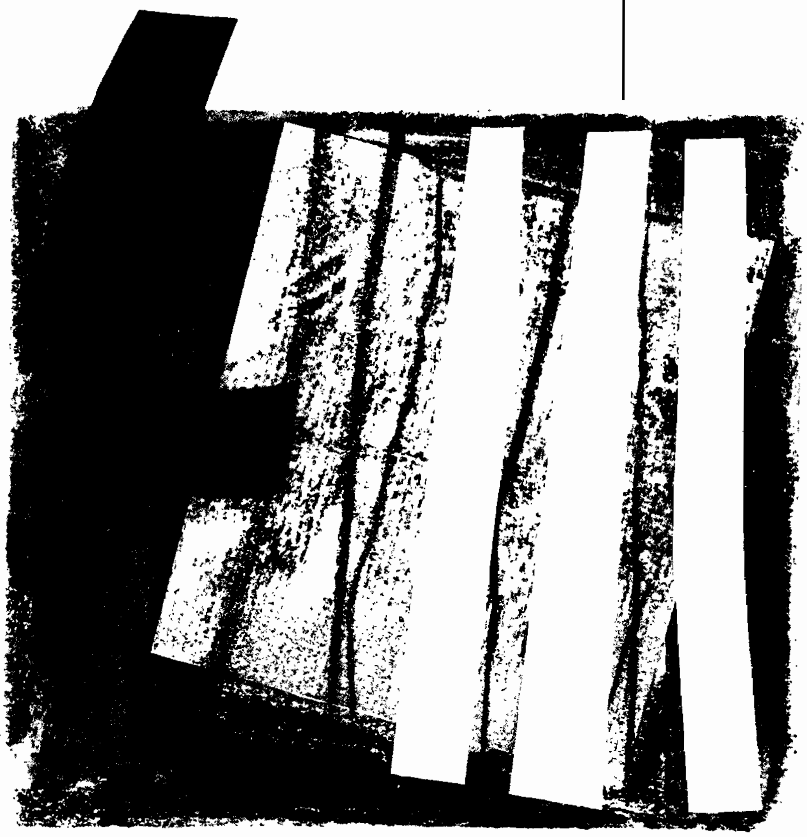
THE VALUE OF SKU RATIONALIZATION: THE POOLING EFFECT UNDER SUBOPTIMAL INVENTORY POLICIES

José A. Alfaro and Charles J. Corbett

00-34



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 00-34
Business Economics Series 7
May 2000

Departamento de Economía de la Empresa
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 6249608

THE VALUE OF SKU RATIONALIZATION: THE POOLING EFFECT UNDER SUBOPTIMAL INVENTORY POLICIES

Jose A. Alfaro[1] and Charles J. Corbett [2]

Abstract

Managing product variety is a widely recognized challenge. Several approaches to this rely on the "pooling effect", the reduction of uncertainty that occurs when individual demands are aggregated. This can occur through reduction of number of products or SKUs, through postponement of differentiation, or in other ways. These approaches are by now well-known and widely applied in practice. However, theoretical analyses of the pooling effect always assume that one has an optimal inventory policy before and after pooling. If this is not the case, how does that affect the value of pooling? This paper analyses the benefits of pooling in terms of costs and service level under optimal and suboptimal policies and proposes a simple framework to analyze the trade-off between implementing pooling and improving inventory policy. We show there is always a range of current inventory levels within which pooling is better and beyond which optimizing inventory policy is better. We analyze how this range varies with the problem parameters and illustrate these findings using highly erratic empirical demand data.

Keywords: pooling, SKU rationalization, product variety, suboptimal inventory policy, service level, fill rate.

[1]Universidad Carlos III de Madrid. Dept. Economía de la Empresa, C/ Madrid, 126, 28903 Getafe (Madrid), Spain. Phone: 34-91-624.95.73. E-mail: jaalfaro@emp.uc3m.es.
[2] Anderson Graduate School of Managament. The University of California

1. INTRODUCTION

The complexity of supply chains can be quite staggering, entailing high costs combined with incessant customer service problems. Increasing variety is a critical ingredient of this complexity, and managing this variety is becoming a field of research in its own right, as witnessed by (among others) the recent book edited by Ho and Tang (1998). Roughly speaking, variety can be induced by having to serve multiple geographic locations, or stocking multiple products or SKUs (stock keeping units). A growing body of literature exists on how to deal with both of these, sometimes focusing on reducing lead times, sometimes on various types of aggregation of uncertainty. This in turn can occur through geographical "pooling" of inventories or through some form of rationalization of product lines. In both cases, the general mechanism underlying pooling or SKU rationalization is the ability to exploit "statistical economies of scale" by aggregating uncertainties, which reduces the total uncertainty one needs to deal with. The analysis and discussion here is presented in terms of SKU rationalization, but the arguments are highly similar in the case of geographical pooling; we use the term *pooling* to refer to both.

Sometimes, pooling inventories is easy to do and (almost) costless; in such situations, little discussion is needed, and companies should seek and exploit such opportunities. A typical example would be shipping kits with a single set of assembly instructions in multiple languages rather than have to distinguish between kits intended for different markets. Often, however, the costs involved are more significant. Some people will argue that reducing product variety, for instance by limiting the number of sizes in which toothpaste is sold, leads to marketing disadvantage. In other cases, rationalizing product lines may directly increase unit manufacturing costs, if multipurpose and more costly components or process changes are required. A well-known example of this is the HP DeskJet printer (Lee, Billington and Carter 1993): the redesigned product could be localized at DCs across the world, but final assembly of the power supply had to be made easier as it would no longer be performed at HP's central facility in Vancouver, leading to higher component and assembly costs. However, these added costs were easily outweighed by the inventory and customer service advantages HP gained. Lee and Tang (1997) develop models to assist in the trade-off between such delayed differentiation and increased manufacturing costs. In such cases, more careful evaluation of the benefits of pooling and of possible alternative strategies is needed. That is precisely the aim of this paper.

Typically, a company's interest in the various types of pooling available is sparked by excessive inventory costs and/or constant customer service problems. Although pooling should benefit such a company in most cases, these may merely be symptoms of poor inventory management to begin with. Indeed, inventory policies in practice are often suboptimal, sometimes dramatically so. Should such a company even consider pooling, or should they get their inventory management sorted out first? In this paper, we address this dilemma by asking and answering the following questions:

- Does having a suboptimal inventory policy increase or reduce the value of pooling?
- Should a company suffering high inventory costs and low service level start by improving its inventory policy, or by implementing pooling? Ideally, obviously, one

would do both, but given that practitioners are often hard-pressed to find enough time to even implement one of the two, it is important to know where to start.

- How can these questions be answered in a context where demand does not follow any recognizable distribution?

Much of the literature related to the pooling effect assumes either identical or independent demand distributions (or both). We use a simple reformulation of the general multivariate normal demand case from which the identically and/or independently distributed cases follow easily as special cases, thus allowing us to dispense with the i.i.d. assumption for part of our analysis. We show how the level of concentration of uncertainty among SKUs, defined later, is a key driver of the value of pooling.

In many supply chains, pooling is implemented as much in order to improve customer service as to reduce costs; therefore, we also evaluate the effects on fill rate and service level. Throughout, we focus explicitly on the inventory issues; benefits of pooling due to, e.g., manufacturing simplification can be (as) significant, but we do not consider those here. Section 2 reviews a selection of relevant literature on various types of pooling. Section 3 presents the framework and basic model for analysis. In Section 4 we characterize the value of pooling under optimal and suboptimal inventory policies and the value of improving the inventory policy before and after pooling. We compare these effects in Section 5 and find that although improving a suboptimal policy is always beneficial, there is a wide range of cases in which it may be better to focus on pooling instead. Section 6 provides a practical illustration of how to determine the value of pooling, using highly erratic but characteristic demand data from a chemical manufacturer. Section 7 contains our conclusions.

2. LITERATURE REVIEW

The number of stock keeping units (SKUs) is the standard measure of product variety in a company. Each of these SKUs may be stored in multiple locations, leading to geographical variety. Reducing either product variety or geographical variety will generally allow a company to maintain current customer service at lower cost (or improve service without incurring extra costs). The fundamental mechanism underlying both approaches is equivalent, even if there are significant differences between them. In both cases, pooling can be defined as a business strategy that consists of aggregating independent demands (Gerchak and Mossman, 1992). Below we first review the geographical context, where the concept of the pooling effect originated, then various approaches and models for SKU rationalization and postponement of differentiation.

Geographical variety

The concept of pooling originated with Eppen's (1979) and Eppen and Schrage's (1981) work on multi-echelon inventory systems with geographically dispersed stocking locations. A typical scenario includes a central depot which supplies N locations (or retailers or warehouses) where exogenous, random demands for a single commodity must be filled. The assumptions made there and in much subsequent work include that demand at each location i in period t is assumed to be independent and normally distributed over time; that all locations have identical linear holding and penalty costs; that the coefficient of variation is negligible so that the probability of negative demand can be ignored; that items are allocated to locations so as to equalize probability of stockout; that unsatisfied demand is backlogged; and that material is non-perishable. Eppen (1979) was the first to show how centralization of inventory could reduce expected costs in a multi-location single-period newsboy problem. He called this "statistical economies of scale", and concludes that the expected holding and penalty costs in a decentralized system exceed those in a centralized system. If demands are identical and uncorrelated, costs increase as the square root of the number of locations; as demand correlation increases, this effect is reduced. Schwarz (1989) defines this as the "risk pooling incentive" for centralizing inventories.

Eppen and Schrage (1981) develop a depot-warehouse system, in which the depot orders from the supplier and allocates products to the warehouse each period or every *m* periods. They show that the reduction of total inventory from pooling is greater as the depot's review period increases and is proportional to the number of products. Erkip et al. (1990) analyze the effects of correlation among products and among successive time periods, assuming the coefficient of variation is equal for all products. They conclude that high positive correlation (around 0.7) results in significantly higher safety stock than the no-correlation case. Schwarz (1989) focuses on lead times and defines the price of risk pooling as cost of the pipeline inventory caused by the internal lead-time. Jönsson and Silver (1987) present an exhaustive study of the impact of changing input parameters on system performance; the redistribution system is more advantageous in situations with high demand variability, a long planning horizon, many locations and short lead times. Jackson and Muckstadt (1989) show that allocating a centrally held inventory to retailers leads to a more balanced distribution of stock in the system. Tagaras (1992) shows that allowing transshipments between retailers leads to similar results as including a distribution center as in Eppen and Schrage's (1981) model. Gerchak and Mossman (1992) show how the order quantity and associated costs depend on the randomness parameter in a simple and highly interpretable manner. Federgruen and Zipkin (1984) extend Eppen and Schrage's (1981) model in three important ways: finite horizon, other-than-normal demand distributions (including exponential and gamma), and non-identical retailers. To summarize, the pooling effect yields a reduction in expected costs, but might increase inventory level.

Product variety

The benefits of delayed product differentiation are quite similar to those of the pooling effect in multi-echelon inventory systems. In fact, most work on postponement has drawn upon this body of research (Garg and Lee, 1999), referring to multiple products instead of multiple locations. Empirical research suggests that appropriate policies can facilitate absorption of higher levels of product variety without significant detrimental effects on costs (Kekre and Srinivasan, 1990 and McDuffie et al., 1996). Figure 1 displays the main strategies that have been developed to reduce the effects of SKU proliferation. Product variety strategies can be categorized into two broad classes (Garg and Lee, 1999): those that do not reduce total lead time but delay or eliminate product differentiation and thus reduce complexity, and those that reduce lead times and thus reduce uncertainty. Detailed surveys of these strategies can be found in Garg and Lee (1999) and Aviv and Federgruen (1999).

Early research studied standardization of components. Collier (1982) and Baker et al. (1986) designed models to minimize aggregate safety stock levels by using component standardization, subject to a service level constraint. Lately research has focused more on various types of postponement. Lee and Tang (1997) define two policies: while *modular design* refers to decomposing the complete product into submodules that can be easily assembled, *process sequencing* refers to resequencing process steps, either through operations reversal or postponement of an operation. Lee and Tang (1998) and Kapuscinski and Tayur (1999) show that reversing two consecutive stages of the process could lead to reduction in variance or standard deviation, thereby improving the performance of the process. Swaminathan and Tayur (1998) propose an integer programming approach to determining the optimal configurations and inventory levels for submodules.

Postponement of an operation consists of delaying it until a later stage in the process, thus delaying the point of differentiation (Aviv and Federgruen, 1999). Lee (1996) describes two models that capture the inventory reduction of delayed product differentiation, assuming that no buffer stocks are held until the end of production process. It shows the savings are greater when demands for the different products are negatively correlated. Lee and Tang (1997) extend this scenario in two ways: allowing for inventories at different points of the process and adding other factors that are normally affected by delayed product differentiation, such as lead times and design cost, processing cost and inventory cost at intermediate stages. Initially, all operations are independent for the two products, but with delayed differentiation the first k operations become common.

This is just a selection of the extensive theoretical literature on reducing product variety, showing how pooling can be implemented in different ways and can be beneficial in a wide range of cases. From a practical perspective, this literature suggests several questions. First, what is the impact of pooling on service level, rather than on inventory and cost? Tagaras (1992) is one of the few to explicitly address this. Second, if a company's inventory policy is not optimal to begin with, how does that affect the value of pooling? There are papers on practical applications of pooling (such as Lee et al. 1993), but this effect of suboptimal inventory policies on the value of pooling has not yet been studied.

3. FRAMEWORK AND BASIC MODEL

The literature on the various manifestations of the pooling effect makes the assumption that the underlying inventory policy is optimal, i.e. that it minimizes total cost. However, in practice, an "optimal" inventory policy can be very hard to determine, partly because the exact costs and demand distributions are not known with certainty. A practitioner, faced with high variety and a suboptimal inventory policy, may well ask how useful pooling will be in such a context. Figure 2 summarizes the situation: starting from the upper-left scenario 1, should the practitioner first improve his inventory policy (scenario 3), or should he first implement pooling (scenario 2)? Of course, scenario 4 is always the most desirable, but it is not clear which of the two paths (1-3-4 or 1-2-4) is preferable. Within this framework, one can ask the questions already posed in the introduction:

1) How does the value of pooling change when the company does not follow an optimal inventory policy, i.e. how does path (1-2) compare with (3-4)?

2) Starting in scenario 1, which is more desirable, pooling (scenario 2) or improved inventory policy (scenario 3)?

In this section we examine each of the four scenarios, largely following Eppen (1979). Assume the company initially produces N SKUs; pooling means consolidating all products into one, whose total demand is the sum of all the original SKUs. We keep the theoretical models relatively simple in order to focus on the qualitative insights; in Section 6 we investigate how one can apply the concepts studied here when most of these assumptions, notably those concerning demand pattern, are violated. Notation is standard and summarized in Table 1. We consider a single planning period. There are initially N SKUs, whose demands z_i follow a multivariate normal distribution $F(\mathbf{z}) \sim N(\mu, \Sigma)$. Write $\mu = \frac{1}{N}\sum_{i=1}^{N} \mu_i$ and $\sigma = \frac{1}{N}\sum_{i=1}^{N} \sigma_i$, the "average" mean and standard deviation, and define a vector $\boldsymbol{x}$ of x_i such that $\sigma_i = x_i \sigma$ for all i and $\sum_{i=1}^{N} x_i = N$. If $x_i = 1$ for all i, then all SKUs contribute equally to total uncertainty. Covariances are $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij} = x_i x_j \rho_{ij} \sigma^2$, so that $\Sigma = \sigma^2\, \boldsymbol{x'Rx}$ with $\boldsymbol{R}$ the correlation matrix. Assume the coefficients of variation of demand (σ_i / μ_i) are small so that the probability of negative demand can be ignored. Unsatisfied demand is backlogged. Correlation over time is negligible.

We analyze the effect of SKU rationalization on total expected inventory cost TC and on two measures of customer service. Given initial inventory y_i for product i, total expected cost is given by:

$$TC(y_i) = \int_0^{y_i} h(y_i - z_i)dF(z_i) + \int_0^{y_i} p(z_i - y_i)dF(z_i) \qquad (1)$$

This can be expressed as:

$$TC(y_i) = h(y_i - \mu_i) + (p+h)\int_y^{\infty}(z_i - y_i)dF(z_i) = \left(hk_i + (p+h)I_n(k_i)\right)\sigma_i \qquad (2)$$

where $k_i = \dfrac{y_i - \mu_i}{\sigma_i}$ and where I_n is the unit normal loss function (Eppen and Schrage, 1991): $I_n(k) = \int_k^{\infty}(z-k)\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz = \frac{1}{\sqrt{2\pi}}e^{-\frac{k^2}{2}} - k[1-\Phi(k)]$ (3)

$I_n(k)$ is convex decreasing in k, and $\lim_{k\downarrow -\infty} I_n(k) = \infty$ and $\lim_{k\uparrow +\infty} I_n(k) = 0$. For service level, we use fill rate and probability of no stock outs. Define fill rate as $\beta_i = \dfrac{E[\min\{y_i, z_i\}]}{E[z_i]}$. The probability of no stockouts is given by $\gamma_i = P\{z_i \le y_i\}$.

4. THE VALUE OF POOLING AND OF IMPROVING INVENTORY POLICY

In this section, we analyze the three performance measures under the four scenarios in Figure 2, i.e. before pooling and after pooling and with a suboptimal and an optimal inventory policy. Total expected cost before pooling (BP) is minimized at $y_i = y_i^*$, defined by $F_i(y_i^*) = \frac{p}{p+h} = \alpha$ or $y_i^* = F_i^{-1}(\frac{p}{p+h})$. Here, α is the critical fractile and represents the optimal probability of not stocking out. The optimal inventory level can be expressed as $y_i^* = \mu_i + k_i^*\sigma_i$ where $k_i^* = \Phi^{-1}(\alpha)$ is the optimal safety factor of product i, with Φ the cumulative standard normal distribution. We define suboptimal policies in terms of k_i: the larger the deviation from k_i^*, the more suboptimal the policy. Any inventory level y_i can be expressed as $y_i=\mu_i + k_i\sigma_i$ for some k_i, so k_i gives us a normalized measure of suboptimality. At optimum, total expected costs for product i are:

$$TC_i^* = \left[hk_i^* + (p+h)I_n(k_i^*)\right]\sigma_i = \left[hk_i^* + (p+h)I_n(k_i^*)\right]x_i\sigma \tag{4}$$

The optimal safety factor is the same for each product: $k_i^* = k^*$. Lemma 1 is well-known:

Lemma 1: *after pooling N SKUs into 1, demand for the consolidated product is* $N(\mu_{AP},\sigma_{AP}^2)$, *where* μ_{AP} *=*$N\mu$ *and, letting* ι *be the N-dimensional vector with all elements equal to one,*

$$\sigma_{AP} = \sqrt{\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sigma_i\sigma_j\rho_{ij}} = \sqrt{\iota'\Sigma\iota} = \sigma\sqrt{\mathbf{x}'\mathbf{R}\mathbf{x}}. \tag{5}$$

Two special cases help us interpret this expression:

- When $\rho_{ij}=0$ for all non-diagonal elements, so that $\boldsymbol{R}=\boldsymbol{I}$ (the N-dimensional unit matrix), then $\sigma_{AP} = \sigma\sqrt{\mathbf{x}'\mathbf{x}} = \sigma\sqrt{N(\sigma_x^2+1)}$, where $\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N}x_i\right)^2$. σ_x^2 is a measure of the concentration of uncertainty: the lower σ_x^2, the more evenly total uncertainty is spread out among all SKUs. Clearly, σ_{AP} increases in σ_x^2; it is minimized at $\mathbf{x}=\iota$, in which case $\sigma_x^2=0$ and $\sigma_{AP}=\sigma\sqrt{N}$.
- If all SKUs are identically distributed, i.e. $\mathbf{x}=\iota$, then $\sigma_{AP} = \sigma\sqrt{N+\sum_{i\neq j}\rho_{ij}}$. Now, σ_{AP} is increasing in all ρ_{ij} as expected, and $\boldsymbol{R}=\boldsymbol{I}$ gives $\sigma_{AP}=\sigma\sqrt{N}$.

In general, $0 \le \sigma_{AP} \le N\sigma$. Later, when discussing sensitivity with respect to *N*, we will have to assume ***R=I*** and $\mathbf{x} = \iota$, as otherwise $\sqrt{\mathbf{x'Rx}}$ could increase or decrease in *N*.

Lemma 2: *for any safety factor k, total inventory level, expected cost, fill rate and non-stockout probability for the system with N SKUs before pooling are given by:*

$$y_{BP} = N\mu + Nk\sigma \tag{6}$$

$$TC_{BP} = N\sigma\left[hk + (p+h)I_n(k)\right] \tag{7}$$

$$\beta_{BP} = \frac{\sum_{i=1}^{N} E[\min\{y_i, z_i\}]}{\sum_{i=1}^{N} E[z_i]} = 1 - \frac{\sigma}{\mu} I_n(k) \tag{8}$$

If ***R=I***, then $\gamma_{BP} = P\{z_i \le y_i \ \forall i\} = \Phi(k)^N$ (9)

To find system performance in optimum, substitute $k_i^* = \Phi^{-1}(\alpha) = \Phi^{-1}(\frac{p}{p+h})$ for *k*.

Proof: for β_{BP}, note that $E[\min\{y_i, z_i\}] = \int_0^{y_i} z_i dF_i(z_i) + \int_{y_i}^{\infty} y_i dF_i(z_i) = \mu_i - \sigma_i I_n(k_i)$, and $\sum_{i=1}^{N} \mu_i = N\mu$; the other statements are well-known.

Lemma 3: *for any safety factor k, total inventory level, expected cost, fill rate and non-stockout probability for the system after pooling N SKUs into 1 are given by:*

$$y_{AP} = N\mu + k\sigma\sqrt{\mathbf{x'Rx}} \tag{10}$$

$$TC_{AP} = \sigma\sqrt{\mathbf{x'Rx}}\left[hk + (p+h)I_n(k)\right] \tag{11}$$

$$\beta_{AP} = \frac{E`[min\{y_{AP}, \sum_{i=1}^{N} z_i\}]}{E[\sum_{i=1}^{N} z_i]} = 1 - \frac{\sigma_{AP}}{N\mu} I_n(k) = 1 - \frac{\sigma}{\mu} \frac{\sqrt{\mathbf{x'Rx}}}{N} I_n(k) \tag{12}$$

$$\text{If } \boldsymbol{R=I}\text{, then } \gamma_{AP} = P\{z \leq y_{AP}\} = \Phi(k) \tag{13}$$

System performance in optimum is found by substituting $k_i^* = \Phi^{-1}(\alpha) = \Phi^{-1}(\frac{p}{p+h})$ for k.

We can now quantify the effects of pooling under an optimal or suboptimal inventory policy, and of improving inventory policy before and after pooling. Actually arriving at the optimal policy is hard to do in practice, so using the optimal policy as a benchmark gives an upper bound on the benefits of improving inventory policy.

Proposition 1: *the effects of optimizing inventory policy before and after pooling (going from scenario 1 to 3 and from scenario 2 to 4 in Figure 2) are given by:*

$$y_{BP} - y_{BP}^* = N\sigma(k - k^*) \geq \sigma\sqrt{\mathbf{x'Rx}}(k - k^*) = y_{AP} - y_{AP}^* \tag{14}$$

$$\begin{aligned} TC_{BP} - TC_{BP}^* &= N\sigma\left[h(k - k^*) + (p + h)\left[I_n(k) - I_n(k^*)\right]\right] \geq \\ &\geq \sigma\sqrt{\mathbf{x'Rx}}\left[h(k - k^*) + (p + h)\left[I_n(k) - I_n(k^*)\right]\right] = TC_{AP} - TC_{AP}^* \end{aligned} \tag{15}$$

$$\beta_{BP}^* - \beta_{BP} = -\frac{\sigma}{\mu}\left[I_n(k^*) - I_n(k)\right]; \qquad \beta_{AP}^* - \beta_{AP} = -\frac{\sigma}{\mu}\frac{\sqrt{\mathbf{x'Rx}}}{N}\left[I_n(k^*) - I_n(k)\right] \tag{16}$$

$$\text{If } \boldsymbol{R=I}\text{, then } \gamma_{BP}^* - \gamma_{BP} = \Phi(k^*)^N - \Phi(k)^N \text{ and } \gamma_{AP}^* - \gamma_{AP} = \Phi(k^*) - \Phi(k) \tag{17}$$

Proposition 1 shows that the effects of improving inventory policy depend primarily on the differences $(k - k^*)$ and $I_n(k) - I_n(k^*)$, where k^* itself depends on $\frac{p}{p+h}$. Due to the convexity of TC we know that the cost of suboptimality is convex increasing in $|k-k^*|$. The value of improving inventory policy before pooling is also increasing in N and σ, as expected. β_{BP} can increase or decrease after improving inventory policy; the change in fill rate $\beta^*_{BP}-\beta_{BP}$ is concave increasing in $(k-k^*)$, and the absolute value $|\beta^*_{BP}-\beta_{BP}|$ increases with σ, decreases with but does not depend on N. After pooling, however, the effect on fill rate depends on N through $\frac{1}{N}\sqrt{\mathbf{x'Rx}}$ which for $\boldsymbol{R}=\boldsymbol{I}$ and $\mathbf{x}=\iota$ is equal to $1/\sqrt{N}$. In this case, as N increases, fill rate after pooling tends to 1. Service level γ can also increase or decrease: the differences $|\gamma^*_{BP}-\gamma_{BP}|$ and $|\gamma^*_{AP}-\gamma_{AP}|$ increase with $|k-k^*|$. The effect of pooling is ambiguous, as $|\gamma^*_{AP}-\gamma_{AP}|-|\gamma^*_{BP}-\gamma_{BP}|$ may be greater or smaller than 0. For N sufficiently large, the difference is always positive. In other words, if avoiding stockouts is a key criterion and N is large, the benefits of improving inventory policy can be significantly increased by also implementing pooling.

Proposition 2: *the effects of pooling under a suboptimal inventory policy (going from scenario 1 to 2 in Figure 2) are given by:*

$$y_{BP}-y_{AP}=\left(N-\sqrt{\mathbf{x'Rx}}\right)k\sigma \tag{18}$$

$$TC_{BP}-TC_{AP}=\left(N-\sqrt{\mathbf{x'Rx}}\right)\sigma\left[hk+(p+h)I_n(k)\right]\geq TC^*_{BP}-TC^*_{AP} \tag{19}$$

$$\beta_{AP}-\beta_{BP}=\left(1-\frac{\sqrt{\mathbf{x'Rx}}}{N}\right)\frac{\sigma}{\mu}I_n(k) \tag{20}$$

If $\boldsymbol{R}=\boldsymbol{I}$, then $\gamma_{AP}-\gamma_{BP}=\Phi(k)-\Phi(k)^N$ (21)

The effects of pooling under an optimal policy (going from scenario 3 to 4) can be found by substituting k^* for k in (18)-(21).

Returning to the two special cases earlier confirms what is known (intuitively or explicitly) about the effects of pooling. If ***R=I***, then the cost reduction depends on $\left(N-\sqrt{\mathbf{x'x}}\right)\sigma = \left(N-\sqrt{N(\sigma_x^2+1)}\right)\sigma$, which is decreasing in $\sigma_X{}^2$, and maximized at $\sigma_X{}^2=0$, or $\mathbf{x}=\iota$. The more evenly uncertainty is spread out among the SKUs, the greater the benefits of pooling. Similarly, if $\mathbf{x}=\iota$, the cost reduction depends on $\left(N-\sqrt{N+\sum_{i\neq j}\rho_{ij}}\right)\sigma$, which is decreasing in ρ_{ij}. The more positive correlation among demands, the lower the value of pooling. Table 2 summarizes sensitivity analyses for all parameters and all performance measures, which agree with results obtained elsewhere (Jönsson and Silver, 1987, and Gerchak and Mossman, 1992).

5. COMPARING POOLING AND IMPROVING INVENTORY POLICY

Returning to the framework in Figure 2, we asked whether, starting with no pooling and a suboptimal inventory policy (scenario 1), it was better to implement pooling (scenario 2) or to improve inventory policy (scenario 3). The initial costs are given by $TC_{BP}(k)$, the two alternatives carry costs of $TC_{AP}(k)$ and $TC^*_{BP}=TC_{BP}(k^*)$ respectively, given in Propositions 1 and 2. Comparing these, we get:

Proposition 3: *in a system with N SKUs and suboptimal safety factor k, pooling without improving inventory policy is better than improving inventory policy without pooling for all k in [k_1,k_2], where the k_i are the two solutions to $TC_{AP}(k) = TC_{BP}(k^*)$, i.e.*

$$h\left(k - \frac{\sqrt{\mathbf{x'Rx}}}{N}k^*\right) + (p+h)\left(I_n(k) - \frac{\sqrt{\mathbf{x'Rx}}}{N}I_n(k^*)\right) = 0 \qquad (22)$$

*k_1 and k_2 do not depend on σ; if **R=I** and $\mathbf{x} = \iota$, then k_1 decreases and k_2 increases in N.*

Proof: for any k, we know that $TC_{BP} >= TC_{AP}$, and because both costs are convex in k and will grow without limit as k tends to plus or minus infinity, we know there must be unique k_1 and k_2 such that k^* lies in $[k_1,k_2]$ and $TC^*_{BP}=TC_{AP}(k_1)$ and $TC^*_{BP} = TC_{AP}(k_2)$. TC_{BP} increases linearly in N, so the same is true for TC^*_{BP}, and TC_{AP} increases linearly in $\sqrt{\mathbf{x'Rx}} = \sqrt{N}$ when ***R=I*** and $\mathbf{x} = \iota$. So, let k_1 and k_2 satisfy $TC^*_{BP} = TC_{AP}(k_i)$, for some given N. Now take some other $N'>N$. It is easy to verify that:

$$TC^*_{BP}(N') = \frac{N'}{N}TC^*_{BP}(N) = \frac{N'}{N}TC_{AP}(k_2, N) > \sqrt{\frac{N'}{N}}TC_{AP}(k_2, N) = TC'_{AP}(k_2, N') \qquad (23)$$

This implies that for this new N', the solution to $TC^*_{BP} = TC'_{AP}(k_2)$ must be some $k_2' > k_2$, as $TC'_{AP}(k)$ is increasing for $k > k^*$ and therefore also at k_2. Similarly, one can show that $k_1' < k_1$, so that the "pooling" range [k_1',k_2'] for N' fully contains the range [k_1,k_2] with the original N. In the limit, as N tends to infinity, k_1 goes to $-\infty$ and k_2 goes to $+\infty$.

This is graphically illustrated in Figure 3, which shows total expected cost before and after pooling when ***R=I*** and $\mathbf{x} = \iota$. To study the effect of p and h, we set $p=h=2$ in the upper graph, $p=20$ and $h=2$ in the middle graph, and $p=2$ and $h=20$ in the lower graph. In all three cases, $N=10$ and $\sigma = 10$. The graphs immediately illustrate that for k in the range $[k_1,k_2]$, it is better to implement pooling without improving inventory policy, whereas for k outside that range it is better to implement an optimal inventory policy with no pooling. Analyzing sensitivity of the range $[k_1,k_2]$ with respect to p and h is more complex. Fortunately, though, we only need analyze one of these parameters, due to some easily verifiable symmetry properties.

Lemma 3: *if we exchange p and h, then $TC(k; p,h) = TC(-k; h,p)$ for all k. With $k^* = \Phi^{-1}(\frac{p}{p+h})$, $TC(k^*)$ is symmetric in p and h, i.e. $TC(k^*; p,h) = TC(k^*; h,p)$.*

Proof: use $I_n(-k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{k^2}{2}} + k[1-\Phi(-k)] = \frac{1}{\sqrt{2\pi}} e^{-\frac{k^2}{2}} + k\Phi(k) = I_n(k) + k$.

The following lemma is then immediate:

Lemma 4: *if $[k_1,k_2]$ is the "pooling" range associated with parameters (p,h), then $[-k_2,-k_1]$ is the pooling range associated with (h,p).*

This is borne out by Figure 3: the middle case with $p=20$ and $h=2$ is symmetric to the lower case with $p=2$ and $h=20$. Therefore, we need only examine the effect of p, by comparing the upper and middle graphs. In the upper graph, the "pooling" range $[k_1,k_2]$ covers a broad area on both sides of the optimal k^*. In this case, improving inventory policy rather than pooling should be recommended only to companies with excessively high or low inventories. As p increases, α tends to 1, and the penalty associated with a safety factor that is too low grows faster than the penalty associated with too high a safety factor. Now, the "pooling" range $[k_1,k_2]$ seems to shift to the right relative to the upper graph, meaning that inventory need be less far below optimal before improving inventory policy becomes better than pooling. In this particular case, pooling is better for all k between approximately 0 and 4, which spans most situations one might expect to find in practice; this range will increase further if we increase N above 10. Only for k negative or very large should this company be more concerned about improving inventory policy than about pooling.

Table 3 computes k_1 and k_2 for a range of parameter values for $\boldsymbol{R}=\boldsymbol{I}$ and $\mathbf{x}=\iota$. Tentatively, these analyses and examples suggest that inventory policies need to be seriously suboptimal before improving inventory policy should be preferred above pooling. That this is true for large N is no surprise, given that $[k_1,k_2]$ grows without bound as N goes to infinity. More surprising is how large this range is for small N. Even for N=5, current safety factor needs to be more than 1 more or less than optimal before pooling becomes less attractive than optimizing inventory policy. At N=10 and beyond, the range is already wide enough that it seems unlikely that a company with even a vague sense of the relative importance of stockout and holding costs would find itself outside that range. Though the table assumes $\boldsymbol{R}=\boldsymbol{I}$ and $\mathbf{x}=\iota$, it is easy to compute the $[k_1, k_2]$-ranges for any other $\boldsymbol{R}$ and $\boldsymbol{x}$. In Section 6, we compute these ranges for a case where $\mathbf{x}\neq\iota$.

6. SKU RATIONALIZATION IN PRACTICE

So far, in comparing the value of pooling relative to the value of improving a suboptimal inventory policy, we have assumed (multivariate) normal demand. Federgruen and Zipkin (1984) show how their approximations are still valid under a considerably broader class of distribution functions. Especially in industrial settings further upstream in supply chains, though, demand patterns can become truly erratic, due to the collection of mechanisms known as the bullwhip effect (Lee, Padmanabhan and Whang, 1997). The question naturally arises: how does the usefulness of the concepts and results derived above and elsewhere in the literature on the pooling effect change as a result? How can a practitioner construct a quick and rough estimate of the value of pooling, before spending too much time on feasibility and implementation studies? Though such questions are clearly impossible to answer in general, we illustrate how one can perform simple analyses to arrive at an initial estimate of the expected benefits of SKU rationalization even when demand distributions for each product do not follow any recognizable distribution. To do so, we use two years of demand data (from 1993-1994) from Pellton International, a chemical manufacturer acting as a supplier to automotive suppliers. Below, we first briefly describe the context and the demand structure, then the analyses and experiments we did using those data, and compare our findings with the theoretical results derived earlier. For reasons of confidentiality, the true company name is disguised; we have also omitted details of lead times in these illustrations. Overall, the analysis has been kept simple, and should be thought of as the first step beyond a "back-of-the-envelope" estimate rather than a final, accurate prediction; for that, a simulation model with more detail would be needed.

Pellton International

A more detailed description of Pellton International can be found in Corbett, Blackburn and Van Wassenhove (1997,1999). Pellton supplies rolls of plastic to automotive suppliers, ranging in width from 60 to 130cm wide, in 1cm increments; moreover, there are several different chemical formulations, colors, etc, leading to a total of over 2000 SKUs. We focus on their two main formulations, grade S1 and S2, and only on the clear (uncolored) plastic. Within grade S2, there is a special formulation for a key customer, grade S2/P. The plastic is produced in rolls of 320cm wide, and slit into rolls of the desired width; this slitting is an integrated part of the production line. Due to the inflexibility of the process, Pellton needs to make to stock. They constantly experienced stockouts and excessive inventory costs, and were interested in applying some form of SKU rationalization. There were three options:

- A program called "mastersizes", in which rolls would only be produced in 5cm increments; the customers would then have to trim the excess material themselves. (This was not a problem, as the customers have to trim the plastic to shape anyway.) This would potentially reduce the number of SKUs by approximately a factor 5.
- Eliminating the special grade S2/P by helping that one customer to switch to the standard grade S2; this would eliminate all sizes of S2/P currently produced.
- Postponement, separating the slitting operation from the production line. This would allow Pellton to produce rolls of 320cm and to postpone slitting to size until firm orders were received. This option would clearly be preferable but also required more capital investment, whereas the other two options were relatively easy to implement.

How does one go about constructing an initial estimate of the value of these options? We should emphasize that it is not our intention to provide an estimate of the value to Pellton of pooling, as more specifics of their processes would need to be taken into account. Rather, we merely use the demand patterns observed by Pellton for illustrative purposes.

Demand structure

Table 4 lists the number of SKUs produced and shipped in the two-year period considered for each grade, and several measures of concentration of demand and of uncertainty. Some individual SKUs represented over 15% of total demand for that grade, others were ordered only a few times. However, for studying the value of pooling, concentration of uncertainty is more relevant than that of demand. Recall that if the SKUs are closer to having identical standard deviations, σ_X^2 will decrease to 0. The result of this concentration of uncertainty is clear from rows 4 and 5: $\sigma_{AP} = \sigma\sqrt{\mathbf{x'x}}$ is between 50 and 100% larger than if $\mathbf{x} = \iota$. Figure 4 shows weekly orders for the highest-volume S1 SKU, and the corresponding frequency distribution; aggregate weekly demand is similarly erratic. Incidentally, this is an excellent illustration of the bullwhip effect: automotive assembly schedules do not vary drastically from week to week, but the graphs show that demand two levels upstream in the supply chain can be totally distorted. For almost all SKUs, correlation between weekly demands is in the (–0.1, +0.1) interval, so we can ignore it. The demand data could not be adequately explained by any reasonable distribution, after extensive goodness-of-fit tests (using Crystal Ball and Maple). We used three standard tests: the chi-squared *(χ^2)* and the empirical distribution function-based (EDF) tests, the Kolmogorov-Smirnov *(K-S)* and Anderson-Darling *(A-D)* tests. Stephens (1984) concludes that *A-D* is the recommended omnibus test statistic for an EDF with unknown parameters, especially when tail behavior of the distribution is important, as when studying inventory policies. The *A-D* statistic is a variation on the standard Kolmogorov-Smirnov test, and for n observations is defined as (Anderson and Darling, 1954):

$$A^2 = -\frac{1}{n}\left\{\sum_{i=1}^{n}(2i-1)\left[\ln z_i + \ln\left(1 - z_{n+1-i}\right)\right]\right\} - n \qquad (24)$$

We also decomposed observed demand into two components: a Bernoulli distribution with empirically determined probability p_i for each SKU i of having positive demand in any given week, followed by a continuous distribution of that demand. We fitted a wide range of distributions, and found that gamma, Weibull, and lognormal performed the best, but even they were more often rejected than accepted and none performed well for a large number of SKUs. The Pellton data meet our requirement of being truly poorly behaved, so we have to use the empirical distribution functions for our analyses. In fact, for a simple initial estimate of the value of various types of pooling, a retrospective analysis using the empirical demand data is also far more convenient.

Pooling or improving inventory policy with empirical demand distributions

The stockout and holding costs were estimated by Pellton to be p=0.0769 and h=0.00278 ECU/m^2 per week respectively, leading to a critical fractile α=0.965. (At that time, 1 ECU was US $ 1.3.) Using these costs, we simulated total holding and shortage costs over the two-year period using the empirical demand data for a range of hypothetical inventory policies specified by $y_i = \mu_i + k\ \sigma_i$ for each SKU i. We repeated this exercise after aggregating weekly demand for all SKUs of the same grade into 5cm increments (the "mastersizes" program), and after aggregating weekly demand for all SKUs into a single SKU (slitting to order, or postponement of customization). Table 4 shows how the mastersizes and postponement strategies reduce the number of SKUs.

Because the demand distributions are poorly behaved, the optimal policy can not be expressed in the form $y_i = \mu_i + k(\alpha)\ \sigma_i$ for all i. We first calculated how total costs depend on k, as shown in Figure 5, the empirical counterpart to Figure 3. Because $p >> h$, the graphs most closely resemble the middle graph in Figure 3. We also constructed $[k_1,k_2]$ ranges as before, for the mastersizes and postponement strategies, as shown in Table 5. If Pellton's current safety factor k lies within $[k_1,k_2]$, pooling (mastersizes or postponement respectively) is better than improving inventory policy; for k outside that range, the reverse is true. Table 5 provides an interesting counterpart to the theoretically derived Table 3. To compare them, note that Pellton's cost structure with α=0.965 corresponds to p=0.965 and h=0.035 in Table 3. Mastersizes applied to grade S1 would reduce SKU count from 35 to 11, equivalent to $N \approx 3$ in Table 3; for postponement, one would compare with N=35. Even though the empirical demand underlying Table 5 bears no resemblance to the i.i.d. normal demand assumed in Table 3, the ranges $[k_1,k_2]$ increase in N in both cases, and interpolating the theoretical values in Table 3 sometimes gets us quite close to the empirical ones in Table 5. The worse the inventory policy gets, the bigger the benefits of pooling, but at the same time, the more likely it will be better to focus on improving the inventory policy instead. Table 5 goes beyond Table 3 by illustrating the impact of concentration of uncertainty: S1 and S2/P have almost the same number of SKUs (N=35 and N=34 respectively), but S2/P has higher concentration of uncertainty ($\sigma_X{}^2$=1.65 against $\sigma_X{}^2$=1.39 for S1), and the corresponding $[k_1,k_2]$ ranges are indeed somewhat smaller for S2/P, so that the value of pooling is lower.

For any given current inventory level, the expected benefits of mastersizes and of postponement are easy to determine from Figure 5. (Of course, in practice one would not construct this graph; we only do so here to compare it with Figure 3). If their current safety factor is k=2, the cost savings of mastersizes are approximately 77% for grade S1 and 75% for grade S2. Eliminating grade S2/P altogether, however, only gives an 11% reduction; upon inspection, this is (partly) because the SKUs of grade S2/P hardly overlap with those of S2, so combining them (without pooling) does little to reduce uncertainty.

6. CONCLUSIONS

In this paper we have provided a simple framework for comparing the relative value of implementing some form of pooling against that of improving a suboptimal inventory policy. We study the general multivariate normal case and show how the combination of correlation and concentration of uncertainty affects the value of pooling. We find that the value of pooling increases as current inventory policy is more suboptimal, but obviously the value of improving that policy also increases. There is always a uniquely defined interval within which pooling leads to greater cost reduction than optimizing inventory policy; outside that range, the reverse is true. If demands are i.i.d., this range expands with the number of SKUs N; we also show how it depends on stockout and holding costs, and conclude the range is generally (very) large, so that in most practical circumstances, pooling, even under a distinctly suboptimal inventory policy, can be preferred over improving that inventory policy. We illustrate how one might construct an initial estimate of the value of pooling using highly erratic demand data from a chemical manufacturer, and find that the theoretical results are corroborated.

The managerial implications are significant. The paper delves more deeply into how and when might expect benefits, either in terms of cost or service level, to result from pooling. Above all, this work adds strength to the argument in favour of implementing some form of SKU rationalization, even when one is fully aware that the inventory policy in place is poor: the benefits will generally outweigh those of optimizing the inventory policy. Moreover, an optimal policy is normally impossible to find, as the demand distributions and cost parameters are not known with any precision. Pooling can reduce the complexity to the point that this suboptimality becomes far less costly. The estimates in Section 6 are very easy to perform in any spreadsheet such as Excel using historical demand data; before implementing pooling, one should at the very least conduct an analysis of this type to construct a rough estimate of the potential benefits.

This work also poses several important questions for future research. We provide a framework for gaining better understanding of how the value of pooling varies with the structure of demand, for instance the concentration of uncertainty among a limited set of SKUs. This needs to be studied further, as does the impact of the type of distribution involved, etc. Both theoretical and extensive numerical analysis would help in this respect. Similar studies looking at other potential costs and benefits of pooling, in manufacturing and elsewhere, would complement this one.

7. REFERENCES

Anderson, T.W., D.A. Darling. 1954. A Test of Goodness of Fit. *American Statistical Association Journal*, December 765-769.

Aviv, Y., A. Federgruen. 1999. The Benefits of Design for Postponement, in S. Tayur, R. Ganeshan, M. Magazine (eds.) *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers 553-584.

Baker, K.R., M. Magazine, H. Nuttle. 1986. The Effect of Commonality on Safety Stock in a Simple Inventory Model. *Management Science*, **32** (8) 982-988.

Collier, D.A. 1982. Aggregate Safety Stock Levels and Component Part Commonality. *Management Science* **28** (11) 1296-1303.

Corbett, C.J., J.D. Blackburn and L.N. Van Wassenhove. 1997. Pellton International: Partnerships or Tug of War? Parts A,B,C. Teaching cases, INSEAD.

Corbett, C.J., J.D. Blackburn and L.N. Van Wassenhove. 1999. Partnerships to Improve Supply Chains. *Sloan Management Review* **40** (4) 71-82.

Eppen, G. 1979. Effects of Centralization on Expected Costs in a Multi-Location Newsboy Problem. *Management Science*, **25** (5) 498-501.

Eppen, G., L. Schrage. 1981. Centralized Ordering Policies in a Multi-Warehouse System with Leadtimes and Random Demand, in L.B. Scharwz, (eds.) *Multi-Level Production/Inventory Systems: Theory and Practice*, North-Holland, New York, 51-67.

Erkip, N., W.H. Hausman, S. Nahmias. 1990. Optimal Centralized Ordering Policies in Multi-Echelon Inventory Systems with Correlated Demands. *Management Science*, **36** (3) 381-392.

Federgruen, A., P. Zipkin. 1984. Approximations of Dynamic, Multilocation Production and Inventory Problems. *Management Science*, **30** (1) 69-84.

Garg, A., H.L. Lee. 1999. Managing Product Variety: an Operations Perspective, in S. Tayur, R., M. Ganeshan, M. Magazine (eds.) *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, 467-490.

Gerchak, Y., D. Mossman. 1992. On the Effect of Demand Randomness on Inventories and Costs. *Operations Research* **40** (4) 804-807.

Ho, T-H., C.S. Tang. 1998. Product Variety Management: Research Advances, Kluwer Academic Publishers, Boston.

Jackson, P.L., J.A. Muckstadt. 1989. Risk Pooling in a Two-Period, Two-Echelon Inventory Stocking and Allocation Problem. *Naval Research Logistics*. **36** 1-26.

Jönsson, H., E.A. Silver. 1987. Analysis of a Two-Echelon Inventory Control System with Complete Redistribution. *Management Science*, **33** (2) 215-227.

Kapuscinski, R., S. Tayur. 1999. Variance vs Standard Deviation: Variability Through Operations Reversal. *Management Science*, **45** (5) 765-767.

Kekre, S., K. Srinivasan. 1990. Broader Product Line: A Necessity to Achieve Success? *Management Science*, **36** (10) 1216-1237.

Lee, H.L, C.A. Billington, B. Carter. 1993. Hewlett-Packard Gains Control of Inventory and Service through Design for Localization. *Interfaces* **23** (4) 1-11.

Lee, H.L. 1996. Effective Inventory and Service Management through Product and Process Redesign. *Operations Research*, **44** (1) 151-159.

Lee, H.L., V. Padmanabhan, S. Whang. 1997. The Bullwhip Effect in Supply Chains. *Sloan Management Review*, **Spring** 93-102

Lee, H.L., C.S. Tang. 1997. Modeling the Costs and Benefits of Delayed Product Differentiation. *Management Science* **43** (1) 40-53.

Lee, H.L., C.S. Tang. 1998 Variability Reduction through Operations Reversal. *Management Science*, **44** (2) 162-172.

McDuffie, J.P., K. Sethuraman, M.L. Fisher. 1996. Product Variety and Manufacturing Performance: Evidence from the International Automotive Assembly Plan Study. *Management Science*, **42** (3) 350-369.

Schwarz, L.B. 1989. A Model for Assesing the Value of Warehouse Risk-Pooling over Outside-Supplier Leadtimes. *Management Science*, **35** (7) 828-842.

Stephens, M.A. 1984. Tests Based on EDF Statistics. R.B. D'Agostino, M.A. Stephens, (eds.) *Goodness-of-fit Techniques*. Dekker, Inc. 97-185.

Swaminathan, J.M., S.R. Tayur. 1998. Managing Broader Product Lines through Delayed Differentiation Using Vanilla Boxes. *Management Science*, **44** (12) 161-S172.

Tagaras, G. 1992. Pooling in Two-Location Inventory Systems with Non-Negligible Replenishment Lead Times. *Management Science*, **38** (2) 1067-1083.

Table 1: Summary of notation

TC	expected total cost
β	fill rate
γ	probability of no stockout
p,h	unit shortage and holding cost per period
α	critical fractile, $\alpha=p/(p+h)$
y	stock level after ordering
z_i	demand, normally distributed with c.d.f. $F(z_i)$, with mean μ_i and standard deviation σ_i
x_i	constants such that $\mu_i=x_i\mu$ and $\sigma_i=x_i\sigma$ for all i
ι	N-dimensional vector with all elements equal to one
Φ, ϕ	unit normal c.d.f. and p.d.f.
Σ, $\boldsymbol{R}$	covariance and correlation matrix of the demands z_i with elements σ_{ij} and ρ_{ij} respectively
k	safety factor; k^* is the optimal safety factor $k^* = \Phi^{-1}(\alpha)$
$I_n(k)$	unit normal loss function
N	number of SKUs before pooling

Table 2: sensitivity analysis of the effects of pooling

	h (1)	p (1)	σ	μ	N (2)
y_{BP}	↓	↑	↑ k^*>0, ↓ k^*<0	↑	↑ k^*>0, ↓ k^*<0
y_{AP}	↓	↑	↑ k^*>0, ↓ k^*<0	↑	↑ k^*>0, ↓ k^*<0
$y_{BP} - y_{AP}$	↓	↑	↑ k^*>0, ↓ k^*<0	=	↑ k^*>0, ↓ k^*<0
TC_{BP}	↑	↑	↑	=	↑
TC_{AP}	↑	↑	↑	=	↑
$TC_{BP} - TC_{AP}$	↑	↑	↑	=	↑
β_{BP}	↓	↑	↓	↑	=
β_{AP}	↓	↑	↓	↑	↑
$\beta_{AP} - \beta_{BP}$	↑	↓	↑	↓	↑
γ_{BP}	↓	↑	=	=	↓
γ_{AP}	↓	↑	=	=	=
$\gamma_{AP} - \gamma_{BP}$	↑ (3)	↓ (3)	=	=	↑

Note: the table shows how the performance measures and the differences between the before- and after-pooling situations vary with the problem parameters, assuming no correlation and equal distributions for all SKUs.

(1) The results for h and p only apply in optimum, those for the other parameters apply for any k. For suboptimal policies, γ, β and y do not depend on p and h, and TC is increasing in both p and h.

(2) The results for N assume $\boldsymbol{R}=\boldsymbol{I}$ and $\mathbf{x} = \iota$. In fact, more general results can be obtained for other $\boldsymbol{R}$ but we omit these here as they are not particularly insightful.

(3) $\gamma_{AP} - \gamma_{BP}$ is increasing in h and decreasing in p if and only if $Np^{N-1} \geq (p+h)^{N-1}$.

Table 3: ranges within which pooling is better than optimizing inventory policy, for given values of p , h, and N

			N											
			2		5		10		25		100		1000	
p	**h**	**k***	**k1**	**k2**	**k1**	**k2**	**k1**	**k2**	**k1**	**k2**	**k1**	**k2**	**k1**	**k2**
0.999	**0.001**	**3.090**	1.142	3.540	1.109	5.324	1.076	7.529	1.018	11.904	0.896	23.809	0.583	75.290
0.995	**0.005**	**2.576**	1.038	3.051	0.940	4.573	0.854	6.467	0.722	10.225	0.480	20.449	-0.080	64.666
0.990	**0.010**	**2.326**	0.950	2.815	0.809	4.215	0.693	5.960	0.523	9.423	0.215	18.846	-0.509	59.596
0.950	**0.050**	**1.645**	0.588	2.177	0.324	3.270	0.123	4.612	-0.167	7.293	-0.717	14.586	-2.431	46.124
0.900	**0.100**	**1.282**	0.340	1.841	0.012	2.791	-0.239	3.925	-0.614	6.205	-1.387	12.410	-4.360	39.243
0.800	**0.200**	**0.842**	0.007	1.439	-0.400	2.239	-0.723	3.133	-1.240	4.949	-2.479	9.898	-7.825	31.301
0.750	**0.250**	**0.674**	-0.126	1.288	-0.565	2.038	-0.920	2.848	-1.510	4.494	-2.997	8.988	-9.474	28.423
0.667	**0.333**	**0.431**	-0.326	1.070	-0.812	1.752	-1.222	2.450	-1.940	3.857	-3.857	7.713	-12.196	24.391
0.600	**0.400**	**0.253**	-0.475	0.913	-0.997	1.550	-1.453	2.174	-2.284	3.415	-4.553	6.830	-14.398	21.597
0.500	**0.500**	**0.000**	-0.692	0.692	-1.269	1.269	-1.802	1.802	-2.824	2.824	-5.642	5.642	-17.841	17.841

Note: the values for k_1 and k_2 are the solutions to equation (22) with i.i.d. demand in Proposition 3 in the text, computed using Maple, the symbolic manipulation language. As the pooling range does not depend on μ or σ, it is sufficient to calculate the range for values of p and h normalized such that $p+h=1$. Moreover, as explained in the text, the "pooling ranges" are symmetric with respect to $k=0$ if one exchanges p and h.

Table 4: number of SKUs and concentration of demand and uncertainty at Pellton

	S1	S2	S2/P
# of SKUs before pooling	35	80	34
# of SKUs accounting for 90% of demand	13	19	10
σ_x^2	1.39	2.70	1.65
$\sqrt{\mathbf{x'x}}$	9.15	17.16	9.49
$\sqrt{N}$	5.92	8.94	5.83
# of SKUs with mastersizes	11	34	9
# of SKUs after postponement	1	1	1

Table 5: ranges for which pooling is better than optimizing inventory policy at Pellton

	k_1	k_2	k^*
S1- mastersizes	0.7	4	2.1
S1 – postponement	-0.1	8.5	2.1
S2- mastersizes	0.9	5.2	2.4
S2 – postponement	0.1	10.5	2.4
S2/P- mastersizes	1.1	4.1	2.2
S2/P – postponement	0.6	5.4	2.2

Note: k^* is the safety factor at which total costs before pooling would be minimized for the given empirical demand data, assuming a common safety factor for all SKUs. As the empirical demand distributions are far from symmtric, the implied optimal k^* will generally not be equal for all SKUs. We did also calculate k^* for each SKU independently and, in most cases, found values close to those in the Table.

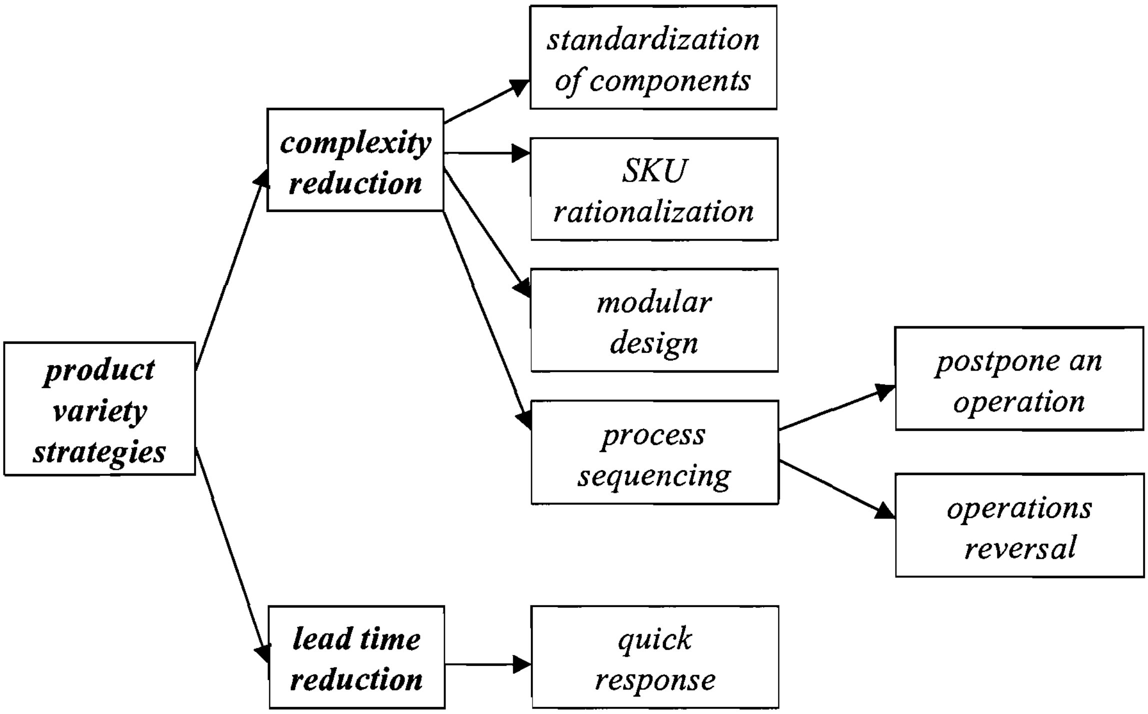


Figure 1: Strategies to reduce uncertainty in product variety

Note: adapted from Garg and Lee (1999). SKU rationalization refers to reducing the number of finished goods SKUs; several of the other strategies can of course lead to reduction of the number of component SKUs kept in stock. Product variety strategies in practice are often combinations of complexity reduction and lead-time reduction.

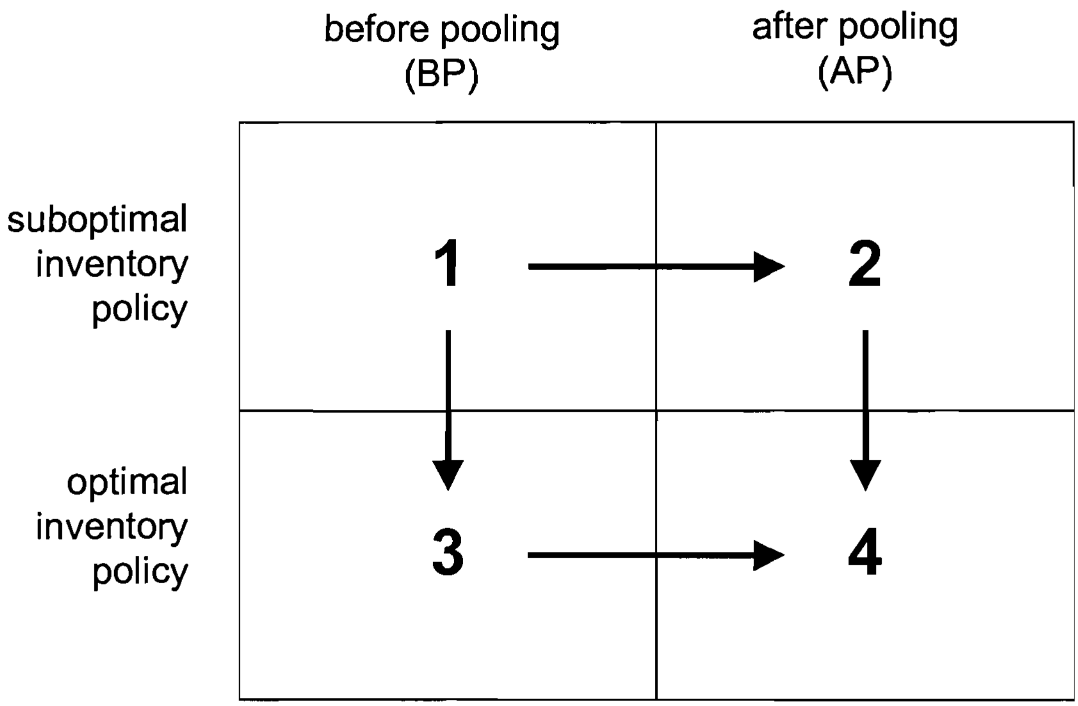


Figure 2: Approaches to managing variety through pooling or inventory policy

Note: many companies have a broad product line and a suboptimal inventory policy, putting them in scenario 1. To improve matters they can either improve their inventory policy (to get to scenario 3) or implement pooling (to get to scenario 2). Obviously, scenario 4 is the most desirable.

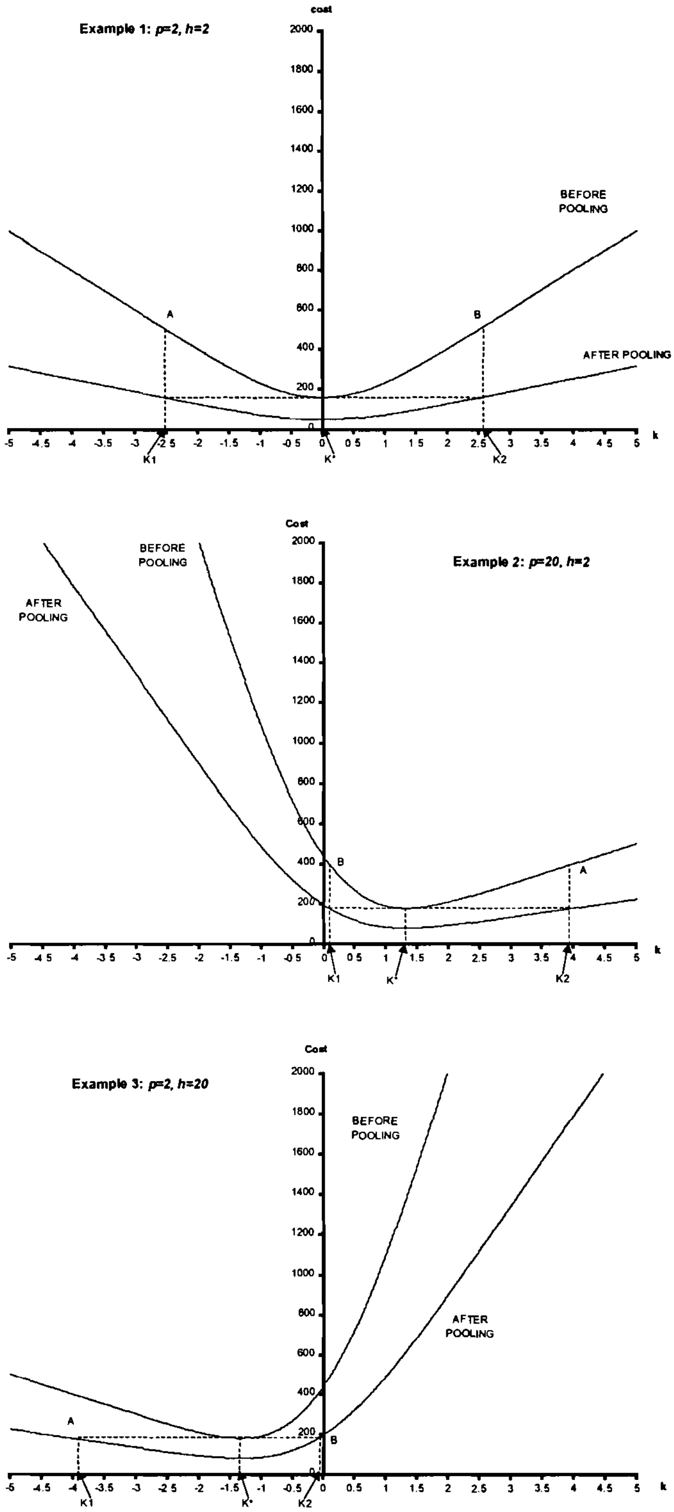


Figure 3: Analysis of pooling and optimal strategies

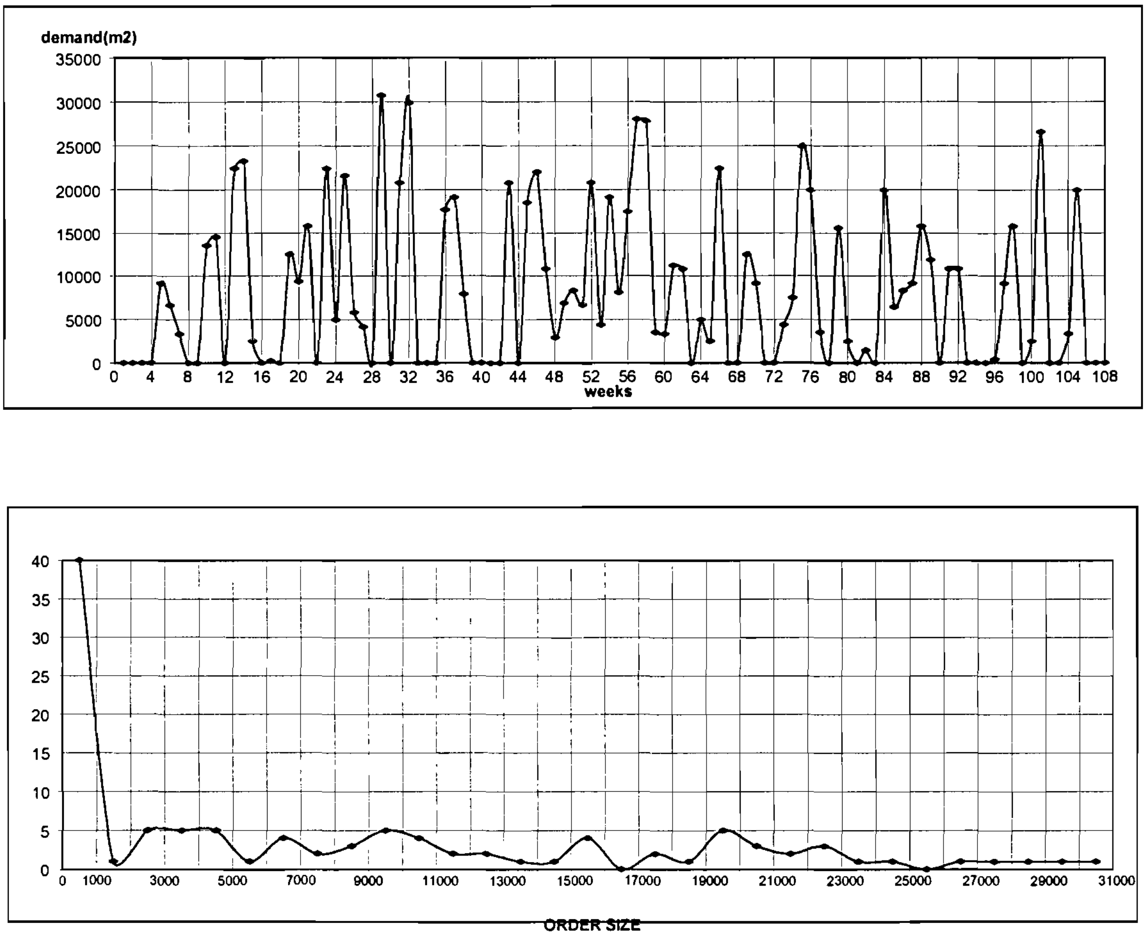


Figure 4: weekly demand pattern and frequency distribution for major product

Note: the top graph shows weekly demand for the highest-volume SKU, accounting for 17.4% of total demand for grade S1. The lower graph gives the frequency distribution.

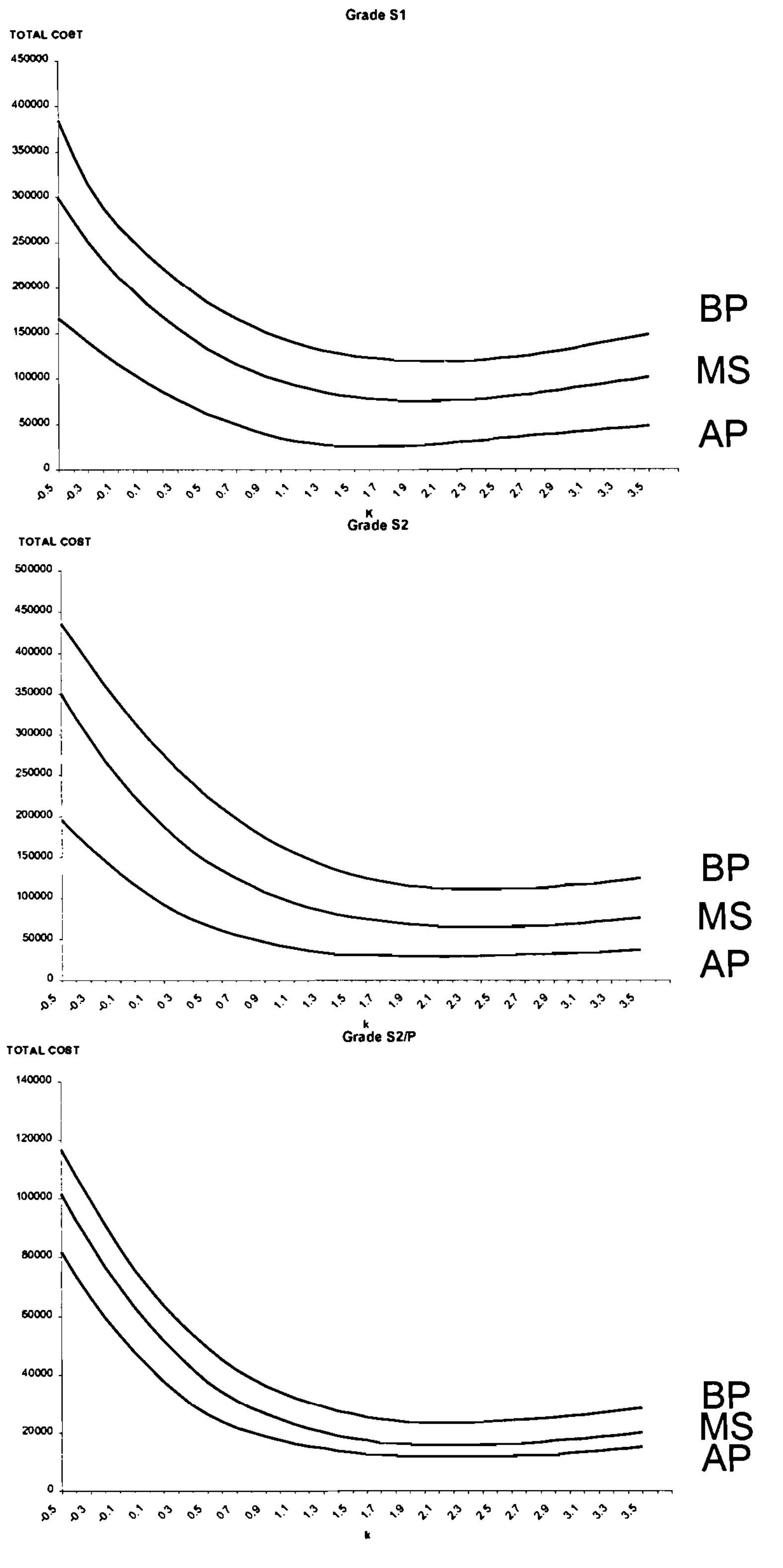


Figure 5: comparing total costs before pooling, with "mastersizes", and after pooling under a range of inventory policies for the three product groups