requires setting the base-stock level equal to $I_S^* = \Phi_S^{-1}(\rho_S)$. The expected s-inventory cost (per period), which we denote by C_S, is then

$$C_S = h_S \int_0^{I_S^*} (I_S^* - s)\phi(s)\,ds + e\int_{I_S^*}^{\infty} (s - I_S^*)\phi(s)\,ds.$$

If $\Lambda_S(I) = \int_I^{\infty} (s - I)\phi_S(s)\,ds$ denotes the loss function when the supplier stocks I units, the supplier's expected cost per period simplifies to

$$C_S = h_S\big(\Phi_S^{-1}(\rho_S) - \mu_S\big) + (e+h_S)\Lambda_S\big(\Phi_S^{-1}(\rho_S)\big). \tag{4.1}$$

If consumer demand X_t is normally distributed, then for any desired fractile value ρ, let $z(\rho)$ denote the corresponding standard normal deviate, and let $L(\rho)$ be the value of the standard normal loss function at $z = z(\rho)$. Then, the supplier's expected cost per period is

$$C_S = (h_S(z(\rho_S) + L(\rho_S)) + eL(\rho_S))\sigma_S. \tag{4.2}$$

Henceforth, we assume that consumer demand follows a normal distribution, and therefore we do not show cost expressions in the form (4.1), which applies for general distributions.

Service capacity decision: The service capacity decision problem entails deciding how many units of in-house service capacity to acquire to minimize the total fixed and (expected) variable costs of acquiring and using these resources, plus the expected overflow cost for outside resources when total order quantity Q_t exceeds the in-house capacity. If $\rho_Q \stackrel{\Delta}{=} (g-(f+v))/(g-v)$ denotes the *critical capacity fractile* value, the optimal level of in-house capacity is $T^* = \Phi_Q^{-1}(\rho_Q)$ (Ernst and Pyke 1993 use a similar approach to model transportation capacity decisions). Therefore, the expected capacity cost per period is $(f+v)\mu_Q + (fz(\rho_Q) + (g-v)L(\rho_Q))\sigma_Q$. Because the first term $(f+v)\mu_Q$ in this expression does not depend on the total-order variability σ_Q, whereas we focus on variability-dependent costs, we define the *expected service capacity cost per period* C_Q exclusive of this term, i.e.,

$$C_Q = (fz(\rho_Q) + (g-v)L(\rho_Q))\sigma_Q. \tag{4.3}$$

Note that under the order-smoothing regime the service capacity T^* is optimal regardless of which supply chain partner owns the service resources or incurs the service costs.

Retailer's inventory decision: Recall that if I_0 denotes the retailer's starting inventory level at time 0, then the retailer's ending inventory level at time t is $I_t = I_0 - R_t$, where R_t is the retailer's effective demand. To minimize expected holding and backorder costs, the retailer will set the starting inventory level I_0 equal to $I_0^* = \Phi_R^{-1}(\rho_R)$, where $\rho_R \triangleq b/(b+h_R)$ is the *retailer's critical fractile* value. Therefore, the *expected r-inventory cost per period* is

$$C_R = \big(h_R(z(\rho_R)+L(\rho_R))+bL(\rho_R)\big)\sigma_R. \tag{4.4}$$

For a particular smoothing policy P, we denote the three variability-dependent costs for s-inventory, capacity, and r-inventory as C_Y^P, for $Y=S$, Q, and R. Expressions (4.2), (4.3), and (4.4) show that the three cost components increase linearly with the respective variability drivers σ_S^P, σ_Q^P, and σ_R^P. Moreover, because $\sigma_S^0=\sigma_Q^0=\sigma_R^0=\sigma$ for the baseline policy ($P=0$) with no smoothing, we see that $C_Y^P=(C_Y^0/\sigma)\sigma_Y^P$, for $Y=S$, Q, and R. For any given smoothing policy P and each derived random variable Y, we have defined the *variability multiplier* θ_Y^P (see Table 1) as the ratio σ_Y^P/σ of the standard deviation σ_Y^P of Y under policy P to the consumer demand standard deviation σ. In terms of the variability multipliers from Table 1, we can express the cost components as

$$C_Y^P = C_Y^0\theta_Y^P \quad \text{for } Y=S, Q, \text{ and } R. \tag{4.5}$$

That is, each of the cost components is equal to the corresponding baseline policy cost multiplied by the variability cost multiplier.

4.3. Characterizing the Supply Chain Cost Structure

Let $TC^P = C_S^P + C_Q^P + C_R^P$ denote the expected systemwide *variability-dependent* cost if we use policy P. Because this cost excludes the variability-independent capacity cost term $(f+v)\mu_Q$, we refer to it as the *net system cost*. Using Equation (4.5), we express this cost as

$$\begin{aligned} TC^P &= C_S^P + C_Q^P + C_R^P = C_S^0\theta_S^P + C_Q^0\theta_Q^P + C_R^0\theta_R^P \\ &= \left(\theta_R^P + \frac{C_S^0}{C_R^0}\theta_S^P + \frac{C_Q^0}{C_R^0}\theta_Q^P\right)C_R^0. \end{aligned} \tag{4.6}$$

Define the *s-inventory cost index* δ as the ratio of the baseline policy's s-inventory and r-inventory costs, i.e., $\delta \triangleq C_S^0/C_R^0$. Similarly, let the *capacity cost index* τ be the ratio of the baseline policy's capacity and r-inventory costs, i.e., $\tau \triangleq C_Q^0/C_R^0$. Surprisingly, as we will show, these two indices suffice to fully characterize the supply chain cost structure and savings due to order smoothing. High values of the *s-inventory cost index* δ indicate high s-inventory costs relative to r-inventory costs (in the baseline case). For instance, if holding costs are nearly the same for both suppliers and retailers and the cost of expediting items from the manufacturer is high (compared to the backorder cost), then δ will be large. We, therefore, refer to supply chain scenarios and products with high δ values as *inflexible supply* (versus *responsive* supply) scenarios. High values of the *capacity cost index* τ signify high service capacity costs relative to the r-inventory costs. This situation occurs when transportation costs are high relative to holding costs (e.g., for items with low value-to-weight ratios such as some building products, or items requiring specialized transportation and hence high capacity outsourcing costs). We refer to products with high values of the capacity cost index as *service-intensive* products.

The two cost indices permit us to express the net system cost function in a convenient way. Specifically, substituting $\delta = C_S^0/C_R^0$ and $\tau = C_Q^0/C_R^0$ in the cost function (4.6) yields

$$TC^P = \big(\theta_R^P + \delta\theta_S^P + \tau\theta_Q^P\big)C_R^0. \tag{4.7}$$

Interestingly, by defining the two key cost indices δ and τ, and the variability multipliers θ_Y^P, we have separated (as multiplicative factors) the influence of the smoothing policy P from the underlying supply chain cost structure. Thus, the cost indices δ and τ are ideally suited for characterizing costs *as a function of variability*. For the baseline policy with no smoothing, the net system cost (4.7) reduces to

$$TC^0 = (1+\delta+\tau)C_R^0. \tag{4.8}$$

For the ES policy, substituting the corresponding variability multipliers from Table 1 yields the net cost

$$TC^E = \left(\frac{1}{\sqrt{\alpha(2-\alpha)}} + \delta\alpha + \tau\sqrt{\frac{\alpha}{(2-\alpha)}}\right)C_R^0. \tag{4.9}$$

Similarly, for the BMA policy, the net system cost is

$$\begin{aligned} TC^{B(W)} = \Bigg(&\sqrt{1+\frac{(1-\alpha)^2W(2W-1)}{6(W-1)}} + \delta\alpha \\ &+ \tau\sqrt{\alpha^2+\frac{(1-\alpha)^2}{(W-1)}}\Bigg)C_R^0. \end{aligned} \tag{4.10}$$

Let us briefly address the convexity of these two cost functions. For the BMA policy with a given window size W, the cost function (4.10) is convex in α, and so we can obtain the optimal value of the smoothing parameter by setting the first derivative of $TC^{B(W)}$ equal to zero. For the ES policy, the second derivative of TC^E is nonnegative, and hence the cost function is convex in α if and only if $2(1+\tau)\alpha^2-(4+\tau)\alpha+3\geq 0$ for all $\alpha\in[0,1]$. Because the minimum value of the left-hand side of this inequality is $8+16\tau-\tau^2$ (at $\alpha=(1+\tau/4)/(1+\tau)$), the inequality holds and the ES policy's cost function is convex if $8+16\tau-\tau^2\geq 0$, or $\tau\leq 8+\sqrt{72}\approx 16.48$. In practice, the expected variability-dependent capacity cost with no smoothing is likely to be far less than 16 times the r-inventory cost, and so we henceforth assume that this relatively mild condition holds.

We measure the effectiveness of any smoothing policy P in terms of the system cost savings that this policy generates relative to the baseline policy's *net* (variability-dependent) system cost.[4] For this purpose, we define policy P's *percentage savings* as $\omega^P=1-(TC^P/TC^0)$. Substituting for TC^P and TC^0 from Equations (4.7) and (4.8), we obtain

$$\omega^P=1-\frac{\theta_R^P+\delta\theta_S^P+\tau\theta_Q^P}{1+\delta+\tau}. \tag{4.11}$$

This metric depends solely on the underlying supply chain cost structure (captured by the two indices δ and τ) and the smoothing characteristics of policy P (captured by the variability multipliers θ_Y^P).

4.4. Closed-Form Solutions and Best Policy for Special Cases

To evaluate the value of order smoothing using the savings metric ω^P, we begin by considering the two special cases—3PS and cross-docking—that we introduced in §3.

[4] Dividing cost savings by the baseline policy's net system cost, rather than total system cost, permits us to develop analytical expressions and insights on the behavior of the savings metric. Later, when we discuss an illustrative application of the model in §5, we measure savings relative to the baseline policy's total system cost.

Variability-Independent Capacity Costs: The 3PS Scenario. For the 3PS scenario, because the costs of service capacity are proportional to actual usage, the capacity cost index τ is zero. In this case, because the TMA policy is optimal (Proposition 3), we minimize system costs by selecting the TMA policy's smoothing parameter α to minimize

$$TC^T\big|_{\tau=0}=\left(\theta_R^T+\delta\theta_S^T\right)C_R^0=\left(\sqrt{1+(1-\alpha)^2}+\delta\alpha\right)C_R^0,$$

which is convex in α. Setting the derivative of $TC^T|_{\tau=0}$ with respect to α equal to zero, we obtain the following optimal smoothing parameter value:

$$\alpha_{\text{opt}}^T\big|_{\tau=0}=1-\frac{\delta}{\sqrt{1-\delta^2}}\quad\text{if } \delta<\frac{1}{\sqrt{2}},\ \text{and 0 otherwise.} \tag{4.12}$$

The corresponding minimum net system cost is $(\sqrt{1-\delta^2}+\delta)C_R^0$ if $\delta<1/\sqrt{2}$, and $\sqrt{2}C_R^0$ otherwise. Consequently, the TMA policy's percentage savings is $\omega^T=(1-\sqrt{1-\delta^2})/(1+\delta)$ if $\delta<1/\sqrt{2}$, and $\omega^T=(1+\delta-\sqrt{2})/(1+\delta)$ otherwise. Figure 2(a) shows the values of the optimal smoothing parameter and percentage savings as functions of the s-inventory cost index δ.

As Figure 2(a) illustrates, order smoothing significantly reduces supply chain costs, even when the supplier's inventory costs are relatively small (i.e., for small δ). For instance, even when δ is only 1/2 (i.e., without smoothing, the expected s-inventory cost is just 50% of the r-inventory cost), order smoothing

Figure 2a 3PS Scenario with $\tau=0$; TMA Policy Is Optimal

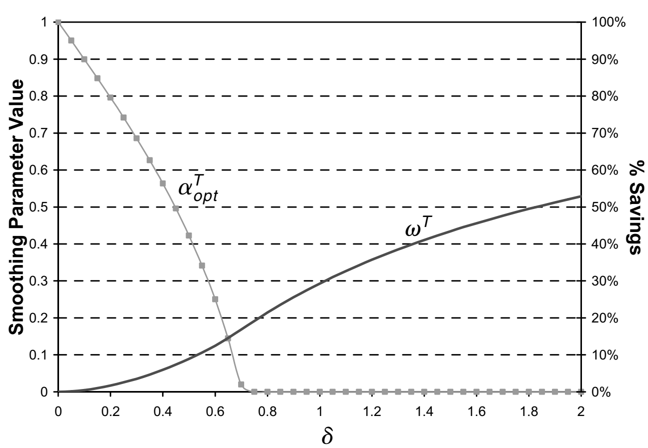
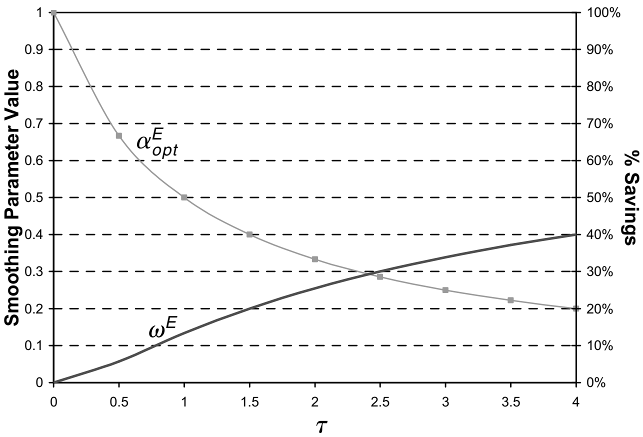


Figure 2b Cross-Docking Scenario with $\delta = 0$; ES Policy Is Best



using the optimal TMA policy provides nearly 10% savings in net system costs over the baseline policy. Notice also that as δ increases, the optimal value of the first-period smoothing parameter α decreases, and hence the desired amount of smoothing increases. From Equation (4.12), we see that when the supplier's expected inventory cost exceeds 71% of the retailer's cost (i.e., if $\delta \geq 1/\sqrt{2} = 0.707$), full smoothing ($\alpha = 0$) is optimal. Thus, we find that moving inventory *downstream* in the supply chain can be advantageous, as this shift provides countervailing variability-reduction benefits upstream.

Variability-Independent *s*-Inventory Costs: The Cross-Docking Scenario. In the cross-docking scenario, the supplier holds no inventory, and so $\delta = 0$. As shown in Proposition 4, the ES policy is best in this case. By setting the derivative of the right-hand side of Equation (4.9) equal to zero, with $\delta = 0$, we obtain the optimal value ES smoothing parameter value

$$\alpha^E_{\text{opt}}\big|_{\delta=0} = \frac{1}{1+\tau}. \tag{4.13}$$

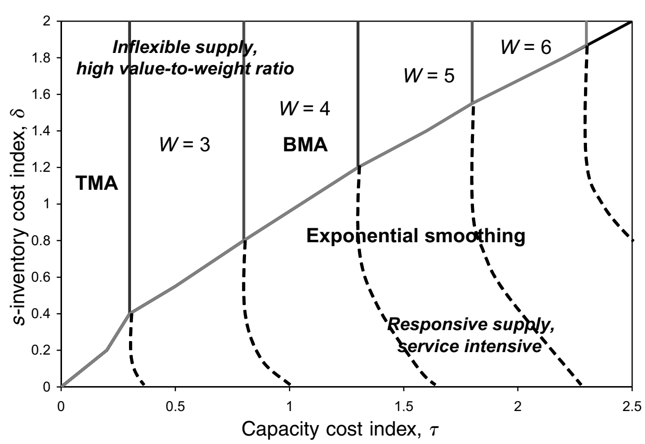
The corresponding minimum net system cost is $TC^E = (\sqrt{2\tau+1})C^0_R$, and so the ES policy's percentage cost savings is $\omega^E = 1 - \sqrt{2\tau+1}/(1+\tau)$. Note that the optimal value of α is always less than one and strictly positive. Figure 2(b) shows how the percentage savings and optimal smoothing parameter value vary as the capacity cost index τ increases. For instance, when the baseline r-inventory cost and capacity cost are equal ($\tau = 1$), exponential smoothing reduces net system costs by around 10%.

Comparing Figures 2(a) and 2(b), we see that as the relative supplier cost (τ or δ) increases, the cross-docking scenario's smoothing parameter decreases more gradually and its rate of increase of savings is lower than for the 3PS setting. Because the s-inventory cost decreases proportionally with the first-period smoothing parameter value α, whereas the service capacity cost does not, a small amount of smoothing reduces upstream order variability more substantially in 3PS systems than for cross-docking.

4.5. Insights for General Cost Scenarios

We now address the general supply chain context in which both s-inventory and capacity costs increase with variability. Given the relative cost indices τ and δ for a supply chain, we can find the optimal smoothing parameters for the ES and BMA policies by minimizing the cost functions (4.9) and (4.10); for these parameter values, Expression (4.11) provides the percentage savings in net system costs. We are interested in characterizing the combinations of τ and δ, representing different product types and supply chain scenarios, for which the ES policy is preferable to the BMA policy. Figure 3 shows a *policy map*, indicating the best policy for different values of τ and δ. To develop this map, we considered values of τ and δ in increments of 0.1. For each (τ, δ) pair, we evaluated the optimal smoothing value and total system cost for the ES policy and BMA policy with different W values, and chose the best policy.

Figure 3 Best Policy Over τ-δ Supply Chain Scenarios



In Figure 3, the diagonal line extending from the origin to the upper-right corner of the policy map demarcates the boundary of the region in which the ES policy is better than the BMA policy. For any (τ, δ) falling below the boundary (i.e., in the lower right portion of the map), the ES policy provides greater system cost savings than the BMA policy, and vice versa. Moreover, as we might expect, starting with the optimal TMA policy (with $W = 2$) at $\tau = 0$, the optimal smoothing window size W increases as the service capacity costs (i.e., the τ index) increases. Observe that for a fixed value of τ, when δ is sufficiently large, the s-inventory cost dominates, thus driving the first-period smoothing parameter α to zero; further increases to δ will not affect the choice of W. For a fixed τ, as δ decreases, the r-inventory (retailer) cost as a percentage of net system cost increases. Because r-inventory costs decrease with W, as we decrease δ, the BMA policy achieves better performance with smaller window sizes, causing the rightwards curvature in the region boundaries for successive W values at low δ values.

Let us now examine the practical contexts that might lead to different relative values of τ and δ. As we noted in §4.3, because δ measures the ratio of s-inventory holding and expediting costs to r-inventory costs, high δ values correspond to systems with inflexible supply (from manufacturer to supplier), whereas low δ represents responsive supply chains. Similarly, relatively high service capacity costs and outsourcing costs imply high values of τ. Items with low value-to-weight ratios, for example, are service-intensive products with large τ values. As Figure 3 illustrates, the ES policy is appealing for service-intensive products with responsive supply, whereas the TMA or more general BMA policy is more effective for supply chain scenarios with inflexible supply and high value-to-weight ratios. We later introduce another dimension, risk pooling, through demand aggregation and capacity sharing in multiproduct and multiretailer systems, to highlight product and supply chain characteristics where order smoothing can be effective.

We next examine the behavior of percentage savings for the best among the ES and BMA policies as τ and δ vary. Let $\omega^* \overset{\Delta}{=} \max\{\omega^E, \omega^B\}$ denote the percentage reduction in net system cost for the best policy.

Figure 4 **Percentage Savings from Best Order-Smoothing Policy**

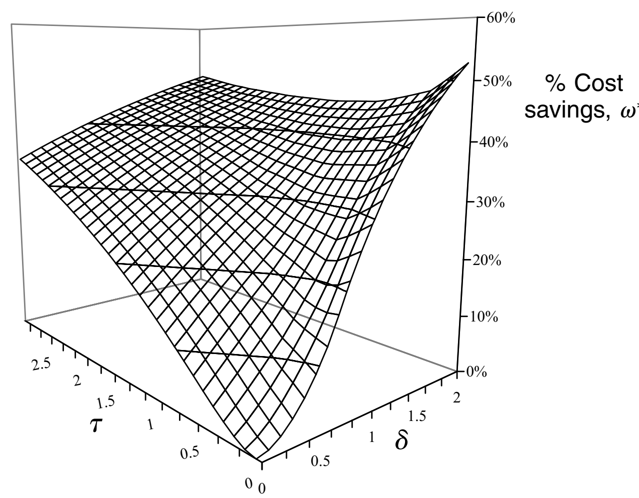


Figure 4 shows the magnitude and variation of ω^* as a function of τ and δ. We see that the savings from order smoothing increases quickly in the values of τ and δ, and order smoothing has the potential to significantly reduce supply chain costs over a wide range of scenarios. For example, when $\tau = \delta = 1$ (i.e., if s-inventory, r-inventory, and capacity costs are equal in the baseline case), order smoothing reduces the net system cost by 29.5%. For this specific example, the BMA policy with a four-period window size is best, and this policy uses a first-period smoothing value of $\alpha = 0.205$, indicating substantial smoothing. Even with $\tau = \delta = 1/2$, the best policy (ES) selects $\alpha = 0.465$ and provides savings of 15.4%. These examples demonstrate the benefit of selecting the correct level of order smoothing, rather than selecting extreme cases of zero or full smoothing.

Next, we show that introducing some degree of variability dampening always reduces total system cost. The baseline (i.e., no smoothing) policy is a special case of both the ES and BMA policies with $\alpha = 1$. Therefore, if the optimal value of α for either the ES or BMA policy is less than one, the total system cost under this policy must be less than the baseline cost, and variability dampening is effective. For the ES policy, if $\delta = 0$, the optimal smoothing parameter value from Expression (4.13) is $\alpha^E_{\text{opt}}|_{\delta=0} = 1/(1+\tau)$. As we increase δ from zero (keeping τ fixed), the increasingly significant s-inventory cost causes α to progressively decrease because the s-inventory cost is proportional to α. Therefore, for any τ and δ, $\hat{\alpha} = 1/(1+\tau)$ is an *upper bound* on the optimal value of α for the ES policy. Because $\hat{\alpha} < 1$ for $\tau > 0$, the optimal ES policy always favors using some degree of order smoothing rather than setting $\alpha = 1$. In Appendix C, we improve the upper bound on the optimal smoothing value for the ES policy (by developing an alternate bound in terms of δ), and show that the BMA policy also selects smoothing parameter values less than one for all scenarios. Thus, some order smoothing is always cost effective.

5. Order Smoothing with Multiple Products and Retailers

For expositional clarity, we have thus far focused on a single-product, single-retailer supply chain setting. In this section, we demonstrate how the previous analysis and insights extend readily to multiproduct, multiretailer contexts. As we noted in §2, order smoothing provides *temporal risk pooling* of consumer demands. In practice, two other risk-pooling opportunities are often available, demand aggregation pooling and shipment consolidation pooling. In distribution contexts, for example, demand aggregation pooling occurs when a supplier serves multiple retailers. The reduced *coefficient of variation* (COV) of the pooled retailers' orders relative to the COV of consumer demand leads to inventory cost savings for the supplier. Because this pooling effect reduces inventory costs, we refer to it as *inventory pooling*. Shipment consolidation pooling occurs when each retailer orders multiple products from the supplier, and these products are jointly shipped; the lower COV of the shipment size compared to the COV of individual product orders leads to lower variability-dependent capacity cost. Because this type of pooling occurs when multiple products share the same capacity, we refer to it as *capacity pooling*.

By considering systems with multiple products and retailers, we next investigate the effectiveness of order smoothing (with its temporal pooling) in scenarios where the supply chain already benefits from existing inventory and capacity pooling. To provide intuition and generate insights on different risk-pooling options, we first consider a system with multiple *identical* retailers and products, and later extend the analysis to nonidentical retailers and products. Our discussions assume that the supplier and retailers agree on a common smoothing policy for all products. This assumption permits us to develop insights on the interaction between order smoothing and other pooling effects, and is reasonable if retailers and products are relatively similar.

5.1. Identical Products and Retailers

Consider a supply chain in which a supplier carries m products and serves n retailers. We initially assume that all retailers and products are identical, i.e., the consumer demands for each product at every retailer are i.i.d. normal distributions with mean μ and standard deviation σ, and all retailers have the same cost structure. In each period, the supplier consolidates the replenishment quantities for all products from a retailer into a single shipment to that retailer. Let C_Y^0, for $Y = S, Q$, and R, denote the baseline s-inventory, capacity, and r-inventory costs for the single-product, single-retailer system. With multiple products and retailers, under the baseline policy, each retailer's r-inventory cost for every product is C_R^0, so the total systemwide r-inventory cost is mnC_R^0. For each product, the distribution of total replenishment orders (which is the same as net orders under the baseline policy) from all retailers has mean $n\mu$ and standard deviation $\sqrt{n}\sigma$. Thus, the baseline s-inventory cost for *each* product equals $\sqrt{n}C_S^0$. For each retailer, the total delivery quantity, with no smoothing, has mean $m\mu$ and standard deviation $\sqrt{m}\sigma$. Therefore, the baseline service capacity cost equals $\sqrt{m}C_Q^0$ per retailer. Letting τ_{mn} and δ_{mn} denote the capacity cost index and s-inventory cost index for an m-product, n-retailer system, we have

$$\delta_{mn} = \frac{m\sqrt{n}C_S^0}{mnC_R^0} = \frac{\delta_{11}}{\sqrt{n}}, \quad \text{and}$$

$$\tau_{mn} = \frac{n\sqrt{m}C_Q^0}{mnC_R^0} = \frac{\tau_{11}}{\sqrt{m}}, \tag{5.1}$$

and

$$TC^P = \left(\theta_R^P + \delta_{mn}\theta_S^P + \tau_{mn}\theta_Q^P\right)mnC_R^0. \tag{5.2}$$

Note that the variability multipliers θ_R^P, θ_S^P, and θ_S^P are the same as those for the single-product, single-retailer case (shown in Table 1).

Expression (5.1) shows that increasing the number of retailers from one to n reduces the s-inventory cost index by a factor of $\sqrt{n}$, and increasing the number of products per retailer reduces the capacity cost index by a factor of $\sqrt{m}$.[5] These adjustments to δ and τ illustrate the natural risk-pooling effects of multiple retailers and multiple products in reducing upstream inventory cost. Given these adjusted values of δ and τ, because the cost function (5.2) has the same structure as before, our previous results on smoothing effects and percentage savings (e.g., Expression (4.11)) still apply. From Figure 4, we see that dividing δ by $\sqrt{n}$ for $n \geq 2$ and dividing τ by $\sqrt{m}$ implies a (τ, δ) pair closer to the origin. If the supply chain already provides substantial inventory and capacity pooling, with large n and m, the figure suggests that the savings from order smoothing are substantially reduced. Thus, suppliers of narrow product lines that serve a small number of retail outlets—as might be the case with specialty brand appliances or upscale electronics products—clearly benefit most from order smoothing.

5.2. Nonidentical Products and Retailers

We now extend the previous analysis by relaxing the assumption that all retailers and products are identical. Let i and j denote retailer and product indices for a set of n retailers and m products. Denote the standard deviation of weekly demand for product j at retailer i as σ_{ij}; we assume for convenience that, for any product, the unit holding and backorder costs are the same at all retailers. We define the *standardized* r-inventory and s-inventory cost coefficients for product j as

$$\beta_S^j = h_j^S(z(\rho_s) + L(\rho_S)) + e_j L(\rho_S) \quad \text{and}$$
$$\beta_R^j = h_j^R(z(\rho_R) + L(\rho_R)) + p_j L(\rho_R).$$

These β_Y^j parameters correspond to the coefficients of the σ_Y values in Equations (4.2) and (4.4), with the added product-specific superscript j; they represent the expected retailer and supplier inventory cost per unit variation of net orders and effective demand, respectively. Under the baseline policy, the expected inventory-related cost for SKU j at retailer i equals $\beta_R^j \sigma_{ij}$, while the supplier's expected inventory cost for product j equals $\beta_S^j \sqrt{\sum_{i=1}^n \sigma_{ij}^2}$, where the square-root of the sum of variances reflects the demand-pooling benefits for the supplier's inventory decisions. Let us now examine the service capacity costs. By selecting the unit for measuring demand for each product so that the same service capacity is needed for one unit of any product, we can assume without loss of generality that the capacity cost parameters—fixed, variable, and overflow costs—do not depend on the product. However, we permit these parameters to vary by retailer. Letting ρ_Q^i denote the critical capacity fractile for retailer i, we define

$$\beta_Q^i \triangleq fz(\rho_Q^i) + (g_i - v_i)L(\rho_Q^i)$$

as the *standardized* capacity coefficient for retailer i (β_Q^i is the coefficient of σ_Q in Equation (4.3), with the retailer superscript i). We can thus express the expected variability-dependent transportation (service capacity) costs for supplying retailer i under the baseline policy as $\beta_Q^i \sqrt{\sum_{j=1}^m \sigma_{ij}^2}$; again, the square-root function accounts for the pooling due to capacity sharing.

Hence, the net system cost under the *baseline* policy is

$$TC^0 = \sum_{j=1}^m \beta_R^j \left(\sum_{i=1}^n \sigma_{ij} \right) + \sum_{j=1}^m \beta_S^j \sqrt{\sum_{i=1}^n \sigma_{ij}^2} + \sum_{i=1}^n \beta_Q^i \sqrt{\sum_{j=1}^m \sigma_{ij}^2}. \tag{5.3}$$

As before, we define the capacity cost index τ and the s-inventory cost index δ as the ratios of baseline service capacity and s-inventory costs, respectively, to the r-inventory cost, i.e.,

$$\tau = \frac{\sum_{i=1}^n \beta_Q^i \sqrt{\sum_{j=1}^m \sigma_{ij}^2}}{\sum_{j=1}^m \beta_R^j (\sum_{i=1}^n \sigma_{ij})} \quad \text{and}$$
$$\delta = \frac{\sum_{j=1}^m \beta_S^j \sqrt{\sum_{i=1}^n \sigma_{ij}^2}}{\sum_{j=1}^m \beta_R^j (\sum_{i=1}^n \sigma_{ij})}.$$

We can now express the net system cost for any smoothing policy P in the same form as for the

[5] Combining the deliveries to multiple retailers into delivery routes permits additional capacity pooling, thus reducing the capacity cost index further. Our model can readily incorporate these capacity-pooling features.

single-item, single-retailer case, i.e., $TC^P = \hat{\beta}_R(\theta_R^P + \theta_S^P\delta + \theta_Q^P\tau)$, where $\hat{\beta}_R \triangleq \sum_{j=1}^m \beta_R^j(\sum_{i=1}^n \sigma_{ij})$. Remarkably, by structuring the analysis in this way, we can now use our previous analysis of percentage savings (ω^*) values (refer to Equation (4.11) and Figure 4) to determine the variability-dependent cost savings in a general multi-item, multiretailer context.

5.3. Illustrative Application of Order Smoothing

To assess the potential supply chain savings due to order smoothing in practice, we consider application of the model to a distribution system for specialty brands of "white" appliances (e.g., dishwashers and washing machines). This setting has several features that make it an attractive context for using order smoothing. First, the value-to-weight ratio for appliances is relatively low, and so capacity costs are significant relative to inventory holding costs. For instance, a typical (large) tractor trailer can accommodate only around 100 units of dishwashers and washing machines, so that the per-unit fixed and variable capacity costs are both relatively high. Second, large appliances, especially specialty brands, are typically stored in a limited number of dealerships (or retail distribution centers), and are delivered to customer premises directly from the dealership (even for standard appliances, retailers often maintain only display units at store locations). Because these dealers represent the "retailers" in our model, their limited number precludes significant inventory pooling for the supplier. Third, specialty brand appliance manufacturers offer a limited product line, and so capacity sharing opportunities are also modest.

To apply our order-smoothing model, we generated representative values for various supply chain parameters for the specialty brand appliance distribution setting using publicly available data. We consider a situation in which the manufacturer's distribution center supplies 5 different products to 10 authorized dealers. Using statistics on annual domestic appliance sales and assuming that dealers have similar sales volumes, we estimated the weekly demand for each product and dealership. Based on standard costs for transportation resources (e.g., around \$800 per week to lease a tractor trailer, \$20 per hour for the driver, and \$0.33 per mile operating cost), published third-party carrier rates, and assumed driving distances from the supplier to the dealerships, we estimated the fixed, variable, and overflow costs per appliance. For inventory costs, we assumed a holding cost rate of 20% of the product cost (\$400 to \$500) for both the supplier and dealer.

For this system, we wish to determine the extent to which order smoothing can reduce distribution operations costs relative to uncoordinated decision making without smoothing. For our computational experiments, we consider a base-case scenario with an estimated (moderate) COV of 1/3 in weekly demand for each product at each dealership, and service levels of 90% for the dealers ($\rho_R = 0.90$) and 95% for the supplier ($\rho_S = 0.95$). The supplier's higher service-level value reflects high costs for expediting. From these service-level values, we can infer the dealers' backlogging cost and supplier's expediting costs, and compute the relative cost indices τ and δ. The base-case parameter values imply a capacity cost index τ of 1.23 and an s-inventory cost index δ of 0.37. Due to the relatively high value of τ and the low value of δ, the ES policy is superior to the BMA policy, and its optimal smoothing parameter value α^* is 0.37. This smoothing policy generates a savings of $\omega = 22.7\%$ in net system costs. If we measure savings relative to the baseline policy's *total* system cost, obtained by adding the variability-independent cost of service capacity (see §4.2) to the baseline net system cost, the resulting *total cost savings*, which we denote as Ω, equals 7.8%. Thus, we find that order smoothing can provide notable supply chain savings for this context.

Next, we consider the sensitivity of our results to changes in the key base-case parameter values.[6] To explore the impact of capacity and inventory pooling on the savings from order smoothing, we increased the number of products and authorized dealers by around 50% (specifically, we increased the number of products from 5 to 8, and the number of authorized dealers from 10 to 15). When the number of products increases to eight, τ decreases by 0.26; consequently, the degree of smoothing decreases (α^* increases to 0.41), and the total cost savings Ω reduces by 1.5% to 6.3%. Increasing the number of dealers by 50%

[6] For all the new cases, the ES policy continued to outperform the BMA policy.

reduces δ by 0.07 and Ω to 7.3%. When we simultaneously increase both of these factors (creating a 15-dealer, 8-product network), Ω is 5.8%.

We next consider how the dealer and supplier service levels impact the potential savings from order smoothing. As we increase the supplier's service level from 0.95 to 0.975, e.g., when expediting costs are high, the degree of smoothing increases (α^* decreases to 0.36) and total cost savings Ω increases by 0.4% to 8.2%. On the other hand, increasing the dealers' service level to 0.95 to reflect higher backorder costs reduces δ by 0.06 and τ by 0.19, thereby decreasing Ω to 7.0%. Finally, increasing the demand COV to 0.5 provides a substantial increase in total cost savings to 10.1%, while reducing COV to 0.25 lowers total cost savings to 6.5%.

Thus, the appliance distribution example illustrates the potential benefits of order smoothing in supply chains with somewhat limited inventory- and capacity-pooling opportunities. Moreover, our computations show that smoothing benefits are relatively robust to parameter variations.

6. Conclusions

Operations managers have long recognized the detrimental effects of supply chain demand variability. Variability imposes significant costs on upstream supply stages because a supplier must not only carry safety-stock inventory to buffer against demand uncertainties, but also (typically) must invest in service capacity to fulfill retail orders. In this paper, we have proposed a variability-centric viewpoint for coordinating supply chains. Order smoothing by downstream stages serves as an effective mechanism to control and dampen variability. This approach not only yields temporal risk-pooling benefits, but also reduces the supplier's effective order uncertainty by providing advance order information. By tuning the smoothing parameters, the system can achieve the "right" level of dampening to reduce upstream inventory and capacity costs while controlling downstream inventory cost increases. As our results have shown, with an intermediate level of smoothing the total system cost decreases relative to independent, autonomous decisions at each stage.

Our analysis has also provided new insights on understanding supply chain cost structures and their drivers. We have shown how the three cost components depend on distinct facets of the supply chain system's variability: Downstream inventory costs depend on the retailer's *effective demand*, upstream inventory costs depend on the *net orders* observed by the supplier, and service-capacity cost relates to the *total order* variability. The two key relative (supplier) cost indices δ and τ that we developed, together with the variance-dampening characteristics reflected in the policy variability multipliers, determine the best policy and smoothing level for any supply chain setting. Our policy map provides a succinct guide for managers to decide the best strategy for their supply chain. Although we initially focused on a single-retailer, single-item scenario, our approach extends seamlessly to more general multiretailer, multi-item contexts. In exploiting the extensibility of this modeling paradigm, we have leveraged on our assumptions that consumer demand follows a normal demand distribution and that the supplier expedites (at extra cost) shipments from the manufacturer to satisfy retailer orders. Applying the model to representative data for a specialty brand appliance product distribution setting demonstrates that order smoothing can be effective in systems with modest amounts of demand and capacity pooling. For our analysis of order smoothing and assessment of its benefits, we consider contexts in which retailers focus on managing their inventories, while upstream stages manage and incur the costs for service capacity and warehouse inventories. Consequently, retailers prefer to propagate the full demand variability upstream, whereas upstream stages prefer low variability in order quantities, creating a tension between the supply chain partners. In supply chain contexts where the retailer also manages service capacity, this tension will diminish (because the retailer's resulting autonomous policy might propagate less variability upstream than the simple base-stock policy), thereby reducing, but not eliminating, the benefits of order smoothing.

Because order smoothing can benefit all upstream stages in a supply chain, extending the modeling approach in this paper beyond two-tier systems represents an interesting avenue for future work. Tackling this problem extension requires characterizing the exact linkages between the relevant costs and sources of variability further upstream in the supply chain. Future work might also address alternative policies for dampening variability propagation.

Acknowledgments
The authors thank the editors and reviewers for their insightful comments and suggestions that have led to improvements in the paper. This work was partially supported by NSF Grant DMI-0115304 and a grant from the Institute for the Study of Business Markets (ISBM) at Penn State University.

Appendix A

Proof of Proposition 2. Suppose a given smoothing policy does *not* satisfy the nonincreasing condition $\alpha_k \geq \alpha_{k+1}$ for $k \geq 2$. We will show that, by interchanging two of the smoothing coefficients, we can achieve a lower total cost, contradicting the optimality of the given policy. For the given policy, let l' and l'' be the smallest indices such that $l'' > l' > 2$ and $\alpha_{l''} > \alpha_{l'}$, and consider the new policy obtained by setting $\hat{\alpha}_{l'} = \alpha_{l''}$ and $\hat{\alpha}_{l''} = \alpha_{l'}$ (and keeping all other smoothing coefficients the same). Because $\sigma_S = \alpha_1 \sigma$, the s-inventory cost is unaffected by α_k values for $k \geq 2$, and hence by the coefficient-interchange operation. Furthermore, because $\sigma_Q = \sigma\sqrt{\sum_{k=1}^{\infty}\alpha_k^2}$, interchanging the coefficients does not change the total-order variability, and hence the supplier's capacity cost. Therefore, the interchange operation reduces total system cost if the standard deviation σ_R of effective demand is lower for the new policy. Recall that, for the general linear smoothing policy, the proportion of period $(t-k)$'s unfilled demand at period t is $u_k = 1 - \sum_{l=1}^{k}\alpha_l$. For $1 \leq k < l'$ and $k \geq l''$, the unfilled proportion u_k remains unaffected by the coefficient interchange operation. For $l' \leq k < l''$, let $\hat{u}_k$ be the unfilled proportion for the new policy (after the coefficient interchange). Because $\hat{\alpha}_{l'} = \alpha_{l''} > \alpha_{l'} = \hat{\alpha}_{l''}$, we have

$$\hat{u}_k = 1 - \sum_{l=1,\, l\neq l'}^{k}\alpha_l - \hat{\alpha}_{l'} < 1 - \sum_{l=1}^{k}\alpha_l = u_k \quad \text{for all} \quad l' \leq k < l''.$$

Therefore, because $\sigma_R^2 = \sigma^2(1 + \sum_{k=1}^{\infty} u_k^2)$, the interchange reduces the standard deviation of effective demand. □

Appendix B

Proof of Proposition 4. We show that for any given target capacity cost, the ES policy achieves lower retailer inventory cost than the BMA policy for all W. Thus, given the optimal value of target capacity cost, if s-inventory costs are variability independent, the ES policy must outperform the BMA policy for all W.

Let us define $\Psi = \sigma_Q^2/\sigma^2$. Then, for the ES model, $\Psi = \alpha/(2-\alpha)$, $\alpha = 2\Psi/(\Psi+1)$, and $(\sigma_R^E)^2/\sigma^2 = (\Psi+1)^2/4\Psi$. Similarly, for the BMA policy, $\Psi = \alpha^2 + (1-\alpha)^2/(W-1)$,

$$\alpha_{B(W)} = \frac{1+\sqrt{(W-1)(\Psi W-1)}}{W},$$

and

$$\frac{\left(\sigma_R^{B(W)}\right)^2}{\sigma^2} = 1 + \left(W + W\Psi - 2 - 2\sqrt{(W-1)(W\Psi-1)}\right)\frac{2W-1}{6W}.$$

The ES policy will yield lower retailer costs than the BMA if, for all W and $\Psi \geq 1/W$, the inequality $(\sigma_R^E)^2 \leq (\sigma_R^{B(W)})^2$ holds, i.e., if

$$\frac{(\Psi+1)^2}{4\Psi} \leq 1 + \left(W + W\Psi - 2 - 2\sqrt{(W-1)(W\Psi-1)}\right)\frac{2W-1}{6W}.$$

After rearranging terms and simplifying, this inequality reduces to:

$$\Psi^2\left(\frac{2(2W-1)}{3} - 1\right) + \Psi\left(2 + \frac{2(2W-1)}{3} - \frac{4(2W-1)}{3W} \cdot \left(1 + \sqrt{(W-1)(W\Psi-1)}\right)\right) \geq 1. \quad \text{(B1)}$$

We establish Condition (B1) by showing that the minimum value of the left-hand side (LHS) function is one, achieved at $\Psi = 1$. The second derivative of the LHS function is

$$\frac{4(2W-1)}{3} - 2 + \frac{2W-1}{3}\sqrt{\frac{W-1}{W\Psi-1}}\left(\frac{W\Psi}{W\Psi-1} - 4\right). \quad \text{(B2)}$$

For $W \geq 2$, $4(2W-1)/3 - 2$ is greater than or equal to 2. The second term in (B2) is increasing in W for any Ψ, and decreasing in Ψ for any W; because the term is equal to -2 for $W = 2$ and $\Psi = 1$, the second derivative (B2) must therefore be nonnegative for all $W \geq 2$ and all $\Psi \in [1/W, 1]$. Hence, the LHS function is convex in Ψ. To find its minimum value, we solve the first-order condition

$$2\left(\frac{2(2W-1)}{3} - 1\right)\Psi + 2 + \frac{2(2W-1)}{3} - \frac{4(2W-1)}{3W} \cdot \left(1 + \sqrt{(W-1)(W\Psi-1)}\right) - \frac{2\Psi(2W-1)(W-1)}{3\sqrt{(W-1)(W\Psi-1)}} = 0. \quad \text{(B3)}$$

The value $\Psi = 1$ satisfies this condition, giving a minimum LHS function value of one. □

Appendix C. Upper Bounds on First-Period Smoothing Parameter Value α

For the ES policy, §4.5 provided the upper bound $\hat{\alpha} = 1/(1+\tau)$ on the first-period smoothing parameter value by considering the extreme case with $\delta = 0$. Similarly, if $\tau = 0$, the net system cost $TC^E = \delta\alpha + 1/\sqrt{\alpha(2-\alpha)}$ is minimized at $\hat{\alpha}(\delta) = \sqrt{1-\hat{\beta}}$, where $\hat{\beta}$ is the solution to the cubic equation $\delta^2(1-\beta)^3 = \beta$. Because increasing τ should progressively decrease α, the value $\hat{\alpha}(\delta)$ is also an upper bound on the optimal value of α for the ES policy. Thus, $\alpha^*_{\text{ES}} \leq \min\{1/(1+\tau),\ \hat{\alpha}(\delta)\}$ for all τ and δ.

For the BMA policy with a given window size W, first consider the special case with $\tau = 0$. The net system cost $TC^{B(W)} = \delta\alpha + \sqrt{1+(1-\alpha)^2 w}$, with $w = W(2W-1)/(6(W-1))$, is convex in α. Solving the first-order condition $dTC^{B(W)}/d\alpha = 0$, we obtain the optimal smoothing parameter value $\alpha^{B(W)}_{\text{opt}}|_{\tau=0} = 1 - \delta/\sqrt{w(w-\delta^2)}$ for the BMA policy. Starting with $\tau = 0$, suppose we progressively increase τ. Because the supplier's capacity cost prefers to have equal weights for a fixed W, increasing τ should progressively decrease α if $\alpha^{B(W)}_{\text{opt}}|_{\tau=0} > 1/W$. Otherwise, if $\alpha^{B(W)}_{\text{opt}}|_{\tau=0} < 1/W$, α increases to a maximum of $1/W$. Hence, $\hat{\alpha}_{B(W)} = \max\{1 - \delta/\sqrt{w(w-\delta^2)}, 1/W\}$ is an upper bound on α for the BMA policy with window size W, for all τ and δ.

References

Anderson, E. G., C. H. Fine, G. G. Parker. 2000. Upstream volatility in the supply chain: The machine tool industry as a case study. *Production Oper. Management* **9**(3) 239–261.

Baganha, M. P., M. A. Cohen. 1998. The stabilizing effect of inventory in supply chains. *Oper. Res.* **S46**(3) S72–S83.

Balakrishnan, A., J. Geunes, M. S. Pangburn. 2002. A constrained replenishment model for upstream variability control in two-stage distribution systems. Working paper, Industrial and Systems Engineering Department, University of Florida, Gainesville, FL.

Balakrishnan, A., J. Geunes, M. S. Pangburn. 2003. Impact of variability control agreements on distribution chain performance. Working paper, Industrial and Systems Engineering Department, University of Florida, Gainesville, FL.

Bassok, Y., R. Anupindi. 1997. Analysis of supply contracts with total minimum commitment. *IIE Trans.* **29**(5) 373–381.

Bourland, K. E., S. G. Powell, D. F. Pyke. 1996. Exploiting timely demand information to reduce inventories. *Eur. J. Oper. Res.* **92** 239–253.

Cachon, G. P. 1999. Managing supply chain demand variability with scheduled ordering policies. *Management Sci.* **45**(6) 843–856.

Cachon, G. P., M. Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Sci.* **46**(8) 1032–1048.

Cachon, G. P., P. H. Zipkin. 1999. Competitive and cooperative inventory policies in a two-stage supply chain. *Management Sci.* **45**(7) 936–953.

Chen, F. 1999. Decentralized supply chains subject to information delays. *Management Sci.* **45**(8) 1076–1090.

Chen, F. 2001. Market segmentation, advanced demand information, and supply chain performance. *Manufacturing Service Oper. Management* **3**(1) 53–67.

Chen, F., Z. Drezner, J. K. Ryan, D. Simchi-Levi. 2000. Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Sci.* **46**(3) 436–443.

Cohen, M. A., H. L. Lee. 1988. Strategic analysis of integrated production-distribution systems: Models and methods. *Oper. Res.* **36**(2) 216–228.

Corbett, C. J., G. A. Decroix. 2001. Shared-savings contracts for indirect materials in supply chains: Channel profits and environmental impacts. *Management Sci.* **47**(7) 881–893.

Cruickshanks, A. B., R. D. Drescher, S. C. Graves. 1984. A study of production smoothing in a job shop environment. *Management Sci.* **30**(3) 368–380.

Dejonckheere, J., S. M. Disney, M. R. Lambrecht, D. R. Towill. 2002. Transfer function analysis of forecasting induced bullwhip in supply chains. *Internat. J. Production Econom.* **78** 133–144.

Dejonckheere, J., S. M. Disney, M. R. Lambrecht, D. R. Towill. 2003. Measuring and avoiding the bullwhip effect: A control theoretic approach. *Eur. J. Oper. Res.* **147** 567–590.

Ernst, R., D. F. Pyke. 1993. Optimal base stock policies and truck capacity in a two-echelon system. *Naval Res. Logist.* **40** 879–904.

Gallego, G., Ö. Özer. 2001. Integrating replenishment decisions with advance demand information. *Management Sci.* **47**(10) 1344–1360.

Gavirneni, S., R. Kapuscinski, S. Tayur. 1999. Value of information in capacitated supply chains. *Management Sci.* **45**(1) 16–24.

Graves, S. C., S. P. Willems. 2000. Optimizing strategic safety stock placement in supply chains. *Manufacturing Service Oper. Management* **2**(1) 68–83.

Graves, S. C., D. B. Kletter, W. B. Hetzel. 1998. A dynamic model for requirements planning with application to supply chain optimization. *Oper. Res.* **46**(S3) S35–S49.

Hariharan, R., P. H. Zipkin. 1995. Customer-order information, leadtimes, and inventories. *Management Sci.* **41**(10) 1599–1607.

Henig, M., Y. Gerchak, R. Ernst, D. F. Pyke. 1997. An inventory model embedded in designing a supply contract. *Management Sci.* **43**(2) 184–189.

Lee, H. L., S. Whang. 1999. Decentralized multi-echelon supply chains: Incentives and information. *Management Sci.* **45**(5) 633–640.

Lee, H. L., V. Padmanabhan, S. Whang. 1997. Information distortion in a supply chain: The bullwhip effect. *Management Sci.* **43**(4) 546–558.

Lee, H. L., K. C. So, C. S. Tang. 2000. The value of information sharing in a two-level supply chain. *Management Sci.* **46**(5) 626–643.

Moinzadeh, K., S. Nahmias. 2000. Adjustment strategies for a fixed delivery contract. *Oper. Res.* **48**(3) 408–423.

Nahmias, S. 2001. *Production and Operations Analysis*, 4th ed. McGraw-Hill, Chicago, IL.

Porteus, E. L. 1990. Stochastic inventory theory. D. P. Heyman, M. Sobel, eds. *Handbooks in OR & MS*, Vol. 2. North-Holland, Elsevier Publishers B.V., New York.

Stalk, G., P. Evans, L. E. Shulman. 1992. Competing on capabilities: The new rules of corporate strategy. *Harvard Bus. Rev.* **70**(2) 57–69.

Sterman, J. D. 1989. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Sci.* **35**(3) 321–339.

Tsay, A. A., S. Nahmias, N. Agrawal. 1999. Modeling supply chain contracts: A review, Chapter 10. S. Tayur, R. Ganeshan, M. Magazine, eds. *Quantitative Models for Supply Chain Management.* Kluwer Academic Publishers, Boston, MA.