Inventory Pooling under Heavy-Tailed Demand

Kostas Bimpikis
Graduate School of Business, Stanford University, Stanford, CA 94305, USA.

Mihalis G. Markakis
Department of Economics and Business, Universitat Pompeu Fabra
& Barcelona School of Management.

Risk pooling has been studied extensively in the operations management literature as the basic driver behind strategies such as transshipment, manufacturing flexibility, component commonality, and drop-shipping. This paper explores the benefit of risk pooling in the context of inventory management using the canonical model first studied in Eppen (1979). Specifically, we consider a single-period multi-location newsvendor model, where n different locations face independent and identically distributed demands and linear holding and backorder costs. We show that Eppen's celebrated result, i.e., that the expected cost savings from centralized inventory management scale with the square root of the number of locations, depends critically on the "light-tailed" nature of the demand uncertainty. In particular, we establish that the benefit from pooling relative to the decentralized case, in terms of both expected cost and safety stock, is equal to $n^{\frac{\alpha-1}{\alpha}}$ for a class of heavy-tailed demand distributions (stable distributions), whose power-law asymptotic decay rate is determined by the parameter $\alpha \in (1, 2)$. Thus, the benefit from pooling under heavy-tailed demand uncertainty can be significantly lower than $\sqrt{n}$, which is predicted for normally distributed demands. We discuss the implications of this result on the performance of periodic-review policies in multi-period inventory management, as well as for the profits associated with drop-shipping fulfilment strategies. Corroborated by an extensive simulation analysis with heavy-tailed distributions that arise frequently in practice, such as power-law and log-normal, our findings highlight the importance of taking into account the shape of the tail of the demand uncertainty when considering a risk pooling initiative.

Key words: inventory management; inventory pooling; demand uncertainty; heavy-tailed distributions.

1. Introduction

One of the primary goals of an inventory management system is to protect the firm against demand uncertainty. Inventory pooling, i.e., the practice of serving multiple markets from a single stock of inventory, has received a lot of attention both among academics and practitioners. By and large, the existing literature highlights the benefit from pooling, arising from statistical economies of scale, when the underlying demand distribution is light-tailed. A central feature of the present paper and the main point of departure from previous literature, is that we consider *heavy-tailed* demand distributions. Our results quantify the relation between the benefit from pooling and how heavy the tail of the demand distribution is, pointing to the fact that the value of pooling, in terms of savings in both expected cost and safety stock, decreases as the tail of the underlying demand distribution becomes heavier, i.e., the probability of "extreme events" becomes higher.

The operational benefits from inventory pooling are well-known and fairly intuitive: serving independent demand streams from a single stock of inventory allows a firm to reduce the risk associated with demand uncertainty, as the stochastic fluctuations of different demands cancel out to some extent. Indeed, Eppen (1979) established that pooling n independent and identically distributed demand streams is $\sqrt{n}$ times less costly than serving each stream independently. However, this important insight relies on two assumptions: (i) there is no additional cost in serving local demands from a central location; and (ii) the demands are normally distributed, i.e., the underlying demand distribution has a "light" tail.

In certain settings these assumptions may not be justified. Although detecting and accurately estimating the parameters of heavy-tailed distributions is quite challenging due to the relative infrequency of "extreme events" that characterize them, several recent studies provide strong evidence of heavy-tailed phenomena in industries of interest. Additionally, we provide empirical evidence of heavy-tailed demand for movies at Netflix, and for shoes at a major North American retailer. It should be noted that, typically, inventory pooling studies ignore additional costs associated with serving multiple locations from a single location. When the benefits from a pooling initiative are not as large as expected, e.g., $\sqrt{n}$, then, of course, these costs become even more relevant.

The discussion above provides a solid motivation to re-examine the benefit from inventory pooling in relation to the shape of the tail of the underlying demand streams. This is precisely what we set out to do in the present paper by studying a model that follows closely the one analyzed by Eppen (1979) and subsequent literature. In particular, we consider a single-period multi-location newsvendor, with independent and identically distributed stochastic demands and linear holding and backorder costs. Our main point of departure from earlier work is that we assume that the underlying demand distribution is heavy-tailed, i.e., its *tail decays at a subexponential rate*. Prime examples of heavy-tailed distributions are power-laws, i.e., Cumulative Distribution Functions of the form:

$$F(x) \sim \frac{1}{x^{\gamma}}, \qquad \gamma > 0.$$

For distributions in this class, the probability of *extreme events* decays at a polynomial rate, and the exponent γ captures the level of uncertainty: lower values for the exponent imply a higher probability of large demand realizations. Another example of heavy-tailed distributions is the family of log-normal distributions, whose tails decay faster than those of power-laws, but still at a subexponential rate.

The main contribution of our work is to provide simple closed-form expressions that characterize the benefit from pooling, in terms of both the expected cost and the safety stock, as a function of the number of locations and the distribution's tail exponent. In particular, we show that as the tail exponent decreases, i.e., the probability of extreme events increases, the benefit from pooling decreases. Thus, our findings highlight the importance of properly accounting for the tail of the underlying demand distribution when deciding on how to manage a firm's inventory, or generally on adopting a risk-pooling initiative. [1]

Moreover, we discuss the implications of our findings on the performance of periodic-review policies in multi-period inventory management. In particular, we consider a single-location multi-period newsvendor where inventory is replenished according to a periodic-review policy, i.e., the firm is constrained to make orders every ℓ time periods. We view such a policy as "pooling in time": the firm satisfies demands received over multiple time periods from a single order. In this setting, frequent inventory replenishment reduces the newsvendor part of the cost as the firm is better able to match supply with demand, whereas infrequent replenishment reduces fixed ordering costs. We provide an upper and a lower bound on the optimal review period that highlight the latter's dependence on the (heavy-tailed) distribution's tail exponent. Specifically, both bounds are monotonically increasing in the tail exponent, suggesting that a demand distribution with a heavier tail leads to a shorter optimal review period. This fact provides an additional illustration of our main insight: pooling, whether "in space" or "in time," is less appealing in the presence of heavy-tailed demand.

There is a sizable literature on inventory pooling. As mentioned above, Eppen's seminal paper (Eppen (1979)) establishes that if n individual markets are pooled together and their demands are *normally distributed* and *uncorrelated*, then the benefit from inventory pooling is $\sqrt{n}$. The same paper shows that this benefit decreases if the (normally distributed) demands are positively correlated, and vanishes in the extreme case of perfect correlation. Federgruen and Zipkin (1984) consider a centralized depot that serves multiple markets, and provide a systematic way to approximate this model by a much simpler single-location inventory system. In the process, they demonstrate the benefit from pooling for different classes of distributions, most notably exponential and Gamma. Corbett and Rajaram (2006) and Mak and Shen (2014) extend Eppen's result regarding the benefit from pooling under correlated demands to more general distributions. Benjaafar et al. (2005) study the benefit from pooling in production-inventory systems, and establish that it decreases as the system's utilization increases due to an increase in the correlation among lead-time demands.

[1] As an additional example of the applicability of our results, in Appendix B we consider the practice of drop-shipping in online retailing (see Netessine and Rudi (2006) and Randall et al. (2006)), where a wholesaler stocks all inventory and ships products directly to customers at the retailers' request.

Closer in spirit to our work, Gerchak and Mossman (1992) and Gerchak and He (2003) study the impact of "demand randomness " on the optimal inventory levels and newsvendor cost through a simple mean-preserving transformation, and provide examples where pooling might neither reduce inventories nor move them closer to the mean or median demand. Berman et al. (2011) report simulation results for a number of distributions and argue that the normal distribution does not always provide accurate estimates with regards to the benefit from pooling as a function of the variance of the distribution. It appears that this is an artifact of the fact that in their simulations the order-up-to level approaches zero as the variance of the distribution grows. They conclude that more theoretical analysis is needed to explore the relation between demand uncertainty and the benefits from pooling. Song (1994) and Ridder et al. (1998) study the effect of demand uncertainty in the context of a single-period single-location newsvendor model, albeit for light-tailed distributions, and provide examples that increasing variability may increase or decrease inventory costs. Finally, although the literature on periodic-review policies is vast (e.g., see Atkins and Iyogun (1988), Eynan and Kropp (1998), Graves (1996), Rao (2003), Shang and Zhou (2010)), there do not seem to exist results that explore the relation between the tail of the demand distribution and the optimal review period.

Related to inventory pooling are also the strands of literature that explore the benefit from investing in flexible manufacturing capacity or resources (Fine and Freund (1990), Van Mieghem (1998), Netessine et al. (2002), Van Mieghem (2003), Bassamboo et al. (2010)), component commonality (Bernstein et al. (2007), Van Mieghem (2004), Bernstein et al. (2011)) , and delayed product differentiation (Lee and Tang (1997), Feitzinger and Lee (1997), Anand and Girotra (2007)). Anupindi and Bassok (1999) explore the interplay of inventory pooling and consumer behavior in an environment that consumers may search for available inventory in other locations in the case of a stock-out. On another note, Swinney (2012) considers pooling strategies in the presence of forward-looking consumers that may delay the time of a purchase, while anticipating a price decrease.

The rest of the paper is organized as follows. First, to motivate the subsequent analysis, we provide empirical evidence of heavy-tailed demand uncertainty in Section 2. We present the single-period multi-location newsvendor model and provide our main results regarding the benefit from inventory pooling in Section 3. Section 4 discusses the implications of our findings on the performance of periodic-review policies in multi-period inventory management. Section 5 concludes the paper. All proofs are relegated to Appendix A. A brief discussion about the practice of drop-shipping and the implications of our results in that context can be found in Appendix B.

2. Empirical Evidence of Heavy-Tailed Demand Uncertainty

Heavy-tailed distributions are used to model uncertainty when extreme events (i.e., very large or very small outcomes) are relatively likely, e.g., much more likely compared to what a normal or an exponential distribution would have predicted. Arguably, markets for fashion, technology (e.g., tablets and smartphones) and creative goods (e.g., movies and books), where there is significant uncertainty with respect to the number of sales of a new product, fall into this category. Chevalier and Goolsbee (2003) estimate that the empirical distribution of book demand at Amazon has a power-law tail with exponent $\gamma = 1.2$, whereas Gaffeo et al. (2008) report that the tail exponents for book sales in Italy in three broad categories (Italian novels, foreign novels, and essays) lie in the interval $\gamma \in (1, 1.4)$, and conclude that "... the values of γ are always significantly different from 2 and, therefore, the estimated degree of uncertainty in the book publishing market is too high to be compatible with a Gaussian distribution."

Furthermore, we have analyzed Netflix data on 17470 movies[2] focusing on the number of movies that received a given number of ratings. Since, presumably, one has to watch a movie (and thus "consume" the DVD) in order to rate it, it can be argued that the particular dataset provides a good approximation to the demand distribution of a typical movie at Netflix. We estimate that the number of movies per number of distinct ratings follows a power-law distribution with tail exponent equal to $\gamma = 0.6$. Even when we exclude movies that received less than 500 ratings, in order to focus on the tail of the distribution, the tail exponent remains quite small, $\gamma = 1.04$.

Finally, following the methodology described in Clauset et al. (2009), we perform a test to determine whether a heavy-tailed or a light-tailed distribution is a better fit for the demand for a group of similar retail products. In particular, we use data that we obtained from a major North American retailer on the number of sales throughout one year for 626 similar products (sneaker shoes). We compare the respective fit of the empirical distribution of the number of products per number of sales, to an exponential distribution and to a power-law distribution, respectively. Our approach can be summarized as follows:

(i) First, we fit the empirical data to the two candidate models to obtain the best fit within the class of distributions specified by the models. We derive Maximum Likelihood Estimators for the parameter of each model, i.e., the scaling parameter $\hat{\gamma}$ for the power-law, and the rate $\hat{\lambda}$ for the exponential distribution.

(ii) Then, we use a goodness-of-fit test to determine whether any of the hypothesized models explains the empirical data. The test is based on calculating the Kolmogrov-Smirnov statistic (a measure of the distance between two distributions) that corresponds to the empirical data and the

[2] We thank Professor Serguei Netessine for giving us access to this dataset.

hypothesized model, and comparing it with the KS statistic that corresponds to synthetic datasets generated by the same model.

(iii) Finally, a measure of the goodness-of-fit is the p-value associated with this experiment, which is defined as the fraction of times that the KS statistic corresponding to synthetic datasets is larger than the one that corresponds to the empirical data. A low p-value (typically, less than 5%) indicates a poor fit between the empirical data and the hypothesized model and, thus, provides strong support against the postulated hypothesis.

In Table 1 we report the p-values associated with our two hypotheses. Clearly, the exponential distribution provides a very poor fit to the empirical distribution and, therefore, can be safely rejected. On the other hand, the power-law distribution seems to provide a reasonable description of the data.[3]

Fitted Distribution	P-Value	Key Parameters
Exponential	Less than 0.0001	$\hat{\lambda} = 0.1247$
Power Law	0.5565	$\hat{\gamma} = 1.3769$

Table 1 Test of goodness-of-fit for the classes of exponential and power-law distributions to the empirical distribution of sales. Parameters $\hat{\lambda}$ and $\hat{\gamma}$ provide the best fit within the classes of exponential distributions and power-laws, respectively. The corresponding p-values indicate the goodness-of-fit of each of the two hypothesized models. The results suggest that the exponential distribution is a poor fit for the dataset, whereas a power law provides a reasonable description of the data

We note that the empirical evidence provided above involves sales figures for different products that belong to the same product category of the form “possible demand realizations Vs how many times each demand is observed in the dataset.” When introducing a new product, in order to decide the optimal inventory level/safety stock and then estimate the potential benefit from a pooling initiative, one needs to form a prior assumption for the demand of the product. The empirical distribution of past sales in the product category provides a natural starting point for forming such a prior. The empirical evidence presented above indicates that heavy-tailed distributions would be appropriate to use as priors for the demand for books at Amazon.com, for DVDs at Netflix, and for shoes at a North American retailer.

[3] We note, however, that the data could also be consistent with other distributions, as such tests can only reject a hypothesis; in particular, that the data is drawn from an exponential distribution. See Clauset et al. (2009) for an excellent discussion on the challenges in identifying and characterizing power-law distributions.

3. The Value of Inventory Pooling under Heavy-Tailed Demand

In this section we provide a detailed description of the standard single-period multi-location newsvendor model, in the context of which we present our main result regarding the benefit from risk pooling in the presence of heavy-tailed uncertainty.

A firm sells a single product in $n > 1$ different locations for a single time period. The demand for the product at location i is random, and denoted by D_i. The random variables $\{D_i;\ i = 1, \ldots, n\}$ are independent and distributed identically to some continuous random variable D, with cumulative distribution function $F(\cdot)$. At the beginning of the period the firm decides its inventory levels. Then, the random demands are realized, and at the end of the period the firm incurs two types of cost: (i) inventory holding cost at rate h per unit, if the inventory turns out to be more than the demand; and (ii) backorder cost at rate p per unit, if the inventory turns out to be less than the demand.

We compare the following two alternatives in terms of managing the firm's inventory: the first is a *fully decentralized* scheme, where inventory is kept at each of the n locations, and the inventory reserved for location i cannot serve demand in any other location $j \neq i$. The second alternative is a *fully centralized* scheme, where the demands from all n locations are satisfied from a single inventory repository. Intuitively, the latter approach should be preferable due to the so-called "statistical economies of scale": when independent demands are aggregated their stochastic fluctuations cancel out to some extent, making the aggregate demand more predictable and, thus, allowing for a reduction in the expected cost. What is more interesting to analyze is the extent that these fluctuations cancel out and, correspondingly, to quantify how much this reduces the expected cost.

A central feature of our model, and the main point of departure from previous literature, is that we consider *heavy-tailed* demands, i.e., probability distributions with subexponential tail asymptotics. More specifically, we derive most of our analytical results assuming that the underlying demand D belongs to the class of *stable distributions*. In general, stable distributions are characterized by four parameters: the stability parameter $\alpha \in (0, 2]$, the skewness parameter $\beta \in [-1, 1]$, the scale parameter $\sigma > 0$, and the location parameter $\mu \in \mathbb{R}$. Henceforth, we denote by $S_\alpha(\sigma, \beta, \mu)$ a stable distribution with stability α, skewness β, location μ, and scale σ.

The stability parameter α determines the shape of the tail of the distribution. In the special case of $\alpha = 2$ we retrieve the normal distribution. In contrast, for $\alpha < 2$ the distribution has infinite variance and features power-law asymptotics with tail exponent equal to α, i.e.,

$$\mathbb{P}(D > x) \sim \frac{1}{x^\alpha},$$

for large values of x. Note that smaller values of the stability parameter correspond to heavier tails and, thus, to higher probabilities of extreme events. In the remainder of the paper we focus on

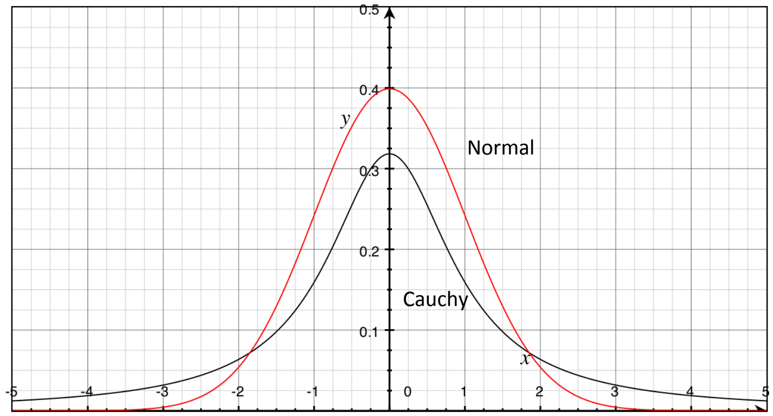

Figure 1 **The Cauchy distribution ($\alpha = 1$) has significantly heavier tails compared to the normal distribution ($\alpha = 2$) (figure taken from http://fukakai.blogspot.com).**

demands following stable distributions with stability parameter $\alpha > 1$, so that their expectations are finite and the corresponding newsvendor costs well-defined.

The location parameter μ is equivalent to the mean of the distribution, whenever this mean is finite, i.e., if $\alpha > 1$. The skewness parameter β determines how skewed is the distribution, and to which direction. Finally, the scale parameter is equivalent to the standard deviation of the distribution, whenever this is finite, i.e., in the case of normal distributions.

In this paper, we derive analytical results on the relative benefits to inventory pooling for stable distributions with $\alpha \leq 2$, and complement those results with a simulation study focusing on power-law and log-normal distributions. While power laws and stable distributions have the same asymptotic behavior, i.e., their tails decay at roughly the same (polynomial) rate, comparing the bodies of these distributions is a challenging task. This is because the probability density functions of stable distributions do not admit explicit expressions (typically, stable distributions are defined through their characteristic functions). That said, one of the main insights of the paper is that it is the tail of the demand distribution that plays a critical role in determining the benefit from pooling. So for our purposes the analytical results derived for stable distributions provide a reasonably good approximation for the relative benefits of pooling under power-laws.

Our analytical results hinge on the following well-known property of stable distributions.

LEMMA 1. *Let the random variables $X_i,\ i = 1, \ldots, n,$ be mutually independent and distributed identically to the stable distribution $S_\alpha(\sigma, \beta, \mu),\ \alpha \in (1, 2]$. Then,*

$$\sum_{i=1}^{n} X_i \stackrel{d}{=} n^{\frac{1}{\alpha}} X_1 + \mu\big(n - n^{\frac{1}{\alpha}}\big).$$

Lemma 1, which captures the essence of stable distributions, states that if i.i.d. stably distributed random variables are added, then (a linear transformation of) the sum is also stably distributed with the same parameters; equivalently, that stable distributions are closed under convolution. The proof of this result utilizes the definition of stable distributions, through their characteristic functions, and the fact that the characteristic function of the sum of independent random variables is equal to the product of their characteristic functions. In the context of this paper, Lemma 1 allows us to express in closed-form the distribution of the aggregate demand in the case of inventory pooling, thus allowing for tractable analysis. For a more extended discussion of stable distributions and their properties, the reader is referred to Zolotarev (1986).

As a final remark regarding our modeling assumptions, stable distributions can take both positive and negative values whereas the demand for a product is clearly nonnegative. An implicit assumption in the paper (which is a "standard" assumption in related literature, and often the case in practice) is that the expectation of the distribution is large. Then, the probability of the demand taking negative values is negligible, and its impact is insignificant.

3.1. Background: the Single-Location Newsvendor Model

We start by reviewing the widely studied single-period single-location newsvendor model, i.e., the case where $n = 1$, which provides some necessary background for the subsequent analysis. Obviously, there is no notion of pooling in this case, and the two inventory management schemes (decentralized and centralized) are equivalent.

The optimization problem that the firm has to solve in the single-location case can be written as follows:

$$\min_{q} \ \mathbb{E}\Big[h\big(q-D\big)^{+} + p\big(D-q\big)^{+}\Big],$$

where the expectation is taken with respect to the distribution of D. It is well-known that if an optimal solution to the above problem, q^*, exists then the following is true:

$$q^* = \inf\left\{q \geq 0 : \mathbb{P}(D \leq q) = F(q) \geq \frac{p}{p+h}\right\}. \tag{1}$$

Note that stable distributions have continuous CDFs, which implies that an optimal solution exists, and satisfies the well-known *critical fractile* condition:

$$F(q^*) = \frac{p}{p+h}. \tag{2}$$

The quantity $(q^* - \mu)$ is often interpreted as the *safety stock* that the newsvendor needs to keep in excess of the mean demand (note that safety stocks can take negative values).

Thus, the optimal expected cost of the single-period single-location newsvendor problem is equal to:

$$C(1) \equiv \mathbb{E}\Big[h\big(q^* - D\big)^+ + p\big(D - q^*\big)^+\Big]. \tag{3}$$

Let us introduce the following shorthand notation: if X is a random variable and E is an event on the underlying probability space, we denote by $\mathbb{E}[X;\ E]$ the expectation $\mathbb{E}[X \cdot \mathbf{1}_E]$, where $\mathbf{1}_E$ is the indicator variable of E. Using this notation, we can rewrite Equation (3) as follows:

$$\begin{aligned}
C(1) &= h\mathbb{E}[q^* - D;\ D \le q^*] + p\mathbb{E}[D - q^*;\ D > q^*] \\
&= h\mathbb{E}[q^*;\ D \le q^*] - h\mathbb{E}[D;\ D \le q^*] + p\mathbb{E}[D;\ D > q^*] - p\mathbb{E}[q^*;\ D > q^*] \\
&= hq^*\mathbb{P}(D \le q^*) - h\mathbb{E}[D;\ D \le q^*] + p\mathbb{E}[D;\ D > q^*] - pq^*\mathbb{P}(D > q^*) \\
&\overset{(2)}{=} hq^*\frac{p}{p+h} - h\mathbb{E}[D;\ D \le q^*] + p\mu - p\mathbb{E}[D;\ D \le q^*] - pq^*\frac{h}{p+h} \\
&= p\mu - (h+p)\mathbb{E}[D;\ D \le q^*].
\end{aligned} \tag{4}$$

3.2. The Value of Inventory Pooling

Coming back to the multi-location newsvendor problem, we first look into the *decentralized inventory management* scheme. In this case, the firm solves the basic newsvendor problem of Section 3.1 independently at each location:

$$\min_{q_i}\ \mathbb{E}\Big[h\big(q_i - D_i\big)^+ + p\big(D_i - q_i\big)^+\Big], \qquad i = 1, \ldots, n,$$

where the expectation is taken with respect to the distribution of D_i. We let $C^d(n)$ denote the optimal expected cost associated with managing n locations in a decentralized manner, i.e.,

$$C^d(n) \equiv \sum_{i=1}^{n} \min_{q_i}\ \mathbb{E}\Big[h\big(q_i - D_i\big)^+ + p\big(D_i - q_i\big)^+\Big].$$

Since the random variables $\{D_i;\ i = 1, \ldots, n\}$ are independent and distributed identically to D, we have that the optimal inventory level for each of the individual newsvendors is q^*, given by Equation (2). Therefore, the optimal aggregate inventory level in the decentralized case, which we denote by $q^d(n)$, is simply equal to $q^d(n) = nq^*$, and the optimal expected cost:

$$C^d(n) = nC(1), \tag{5}$$

where $C^(1)$ is the optimal expected cost in the single-period single-location newsvendor problem.

Next, we turn our attention to the *centralized inventory management* scheme. In this case, the firm solves a single optimization problem that is equivalent to that of the single-location newsvendor of Section 3.1, when the underlying demand is replaced by the sum of the demands in the n locations. Let $C^c(n)$ denote the optimal expected cost associated with managing n locations in a centralized manner, i.e.,

$$C^c(n) \equiv \min_q \ \mathbb{E}\Big[h\Big(q - \sum_{i=1}^n D_i\Big)^+ + p\Big(\sum_{i=1}^n D_i - q\Big)^+\Big],$$

where the expectation is taken with respect to the distribution of $\sum_{i=1}^n D_i$.

Equation (1) implies that the optimal inventory level in the centralized case, which we denote by $q^c(n)$, is equal to

$$q^c(n) = \inf\Big\{q \geq 0 \ : \ \mathbb{P}\Big(\sum_{i=1}^n D_i \leq q\Big) \geq \frac{p}{p+h}\Big\}.$$

Therefore,

$$C^c(n) = \mathbb{E}\Big[h\Big(q^c(n) - \sum_{i=1}^n D_i\Big)^+ + p\Big(\sum_{i=1}^n D_i - q^c(n)\Big)^+\Big].$$

The following theorem provides a simple relation between the optimal expected costs of the centralized and decentralized inventory management schemes, as a function of the stability parameter of stable distributions and the number of locations.

Theorem 1. *Consider the single-period multi-location newsvendor problem with mutually independent demands that are distributed identically to the stable distribution $S_\alpha(\sigma, \beta, \mu)$, $\alpha \in (1, 2]$. The relative benefit from inventory pooling is equal to*

$$\frac{C^d(n)}{C^c(n)} = n^{\frac{\alpha-1}{\alpha}}.$$

Moreover, the relative benefit in terms of the safety stock is the same,

$$\frac{q^d(n) - n\mu}{q^c(n) - n\mu} = n^{\frac{\alpha-1}{\alpha}}.$$

Proof: See Appendix A.

Theorem 1 captures the relative benefit from pooling, i.e., the savings in both expected cost and safety stock, relative to the benchmark of a fully decentralized system. Moreover, it illustrates the main insight of the paper: *the benefit from inventory pooling is inherently dependent on the tail of the underlying demand; in particular, it decreases as the tail of the demand becomes heavier.*

We note that the scope of Theorem 1 is quite broad, since it holds for all feasible values of the skewness, scale, and location parameters, which determine the "body" of the distribution. Let us now look at the implications of Theorem 1 in two special cases:

(i) *Normally distributed demands:* When $\alpha = 2$ the underlying demand distribution is normal, and we obtain that

$$C^d(n) = \sqrt{n}C^c(n) \quad \text{and} \quad q^d(n) - n\mu = \sqrt{n}\big(q^c(n) - n\mu\big),$$

recovering the result of Eppen (1979);

(ii) *Cauchy distributed demands:* When $\alpha \to 1$ the underlying demand distribution is approximately the Cauchy distribution, and we obtain that

$$C^d(n) \approx C^c(n) \quad \text{and} \quad q^d(n) \approx q^c(n),$$

namely there is essentially no benefit from pooling.

Finally, we provide some intuition on the findings of Theorem 1. Consider the single-period single-location newsvendor model. It is relatively clear that a more "predictable" demand can lead to lower expected inventory costs. At the extreme case of deterministic demand, the firm simply holds inventory equal to the demand and incurs zero costs. This is also suggested by Equation (4), since a more predictable demand would correspond to a distribution that has little probability mass in the tail.

The value of inventory pooling lies in the fact that when n i.i.d. demand streams are aggregated their stochastic fluctuations cancel out to some extent, making the total demand more predictable relative to the mean $n\mu$. To quantify the benefit from pooling, though, one needs to obtain a good understanding on the extent to which these fluctuations cancel out. In mathematical terms, the main determinant of the benefit from pooling is the order of magnitude of the fluctuations of $\sum_{i=1}^n D_i$ around $n\mu$.

When the demand distribution is light-tailed, then the Central Limit Theorem implies that the fluctuations are of order $O\big(\sqrt{n}\big)$. In turn, this implies that the expected holding and backorder costs are of order $O\big(\sqrt{n}\big)$. In particular, when the demands are normally distributed, it can be shown that the optimal expected cost in the centralized case is equal to $\sqrt{n}C(1)$ (Eppen (1979)). Thus, the relative benefit from pooling is in this case:

$$\frac{C^d(n)}{C^c(n)} = \frac{nC(1)}{\sqrt{n}C(1)} = \sqrt{n}.$$

On the other hand, when the demand distribution is heavy-tailed, then the order of magnitude of these fluctuations is greater: the Generalized Central Limit Theorem implies that the fluctuations are of order $O\big(n^{\frac{1}{\alpha}}\big)$, where $\alpha \in (0,2)$ is the tail exponent of the demand distribution. This implies that the aggregate demand is less predictable and, thus, the benefit from pooling should be smaller. In particular, when the demands follow a stable distribution with $\alpha \in (1,2)$, we prove that the optimal expected cost in the centralized case is equal to $n^{\frac{1}{\alpha}}C(1)$, which implies that the relative benefit from pooling is now:

$$\frac{C^d(n)}{C^c(n)} = \frac{nC(1)}{n^{\frac{1}{\alpha}}C(1)} = n^{\frac{\alpha-1}{\alpha}}.$$

At this point, we should note that the stochastic variability of stable distributions depends not only on the stability parameter α, but also on the scale parameter σ, which is equivalent to the standard deviation in the case of normal distributions. As it is apparent from Theorem 1, the scale parameter does not affect the relative benefit from pooling, i.e., $C^d(n)/C^c(n)$. Another metric of interest, though, is the absolute benefit from pooling, i.e., $C^d(n) - C^c(n)$. The following corollary expresses this absolute benefit as a function of the demand variability and the number of demands pooled.

Corollary 1. *Consider the single-period multi-location newsvendor problem with mutually independent demands that are distributed identically to the stable distribution $S_\alpha(\sigma,\beta,\mu)$, $\alpha \in (1,2]$. The absolute benefit from inventory pooling is equal to*

$$C^d(n) - C^c(n) = \left(n - n^{\frac{1}{\alpha}}\right)C(1),$$

where $C(1)$ is the optimal expected cost in the single-period single-location newsvendor problem.

Proof: This result follows directly from the proof of Theorem 1.

Clearly, the term $\left(n - n^{\frac{1}{\alpha}}\right)$ decreases as α decreases, namely as the tail of the demand distribution becomes heavier. On the other hand, Equation (4) suggests that a decrease in α has the opposite effect on $C(1)$: the only way that the mean of a distribution stays the same while the positive tail becomes heavier is when probability mass from the body of the distribution moves towards lower values. Thus, the term $\mathbb{E}[D;\ D \le q^*]$ is expected to decrease, which would imply that $C(1)$ increases as α decreases. That said, any potential increase in $C(1)$, irrespective of whether it is attributed to the stability or to the scale parameter, does not scale with n. Hence, the net effect should be a decrease in the absolute benefit from pooling when the number of demand streams that are pooled together is large enough (which is actually the case of interest). This insight is corroborated by simulation results towards the end of the section.

Theorem 1 and Corollary 1 determine the benefit from pooling when the demand distribution has a significantly heavy tail (power-law tail with exponent less than two) but provide no information about heavy-tailed demands with a somewhat lighter tail, e.g., power laws with tail exponent greater than or equal to two, or log-normal distributions. Obtaining analytical results for these distributions turns out to be a challenging task because they are not closed under convolution.

Without this property the analysis of the centralized system becomes quite challenging, since quantiles and expectations have to be computed with respect to a multidimensional convolution of densities for which, to the best of our knowledge, there are no closed-form analytical expressions that can lead to tractable analysis.[4]

On the positive side, in contrast to stable distributions, it is straightforward to simulate power-law and log-normal distributions. The numerical experiments presented below, for a variety of distributions and parameter values, corroborate the main insight of Theorem 1 and Corollary 1, i.e., that the benefit from pooling reduces as the tail of the demand distribution becomes heavier. More specifically, in Tables 2a–3b we report the ratios of the optimal expected costs of decentralized over centralized inventory management for $n = 10$, 20, 40, and 50 locations, where the demand is modeled as a power-law distribution with exponents ranging from 1.1 to 15. We follow Clauset et al. (2009) and assume that the demand has the following probability density function:

$$f(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\gamma}, \qquad x \in [x_{min}, \infty), \tag{6}$$

where the tail exponent of the probability distribution function in this case is equal to $(-\gamma + 1)$. We use $x_{min} = 1$ and we normalize the mean to $\mu = 100$ for the ratios reported in Tables 2a–3b. In the tables we also report the ratios for a log-normal distribution with scale parameter equal to $\sqrt{2}$ and variance equal to 50, as well as for a normal distribution with the same variance. Log-normal distributions, unlike stable distributions with $\alpha < 2$, have finite variances but their tails are significantly heavier than the ones of normal distributions (the probability of extreme events decays at a subexponential rate).

In all cases, the benefit from centralized inventory management decreases significantly as the tail of the underlying demand distribution becomes heavier. As expected, the values of the ratios do not match exactly our theoretical findings, wherever the comparison is valid, since power-law distributions are not stable distributions (although they do have the same asymptotic behavior). However, we emphasize that for relatively small values of the exponents, i.e., for demand distributions with sufficiently heavy tails, the predictions of Theorem 1 are much closer to the simulation results compared to the standard square-root rule. Furthermore, even though the log-normal and normal distributions have the same variance, they lead to considerably different relative benefits from pooling. This implies that the variance may not be the most suitable quantity to study when estimating the relative benefit from pooling. Instead, the shape of the tail seems to be more appropriate.

[4] Refer to Ramsay (2006) for an overview of what is known with respect to the convolution of power-law distributed random variables.

Tail Exponent γ	Ratio	Predicted Ratio
1.1	1.25	1.23
1.2	1.50	1.47
1.5	1.61	2.15
1.7	1.73	2.58
1.9	1.84	2.97
2	1.88	3.16
2.5	2.07	
5	2.45	
10	2.63	
15	2.69	
Log-normal	1.93	

(a) Number of locations: 10

Tail Exponent γ	Ratio	Predicted Ratio
1.1	1.35	1.31
1.2	1.71	1.65
1.5	1.94	2.71
1.7	2.14	3.43
1.9	2.32	4.13
2	2.39	4.47
2.5	2.72	
5	3.37	
10	3.66	
15	3.75	
Log-normal	2.53	

(b) Number of locations: 20

Table 2 Ratio of the costs of decentralized over centralized inventory management for power-law distributions with different values for the tail exponent γ, as well as log-normally distributed demand, when the number of locations is 10 and 20. For the power-laws we use $x_{min} = 1$ and we normalize the mean to $\mu = 100$. The scale parameter for the log-normal distribution is set to $\sqrt{2}$ and its variance is equal to 50. The third column in each of the sub-tables reports the ratio predicted from Theorem 1.

Tail Exponent γ	Ratio	Predicted Ratio
1.1	1.47	1.40
1.2	1.96	1.85
1.5	2.35	3.42
1.7	2.67	4.57
1.9	2.97	5.74
2	3.07	6.33
2.5	3.63	
5	4.68	
10	5.14	
15	5.26	
Log-normal	3.37	

(a) Number of locations: 40

Tail Exponent γ	Ratio	Predicted Ratio
1.1	1.51	1.43
1.2	2.04	1.92
1.5	2.51	3.68
1.7	2.88	5
1.9	3.22	6.38
2	3.36	7.07
2.5	3.99	
5	5.21	
10	5.73	
15	5.88	
Log-normal	3.72	

(b) Number of locations: 50

Table 3 Ratio of the costs of decentralized over centralized inventory management for power-law distributions with different values for the tail exponent γ, as well as log-normally distributed demand, when the number of locations is 40 and 50. For the power-laws we use $x_{min} = 1$ and we normalize the mean to $\mu = 100$. The scale parameter for the log-normal distribution is set to $\sqrt{2}$ and its variance is equal to 50. The third column in each of the sub-tables reports the ratio predicted from Theorem 1.

The power-law distribution of Equation (6) is also convenient for gaining additional insight (on top of what is provided by the latter part of Theorem 1) on how the optimal order quantity depends on the tail of the demand distribution. For simplicity, let us focus on the single-location case with $x_{min} = 1$. The optimal order quantity satisfies Equation (2), which for the particular distribution translates to:

$$\int_{q^*}^{\infty} (\gamma - 1)x^{-\gamma} dx = \frac{h}{p+h}.$$

Therefore,

$$q^* = \left(\frac{p+h}{h}\right)^{\frac{1}{\gamma-1}}, \qquad \gamma > 1,$$

which implies that the optimal order quantity in the single-location case q^* is decreasing in γ, i.e., increasing in the probability of tail events.

Finally, it is worthwhile mentioning that although we are mostly interested in the relative benefit from inventory pooling, and how the latter scales as a function of the number of locations, we verify numerically that in the presence of heavy-tailed demand the absolute benefit from pooling is also much more modest. More specifically, we compare the absolute benefit from pooling, $C^d(n) - C^c(n)$, in the following two cases:

(i) when the demand has probability density function that is given by Expression (6) with $\gamma = 2.1$, i.e., the tail exponent is 1.1, and mean equal to 10.

(ii) when the demand is exponentially distributed with rate $\lambda = 0.1$ (so that the average demand is again equal to 10).

When the holding and backorder costs are both equal to one, we obtain that $C^d(50) - C^c(50) = 290$ for exponentially distributed demand, as opposed to $C^d(50) - C^c(50) = 107$ for power-law demand. Thus, if a manager assumed incorrectly that the demand is exponentially distributed, she would overestimate the absolute benefit from pooling by a factor of 2.7.

3.3. Demands with Bounded Support

Typically, datasets that support the heavy-tail hypothesis, such as the ones presented in Section 2, provide evidence of a heavy tail up to a certain point, beyond which there is simply no data. This raises the following issue regarding our modeling choice of stable distributions: despite the fact that they may fit the data well on certain occasions, stable distributions have unbounded support and infinite variance. In practice, though, datasets always have bounded support and, consequently, finite sample moments. Therefore, one may wonder whether the main insight of Section 3.2 is an artifact of specific properties of stable distributions, rather than a salient feature of heavy-tailed distributions.

Tables 2a–3b provide support of the latter claim, i.e., that the shape of the tail largely determines the relative benefits from pooling, by comparing the ratios for log-normally and normally distributed demands with the same variance. In this section, we argue further that heavy tails are indeed the cause of the decreasing benefit from pooling, irrespective of the support or the variance, by assuming that the underlying demand follows a *truncated* stable distribution. More specifically, the demands at different locations $\left\{D_i^k;\ i=1,\dots,n\right\}$ are independent random variables that are distributed identically to

$$D^k = D \cdot \mathbf{1}_{|D|<k},$$

where $D \sim S_\alpha(\sigma,\beta,\mu)$, $\alpha \in (1,2]$, and $k \in \mathbb{Z}_+$ is the truncation point. We note that truncated distributions have bounded support and, thus, finite variance.

Again, our goal is to compare the optimal expected cost of the decentralized inventory management scheme, $C_k^d(n)$, to the optimal expected cost of the centralized scheme, $C_k^c(n)$. The following proposition summarizes our findings, suggesting that the main intuition from the nontruncated case is preserved.

Proposition 1. *Consider the single-period multi-location newsvendor problem with mutually independent demands that are distributed identically to the stable distribution $S_\alpha(\sigma,\beta,\mu)$, $\alpha \in (1,2]$, truncated at $k \in \mathbb{Z}_+$. The optimal expected cost in the decentralized case, $C_k^d(n)$, and the optimal expected cost in the centralized case, $C_k^c(n)$, are related as follows: for every $\epsilon > 0$ there exists $k_0(\epsilon)$, such that*

$$\left|\frac{C_k^d(n)}{C_k^c(n)} - n^{\frac{\alpha-1}{\alpha}}\right| \le \epsilon, \qquad \forall k \ge k_0(\epsilon).$$

Proof: See Appendix A.

Proposition 1 serves mainly as a proof of concept, to the fact that it is the heavy-tailed nature of the demand distribution that leads to a decrease in the benefit from pooling, rather than the boundedness of the support or the finiteness of the variance. [5]

The simulations that follow provide additional evidence to this point, and empirically verify that truncation points need not be very large for the main insight to hold. In particular, we report in Table 4 the ratio of the optimal expected costs (decentralized over centralized) for $n = 50$ locations, when demand is modeled as a power-law distribution with exponents ranging from 1.1 to 1.9, and truncation points 50, 100, and 150. The power-law distribution is normalized so that the mean is always equal to 10, thus the truncation points 50, 100, and 150 are 5, 10, and 15 times the mean. The monotonicity predicted by Proposition 1, i.e., that the benefit from inventory pooling decreases as the tail of the underlying demand distribution becomes heavier, appears to be robust to demand truncation.

[5] We note that the truncation point $k_0(\epsilon)$ depends also on α, a dependence that has been omitted in Proposition 1 in order to simplify the notation. Typically, we would expect higher probability of tail events (i.e., lower α) to lead to a higher truncation point. However, proving such a result would require additional probabilistic assumptions, exceeding the scope of this paper.

Tail Exponent γ	Ratio			
	Cutoff 50	Cutoff 100	Cutoff 150	No cutoff
1.2	4.29	3.58	2.89	1.72
1.5	4.81	4.05	2.99	2.41
1.7	4.92	4.17	3.16	2.82
1.9	4.97	4.25	3.38	3.18
Ratio for the normal distribution	7.07			

Table 4 Ratio of the costs of decentralized over centralized inventory management for truncated power-law distributions as a function of the distribution's tail exponent γ. The mean of the nontruncated distribution is normalized to 10 and the number of locations is 50. We consider truncation points that are 5, 10, and 15 times the mean of the nontruncated distribution. As a point of reference, in the dataset we obtained from a major North American retailer the ratio of the maximum value to the mean was around 15, whereas in the Netflix dataset it was more than 100.

4. "Pooling in Time": Replenishment under Heavy-Tailed Demand

In this section we consider a single-location multi-period newsvendor model, where inventory is replenished according to a periodic-review (R,T) policy, i.e., the firm reviews its inventory levels every T periods and places an order so as to raise the inventory position up to R. We call this setting "pooling in time" as, essentially, the firm serves demand realized in multiple time periods from a single inventory order. We show that the benefit from pooling in time, i.e., replenishing according to a periodic-review policy instead of every single period, decreases with the probability of extreme events, and we provide an upper and a lower bound on the optimal review period that depend on the tail-shape parameter α.

More concretely, we consider a firm that faces stochastic demand for a single product at discrete time periods $1,\ldots,t$. We denote by D_τ the demand at period τ and assume that the random variables $\{D_\tau;\ \tau=1,\ldots,t\}$ are independent and distributed identically to the stable distribution $S_\alpha(\sigma,\beta,\mu),\ \alpha\in(1,2]$.

The firm incurs three types of cost: (i) inventory holding cost for excess inventory, at rate h per product and per time period; (ii) backorder cost for inventory shortage, at rate p per product and per time period; and (iii) fixed ordering cost, at rate f per order placed. We assume that there is no lead time between placing and receiving an order.

Intuitively, a short review period implies low newsvendor cost, i.e., joint inventory holding and backorder cost, since the firm can better match supply and demand. On the other hand, a long review period implies infrequent placement of orders, so the fixed ordering cost is low. The goal is to achieve the optimal tradeoff between these opposing forces, i.e., to determine the length of the review period that minimizes the total expected cost.

For the rest of the section we denote the length of the review period by ℓ. Moreover, to simplify the analysis, we make the (unrealistic but not crucial) assumption that the length of the horizon t is an integer multiple of ℓ for all different values of the review period.

Let $\bar{q}_\ell$ be a solution to the optimization problem:

$$\min_{q>0} \sum_{i=1}^{\ell} \mathbb{E}\Big[h\Big(q - \sum_{j=1}^{i} D_j\Big)^+ + p\Big(\sum_{j=1}^{i} D_j - q\Big)^+\Big].$$

Since the demands at different periods are i.i.d. random variables, and there is no lead time between placing and receiving an order, the optimal periodic-review policy with period ℓ is "order up to $\bar{q}_\ell$, once every ℓ periods." The total expected cost of this policy can be broken down into "newsvendor" and "fixed ordering" part, as follows:

$$K^{pr}(\ell) = C^{pr}(\ell) + \frac{tf}{\ell},$$

where

$$\begin{aligned} C^{pr}(\ell) &= \sum_{\kappa=1}^{t/\ell} \sum_{i=1}^{\ell} \mathbb{E}\Big[h\Big(\bar{q}_\ell - \sum_{j=1}^{i} D_{(\kappa-1)\ell+j}\Big)^+ + p\Big(\sum_{j=1}^{i} D_{(\kappa-1)\ell+j} - \bar{q}_\ell\Big)^+\Big], \\ &= \frac{t}{\ell} \sum_{i=1}^{\ell} \mathbb{E}\Big[h\Big(\bar{q}_\ell - \sum_{j=1}^{i} D_j\Big)^+ + p\Big(\sum_{j=1}^{i} D_j - \bar{q}_\ell\Big)^+\Big]. \end{aligned}$$

THEOREM 2. *Consider the single-location multi-period newsvendor problem, with mutually independent demands that are distributed identically to the stable distribution $S_\alpha(\sigma, \beta, \mu)$, $\alpha \in (1, 2]$. The optimal review period for inventory replenishment, ℓ^*, satisfies:*

$$\Big\lfloor \Big(\frac{\alpha f}{C(1)}\Big)^{\frac{\alpha}{\alpha+1}} \Big\rfloor \le \ \ell^* \ \le \Big\lceil \Big(\frac{\alpha f + f}{C(1)}\Big)^{\frac{\alpha}{\alpha+1}} \Big\rceil,$$

where $C(1)$ is the optimal expected cost in the single-period single-location newsvendor problem.

Proof: See Appendix A.

Note that the gap between upper and lower bound on the optimal review period is of the order of $f/C(1)$, since their common exponent is less than one. Thus, when the newsvendor cost $C(1)$ is of the order of f or larger, which could very well be the case under heavy-tailed demand uncertainty, Theorem 2 provides a relatively tight characterization of the optimal review period.

Furthermore, both the upper and the lower bound on the optimal review period are monotonically increasing in α, since both the bases and the exponent are monotonically increasing in α. So, the heavier the tails of the stochastic demands, the shorter the optimal review period and, thus, the higher the cost as each replenishment is associated with fixed cost f. In other words, the benefit from "pooling in time," compared to ordering every single time period, decreases as the tail of the demand distribution becomes heavier. This is yet another illustration of our main insight, that the benefit from pooling in the form of centralized inventory management is largely determined by the tail of the demand distribution.

In order to provide clean expressions relating tail exponent and optimal review period, we had to make certain simplifying assumptions, e.g., zero lead time, horizon being a multiple of the review period. If these assumptions are relaxed, the expressions for the upper and lower bounds should be considerably more complicated. However, we believe that the qualitative nature of our findings will remain intact.

5. Concluding Remarks

The goal of this paper was to illustrate the importance of accounting for the tail of the demand distribution when considering investing in centralized inventory management, delayed product differentiation, or flexible capacity and resources. Although it is true, in general, that the above practices lead to a decrease in the associated costs, we showed that the savings are crucially dependent on how variable the demand is, in the sense of how likely extreme events are. More specifically, we provided simple closed-form expressions that relate the savings from centralized inventory management (in terms of both expected inventory cost and investment in safety stock) of a firm that faces stably distributed demands, as a function of the stability parameter, which captures the shape of the tail. Our results indicate that the benefit from inventory pooling decreases as the tail of the underlying demand becomes heavier. Therefore, pooling may not be as operationally attractive as when the underlying uncertainty is "well-behaved." This, in conjunction with empirical evidence supporting heavy-tailed uncertainty in several industries of interest, point to the fact that the decision maker should carefully estimate the relevant parameters before committing to an inventory management (or resource investment) strategy. Ignoring the, potentially, heavy-tailed nature of the demands may overestimate the cost savings of a pooling initiative and, thus, lead to suboptimal decisions.

Our intention was to study the interplay between heavy-tailed demand and pooling benefit in the simplest possible setting, so that our findings can be easily interpreted and directly compared to standard results in the literature. As a next step, several extensions to our benchmark model are worthwhile exploring. For example, consider an environment where demands in different locations are independent but differ in the distribution tail, and serving them from a centralized location involves an additional transportation cost. What is the optimal inventory management structure then? Which of the demand streams should be pooled together?

Similar issues arise when considering the optimal level of investment in flexible capacity or resources, e.g., see Van Mieghem (1998), Netessine et al. (2002). We believe that the analysis in those papers can be appropriately generalized for the class of stable distributions and, thus, we can obtain similar results that relate the tail of the underlying demand distribution to the optimal investment strategy. More generally, accounting for heavy-tailed uncertainty in operations management contexts, even outside resource pooling, is a first-order concern and we hope that the methodology employed in the present paper will be useful in future works.

Appendix A: Proofs

Proof of Theorem 1

Using Lemma 1, we can rewrite $q^c(n)$ as follows:

$$\begin{aligned} q^c(n) &= \inf\left\{q \geq 0 \;:\; \mathbb{P}\Big(n^{\frac{1}{\alpha}} D + \mu\big(n - n^{\frac{1}{\alpha}}\big) \leq q\Big) \geq \frac{p}{p+h}\right\} \\ &= \inf\left\{n^{\frac{1}{\alpha}} x + \mu\big(n - n^{\frac{1}{\alpha}}\big) \geq 0 \;:\; \mathbb{P}(D \leq x) \geq \frac{p}{p+h}\right\} \\ &= n^{\frac{1}{\alpha}} \cdot \inf\left\{x \geq 0 \;:\; \mathbb{P}(D \leq x) \geq \frac{p}{p+h}\right\} + \mu\big(n - n^{\frac{1}{\alpha}}\big) \\ &= n^{\frac{1}{\alpha}} q^* + \mu\big(n - n^{\frac{1}{\alpha}}\big), \end{aligned} \tag{7}$$

where the second equality follows from setting $x = n^{-\frac{1}{\alpha}}\big(q - \mu\big(n - n^{\frac{1}{\alpha}}\big)\big)$. Equation (7), and the fact that $q^d(n) = nq^*$, imply directly the second part of Theorem 1.

Coming to the first part, by definition, the optimal expected cost in the centralized case, $C^c(n)$, is equal to

$$C^c(n) = \mathbb{E}\Big[h\Big(q^c(n) - \sum_{i=1}^n D_i\Big)^+ + p\Big(\sum_{i=1}^n D_i - q^c(n)\Big)^+\Big].$$

Working similarly to Equation (4), we can rewrite the above cost as follows:

$$C^c(n) = np\mu - (h+p)\mathbb{E}\Big[\sum_{i=1}^n D_i \;;\; \sum_{i=1}^n D_i \leq q^c(n)\Big].$$

Lemma 1 and Equation (7) imply that

$$\begin{aligned} C^c(n) &= np\mu - (h+p)\mathbb{E}\Big[n^{\frac{1}{\alpha}} D + \mu\big(n - n^{\frac{1}{\alpha}}\big);\; D \leq q^*\Big] \\ &\overset{(2)}{=} np\mu - n^{\frac{1}{\alpha}}(h+p)\mathbb{E}[D;\; D \leq q^*] - np\mu + n^{\frac{1}{\alpha}} p\mu \\ &\overset{(4)}{=} n^{\frac{1}{\alpha}} C(1). \end{aligned} \tag{8}$$

Theorem 1 follows directly from Equation (8) and the fact that $C^d(n) = nC(1)$.

Proof of Proposition 1

In light of Theorem 1, it suffices to show that $C_k^d(n) \to C^d(n)$ and $C_k^c(n) \to C^c(n)$, as k becomes large. We first prove that the sequence $\{C_k^d(n);\; k \in \mathbb{Z}_+\}$ converges to $C^d(n)$. Let

$$q_k^* = \inf\left\{q \geq 0 : \mathbb{P}\big(D^k \leq q\big) \geq \frac{p}{p+h}\right\}.$$

As $k \to \infty$ the sequence $\{q_k^*;\; k \in \mathbb{Z}_+\}$ converges to $q^* < \infty$, because $D^k \xrightarrow{d} D$ and D is a proper random variable. Moreover, let

$$X^k = h\big(q_k^* - D^k\big)^+ + p\big(D^k - q_k^*\big)^+,$$

and

$$X = h\big(q^* - D\big)^+ + p\big(D - q^*\big)^+.$$

Since $D^k \xrightarrow{d} D$ and $q_k^* \to q^*$, we have that

$$X^k \xrightarrow{d} X.$$

Also, note that the sequence $\left\{q_k^*;\ k \in \mathbb{Z}_+\right\}$ is monotonically nondecreasing. Thus,

$$X^k \leq hq^* + pD, \qquad \forall k \in Z_+,$$

and

$$\mathbb{E}[hq^* + pD] < \infty,$$

since $q^* < \infty$ and $\mathbb{E}[D] < \infty$ (the latter is true because $\alpha > 1$).

Then, the Dominated Convergence Theorem implies that $\mathbb{E}\left[X^k\right] \to \mathbb{E}[X]$. In turn, this implies the desired result because $C_k^d(n) = n\mathbb{E}\left[X^k\right]$ and $C^d(n) = n\mathbb{E}[X]$.

The convergence of the sequence $\left\{C_k^c(n);\ k \in \mathbb{Z}_+\right\}$ can be proved in a similar way.

Proof of Theorem 2

First, let us derive a lower bound on $C^{pr}(\ell)$, which, in turn, will provide us with an upper bound on ℓ^*. We have that

$$\begin{aligned}
C^{pr}(\ell) &= \frac{t}{\ell}\sum_{i=1}^{\ell}\mathbb{E}\left[h\Big(\bar{q}_\ell - \sum_{j=1}^{i} D_j\Big)^+ + p\Big(\sum_{j=1}^{i} D_j - \bar{q}_\ell\Big)^+\right] \\
&\geq \frac{t}{\ell}\sum_{i=1}^{\ell} i^{\frac{1}{\alpha}} C(1) \\
&\geq \frac{t}{\ell}\frac{\alpha}{\alpha+1}\ell^{1+\frac{1}{\alpha}} C(1) \\
&= \frac{\alpha t C(1)}{\alpha+1}\ell^{\frac{1}{\alpha}}.
\end{aligned}$$

The first inequality follows from the proof of Theorem 1. The second inequality is implied by the following simple lemma: for every $\ell \in \mathbb{N}$ and $\alpha \in (0,\infty)$ we have that

$$\sum_{i=1}^{\ell} i^{\frac{1}{\alpha}} \geq \frac{\alpha}{\alpha+1}\ell^{1+\frac{1}{\alpha}}.$$

This result follows from the fact that function $x^{\frac{1}{\alpha}}$ is monotonically increasing. More specifically, we have that

$$i^{\frac{1}{\alpha}} = \int_{i-1}^{i} i^{\frac{1}{\alpha}}\,dx \geq \int_{i-1}^{i} x^{\frac{1}{\alpha}}\,dx,$$

which implies that

$$\sum_{i=1}^{\ell} i^{\frac{1}{\alpha}} \geq \int_0^{\ell} x^{\frac{1}{\alpha}}\,dx = \frac{\alpha}{\alpha+1}\ell^{1+\frac{1}{\alpha}}.$$

Coming back to the proof of Theorem 2, the total expected cost of the periodic-review policy is bounded from below as follows:

$$K^{pr}(\ell) \geq \frac{\alpha t C(1)}{\alpha+1}\ell^{\frac{1}{\alpha}} + \frac{tf}{\ell} \triangleq g_l(\ell).$$

Let us momentarily relax the integrality constraint. The minimum of $g_l(\cdot)$ can be obtained from the first order optimality condition:

$$\frac{\partial g_l}{\partial \ell} = 0 \iff \frac{tC(1)}{\alpha+1}\ell^{\frac{1}{\alpha}-1} - \frac{tf}{\ell^2} = 0 \iff \ell = \left(\frac{\alpha f + f}{C(1)}\right)^{\frac{\alpha}{\alpha+1}}.$$

It can be easily verified that $g_l(\cdot)$ is monotonically decreasing if $\ell < \left(\frac{\alpha f+f}{C(1)}\right)^{\frac{\alpha}{\alpha+1}}$, and monotonically increasing if $\ell > \left(\frac{\alpha f+f}{C(1)}\right)^{\frac{\alpha}{\alpha+1}}$. Thus, the expression above provides the unique global minimum.

Now, the fact that both $C^{pr}(\ell)$ and its lower bound are monotonically increasing functions of ℓ implies that the unique minimizer of $g_l(\cdot)$ is an upper bound on the optimal review period. Taking into account the integrality constraint, we have that

$$\ell^* \leq \left\lceil \left(\frac{\alpha f+f}{C(1)}\right)^{\frac{\alpha}{\alpha+1}} \right\rceil.$$

Let us now derive an upper bound on $C^{pr}(\ell)$, which, in turn, will provide us with a lower bound on ℓ^*. We have that

$$\begin{aligned}
C^{pr}(\ell) &= \frac{t}{\ell}\sum_{i=1}^{\ell}\mathbb{E}\Big[h\Big(\bar{q}_\ell - \sum_{j=1}^{i} D_j\Big)^+ + p\Big(\sum_{j=1}^{i} D_j - \bar{q}_\ell\Big)^+\Big] \\
&= \frac{t}{\ell}\min_q \sum_{i=1}^{\ell}\mathbb{E}\Big[h\Big(q - \sum_{j=1}^{i} D_j\Big)^+ + p\Big(\sum_{j=1}^{i} D_j - q\Big)^+\Big] \\
&\leq \frac{t}{\ell}\min_q \sum_{i=1}^{\ell}\mathbb{E}\Big[h\Big(q - \sum_{j=1}^{\ell} D_j\Big)^+ + p\Big(\sum_{j=1}^{\ell} D_j - q\Big)^+\Big] \\
&= \frac{t}{\ell}\sum_{i=1}^{\ell}\mathbb{E}\Big[h\Big(q^c(\ell) - \sum_{j=1}^{\ell} D_j\Big)^+ + p\Big(\sum_{j=1}^{\ell} D_j - q^c(\ell)\Big)^+\Big] \\
&= \frac{t}{\ell}\sum_{i=1}^{\ell}\ell^{\frac{1}{\alpha}}C(1) \\
&= tC(1)\ell^{\frac{1}{\alpha}},
\end{aligned}$$

where the inequality follows from the proof of Proposition 2 of Corbett and Rajaram (2006). This implies that the total cost of the periodic-review policy is bounded from above as follows:

$$K^{pr}(\ell) \leq tC(1)\ell^{\frac{1}{\alpha}} + \frac{tf}{\ell} \triangleq g_u(\ell).$$

Similarly to the case of $g_l(\cdot)$, if the integrality constraint is relaxed then the unique global minimum of $g_u(\cdot)$ can be obtained from the first order optimality condition:

$$\frac{\partial g_u}{\partial \ell} = 0 \iff \frac{tC(1)}{\alpha}\ell^{\frac{1}{\alpha}-1} - \frac{tf}{\ell^2} = 0 \iff \ell = \left(\frac{\alpha f}{C(1)}\right)^{\frac{\alpha}{\alpha+1}}.$$

The unique minimizer of $g_u(\cdot)$ is a lower bound on the optimal review period. Taking into account the integrality constraint, we have that

$$\ell^* \geq \left\lfloor \left(\frac{\alpha f}{C(1)}\right)^{\frac{\alpha}{\alpha+1}} \right\rfloor.$$

Appendix B: Profits in a Drop-Shipping Arrangement

The practice of drop-shipping is widely adopted in online retailing. In a nutshell, under a drop-shipping arrangement a wholesaler stocks all the inventory and fulfills customer orders at the request of retailers. The latter do not carry any inventory and focus on other functions of the buyer-seller relationship, e.g., marketing. Wholesalers that offer a drop-shipping arrangement trade off a higher wholesale price for the inventory risk that this practice entails. Hence, a drop-shipping arrangement would be preferable when it leads to a significant increase in the aggregate profits, since then the wholesaler would capture a fraction of the additional profits by increasing the wholesale price and, thus, be compensated for the up-front cost of investing in fulfillment services, as well as the additional risk associated with this strategy. The reader is referred to Netessine and Rudi (2006) and Randall et al. (2006) for a theoretical and an empirical treatment of the drop-shipping practice, respectively.

Consider the following simple model of a market with one wholesaler and a single product. Retailer i sells the product to its customers at a market price r (same for all retailers) and faces demand D_i. The random variables $\{D_i;\ i=1,\dots,n\}$ are assumed to be independent and distributed identically to the stable distribution $S_\alpha(\sigma,\beta,\mu),\ \alpha\in(1,2]$. The marginal cost for the wholesaler is equal to $c<r$. Retailers buy the product from the wholesaler at price w^D in a drop-shipping arrangement, where the wholesaler fulfills customer orders directly, and at price w^T in a traditional arrangement, where retailers stock and own their inventory. Finally, let $R^D(n)$ and $R^T(n)$ denote the optimal total expected profits, i.e., the sum of the profits of the wholesaler and the retailers, of a market with n retailers under a drop-shipping arrangement and a traditional approach, respectively.

We have that

$$R(1)\equiv r\mathbb{E}[\min(D,q^*)]-cq^*=R^D(1)=R^T(1),$$

where q^* is the optimal newsvendor quantity, defined as

$$\mathbb{P}(D\le q^*)=\frac{r-c}{c}.$$

In the spirit of Theorem 1, we provide a simple expression for the difference between the total profits in a drop-shipping and a traditional arrangement, as a function of both the stability parameter and the number of retailers in the market.

Proposition 2. *Assume that retailer demands are mutually independent and distributed identically to the stable distribution $S_\alpha(\sigma,\beta,\mu),\ \alpha\in(1,2]$. The difference in optimal expected profits between a drop-shipping and a traditional arrangement is equal to:*

$$\begin{aligned}R^D(n)-R^T(n)&=c\big(n-n^{\frac{1}{\alpha}}\big)\Big(\mathbb{E}[D\mid D\ge q^*]-\mu\Big)\\&=\big(n-n^{\frac{1}{\alpha}}\big)\Big((r-c)\mu-R(1)\Big).\end{aligned}$$

Proof: The quantity $R(1)$ can be expressed as follows:

$$\begin{aligned}R(1)&=r\mathbb{E}[D]-r\mathbb{E}[(D-q^*)^+]-cq^*\\&=r\mu-r\mathbb{E}[D;D\ge q^*]+rq^*\mathbb{P}(D\ge q^*)-cq^*\\&=r\mu-r\mathbb{E}[D;D\ge q^*]+rq^*\frac{c}{r}-cq^*\\&=r\mu-r\mathbb{E}[D;D\ge q^*],\end{aligned}$$

where the third equality follows from the fact that q^* is the optimal inventory level for a newsvendor with price r and cost c.

Clearly, $R^T(n) = nR(1)$. Moreover, following a similar approach to the proof of Theorem 1 (cf. Equations (7) and (8)) we have that

$$q^D(n) = n^{\frac{1}{\alpha}} q^* + \mu\big(n - n^{\frac{1}{\alpha}}\big),$$

and

$$\begin{aligned} R^D(n) &= r\mathbb{E}\Big[\sum_{i=1}^n D_i\Big] - r\mathbb{E}\Big[\Big(\sum_{i=1}^n D_i - q^D(n)\Big)^+\Big] - cq^D(n) \\ &= (r-c)\mu n + c\mu n^{\frac{1}{\alpha}} - rn^{\frac{1}{\alpha}}\mathbb{E}[D; D \geq q^*]. \end{aligned}$$

Therefore,

$$\begin{aligned} R^D(n) - R^T(n) &= \big(n - n^{\frac{1}{\alpha}}\big)\Big(r\mathbb{E}[D; D \geq q^*] - c\mu\Big) \\ &= c\big(n - n^{\frac{1}{\alpha}}\big)\Big(\mathbb{E}[D \mid D \geq q^*] - \mu\Big), \end{aligned}$$

where the second equality comes from the fact that $\mathbb{P}(D \geq q^*) = c/r$ and

$$E[D; D \geq q^*] = P(D \geq q^*)\mathbb{E}[D \mid D \geq q^*].$$

Finally, we can also use the fact that

$$r\mathbb{E}[D; D \geq q^*] - c\mu = (r-c)\mu - R(1),$$

in order to relate the profits of a drop-shipping arrangement to the profits of a single-retailer market. □

Proposition 2 provides an expression for the increase in aggregate profits, as a function of the number of retailers and the stability parameter of the demand distribution. This difference can be viewed as an upper bound on the additional profits that the wholesaler can extract by raising the wholesale price when offering drop-shipping to the retailers. Note that the difference depends on $R(1)$, which is, presumably, known to the wholesaler as she is already operating in the market. Also note that the prices w^T and w^D do not appear in the expressions of Proposition 2 because we are interested in total profits of the market, whereas w^T and w^D are only involved in money transfers between the wholesaler and the retailers.

Finally, we report on simulation results using the same parameters as in Netessine and Rudi (2006). In particular, we assume that the market price is $r = 20$, the wholesale price in the traditional arrangement is $w^T = 8$, the marginal cost of the wholesaler is $c = 3$, and the number of locations is $n = 10$. Our goal is to estimate an upper bound on the *additional* profits that a wholesaler can extract by offering a drop-shipping arrangement. To this end, we compare the profits of a wholesaler under the traditional arrangement, where retailers stock their own inventory which they order from the wholesaler at price w^T, with the profits of a wholesaler that offers a drop-shipping arrangement but guarantees that each retailer makes as much profit as in the traditional arrangement (to ensure that the retailers would take up on drop-shipping). Table 5 summarizes our results. Assuming a light-tailed demand distribution (normal with $\mu = 100, \sigma = 50$) overestimates the potential profits associated with drop-shipping by about an order of magnitude when the true demand is heavy-tailed.

	Upper bound on drop-shipping profits
Power-law ($\mu = 100$, $\alpha = 1.2$)	227
Normal Demand (μ=100, $\sigma = 50$)	2490

Table 5 Additional profits for drop-shipping relative to a traditional arrangement as a function of the tail of the demand distribution.

References

Anand, Krishnan S, Karan Girotra. 2007. The strategic perils of delayed differentiation. *Management Science* **53**(5) 697–712.

Anupindi, Ravi, Yehuda Bassok. 1999. Centralization of stocks: Retailers vs. manufacturer. *Management Science* **45**(2) 178–191.

Atkins, Derek R, Paul O Iyogun. 1988. Periodic versus "can-order" policies for coordinated multi-item inventory systems. *Management Science* **34**(6) 791–796.

Bassamboo, Achal, Ramandeep S Randhawa, Jan A Van Mieghem. 2010. Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. *Management Science* **56**(8) 1285–1303.

Benjaafar, Saif, William L Cooper, Joon-Seok Kim. 2005. On the benefits of pooling in production-inventory systems. *Management Science* **51**(4) 548–565.

Berman, Oded, Dmitry Krass, M Mahdi Tajbakhsh. 2011. On the benefits of risk pooling in inventory management. *Production and Operations Management* **20**(1) 57–71.

Bernstein, Fernando, Gregory A DeCroix, Yulan Wang. 2007. Incentives and commonality in a decentralized multiproduct assembly system. *Operations Research* **55**(4) 630–646.

Bernstein, Fernando, A Gürhan Kök, Lei Xie. 2011. The role of component commonality in product assortment decisions. *Manufacturing & Service Operations Management* **13**(2) 261–270.

Chevalier, Judith, Austan Goolsbee. 2003. Measuring prices and price competition online: Amazon. com and barnesandnoble. com. *Quantitative Marketing and Economics* **1**(2) 203–222.

Clauset, Aaron, Cosma Rohilla Shalizi, Mark EJ Newman. 2009. Power-law distributions in empirical data. *SIAM review* **51**(4) 661–703.

Corbett, Charles J., Kumar Rajaram. 2006. A generalization of the inventory pooling effect to nonnormal dependent demand. *Manufacturing & Service Operations Management* **8**(4) 351–358.

Eppen, Gary D. 1979. Note: Effects of centralization on expected costs in a multi-location newsboy problem. *Management Science* **25**(5) 498–501.

Eynan, Amit, Dean H Kropp. 1998. Periodic review and joint replenishment in stochastic demand environments. *IIE transactions* **30**(11) 1025–1033.

Federgruen, Awi, Paul Zipkin. 1984. Approximations of dynamic, multilocation production and inventory problems. *Management Science* **30**(1) 69–84.

Feitzinger, Edward, Hau L Lee. 1997. Mass customization at hewlett-packard: the power of postponement. *Harvard Business Review* **75** 116–123.

Fine, Charles H, Robert M Freund. 1990. Optimal investment in product-flexible manufacturing capacity. *Management Science* **36**(4) 449–466.

Gaffeo, Edoardo, Antonello E Scorcu, Laura Vici. 2008. Demand distribution dynamics in creative industries: The market for books in italy. *Information Economics and Policy* **20**(3) 257–268.

Gerchak, Yigal, Qi-Ming He. 2003. On the relation between the benefits of risk pooling and the variability of demand. *IIE transactions* **35**(11) 1027–1031.

Gerchak, Yigal, David Mossman. 1992. On the effect of demand randomness on inventories and costs. *Operations Research* **40**(4) 804–807.

Graves, Stephen C. 1996. A multiechelon inventory model with fixed replenishment intervals. *Management Science* **42**(1) 1–18.

Lee, Hau L, Christopher S Tang. 1997. Modelling the costs and benefits of delayed product differentiation. *Management science* **43**(1) 40–53.

Mak, Ho-Yin, Zuo-Jun Max Shen. 2014. Pooling and dependence of demand and yield in multiple-location inventory systems. *Manufacturing & Service Operations Management* **16**(2) 263–269.

Netessine, Serguei, Gregory Dobson, Robert A Shumsky. 2002. Flexible service capacity: Optimal investment and the impact of demand correlation. *Operations Research* **50**(2) 375–388.

Netessine, Serguei, Nils Rudi. 2006. Supply chain choice on the internet. *Management Science* **52**(6) 844–864.

Ramsay, Colin M. 2006. The distribution of sums of certain iid pareto variates. *Communications in StatisticsTheory and Methods* **35**(3) 395–405.

Randall, Taylor, Serguei Netessine, Nils Rudi. 2006. An empirical examination of the decision to invest in fulfillment capabilities: A study of internet retailers. *Management Science* **52**(4) 567–580.

Rao, Uday S. 2003. Properties of the periodic review (r, t) inventory control policy for stationary, stochastic demand. *Manufacturing & Service Operations Management* **5**(1) 37–53.

Ridder, Ad, Erwin Van Der Laan, Marc Salomon. 1998. How larger demand variability may lead to lower costs in the newsvendor problem. *Operations Research* **46**(6) 934–936.

Shang, Kevin H, Sean X Zhou. 2010. Optimal and heuristic echelon (r, nq, t) policies in serial inventory systems with fixed costs. *Operations Research* **58**(2) 414–427.

Song, Jing-Sheng. 1994. The effect of leadtime uncertainty in a simple stochastic inventory model. *Management Science* **40**(5) 603–613.

Swinney, Robert. 2012. Inventory pooling with strategic consumers: Operational and behavioral benefits. *Working paper* .

Van Mieghem, Jan A. 1998. Investment strategies for flexible resources. *Management Science* **44**(8) 1071–1078.

Van Mieghem, Jan A. 2003. Capacity management, investment, and hedging: Review and recent developments. *Manufacturing & Service Operations Management* **5**(4) 269–302.

Van Mieghem, Jan A. 2004. Note: Commonality strategies: Value drivers and equivalence with flexible capacity and inventory substitution. *Management Science* **50**(3) 419–424.

Zolotarev, Vladimir M. 1986. *One-dimensional stable distributions*, vol. 65. American Mathematical Society.