OPERATIONS RESEARCH
Vol. 56, No. 4, July–August 2008, pp. 916–936
ISSN 0030-364X | EISSN 1526-5463 | 08 | 5604 | 0916

DOI 10.1287/opre.1080.0551

Selecting a Portfolio of Suppliers Under Demand and Supply Risks

Awi Federgruen
Graduate School of Business, Columbia University, New York, New York 10027, af7@columbia.edu

Nan Yang
Johnson School of Management, Cornell University, Ithaca, New York 14853, ny38@cornell.edu

We analyze a planning model for a firm or public organization that needs to cover uncertain demand for a given item by procuring supplies from multiple sources. Each source faces a random yield factor with a general probability distribution. The model considers a single demand season. All supplies need to be ordered before the start of the season. The planning problem amounts to selecting which of the given set of suppliers to retain, and how much to order from each, so as to minimize total procurement costs while ensuring that the uncertain demand is met with a given probability. The total procurement costs consist of variable costs that are proportional to the total quantity delivered by the suppliers, and a fixed cost for each participating supplier, incurred irrespective of his supply level. Each potential supplier is characterized by a given fixed cost and a given distribution of his random yield factor. The yield factors at different suppliers are assumed to be independent of the season's demand, which is described by a general probability distribution.

Determining the optimal set of suppliers, the aggregate order and its allocation among the suppliers, on the basis of the exact shortfall probability, is prohibitively difficult. We have therefore developed two approximations for the shortfall probability. Although both approximations are shown to be highly accurate, the first, based on a large-deviations technique (LDT), has the advantage of resulting in a rigorous upper bound for the required total order and associated costs. The second approximation is based on a central limit theorem (CLT) and is shown to be asymptotically accurate, whereas the order quantities determined by this method are asymptotically optimal as the number of suppliers grows. Most importantly, this CLT-based approximation permits many important qualitative insights.

Subject classifications: inventory/production: approximations/heuristics; operating characteristics; reliability: failure models.
Area of review: Manufacturing, Service, and Supply Chain Operations.
History: Received August 2005; revisions received February 2006, September 2006, January 2007; accepted January 2007.

1. Introduction and Summary

Standard supply chain management texts discuss the benefits of consolidating the set of suppliers in the chain. These benefits include economies of scale in the production costs as well as statistical economies of scale due to the pooling of demand risks. Recently, many corporations and governments alike have recognized a variety of risks associated with external disruptions of the supply process. These provide a powerful argument against (maximal) consolidation. Such disruptions may arise because of "natural" disasters, e.g., fires in production plants or the need to shut down a facility because of violations of quality regulations or standards. Disruptions may also occur because of labor strikes, or planned acts of sabotage resulting from terror attacks, among others. Although these disruptions may be rare, their consequences can be catastrophic for an individual firm as well as for a region or a country as a whole.

In the private sector, "planning for disaster" has become one of the foci of supply chain planning; see, e.g., *Longitudes 04* (2004). This conference report describes, e.g., a case study of Ericsson, which, in contrast to Nokia, suffered major and long-term losses in profits and market shares for its cellular phone business due to its unhedged dependence on a single chip supplier in New Mexico and its lack of preparedness to switch to alternative suppliers in response to a major fire disabling this chip supplier. Terrorist-generated disasters targeted at universally critical component suppliers such as chip manufacturers may have a crippling effect, not just on this industry sector, but on many other major sectors of the economy as well.

Similarly, in the Fall of 2004, the United States saw half of its supply of flu vaccines cut out when the Chiron plant in Liverpool had to be closed down because of violations of FDA standards. In a year without vaccine shortages, no fewer than 36,000 deaths—12 times the number of 9/11 victims—and 200,000 hospitalizations are attributed to influenza and its complications. In terms of productivity, between \$11 billion and \$20 billion is lost annually due to influenza. The sudden elimination of one of only two manufacturers and half the national vaccine supply was hardly an unforeseeable or rare event because numerous Senate testimonies and General Accounting Office reports have documented recurring supply problems with this and other vaccines; see, e.g., Heinrich (2001a, b; 2004). In 2004, the Centers for Disease Control and Prevention (CDC) identified a target population of 100 million individuals who should have been vaccinated with the flu vaccine. Remarkably, the United States was dependent on just *two* suppliers, whereas England, with a target population of only 14 million, had its supply spread over *six* suppliers. Moreover, the vulnerability experienced with respect to the flu vaccine is hardly unique. Similar problems have arisen repeatedly over the last decade with respect to other perhaps even more crucial vaccines such as those required to immunize the children's population against highly contagious diseases.[1]

As a final example, oil is arguably the most critical commodity for the functioning of our economy. Its supply is primarily limited by existing refinery capacity. In the past 20 years, as the real-valued U.S. gross domestic product grew by 86.5%, the number of refineries decreased by more than 50%. This consolidation occurred because various types of economies of scale drove smaller refineries out of the market; other refineries identified the abovementioned benefits of pooling capacity and of running refineries at near 100% utilization. (In July and August of 2004, U.S. refineries were operating at 97% of available capacity.) Moreover, in case of a domestic supply disruption, little recourse can be expected from overseas refineries: the push of oil prices to record high levels this year is generally attributed to a lack of global refinery capacity. The Department of Energy predicts that current "financial, environmental, and legal considerations make it unlikely that new refineries will be built in the United States" (see Department of Energy 2005, p. 2). Most ominously, close to half of our capacity is located in a relatively small region on the Gulf Coast; disruption of its refinery and distribution process, the result of Katrina-like hurricanes, for example, could have a crippling effect on our economy. Since the 1950s, all U.S. administrations have intervened in the market by maintaining a stockpile of strategic reserves so as to mitigate the impact of sudden supply problems. However, the stockpile is largely in terms of crude oil, whereas refinery capacity has become the true bottleneck in the system. It is also the most vulnerable part of the oil supply chain because repairs of refinery equipment can take months to years to be completed. This puts into question whether the strategic reserves should not be replaced or complemented by incentives to expand the refinery base in different parts of the country.

In this paper, we study the multisourcing problem with unreliable suppliers, when stochastic demand needs to be covered with at least some prespecified probability. More specifically, we analyze a planning model for a firm or public organization that needs to cover uncertain demand for a given item by procuring supplies from multiple sources. Each source faces a random yield factor with a general probability distribution on the unit interval. An important special case is where this distribution has a positive mass at zero, representing the possibility of a complete shutdown due to an unplanned disruption. The model considers a single demand season. As in the case of the flu vaccine or other items with long production or distribution leadtimes, all supplies need to be ordered before the start of the season. The planning problem amounts to selecting which of the given set of suppliers to retain and how much to order from each, so as to minimize total procurement costs while ensuring that the uncertain demand is met with a given probability. The total procurement costs consist of *variable costs* that are proportional to the total quantity delivered by the suppliers—without loss of generality with a cost rate of one—and a *fixed cost* for each participating supplier, incurred irrespective of his supply level. Each potential supplier is characterized by a given fixed cost and a given distribution of his random yield factor. The yield factors at different suppliers are assumed to be independent of the season's demand, which is described by a general probability distribution. Thus, let:

$N =$ number of all available suppliers;
$K_i =$ fixed cost of operating at supplier i, $i = 1, \ldots, N$;
$X_i =$ random yield factor at supplier i's facility, with c.d.f. $G_i(\cdot)$, mean p_i, variance ς_i^2, and coefficient of variation $\gamma_i = \varsigma_i/p_i$, $i = 1, \ldots, N$;
$D =$ uncertain demand during the season, with a strictly increasing continuous c.d.f. $F(\cdot)$, complementary c.d.f. $\bar{F}(\cdot)$, inverse c.d.f. $F^{-1}(\cdot)$, mean μ, variance σ^2, coefficient of variation $\gamma_D = \sigma/\mu$, and finite moments;
$I^0 =$ initial inventory at the beginning of the season;
$\bar{D} = D - I^0 =$ uncovered demand with coefficient of variation $\gamma_{\bar{D}} = \sigma/(\mu - I^0)$;
$\alpha =$ maximum permitted probability of a shortfall ($\leqslant 0.5$);
$U =$ a standard Normal random variable with c.d.f. $\Phi(\cdot)$, complementary c.d.f. $\bar{\Phi}(\cdot)$, and $z_\alpha = \Phi^{-1}(1-\alpha)$;
$p_{\min}[p_{\max}] = \min_i p_i[\max_i p_i]$.

Without loss of generality, we assume that $\Pr[D > I^0] > \alpha$; otherwise, it is optimal not to order. We number the suppliers in increasing order of the coefficient of variation (c.v.) values of their yield distributions, i.e., $\gamma_1 \leqslant \gamma_2 \leqslant \cdots \leqslant \gamma_N$. We initially assume that the yield factors at the different suppliers are also independent of each other. However, to cover settings like oil refineries in close proximity being potentially hit by a common storm (e.g., Hurricane Katrina) or a common terrorist attack, we discuss in §7 how our results can be generalized to allow for dependent yield factors. Another important assumption is that the variable purchase price is identical for all suppliers. This assumption is appropriate when the suppliers offer (close to) perfect substitutes, as in the case of vaccines and refined oil products, but may fail to hold in other settings.

Determining the optimal set of suppliers, the aggregate order and its allocation among the suppliers on the basis of the exact shortfall probability is prohibitively difficult. We have therefore developed two approximations for the shortfall probability. Although both approximations are shown to be highly accurate, the first, based on a large-deviations technique (LDT), has the advantage of resulting in a rigorous upper bound for the required total order and associated costs. The second approximation is based on a central limit theorem (CLT) and is shown to be *asymptotically accurate*, whereas the order quantities determined by this method are *asymptotically optimal* as the number of suppliers grows. Most importantly, when analyzing the problem under the CLT-based approximation, the following important qualitative insights emerge: First, whether a set of suppliers allows for a *feasible* solution, depends not just on its cardinality but also on each supplier's predictability, as measured by the c.v. of his yield distribution. Defining a (hypothetical) supplier with a c.v. value of one as a "base supplier," a supplier with c.v. $=\gamma$ represents γ^{-2} base supplier equivalents (BSEs). A set of suppliers is feasible under general demands if its total number of BSEs is in excess of a critical number, given by a simple function of the permitted shortfall probability α, only. Under Normal demands, this condition is necessary as well, as long as the starting inventory is below the mean demand. When the initial inventory exceeds the mean demand by s standard deviations of demand, this minimum threshold is reduced by s^2. In particular, whether a set of suppliers is feasible or not does not depend on the shape of the demand distribution or any of its moments, the mean and standard deviation included, as long as the starting inventory is below the mean. (If the starting inventory is above the mean, feasibility of a set of suppliers depends on the demand distribution, only via the single measure s.) It follows that the minimal number of required suppliers is given by the smallest number n for which the total number of BSEs among the first n suppliers exceeds the critical threshold. The number of required suppliers can be reduced by improving their reliability; moreover, as a supplier with c.v. $=\gamma$ contributes γ^{-2} BSE's, the benefits of reductions of the c.v. value of a yield distribution become progressively larger. This gives support to management philosophies such as "Six Sigma," which advocate that companies should strive for near perfection, rather than terminating their quality improvement program when a "reasonable" level of quality or reliability is reached. The allocation scheme that splits the aggregate order in proportion to the suppliers' mean-to-variance ratio of their yield distribution has the best chance of enabling feasibility: if a feasible solution fails to exist under *this* scheme, it fails to exist under *any*.

Additional BSEs beyond the minimum number help to reduce the variable procurement costs. We refer to this surplus as the *suppliers' safety margin*. If the demand distribution is Normal, the minimal procurement cost for a given set of suppliers can be given as a closed-form expression that depends on the set of suppliers and their yield distributions via a single measure, i.e., the number of BSEs. Moreover, this minimum cost value is a convex decreasing function of the number of BSEs that converges to $\mu+z_\alpha\sigma-I^0=\mu(1+z_\alpha\gamma_D)-I^0$, the classical optimal order quantity under a single reliable supplier. To appreciate the price paid for dealing with unreliable suppliers, consider, for example, the case where the starting inventory is below the mean demand. Here, the asymptotic lower bound needs to be adjusted in two ways: first, the mean μ needs to be increased by a factor given by (#BSE's)/(the suppliers' safety margin); second, the coefficient of variation γ_D needs to be increased to $\gamma'_D=\sqrt{\gamma_D^2+(1-z_\alpha^2\gamma_D^2)/(\#\text{BSE's})}$.[2] The correction factor for the mean and the correction term for γ'_D decrease to one and zero, respectively, as the number of BSEs grows.

Under Normal demands, the above allocation scheme, which splits orders in proportion to the suppliers' mean-to-variance ratio, is not just the best guarantee for feasibility, but it minimizes variable procurement costs as well. The choice of the allocation scheme is particularly important when the number of suppliers is small, but its importance vanishes as n tends to infinity. Finally, we show that the problem of selecting a set of suppliers that minimizes fixed plus variable costs, is NP-complete, even under the CLT-approximation. However, we show that a procedure that builds the set of suppliers *greedily* comes very close to achieving the optimum, both in terms of worst-case and average behavior. (The worst-case optimality gap is 36.8%. In a numerical study involving 5,637 instances, the optimum is found in 98.5% of the cases, whereas the average optimality gap is 0.015%.)

The managerial implication is that if a firm or government organization wishes to expand its supplier base gradually, e.g., by adding a supplier in each of several seasons, little, if anything, is lost compared to the situation where the steady-state set of suppliers can be established all at once.

We start with the fully symmetric case where all N potential suppliers have identical fixed costs and yield distributions. We show that, even in this symmetric case, it may, for general demand distributions, fail to be optimal to order equal quantities from the different participating suppliers. Nevertheless, it is, in the symmetric case, reasonable to restrict oneself to the use of equal order sizes. Even so, it is important to invoke the LDT- or the CLT-based approximation, even in the fully symmetric case, and this is for two reasons: First, calculation of the optimal order quantity becomes increasingly more tedious as the number of suppliers grows. Second, the structural dependence of the exact optimal total order quantity—under the above restriction of equal order sizes—with respect to N may exhibit local irregularities that are smoothed out by employing the approximations. For example, although we show that under the above restriction the optimal average order quantity per supplier always decreases when the supply process is shared among additional sources, the optimal total order quantity may sometimes increase. In contrast, we show that the total order quantity always decreases, and convexly so, when determined by the LDT-based approximation. Likewise, the CLT-based approximation is always decreasing in N, and convexly decreasing in N if the demand distribution is Normal. These properties allow for a simple characterization of the optimal number of suppliers, $n^* \leqslant N$, under either one of the two approximations.

We proceed with the general case where firms may have different fixed costs and different yield factor distributions. In this case, it may be with considerable loss of optimality to order equal quantities from each of a given set of suppliers. As a consequence, all possible allocation schemes for the aggregate order quantity must be considered, and it is therefore no longer feasible to determine the optimal order quantities on the basis of the exact shortfall probability. However, in the approximate model that arises, under either one of the two approximations, optimal order quantities can be determined via simple numerical procedures. Moreover, under Normal demands, the CLT-based approximation permits closed-form expressions for the optimal set of orders, and, as mentioned, the total order quantity is again convexly decreasing in the number of BSEs represented by the set of suppliers.

The remainder of this paper is organized as follows: In §2, we review the related literature. In §3, we characterize, for the case of identical suppliers, the required orders on the basis of the exact shortfall probability. The LDT- and CLT-based approximations, as well as the structural properties of the order quantities determined by these approximations, are developed in §§4 and 5, respectively. Section 5 concludes with numerical comparisons of the exact order quantities and those determined by the two approximate methods. The general model with nonidentical suppliers is analyzed in §6. Section 7 summarizes our main conclusions and completes the paper.

2. Literature Review

Yano and Lee (1995) and Grosfeld-Nir and Gerchak (2004) provide surveys of a large literature on inventory systems with random yields. Almost all studies assume a single supplier. Of particular importance is Henig and Gerchak (1990), which characterizes the structure of an optimal ordering policy in a standard single-item periodic-review system, where only a random fraction of any order is usable. Most recently, Lewis et al. (2005) apply a standard single-supplier model with Markov-modulated lead times to study the impact of supplier disruptions. See also Zipkin (2000, §9.4.8).

Several papers have demonstrated the benefits of dual sourcing in the presence of supply uncertainty: Yano (1991) appears to have been the first to demonstrate the potential of dual sourcing as an insurance against catastrophic failures. Gerchak and Parlar (1990) and Parlar and Wang (1993) consider a variant of the economic order quantity (EOQ) model, in which orders can be split among two suppliers, each with a random yield factor. (In the EOQ model, a single item is sold continuously at a constant deterministic rate.) The authors determine the optimal split of the total order quantity between the two suppliers. Parlar and Wang (1993) also consider a variant of our model with two suppliers, supplier-dependent variable costs but no fixed costs. They developed closed-form expressions for an approximate solution that empirically comes within 10% of the optimal objective value. Tomlin (2006) considers a model with one unreliable supplier and one completely reliable but more expensive supplier, which is used during supply disruptions of the former. Most recently, Chopra et al. (2005) consider a single-period model with a known demand quantity and two suppliers, one of which is again completely reliable, whereas the other may either completely fail or deliver a random fraction (possibly larger than one) of the requested order. The authors determine the optimal order quantities when the uncertain yield of the unreliable supplier is correctly characterized as a mixed distribution with a positive mass at zero, as well as when the yield factor is approximated by a single continuous distribution with matching first and second moments.

Anupindi and Akella (1993) and Swaminathan and Shanthikumar (1999) consider the special case of the model treated here, along with multiperiod extensions, where there are $N = 2$ suppliers and the order sizes are determined so as to minimize end of the period overage and underage costs (as opposed to meeting a service-level guarantee). It appears that only Ilan and Yadin (1985) have addressed a model with an arbitrary set of potential suppliers where, as in our paper, the question is how many suppliers to use, in which specific combination, and how to allocate the aggregate supply quantity among them. In reviewing this paper, Yano and Lee (1995, p. 329) write: “This problem is extremely complex, and hence it is difficult to obtain structural results.”

Our work is also related to the literature on multidimensional newsvendor problems and that on assembly systems with random yields for the component manufacturers. For the former, see Harrison and Van Mieghem (1999), Van Mieghem (1998), Rudi and Zheng (1997), and Van Mieghem and Rudi (2002). The latter stream of papers was initiated by Yao (1988) and includes Singh et al. (1990), Gerchak et al. (1994), Lee (1996), and Gurnani et al. (1996, 2000) as important contributions. Yao (1988), for example, considered the problem of minimizing the cost of procuring the components subject to a service-level constraint, i.e., subject to achieving a fixed target quantity from the assembly stage with a minimum probability. Although demand in his model is assumed to be known, the service-level constraint bears similarity to that in our model. In both types of models, supply levels need to be determined for several suppliers, whose products act as complements for the delivery of one or more final consumer goods with uncertain demand. In our model, the different suppliers provide substitutes and the need for operational hedging and diversification arises because of uncertain yields.

3. Identical Suppliers

We first consider the base model where all suppliers share identical cost parameters and yield distributions. We therefore omit the subscript i from the cost parameters and $G(\cdot)$-distributions. Because the suppliers are indistinguishable, assume that we place identical orders of size y with each of a selected set of $n \leqslant N$ suppliers, for a total order $Y = ny$. (Below, we give conditions under which identical order sizes are indeed optimal.) The probability of satisfying the season's demand can thus be expressed by

$$\Pr\left(I^0 + y\sum_{i=1}^{n} X_i \geqslant D\right) = \int_0^n F(I^0 + yu)\, dG^{(n)}(u), \tag{1}$$

where $G^{(n)}(\cdot)$ denotes the n-fold convolution of the $G(\cdot)$ distribution. This probability clearly increases with y, so that the order size that minimizes variable procurement costs is given by

$$y^*(n) = \min\left\{y\colon \int_0^n F(I^0 + yu)\, dG^{(n)}(u) \geqslant 1 - \alpha\right\}. \tag{2}$$

Let $Y^*(n) \stackrel{\text{def}}{=} ny^*(n)$ denote the minimum total order size.

For example, in the important special case[3] where the yield factor is a Bernoulli random variable, i.e., the yield distribution $G(\cdot)$ is concentrated on the values zero and one, with p the probability of a successful completion of an order, (2) may be replaced by

$$y^*(n) = \min\left\{y\colon \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} F(I^0 + iy) \geqslant 1 - \alpha\right\}. \tag{3}$$

For any given value of y, it is possible, although sometimes tedious, to evaluate the left-hand sides of the inequalities in (2) and (3). The optimal order size y^* can thus be computed by a simple bisection search. Clearly, if n is sufficiently small and a source may fail completely with positive probability, i.e., $G(0) > 0$, no order quantity may be large enough to meet the service-level constraint. The following theorem provides a tight lower bound for n, the number of eligible suppliers, to guarantee that the service-level constraint can be met, i.e., an optimal order quantity y^* exists. When $G(0) > \alpha$, because with probability $G(0)$ any given supplier is unable to deliver anything, the service constraint is violated, irrespective of the order sizes, as long as the number of suppliers n is sufficiently small, i.e., $[G(0)]^n \bar{F}(I^0) > \alpha$ or $n < \underline{n} = [\ln\alpha - \ln\bar{F}(I^0)]/\ln[G(0)]$. Thus, $n \geqslant \underline{n}$ is necessary for a feasible solution, and the next theorem also shows that $n > \underline{n}$ is sufficient as well. It also shows that as n grows beyond this minimum number of suppliers, y^* decreases, whereas the minimum total order quantity $Y^*(n) = ny^*(n)$ approaches the limit $[F^{-1}(1-\alpha) - I^0]/p$. In other words, when n is large, the required total order approaches what would be ordered from a single completely reliable supplier who delivers a fraction $p = EX$ of any order, with probability one.

THEOREM 1. (a) *If $I^0 \geqslant F^{-1}(1-\alpha)$, $y^* = 0$.*

(b) *Assume that $I^0 < F^{-1}(1-\alpha)$. Let $\underline{n} \stackrel{\text{def}}{=} [\ln\bar{F}(I^0) - \ln\alpha]/[-\ln G(0)]$ if $G(0) > 0$, and $\underline{n} \stackrel{\text{def}}{=} 0$, otherwise. If $n < \underline{n}$, no optimal order quantity exists. Conversely, if $n > \underline{n}$, an optimal order quantity exists.*

(c) *Assume that $I^0 < F^{-1}(1-\alpha)$. For $n > \underline{n}$, y^* is decreasing in n.*

(d) *Assume that $I^0 < F^{-1}(1-\alpha)$. $\lim_{n\to+\infty} Y^*(n) = [F^{-1}(1-\alpha) - I^0]/p$.*

PROOF. (a) Part (a) is immediate.

(b) If $n < \underline{n}$, the shortfall probability is bounded from below by $G(0)^n\bar{F}(I^0) > \alpha$, and no order quantity is large enough to satisfy the service-level constraint. Let $\xi \stackrel{\text{def}}{=} (X \mid X > 0)$. If $n > \underline{n}$, the shortfall probability is

$$\begin{aligned}
&\Pr\left(I^0 + y\sum_{i=1}^{n} X_i < D\right)\\
&= G(0)^n\bar{F}(I^0) + (1 - G(0)^n)\Pr\left(I^0 + y\sum_{i=1}^{n} X_i < D \,\middle|\, \sum_{i=1}^{n} X_i > 0\right)\\
&\leqslant G(0)^n\bar{F}(I^0) + (1 - G(0)^n)\Pr(I^0 + yX < D \mid X > 0)\\
&= G(0)^n\bar{F}(I^0) + (1 - G(0)^n)E_\xi\bar{F}(I^0 + \xi y) < \alpha
\end{aligned}$$

for y sufficiently large. The first equality follows by conditioning on the event that all n suppliers face a complete breakdown and the complementary event that at least one of them does not. If $\sum_{i=1}^n X_i > 0$, there exists a supplier $1 \leqslant k \leqslant n$, with $X_k > 0$ and $\Pr(y\sum_{i=1}^n X_i + I^0 < D \mid \sum_{i=1}^n X_i > 0) \leqslant \Pr(yX_k + I^0 < D \mid X_k > 0) = E_\xi\bar{F}(\xi y + I^0)$, which explains the first inequality. The second inequality follows from $n > \underline{n}$ and $\lim_{y\to\infty} E_\xi\bar{F}(\xi y + I^0) = E_\xi \lim_{y\to\infty}\bar{F}(\xi y + I^0) = 0$, using the dominated convergence theorem.

(c) $\Pr(I^0 + y\sum_{i=1}^n X_i < D) \geqslant \Pr(I^0 + y\sum_{i=1}^{n+1} X_i < D)$. Thus, (2) implies $y^*(n) \geqslant y^*(n+1)$.

(d) Let $\widehat{F}(\cdot)$ denote the c.d.f. of $\bar{D}/p$ and $\widehat{F}_n(\cdot)$ the c.d.f. of $\bar{D}/((1/n)\sum_{i=1}^n X_i)$. Let $\widehat{F}^{-1}(\cdot)$ denote the inverse c.d.f. of $\widehat{F}$ and $\widehat{F}_n^{-1}(\cdot)$ the generalized inverse of $\widehat{F}_n$, i.e., $\widehat{F}_n^{-1}(u) = \min\{y\colon \widehat{F}_n(y) \geqslant u\}$. It follows from Chow and Teicher (1997, p. 272, Corollary 2) that $\lim_{n\to\infty}\widehat{F}_n(x) = \widehat{F}(x)$ for all x because $(1/n)\sum_{i=1}^n X_i \to p$ a.s. Thus, by the continuity of $\widehat{F}$, it follows (see Thorisson 2000, p. 24) that $\lim_{n\to\infty} Y^*(n) = \lim_{n\to\infty}\widehat{F}_n^{-1}(1-\alpha) = \widehat{F}^{-1}(1-\alpha) = (F^{-1}(1-\alpha) - I^0)/p$. □

REMARK. In general, when $n = \underline{n}$, no feasible order quantity y exists because usually there is a positive probability of a shortfall under any order quantity y, even in the event where at least one supplier avoids a complete breakdown.

The only exception arises when D has bounded support and the yield distribution $G(\cdot)$ is discontinuous in zero, such that a supplier avoiding a complete breakdown can guarantee itself a minimum yield factor $\xi^0 > 0$.

Theorem 1(c) shows that as the number of suppliers increases, $y^*(n)$, the order size per supplier, decreases. One might conjecture that the total order quantity, $ny^*(n)$, also decreases as the supply risks are diversified among additional suppliers. Various numerical examples show, however, that this may fail to hold. The surprising lack of monotonicity of the total supply is explained as follows: contrary to our facile assumption, assigning equal order sizes to different suppliers may fail to minimize the total order size, and hence the costs, even when the suppliers have completely identical characteristics. To illustrate this, consider an example with Normally distributed demand ($\mu = 100$, $\gamma_D = 0.05$), Bernoulli yields with $p = 0.975$, $I^0 = 0$, and $\alpha = 0.05$. It is easily verified, e.g., using (3), that the shortfall probability is exactly 5% when an order for $\mu(1+1.94\gamma_D) = 109.7$ units is placed with a single supplier. However, two suppliers each ordered to deliver half the quantity (54.85 units) will result in a shortfall probability of 7.4%. Thus, even when two suppliers are available, a single source results in a lower order than if the total order is spread equally between the two suppliers. Clearly, if optimal (possibly nonidentical) allocations of the total order are considered, the minimum total supply is nonincreasing in the number of suppliers. (When going from n to $n+1$ suppliers, the option to assign a zero-size order to the $(n+1)$st supplier remains feasible.) Only when the p.d.f. of the demand distribution is nonincreasing are equal order sizes guaranteed to minimize the total order size, i.e., the total order size $Y^*(n)$ is necessarily nonincreasing.

Corollary 1. *Assume that the p.d.f. of D is nonincreasing on $\mathbb{R}^+$. Equal-size orders are optimal and the total order quantity $Y^*(n) = ny^*(n)$ is nonincreasing in n.*

Proof. As explained above, it suffices to show that equal order sizes are optimal. Let y_i be the order placed with supplier i. The probability of satisfying demand is given by: $E_{\{X_1,\dots,X_N\}}F(\sum_{i=1}^n y_iX_i + I^0)$. Let $\hat{\mathbf{y}}$ denote an optimal vector of orders, minimizing the total order quantity $\sum_{i=1}^n y_i$, subject to ensuring that the probability of a shortfall is no more than α. If $\hat{\mathbf{y}}$ has nonidentical components, all permutations of this vector $\{\hat{\mathbf{y}}^{(1)}, \dots, \hat{\mathbf{y}}^{(M)}\}$ are optimal as well, i.e., these permutations also minimize the order quantity and yield a shortfall probability of no more than α. Let $\mathbf{y}^0 = (1/M)\sum_{l=1}^M \hat{\mathbf{y}}^{(l)}$. $E_{\{X_1,\dots,X_N\}}F(\sum_{i=1}^n y_i^0 X_i + I^0) \geqslant (1/M)\sum_{l=1}^M E_{\{X_1,\dots,X_N\}}F(\sum_{i=1}^n \hat{y}_i^{(l)} X_i + I^0) = 1-\alpha$ because F is concave by our assumption regarding the p.d.f. of D. Because $\mathbf{y}^0$ has the same total order size as each of the optimal solution $\{\hat{\mathbf{y}}^{(l)}\}$, $\mathbf{y}^0$, with identical components, is optimal as well. □

Other than the exponentials and uniform distributions on an interval $[0, a]$, very few distributions have nonincreasing p.d.f.s. Moreover, the determination of y^* in (2) may already be somewhat tedious when n is large or when convolutions of the $G(\cdot)$ distributions are hard to compute. Furthermore, to calculate the exact shortfall probability under a given set of nonidentical order sizes—as required for most demand distributions—is considerably harder, let alone to determine the optimal vector of orders. It is for this reason that we now discuss two approximation methods, one based on a large-deviations technique and one based on a central limit theorem.

4. An Approximation Based on a Large-Deviations Technique

We develop an approximation based on a large-deviations technique (LDT). Let $\widehat{X}_i \stackrel{\text{def}}{=} 1 - X_i$ denote the loss factor experienced by supplier i, $i = 1, \dots, N$, each with c.d.f. $\widehat{G}(\cdot)$, mean $q \equiv (1-p)$, and variance ς^2. Let $L(\theta) \stackrel{\text{def}}{=} \mathrm{E}[e^{\theta\widehat{X}}]$ and $l(\theta) \stackrel{\text{def}}{=} \ln \mathrm{E}[e^{\theta\widehat{X}}]$ denote the moment-generating function and cumulant-generating function of the loss factor distribution, which exist for all $\theta > 0$ because $\widehat{G}$ has bounded support. A cumulant-generating function is known to be convex. For example, because[4]

$$l'(\theta) = \frac{E_{\widehat{X}}[\widehat{X}e^{\theta\widehat{X}}]}{E_{\widehat{X}}[e^{\theta\widehat{X}}]}, \tag{4}$$

$l''(\theta) = \{E_{\widehat{X}}[\widehat{X}^2 e^{\theta\widehat{X}}]E_{\widehat{X}}[e^{\theta\widehat{X}}] - E^2_{\widehat{X}}[\widehat{X}e^{\theta\widehat{X}}]\}/E^2_{\widehat{X}}[\widehat{X}e^{\theta\widehat{X}}] \geqslant 0$, by the Cauchy-Schwarz inequality applied to the pair of two random variables $U = \widehat{X}e^{\theta\widehat{X}/2}$ and $V = e^{\theta\widehat{X}/2}$. (The interchange of the differentiation and expectation operators is justified in a similar way as in Endnote 4.) Because D has finite moments, there exists some $0 < \bar{\theta} \leqslant \infty$ such that the moment-generating function $M(\theta) \stackrel{\text{def}}{=} \mathrm{E}[e^{\theta\bar{D}}]$ and the cumulant-generating function $m(\theta) \stackrel{\text{def}}{=} \ln \mathrm{E}[e^{\theta\bar{D}}]$ exist for all $0 < \theta < \bar{\theta}$. $m(\theta) \geqslant \ln\{\Pr[D > I^0]\mathrm{E}[e^{\theta\bar{D}} \mid \bar{D} > 0]\} \geqslant \ln\alpha + \ln \mathrm{E}[e^{\theta\bar{D}} \mid \bar{D} > 0]$, so that $\lim_{\theta\uparrow\bar{\theta}} m(\theta) = \infty$. Thus, because $m(0) = 0$ and $m(\cdot)$, as a cumulant-generating function, is convex, it achieves every value in $\mathbb{R}^+$ exactly once. Thus, let $m^{-1}(\cdot)\colon \mathbb{R}^+ \to [0, \bar{\theta})$ denote its inverse function on the positive half-line.

Allowing for nonidentical order sizes, let: $w_i =$ the fraction of the total supply assigned to supplier i, $i = 1, \dots, N$. For $\theta > 0$ and using Markov's inequality, the shortfall probability can then be bounded by

$$\begin{aligned}
\Pr\left(Y\sum_{i=1}^n w_iX_i < \bar{D}\right) &= \Pr\left(Y - Y\sum_{i=1}^n w_i\widehat{X}_i < \bar{D}\right)\\
&= \Pr\left(\sum_{i=1}^n w_i\widehat{X}_i > 1 - \frac{\bar{D}}{Y}\right)\\
&= E_{\bar{D}}\left[\Pr\left(e^{\theta(\sum_{i=1}^n w_i\widehat{X}_i)} > e^{\theta(1-\bar{D}/Y)} \,\middle|\, \bar{D}\right)\right]\\
&\leqslant E_{\bar{D}}\left\{e^{-\theta}e^{\theta\bar{D}/Y}\right\}\prod_{i=1}^n L(w_i\theta)\\
&= e^{-\theta}M\left(\frac{\theta}{Y}\right)\prod_{i=1}^n L(w_i\theta).
\end{aligned} \tag{5}$$

Unlike the exact probability, this upper bound is minimized for any $Y > 0$ by spreading the total order equally across the suppliers: The problem $\min\{\prod_{i=1}^n L(w_i\theta)$ s.t. $\sum_{i=1}^n w_i = 1;\ w_i \geqslant 0\}$ is equivalent to

$$\min\left\{\sum_{i=1}^{n} l(w_i\theta) \text{ s.t.} \sum_{i=1}^{n} w_i = 1;\ w_i \geqslant 0\right\} = nl\left(\frac{\theta}{n}\right). \tag{6}$$

(Because $l(\cdot)$ is convex, $w_i^* = 1/n$ achieves the minimum in (6).) Thus, for any $\theta > 0$, the upper bound in (5) equals α, if Y satisfies the equation $e^{-\theta}M(\theta/Y)[L(\theta/n)]^n = \alpha$, or, equivalently,

$$m\left(\frac{\theta}{Y}\right) = \theta - nl\left(\frac{\theta}{n}\right) + \ln\alpha. \tag{7}$$

Theorem 2 (Upper-Bound Approximation). *Assume that $n > \underline{n}$. Let $\theta^0(n)$ be the unique root of the characteristic equation*

$$H_n(\theta) \stackrel{\text{def}}{=} \ln\alpha - nl\left(\frac{\theta}{n}\right) + \theta = 0. \tag{8}$$

(a) *For all $\theta > \theta^0(n)$, (7) has a unique root $\overline{Y}(\theta,n) = \theta/[m^{-1}(\theta - nl(\theta/n) + \ln\alpha)]$, which is convexly decreasing in n. Moreover, $Y^*(n) \leqslant \overline{Y}^*(n) \stackrel{\text{def}}{=} \min_{\theta>\theta^0}\overline{Y}(\theta,n) \leqslant \overline{Y}(\theta,n)$ for any $\theta > \theta^0(n)$.*

(b) *$\overline{Y}^*(n)$ is decreasing in n and α.*

(c) *If $G(0) > 0$, $\overline{Y}^*(n) = \overline{Y}(\theta^*, n)$, with θ^* the unique point θ where a sign change occurs for*

$$\frac{1}{\overline{Y}(\theta,n)}m'\left(\frac{\theta}{\overline{Y}(\theta,n)}\right) + l'\left(\frac{\theta}{n}\right) - 1. \tag{9}$$

(d) *Assume that $G(0) = 0$. One of the following two cases prevails*:

(i) *$\overline{Y}^*(n) = \overline{Y}(\theta^*, n)$, with $\theta^* \geqslant \theta^0(n)$, the unique point where a sign change occurs for* (9).

(ii) *$\overline{Y}(\theta,n)$ is monotonically decreasing and $\overline{Y}^*(n) = \lim_{\theta\uparrow\infty}\overline{Y}(\theta,n)$.*

(e) *Assume that D is Normal. For all $\theta > \theta^0(n)$,*

$$\overline{Y}(\theta,n) = \begin{cases} (\mu - I^0)\dfrac{\theta\left(\sqrt{1+2\gamma_{\bar D}^2 H_n(\theta)}+1\right)}{2H_n(\theta)} = (\mu - I^0)\dfrac{\theta\gamma_{\bar D}^2}{\sqrt{1+2\gamma_{\bar D}^2 H_n(\theta)}-1} & \text{if } I^0 < \mu, \\[2ex] \dfrac{\theta\sigma}{\sqrt{2H_n(\theta)}} & \text{if } I^0 = \mu, \\[2ex] (I^0 - \mu)\dfrac{\theta\left(\sqrt{1+2\gamma_{\bar D}^2 H_n(\theta)}-1\right)}{2H_n(\theta)} = (I^0 - \mu)\dfrac{\theta\gamma_{\bar D}^2}{\sqrt{1+2\gamma_{\bar D}^2 H_n(\theta)}+1} & \text{if } I^0 > \mu, \end{cases} \tag{10}$$

which is minimized in the unique point θ^* $(>\theta^0)$ where the function $\gamma_{\bar D}^{-2} + 2H_n(\theta) - \gamma_{\bar D}^{-1}\sqrt{\gamma_{\bar D}^{-2} + 2H_n(\theta)} - \theta + \theta l'(\theta/n)$ changes signs. In particular, $\overline{Y}^*(n)$ is of the form $\overline{Y}^*(n) = |\mu - I^0|\bar\Psi_n(\gamma_{|\bar D|})$, with $\bar\Psi_n(\cdot)$ an increasing function.

Proof. (a) H_n has at most one root because, by (4), $H_n'(\theta) = -l'(\theta/n) + 1 = 1 - E_{\widehat X}[\widehat X e^{(\theta/n)\widehat X}]/E_{\widehat X}[e^{(\theta/n)\widehat X}] > 0$, because $E_{\widehat X}[e^{(\theta/n)\widehat X}] = \widehat G(0) + \int_0^1 e^{(\theta/n)u}d\widehat G(u) \geqslant \widehat G(0) + \int_0^1 ue^{(\theta/n)u}d\widehat G(u) \geqslant E_{\widehat X}[\widehat X e^{(\theta/n)\widehat X}]$ and at least one of the two inequalities must be strict. Moreover, $H_n(0) = \ln\alpha < 0$ and

$$\begin{aligned} H_n(\theta) &\stackrel{\text{def}}{=} \ln\alpha - n[l(\theta/n) - \ln e^{\theta/n}] \\ &= \ln\alpha - n\ln[L(\theta/n)/e^{\theta/n}] \\ &= \ln\alpha - n\ln\left[\left(G(0)e^{\theta/n} + \int_0^1 e^{(\theta/n)u}d\widehat G(u)\right)\Big/ e^{\theta/n}\right] \\ &= \ln\alpha - n\ln\left[G(0) + \int_0^1 e^{(\theta/n)(u-1)}d\widehat G(u)\right]. \end{aligned}$$

Because $\lim_{\theta\uparrow\infty}\int_0^1 e^{(\theta/n)(u-1)}d\widehat G(u) = 0$, by the dominated convergence theorem, we have

$$\begin{aligned} &G(0) = 0 \;\Rightarrow\; \lim_{\theta\uparrow\infty} H_n(\theta) = \infty, \\ &G(0) > 0 \;\Rightarrow\; \lim_{\theta\uparrow\infty} H_n(\theta) = \ln\alpha - n\ln G(0) > 0 \quad \text{for } n > \underline{n}. \end{aligned} \tag{11}$$

Thus, the strictly increasing function $H_n(\theta)$ starts at a negative value and is positive for θ sufficiently large, so that it has a unique root θ^0. For $\theta > \theta^0$, Y satisfies (7) if and only if $m(\theta/Y) = H_n(\theta) \Leftrightarrow \theta/Y = m^{-1}(H_n(\theta)) \Leftrightarrow Y = \theta/m^{-1}(H_n(\theta))$ because $H_n(\theta) > 0$ and $m^{-1}(\cdot)$ is defined on $\mathbb{R}^+$. Next, write $\overline{Y}(\theta,n) = r(m^{-1}(H_n(\theta)))$, with $r(x) = \theta/x$. The left-hand side of (6) is clearly decreasing in n because the solution $w_i = 1/n$, $i = 1,\ldots,n$ and $w_{n+1} = 0$ is optimal when there are n suppliers and feasible when there are $n+1$. This implies that $H_n(\theta)$ is increasing in n, and it is concave in n by the convexity of $l(\cdot)$; see Boyd and Vandenberghe (2004, §3.2.6, p. 89). As shown above, $m(\cdot)$, as a cumulant-generating function, is convex and increasing when positive. Thus, $m^{-1}(\cdot)$ is concave and increasing on $\mathbb{R}^+$, and so is the composition $m^{-1}(H_n(\theta))$ as a function of n. Finally, because $r(\cdot)$ is convex and decreasing on $\mathbb{R}^+$, the composition $\overline{Y}(\theta,n) = r(m^{-1}(H_n(\theta)))$ is convex and decreasing in n as well. Finally, by (5), $\Pr(\overline{Y}(\theta,n)\sum_{i=1}^n w_iX_i < D) \leqslant \alpha$, so that $Y^*(n) \leqslant \overline{Y}(\theta,n)$ for all $\theta > \theta^0(n)$.

(b) Immediate from the monotonicity of $\overline{Y}(\theta,n)$ in n and α for all $\theta > \theta^0(n)$. (The monotonicity of $\overline{Y}(\theta,n)$ in α is immediate because $m^{-1}(\cdot)$ is an increasing function.)

(c) For $\theta \geqslant \theta^0(n)$, $\theta - nl(\theta/n) + \ln\alpha > 0$, and $m^{-1}(\theta - nl(\theta/n) + \ln\alpha)$ is concave in θ because the composition of a concave increasing function and a concave function, and $\overline{Y}(\theta,n) = \theta/[m^{-1}(\theta - nl(\theta/n) + \ln\alpha)]$ is quasi-convex in θ because for all $C > 0$, the level sets $\{\theta: \overline{Y}(\theta, n) \leqslant C\} = \{\theta: m^{-1}(\theta - nl(\theta/n) + \ln\alpha) - \theta/C \geqslant 0\}$ are convex sets because $m^{-1}(\theta - nl(\theta/n) + \ln\alpha) - \theta/C$ is concave in θ. This implies that $\overline{Y}(\theta, n)$ has at most one local minimum in θ. By (11), $\lim_{\theta\uparrow\infty} H_n(\theta) < \infty$, when $G(0) > 0$, so that by part (a), $\lim_{\theta\uparrow\infty} \overline{Y}(\theta, n) = \lim_{\theta\uparrow\infty} \theta/[m^{-1}(H_n(\theta))] = \infty$, while $\lim_{\theta\downarrow\theta^0} \overline{Y}(\theta, n) = \infty$, by (7). Thus, $\overline{Y}(\theta, n)$ has a unique local minimum θ^* ($>\theta^0$). Because $\overline{Y}$ is quasi-convex in θ, θ^* is the unique value where $\partial\overline{Y}/\partial\theta$ changes signs. Because $\theta/\overline{Y} = m^{-1}(\theta - nl(\theta/n) + \ln\alpha)$, $m(\theta/\overline{Y}) - \theta + nl(\theta/n) - \ln\alpha = 0$. By the implicit function theorem,

$$\frac{\partial\overline{Y}}{\partial\theta} = \left[\frac{\overline{Y}^2(\theta, n)}{\theta m'(\theta/\overline{Y}(\theta, n))}\right] \cdot\left[\frac{1}{\overline{Y}(\theta, n)} m'\left(\frac{\theta}{\overline{Y}(\theta, n)}\right) + l'\left(\frac{\theta}{n}\right) - 1\right]. \tag{12}$$

Thus, θ^* is the unique point where (9) changes sign.

(d) $\overline{Y}(\theta, n)$ is quasi-convex in θ, as shown in part (c). Thus, because $\lim_{\theta\downarrow\theta^0} \overline{Y}(\theta, n) = \infty$, either (i) $\overline{Y}(\theta, n)$ has a unique minimum θ^* where (9) changes signs; or (ii) $\overline{Y}(\theta, n)$ is monotonically decreasing in θ and $\overline{Y}^*(n) = \lim_{\theta\uparrow\infty} \overline{Y}(\theta, n)$.

(e) When D is Normal, $m(\theta) = (\mu - I^0)\theta + \sigma^2\theta^2/2$. Substituting this expression in (7), and multiplying both sides with Y^2, we get the quadratic equation $[\theta - nl(\theta/n) + \ln\alpha]Y^2 - (\mu - I^0)\theta Y - \sigma^2\theta^2/2 = 0$ for any given $\theta > \theta^0(n)$. By the proof of part (a), $H_n(\theta) = [\theta - nl(\theta/n) + \ln\alpha] > 0$ for $\theta > \theta^0(n)$, so that the quadratic equation has one positive root: $\overline{Y} = [(\mu - I^0)\theta + \theta\sqrt{(\mu - I^0)^2 + 2H_n(\theta)\sigma^2}]/2H_n(\theta)$. The expressions in (10) follow.

$\overline{Y}(\theta, n)$ cannot be monotonically decreasing in θ, for if it were, there exists a level Y^0 such that $\overline{Y}(\theta, n) \leqslant Y^0$ for all θ sufficiently large and, hence, for all θ sufficiently large, $m'(\theta/\overline{Y}(\theta, n))/\overline{Y}(\theta, n) + l'(\theta/n) - 1 \geqslant (\mu - I^0 + \sigma^2\theta/Y^0)/Y^0 - 1$ because $m'(\theta) = \mu - I^0 + \sigma^2\theta > 0$ for θ sufficiently large, and $l' \geqslant 0$, by (4). The second factor in (12) is therefore positive for all θ sufficiently large; because the first factor is positive for θ sufficiently large, $\partial\overline{Y}/\partial\theta > 0$ for all θ sufficiently large, contradicting the assumption that $\overline{Y}$ is monotonically decreasing in θ. By parts (c) and (d), $\overline{Y}(\theta, n)$ has a unique minimum θ^* ($>\theta^0$), where (9) changes signs. Substituting (10) and $m'(\theta) = \mu - I^0 + \sigma^2\theta$ into (9), we obtain the desired characterization of θ^*. Finally, for $I^0 < \mu$, (10) implies the representation $\overline{Y}^*(n) = (\mu - I^0)\bar{\Psi}_n(\gamma_{\bar{D}})$, with $\bar{\Psi}_n(\cdot) = \min_\theta[\theta(\sqrt{1 + 2\gamma_{\bar{D}}^2 H_n(\theta)} + 1)/(2H_n(\theta))]$ increasing in $\gamma_{\bar{D}}$. For $I^0 > \mu$, (10) shows that $\overline{Y}^*(n) = (I^0 - \mu)\bar{\Psi}_n(|\gamma_{\bar{D}}|)$, with $\bar{\Psi}_n(\cdot) = \min_\theta[\theta(\sqrt{1 + 2\gamma_{\bar{D}}^2 H_n(\theta)} - 1)/(2H_n(\theta))]$ increasing in $|\gamma_{\bar{D}}|$. □

Remark. As with the Normals, a closed-form expression for $\overline{Y}(\theta, n)$ can be obtained from (7) for many other classes of demand distributions. For example, when D has a Gamma(λ, ν) distribution with scale parameter λ and shape parameter ν, $m(\theta) = \nu(\ln\lambda - \ln(\lambda - \theta))$. The resulting Equation (7) can again be solved in closed form. Moreover, because $\lim_{\theta\uparrow\lambda} m'(\theta) = \infty$, it follows from the proof of part (e) that $\overline{Y}(\theta, n)$ has, again, a unique minimum θ^* where the function in (9) changes signs.

5. An Approximation Based on a Central Limit Theorem

A second approximation is obtained by writing the probability of a shortfall

$$\begin{aligned} &P_n(Y, w_1, \ldots, w_n) \\ &\quad \stackrel{\text{def}}{=} \Pr\left(Y - Y\sum_{i=1}^n w_i\widehat{X}_i < \bar{D}\right) = \Pr\left(\sum_{i=1}^n w_i\widehat{X}_i > 1 - \bar{D}\Big/Y\right) \\ &\quad = \Pr\left[\frac{\sum_{i=1}^n (nw_i)\widehat{X}_i - nq}{nW(n)\varsigma} > \frac{(1 - \bar{D}/Y) - q}{\varsigma W(n)}\right], \end{aligned}$$

where $W(n) \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^n w_i^2}$. In Theorem 3 below, we invoke a central limit theorem to replace the left-hand side of the last inequality defining the shortfall event by a random variable U with a standard Normal distribution (independent of D). This gives rise to the approximate shortfall probability: $\widetilde{P}_n(Y, w_1, \ldots, w_n) \stackrel{\text{def}}{=} \Pr[\bar{D}/Y > p - UW(n)\varsigma]$. When dealing with an ever larger set of potential suppliers, the share of the aggregate order that is assigned to any given supplier will, under a given allocation scheme, depend on the total number of suppliers. We therefore, henceforth, write $w_{i,n}$ to represent supplier i's share when considering the set of suppliers $\{1, \ldots, n\}$. For example, the allocation scheme that splits the aggregate order equally among all potential suppliers, has $w_{i,n} = 1/n$ for all $i = 1, \ldots, n$. The next theorem shows that the approximate probability is asymptotically accurate and that the order quantity determined with this approximation is asymptotically optimal.

Theorem 3 (CLT-Based Approximation). (a) *Consider an arbitrary allocation scheme* $\{w_{i,n}\}$ *with*

$$\frac{\max_i w_{i,n}}{\min_i w_{i,n}} \leqslant A \quad \text{for some constant } A, \tag{13}$$

i.e., the ratio of the largest and the smallest order size remains bounded as $n \to \infty$. *For any* $0 < \eta \leqslant \frac{1}{2}$, *there exists a constant* C_η *such that for all* Y,

$$\left|P_n(Y, w_{1,n}, \ldots, w_{n,n}) - \widetilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n})\right| \leqslant C_\eta n^{-\eta}.$$

In particular, $\lim_{n\to\infty}[P_n(Y, w_{1,n}, \ldots, w_{n,n}) - \widetilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n})] = 0$.

(b) *Let* $\widetilde{Y}(n \mid \mathbf{w}) \stackrel{\text{def}}{=} \min\{Y \mid \widetilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n}) \leqslant \alpha\}$. *Then,* $\lim_{n\to\infty} \widetilde{Y}(n \mid \mathbf{w}) = [F^{-1}(1 - \alpha) - I^0]/p = \lim_{n\to\infty} Y^*(n)$ *for any allocation scheme* $\{w_{i,n}\}$ *that satisfies* (13).

Proof. (a) Let $\Xi_i \stackrel{\text{def}}{=} nw_{i,n}(\widehat{X}_i - q)$. Note that $E\Xi_i = 0$ and $E\Xi_i^2 = n^2 w_{i,n}^2 \varsigma^2$, so that $s_n^2 \stackrel{\text{def}}{=} \sum_{i=1}^n E\Xi_i^2 = n^2\varsigma^2 \sum_{i=1}^n w_{i,n}^2 = \varsigma^2 \sum_{i=1}^n (w_{i,n}/n^{-1})^2 = \varsigma^2 \sum_{i=1}^n (w_{i,n}/(1/n)\sum_{j=1}^n w_{j,n})^2 \geqslant n\varsigma^2 A^{-2}$. Also, for any $0 < \delta \leqslant 1$, let $m_{2+\delta} \stackrel{\text{def}}{=} (E|\widehat{X} - q|^{2+\delta})^{1/(2+\delta)} < \infty$ because $\widehat{X}$, with bounded support, has finite moments:

$$\Gamma_n \stackrel{\text{def}}{=} \left(\sum_{i=1}^n E\Xi_i^{2+\delta}\right)^{(2+\delta)^{-1}} = m_{2+\delta}\left(\sum_{i=1}^n (nw_{i,n})^{2+\delta}\right)^{(2+\delta)^{-1}}$$
$$= m_{2+\delta}\left[\sum_{i=1}^n \left(\frac{w_{i,n}}{(1/n)\sum_{j=1}^n w_{j,n}}\right)^{2+\delta}\right]^{(2+\delta)^{-1}}$$
$$\leqslant A m_{2+\delta} n^{1/(2+\delta)}.$$

Then, there exists a constant $\widetilde{C}_\delta$ such that, with $\widetilde{F}(\cdot)$ the c.d.f. of $\bar{D}$,

$$\left|P_n(Y, w_{1,n}, \ldots, w_{n,n}) - \widetilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n})\right|$$
$$\leqslant \int \left| \Pr\left[\sum_{i=1}^n (nw_{i,n})(\widehat{X}_i - q) > n\left(1 - \frac{\bar{D}}{Y}\right) - nq \,\middle|\, \bar{D} = d\right] - \Pr\left[U > \frac{p - \bar{D}/Y}{W(n)\varsigma} \,\middle|\, D = d\right]\right| d\widetilde{F}(d)$$
$$= \int \left|\Pr\left[\sum_{i=1}^n \Xi_i > s_n \frac{p - d/Y}{W(n)\varsigma}\right] - \Pr\left[U > \frac{p - d/Y}{W(n)\varsigma}\right]\right| d\widetilde{F}(d)$$
$$\leqslant \int \sup_x \left|\Pr\left[\sum_{i=1}^n \Xi_i > s_n x\right] - \Pr[U > x]\right| d\widetilde{F}(d)$$
$$\leqslant \widetilde{C}_\delta \frac{\Gamma_n^{2+\delta}}{s_n^{2+\delta}} \leqslant \widetilde{C}_\delta \frac{A^{2+\delta} m_{2+\delta}^{2+\delta} n}{A^{-2-\delta}\varsigma^{2+\delta} n^{1+\delta/2}}$$
$$= \left\{\widetilde{C}_\delta A^{4+2\delta} m_{2+\delta}^{2+\delta} \varsigma^{-2-\delta}\right\} n^{-\delta/2} \stackrel{\text{def}}{=} C_{\delta/2} n^{-\delta/2},$$

where the next-to-last inequality follows from the Berry-Esseen theorem, for example, Theorem 3 on p. 322 in Chow and Teicher (1997), which applies in view of (13).

(b) As in the proof of Theorem 1(d), let $\widehat{F}_n$ be the c.d.f. of $\bar{D}/(\sum_{i=1}^n w_{i,n} X_i)$. $W^2(n) = \sum_{i=1}^n w_{i,n}^2 = \sum_{i=1}^n (1/n^2)(w_{i,n}/(\sum_{j=1}^n w_{j,n})/n)^2 \leqslant A^2/n$, so that $W(n) \leqslant A/\sqrt{n}$. It follows that the standard deviation of $(\sum_{i=1}^n w_{i,n} X_i) = W(n)\varsigma \leqslant A\varsigma/\sqrt{n}$, so that $(\sum_{i=1}^n w_{i,n} X_i) \to p$ in probability, for example, by the Chebycheff inequality. It follows that $\widehat{F}_n$ converges pointwise to $\widehat{F}$, the c.d.f. of $\bar{D}/p$, and $\lim_{n\to\infty} Y^*(n) = \lim_{n\to\infty} \widehat{F}_n^{-1}(1-\alpha) = \widehat{F}^{-1}(1-\alpha) = (F^{-1}(1-\alpha) - I^0)/p$, where the second equality follows, again, from pointwise convergence of the generalized inverse c.d.f.s and the continuity of $\widehat{F}$; see Thorisson (2000, p. 24). The proof of $\lim_{n\to\infty} \widetilde{Y}(n \mid \mathbf{w}) = (F^{-1}(1-\alpha) - I^0)/p$ is analogous, replacing $\widehat{F}_n$ by $\widetilde{F}_n$, the c.d.f. of $\bar{D}/(p - UW(n)\varsigma)$ because $(p - UW(n)\varsigma) \to p$ a.s. $\square$

The approximate shortfall probability $\widetilde{P}_n(Y, \mathbf{w}) = \Pr[\bar{D}/Y + UW(n)\varsigma > p]$ is considerably simpler to calculate than the exact probability because it requires the convolution of only two random variables. The CLT-based approximation thus has the advantage of being asymptotically accurate for any order size and any allocation scheme that satisfies (13). Moreover, the order quantity derived under this CLT approximation is, under any such allocation scheme, asymptotically optimal.[5]

We now show that the CLT approximation allows for a simple characterization of when a set of suppliers is feasible. With Normal demands, it also permits a closed-form expression of the optimal order quantities. We develop these results, here, for the special case where the starting inventory $I^0 = 0$. The general case, with $I^0 \geqslant 0$, is developed in §6, in conjunction with the extension to general, nonidentical suppliers. If $I^0 = 0$, we have that for a general demand distribution, with support on the positive half-line, $\widetilde{P}_n(Y, \mathbf{w})$ is continuously decreasing in Y for any allocation scheme $\mathbf{w}$. If $R(n \mid \mathbf{w}) \stackrel{\text{def}}{=} [\gamma^2(\sum_{i=1}^n w_{i,n}^2)]^{-1} < z_\alpha^2$, no feasible order quantity exists because, by the dominated convergence theorem, $\lim_{Y\to\infty} \widetilde{P}_n(Y, \mathbf{w}) = \Pr[UW(n)\varsigma > p] > \alpha$. The condition is equivalent to $\varsigma\sqrt{\sum_{i=1}^n w_{i,n}^2}/p > z_\alpha^{-1}$, the left-hand side of which represents the c.v. of the aggregate supply. Thus, as long as this c.v. is in excess of z_α^{-1}, no feasible solution exists. At the same time, for any given allocation scheme $\mathbf{w}$, if $R(n \mid \mathbf{w}) > z_\alpha^2$, the optimal total order quantity Y^* is the unique root of the equation

$$\Pr\left[\frac{D}{Y} + \frac{Up}{\sqrt{R(n \mid \mathbf{w})}} > p\right] = \alpha. \tag{14}$$

Conversely, for any given total order quantity Y, all allocation schemes $\mathbf{w}$ result in the same cost value, but only some may be feasible, i.e., satisfy the service constraint. The random variable $\{D/Y + UW(n)\varsigma\}$ has the same mean μ/Y for all allocation vectors $\mathbf{w}$, and its variance $(\sigma^2/Y^2 + W(n)^2\varsigma^2)$ is clearly minimized when splitting the order equally among the suppliers, i.e., under the allocation scheme $\mathbf{w}^*$, with $w_{i,n}^* = 1/n$, $i = 1, \ldots, n$, for which

$$R(n \mid \mathbf{w}^*) = \max_{\mathbf{w}} R(n \mid \mathbf{w}). \tag{15}$$

(The convex function $\sum_{i=1}^n w_{i,n}^2$ attains its minimum on $\mathscr{W} = \{w_{i,n} \geqslant 0\colon \sum_{i=1}^n w_{i,n} = 1\}$ for $w_{i,n} = 1/n$.) However, for a general demand distribution, the variance-minimizing allocation vector does not necessarily minimize the complementary c.d.f. value in the point p. Indeed, nonidentical allocations may minimize the CLT-based shortfall probability $\widetilde{P}_n$, even though Theorem 3(b) shows that equal allocations are asymptotically optimal (along with any other allocation scheme satisfying (13)). A combined optimal order quantity $\widetilde{Y}^*(n)$ and corresponding allocation scheme $\widehat{\mathbf{w}}$ can be found by the following numerical procedure: note that $R(n \mid \mathbf{w})$ may adopt any value in $\gamma^{-2}(1, n]$ because the convex, and hence continuous, function $\sum_{i=1}^n w_{i,n}^2$ has a minimum value of $1/n$ and a maximum value of one on the compact set $\mathscr{W}$. Thus, feasible orders exist for any $R \in (\max(\gamma^{-2}, z_\alpha^2), \gamma^{-2} n]$, and the optimal aggregate order size can be found by minimizing the unique root of (14) when varying R over this interval. Assuming the minimum is achieved for $R = R^*$, an associated allocation scheme $\widehat{\mathbf{w}}$ can be found as follows: Let $\mathbf{w}^1$ be the allocation scheme that assigns the entire order to supplier 1 and $\mathbf{w}^2$ the allocation scheme that spreads the order equally among all n suppliers. Let $\widehat{\mathbf{w}} = \beta\mathbf{w}^2 + (1-\beta)\mathbf{w}^1$, i.e., $\widehat{w}_{1,n} = \beta/n + (1-\beta)$ and $\widehat{w}_{i,n} = \beta/n$, $i = 2, \ldots, n$. It is easily verified that $\sum_{i=1}^n (\widehat{w}_{i,n})^2 = \beta^2/n^2 + (1-\beta)^2 + 2\beta(1-\beta)/n + \beta(n-1)/n^2$. Equating this expression to $1/(\gamma^2 R^*)$ and multiplying both sides by n^2, we obtain after some algebra the quadratic equation $\beta^2 - \beta(2n-1)/(n-1) + n^2(1-\gamma^{-2}R^{*-1})/(n-1)^2 = 0$. The only root between zero and one is $\beta = (2n-1)/[2(n-1)] - \sqrt{4n^2/(\gamma^2 R^*) - 4n + 1}/[2(n-1)]$. (Because $1/n \leqslant 1/(\gamma^2 R^*) < 1$, the discriminant $[4n^2/(\gamma^2 R^*) - 4n + 1]/[(n-1)^2]$ lies between $[1/(n-1)^2, (2n-1)^2/(n-1)^2)$.)

When D is Normal, the random variable $\{D/Y + UW(n)\varsigma\}$ is itself Normal, so that $\widetilde{P}_n(Y, \mathbf{w}) = \bar{\Phi}((p - \mu/Y)/\sqrt{\sigma^2/Y^2 + W(n)^2\varsigma^2})$, which is minimized by minimizing the expression in the denominator of $\bar{\Phi}(\cdot)$ for any Y such that $(p - \mu/Y) > 0$, i.e., for $Y > \mu/p$. (Note that $\widetilde{P}_n(Y, \mathbf{w}) \geqslant 0.5$ if $Y \leqslant \mu/p$; thus, the desired aggregate order quantity $Y > \mu/p$ because $\alpha \leqslant 0.5$.) Thus, with Normal demand, it is optimal, under the CLT-based approximation, to split the aggregate order equally among the suppliers. This permits us to obtain a closed-form expression for $\widetilde{Y}^*(n)$ and to show that it is convexly decreasing in n.

Corollary 2. *Assume that $I^0 = 0$.*

(a) *For any demand distribution $F(\cdot)$ with support on the positive half-line, if $n < z_\alpha^2\gamma^2$, irrespective of which allocation scheme $\mathbf{w}$ is used, there is no $\widetilde{Y}(n \mid \mathbf{w})$ that satisfies the service constraint*

$$\widetilde{P}_n(\widetilde{Y}, w_{1,n}, \ldots, w_{n,n}) \leqslant \alpha; \tag{16}$$

if $n > z_\alpha^2\gamma^2$, $\widetilde{Y}^*(n)$ is decreasing in n.

(b) *Assume that D is Normal. If $n < z_\alpha^2\gamma^2$, irrespective of which allocation scheme $\mathbf{w}$ is used, there is no $\widetilde{Y}(n \mid \mathbf{w})$ that satisfies the service constraint* (16). *If $n > z_\alpha^2\gamma^2$, then equal-size orders are optimal and*

$$\begin{aligned}\widetilde{Y}^*(n) &= \frac{\mu}{p(1 - z_\alpha^2\gamma^2/n)}\left(1 + z_\alpha\sqrt{\gamma_D^2\left(1 - \frac{z_\alpha^2\gamma^2}{n}\right) + \frac{\gamma^2}{n}}\right)\\ &= \mu\widetilde{\Psi}_n(\gamma_D),\end{aligned} \tag{17}$$

with $\widetilde{\Psi}_n(\cdot)$ and $\widetilde{Y}^*(n)$ convex in γ_D, decreasing in n, and convexly decreasing in n if $\gamma_D \leqslant 2\sqrt{3}/z_\alpha$.

Proof. (a) Immediate from the above procedure to calculate $\widetilde{Y}^*(n)$ because the feasible range for R is empty when $n < z_\alpha^2\gamma^2$ and increasing in n when $n > z_\alpha^2\gamma^2$.

(b) Because

$$\widetilde{P}_n(Y, \mathbf{w}) = \bar{\Phi}\frac{p - \mu/Y}{\sqrt{\sigma^2/Y^2 + W(n)^2\varsigma^2}}$$

is decreasing in Y for $Y > \mu/p$, the proof of part (a) can be used to show that no feasible $\widetilde{Y}(n \mid \mathbf{w})$ exists if $n < z_\alpha^2\gamma^2$, and that $\widetilde{Y}^*(n)$ is decreasing in n if $n > z_\alpha^2\gamma^2$. If $n > z_\alpha^2\gamma^2$, because $w_{1,n}^* = \cdots = w_{n,n}^* = 1/n$ (see (15)), we show that a unique value Y exists with

$$\frac{p - \mu/Y}{\sqrt{\sigma^2/Y^2 + \varsigma^2/n}} = \frac{pY - \mu}{\sqrt{\sigma^2 + \varsigma^2(Y^2/n)}} = z_\alpha, \tag{18}$$

and hence $\widetilde{P}_n(Y, \mathbf{w}^*) = \alpha$. Squaring both sides of the equation, we obtain the quadratic equation $(pY - \mu)^2 = z_\alpha^2(\sigma^2 + \varsigma^2 Y^2/n)$, only the larger root of which satisfies the original equation (18). (The smaller root $\widetilde{Y}'(n)$ solves the equation $pY - \mu = -z_\alpha\sqrt{\sigma^2 + \varsigma^2 Y^2/n}$.) Grouping corresponding terms, the quadratic equation can be written as $(p^2 - z_\alpha^2\varsigma^2/n)Y^2 - 2p\mu Y + \mu^2 - z_\alpha^2\sigma^2 = 0$ and

$$\begin{aligned}&\widetilde{Y}^*(n)\\ &= \left[p\mu + \sqrt{z_\alpha^2\varsigma^2\mu^2/n + z_\alpha^2\sigma^2p^2 - z_\alpha^4\varsigma^2\sigma^2/n}\right]\\ &\quad \Big/\left[p^2(1 - z_\alpha^2\gamma^2/n)\right]\\ &= \left[\mu + z_\alpha\sqrt{\gamma^2\mu^2/n + \sigma^2 - z_\alpha^2\gamma^2\sigma^2/n}\right]\Big/\left[p(1 - z_\alpha^2\gamma^2/n)\right]\\ &= \mu\left[1 + z_\alpha\sqrt{\gamma_D^2(1 - z_\alpha^2\gamma^2/n) + \gamma^2/n}\right]\Big/\left[p(1 - z_\alpha^2\gamma^2/n)\right].\end{aligned}$$

Because $n > z_\alpha^2\gamma^2$, $\widetilde{\Psi}_n(\cdot)$ is increasing in γ_D; convexity in γ_D follows by simple calculus: any function of the form $\sqrt{A\gamma_D^2 + B}$ with $A, B > 0$ has a second-order derivative $AB(A\gamma_D^2 + B)^{-3/2} > 0$. Writing

$$\widetilde{Y}^*(n) = \frac{\mu}{p}\left(\frac{1}{1 - z_\alpha^2\gamma^2/n} + z_\alpha\sqrt{\frac{\gamma_D^2}{1 - z_\alpha^2\gamma^2/n} + \frac{\gamma^2/n}{(1 - z_\alpha^2\gamma^2/n)^2}}\right),$$

one verifies that each of the terms, inside and outside the square root, is decreasing in n, so that $\widetilde{Y}^*(n)$ is decreasing in n. We refer to the online appendix for the convexity proof. An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/. □

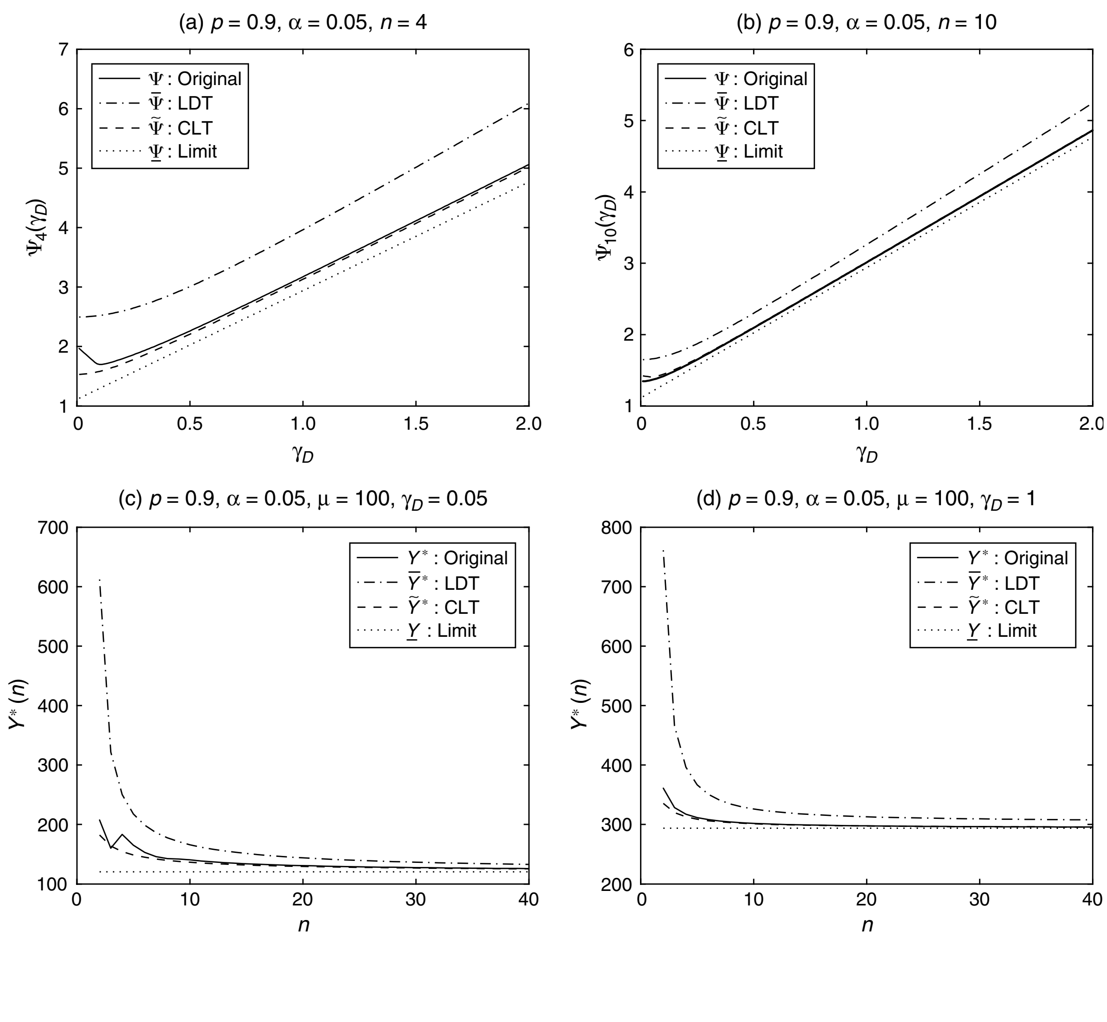
As with the exact analysis (see Theorem 1(a)), Corollary 2(a) identifies a minimum value of suppliers to ensure a feasible solution. At the same time, the condition $n > z_\alpha^2\gamma^2$ is hardly restrictive. After all, in most applications, $p \geqslant 0.5$, so that $\gamma \leqslant \sqrt{(1-p)/p} \leqslant 1$, where the first inequality follows from the fact that for a given mean p, the variance of X is maximized by the two-point distribution with $\Pr[X = 1] = p$ and $\Pr[X = 0] = 1 - p$; see Müller and Stoyan (2002, p. 57, Example 1.10.5). If $\alpha = 0.05$, the condition is satisfied for $n \geqslant 3$, and if $\alpha = 0.025$ for $n \geqslant 4$, even if each supplier is very unreliable with p close to 0.5. The condition $\gamma_D \leqslant 2\sqrt{3}/z_\alpha$ is without any loss of practical generality. Even if the shortfall probability $\alpha = 10^{-3}$, the condition is satisfied as long as $\gamma_D \leqslant 3.4641/3.0902 \approx 1.12$; to avoid a large probability of negative demands, the Normal distribution is only appropriate when its c.v. is significantly below one.

Recall that both $Y^*(n)$ and $\widetilde{Y}^*(n)$ decrease to an asymptotic limit $(\mu + z_\alpha\sigma)/p = \mu(1+z_\alpha\gamma_D)/p$. This reflects an "ideal" situation where one can count on a fraction p of the orders to become available. When the orders are spread over n unreliable suppliers, the required total order is the same as that under a fully reliable yield with a lower yield factor $p' = p(1-z_\alpha^2\gamma^2/n)$ and a higher c.v. for the demand $\gamma_D' = \gamma_D\sqrt{1+(1/\gamma_D^2) - z_\alpha^2(\gamma^2/n)}$ (see Endnote 1). Note that both p' and γ_D' approach p and γ_D as n increases.

We have seen that when the demand distribution is Normal, both the LDT-based and the CLT-based approximations $\overline{Y}^*(n)$ and $\widetilde{Y}^*(n)$ can be written in the form $\overline{Y}^*(n) = \mu\bar{\Psi}_n(\gamma_D)$ and $\widetilde{Y}^*(n) = \mu\widetilde{\Psi}_n(\gamma_D)$. The same applies to the exact order quantity $Y^*(n)$ because the integral in (2) can be written as $\int_0^n \Phi((yu-\mu)/\sigma)\,dG^{(n)}(u) = \int_0^n \Phi((Yu/(n\mu)-1)/\gamma_D)\,dG^{(n)}(u)$. Thus, $\Psi_n(\cdot)$, $\bar{\Psi}_n(\cdot)$, and $\widetilde{\Psi}_n(\cdot)$ all depend on the parameters μ and σ only via their ratio $\gamma_D = \sigma/\mu$. In Figures 1(a) and 1(b), we display these functions, along with the asymptotic limit $\underline{\Psi}(\gamma_D) \stackrel{\text{def}}{=} (1+z_\alpha\gamma_D)/p$, for instances with Bernoulli yields with $p = 0.9$ and $\alpha = 0.05$. Figure 1(a) (1(b)) displays the curves for $n = 4$ (10) suppliers. Note that the CLT-based approximation is remarkably close to the exact solution even for small values of n, and certainly when n is larger, as in Figure 1(b). Convergence of the CLT-based approximation to the asymptotic limit is also quite rapid. The gap for the upper-bound curve $\bar{\Psi}_n(\cdot)$ decreases as n increases, but not necessarily to zero. Figures 1(c) and 1(d) display the exact total order $Y^*(n)$, the two approximations $\overline{Y}^*(n)$ and $\widetilde{Y}^*(n)$ along with the asymptotic limit $\underline{Y} \stackrel{\text{def}}{=} \mu(1+z_\alpha\gamma_D)/p$ for two values of γ_D: $\gamma_D = 0.05$ and $\gamma_D = 1$. The accuracy of, in particular, the CLT-based approximation, again, is remarkable even for small values of n. The above observations are robust across the spectrum of possible parameter combinations. We recommend the CLT-based approximation as the estimate for the optimal order size, but suggest the LDT-approximation to generate a rigorous bound.

Finally, considering total, i.e., fixed plus variable procurement costs, the optimal number of suppliers n^* is easily determined when using the LDT-based upper bound approximation. For any given $\theta > \theta^0$, Theorem 2(a) shows

Figure 1. Total order quantity: Exact and approximate solutions.



that $\bar{Y}(n, \theta)$ is convexly decreasing in n, so that $\bar{n}^*(\theta) = \min\{n: \bar{Y}(n, \theta) - \bar{Y}(n+1, \theta) \leqslant K\}$. (We have observed, numerically, that the best approximation $\bar{Y}^*(n)$ is, similarly, always convex in n, even though a formal proof for this property remains outstanding.) The same simple characterization for the optimal number of suppliers applies under the CLT-based approximation, when the demand distribution is Normal; see Corollary 2(b). In contrast, when the exact solution $Y^*(n)$ is used, under equal order sizes, all possibilities $n = 1, \ldots, N$ need to be evaluated because, as mentioned, $Y^*(n)$ may fail to be monotone, let alone convexly decreasing. (We also show in §3 that the globally optimal exact aggregate order size is decreasing in n, but it is unknown whether this quantity is convexly decreasing.)

6. Supplier-Dependent Yield Distributions and Fixed Costs

In this section, we consider the case where the suppliers are differentiated in terms of their yield distributions $\{G_i\}$ and their fixed-order costs $\{K_i\}$. We again first consider, for a preselected group of n suppliers (without loss of generality, suppliers $1, \ldots, n$), which set of orders minimizes variable procurement costs subject to meeting the service-level constraint. As mentioned, with nonidentical suppliers, it may be with considerable loss of optimality to order equal quantities from each of a given set of suppliers. If all possible allocation schemes need to be considered, it is no longer practical to determine the optimal order quantities on the basis of the exact shortfall probability. However, this can be done when replacing the exact shortfall probability by either the LDT-based or the CLT-based approximation, both appropriately generalized to allow for general nonidentical suppliers. Online Appendix B gives a brief discussion of the former, whereas the remainder of this section is devoted to the latter.

Consider first an arbitrary scheme $\{w_{i,n}\}$ to allocate a given total order Y among the n suppliers. Using the derivation preceding Theorem 3, the CLT-based approximation for the exact shortfall probability $P_n(Y, w_{1,n}, \ldots, w_{n,n})$ is now given by

$$\tilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n}) = \Pr\left[\frac{\bar{D}}{Y} > \sum_{i=1}^{n} p_i w_{i,n} - U\sqrt{\sum_{i=1}^{n} w_{i,n}^2 \varsigma_i^2}\right]. \tag{19}$$

The asymptotic accuracy of this CLT-based approximation and the asymptotic optimality of the order quantity Y determined on the basis of this approximation can both be derived, as a simple generalization of Theorem 3, for any allocation scheme $\mathbf{w}$ that satisfies (13). As a consequence, we confine ourselves to allocation schemes that satisfy (13). We merely assume that the standard deviations of the yield factors are uniformly bounded away from zero. (If any of the suppliers, i, is completely reliable, i.e., $\varsigma_i = 0$, it is, in the absence of fixed costs, clearly optimal to use this supplier as the single source and order a quantity $[F^{-1}(1-\alpha) - I^0]/p_i$ because one only pays for the effectively produced units.) Let $Y^*(n \mid \mathbf{w}) \overset{\text{def}}{=} \min\{Y \mid P_n(Y, w_{1,n}, \ldots, w_{n,n}) \leqslant \alpha\} = \arg\min_Y\{\sum_{i=1}^n p_i w_{i,n} Y \mid P_n(Y, w_{1,n}, \ldots, w_{n,n}) \leqslant \alpha\}$ denote the total order quantity that minimizes variable procurement costs under the service-level constraint, and let $Y_E^*(n \mid \mathbf{w}) \overset{\text{def}}{=} \sum_{i=1}^n p_i w_{i,n} Y^*(n \mid \mathbf{w})$ denote the expected available supply under this set of orders. Let $\tilde{Y}(n \mid \mathbf{w})$ and $\tilde{Y}_E(n \mid \mathbf{w})$ denote the same quantities based on the CLT-approximation $\tilde{P}_n$ for the shortfall probability. Also, let $R(n \mid \mathbf{w}) \overset{\text{def}}{=} (\sum_{i=1}^n w_{i,n} p_i)^2 / (\sum_{i=1}^n w_{i,n}^2 \varsigma_i^2)$. and $\bar{R}(n) \overset{\text{def}}{=} \max\{R(n \mid \mathbf{w}): \sum_{i=1}^n w_{i,n} = 1;\ \mathbf{w} \geqslant 0\}$. In the case of identical suppliers, $R(n \mid \mathbf{w})$ reduces to the function $R(n \mid \mathbf{w}) = [\gamma^2(\sum_{i=1}^n w_{i,n}^2)]^{-1}$, employed in §5. As in §5, let $\mathbf{w}^*$ denote the allocation scheme that maximizes $R(n \mid \mathbf{w})$. In online Appendix C, we show that $\bar{R}(n) = R(n \mid \mathbf{w}^*) = \sum_{i=1}^n (p_i/\varsigma_i)^2 = \sum_{i=1}^n \gamma_i^{-2}$, where

$$w_{i,n}^* = \frac{p_i/\varsigma_i^2}{\left(\sum_{j=1}^n p_j/\varsigma_j^2\right)}, \quad i = 1, \ldots, n. \tag{20}$$

Theorem 4 (CLT-Based Approximation for Nonidentical Suppliers). *Consider an arbitrary allocation scheme* $\mathbf{w} = \{w_{i,n}\}$ *satisfying* (13). *Also, assume uniform positive lower bounds* $\underline{p}$ *and* $\underline{\varsigma}$ *such that* $p_i \geqslant \underline{p}$ *and* $\varsigma_i \geqslant \underline{\varsigma}$ *for all* $i = 1, 2, \ldots$.

(a) $\lim_{n\to\infty} [P_n(Y, w_{1,n}, \ldots, w_{n,n}) - \tilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n})] = 0$ *for all* Y.

(b) $\lim_{n\to\infty} Y_E^*(n \mid \mathbf{w}) = F^{-1}(1-\alpha) - I^0 = \lim_{n\to\infty} \tilde{Y}_E(n \mid \mathbf{w})$.

(c) *Assume that* D *has a general c.d.f.* $F(\cdot)$ *with support on the positive half-line. If* $R(n \mid \mathbf{w}) > z_\alpha^2$, $\tilde{Y}_E(n \mid \mathbf{w})$ *is the unique root of the equation*

$$\Pr\left[\frac{\bar{D}}{\tilde{Y}_E(n \mid \mathbf{w})} + \frac{U}{\sqrt{R(n \mid \mathbf{w})}} > 1\right] = \alpha. \tag{21}$$

(d) *A feasible solution exists, under* some *allocation scheme, if* $\bar{R}(n) > z_\alpha^2$.

Proof. (a) Identical to that of Theorem 3(a), replacing the identities $s_n^2 = n^2 \varsigma^2 \sum_{i=1}^n w_{i,n}^2$ by $s_n^2 \geqslant n^2 \underline{\varsigma}^2 \sum_{i=1}^n w_{i,n}^2$, and $m_{2+\delta} \overset{\text{def}}{=} (E|\hat{X} - q|^{2+\delta})^{1/(2+\delta)}$ by $m_{2+\delta} \overset{\text{def}}{=} \max_i\{(E|\hat{X}_i - q_i|^{2+\delta})^{1/(2+\delta)}\} < \infty$.

(b) $\Pr(\sum_{i=1}^n w_{i,n} X_i Y^*(n \mid \mathbf{w}) \geqslant \bar{D}) = 1 - \alpha$. If the first limit result fails, there exists a subsequence $\{n_k\}_{k=1}^\infty$ such that $F^{-1}(1-\alpha) - I^0 \neq \lim_{k\to\infty} \sum_{i=1}^{n_k} w_{i,n_k} p_i Y^*(n_k \mid \mathbf{w}) = L \leqslant \infty$. W.l.o.g., assume that $L < F^{-1}(1-\alpha) - I^0$. (The case where $L > F^{-1}(1-\alpha) - I^0$ is analogous.) Thus, for some $k^* \geqslant 1$ and for all $k \geqslant k^*$: $\sum_{i=1}^{n_k} w_{i,n_k} p_i Y^*(n_k \mid \mathbf{w}) \leqslant \frac{1}{2}(L + F^{-1}(1-\alpha) - I^0)$, so that

$$\begin{aligned} 1-\alpha &= \Pr\left(\sum_{i=1}^{n_k} w_{i,n_k} X_i Y^*(n_k \mid \mathbf{w}) \geqslant \bar{D}\right) \\ &= \Pr\left(\sum_{i=1}^{n_k} w_{i,n_k} p_i Y^*(n_k \mid \mathbf{w}) \geqslant \bar{D} \Big/ \frac{\sum_{i=1}^{n_k} w_{i,n_k} X_i}{\sum_{i=1}^{n_k} w_{i,n_k} p_i}\right) \\ &\leqslant \Pr\left(\frac{1}{2}(L + F^{-1}(1-\alpha) - I^0) \geqslant \bar{D} \Big/ \frac{\sum_{i=1}^{n_k} w_{i,n_k} X_i}{\sum_{i=1}^{n_k} w_{i,n_k} p_i}\right). \end{aligned} \tag{22}$$

Let $\Xi_i^{(n)} \stackrel{\text{def}}{=} n w_{i,n}(X_i - p_i)$. Note that $(X_i - p_i)^2 \leqslant 1$ a.s., so that

$$\varsigma_i = \sqrt{\mathrm{E}[(X_i - p_i)^2]} \leqslant 1 \tag{23}$$

Thus, $\sum_{i=1}^{\infty} \mathrm{E}[\Xi_i^{(n)}]^2/i^2 = \sum_{i=1}^{\infty} n^2 w_{i,n}^2 \varsigma_i^2/i^2 = \sum_{i=1}^{\infty}(w_{i,n}/(1/n)\sum_{j=1}^{n} w_{j,n})^2\varsigma_i^2/i^2 \leqslant A^2\sum_{i=1}^{\infty} i^{-2} < \infty$. The Strong Law of Large Numbers (see Chow and Teicher 1997, Theorem 1, p. 124), shows that $\lim_{n\to\infty}\sum_{i=1}^{n} w_{i,n}(X_i - p_i) = \lim_{n\to\infty}(1/n)\sum_{i=1}^{n}\Xi_i^{(n)} = 0$ a.s. Because $\sum_{i=1}^{n} w_{i,n}p_i \geqslant \underline{p}$, we have $\lim_{n\to\infty}(\sum_{i=1}^{n} w_{i,n}X_i - \sum_{i=1}^{n} w_{i,n}p_i)/\sum_{i=1}^{n} w_{i,n}p_i = 0$ a.s., i.e., $\lim_{n\to\infty}\sum_{i=1}^{n} w_{i,n}X_i/\sum_{i=1}^{n} w_{i,n}p_i = 1$ a.s. The right side of (22) converges to $F(\frac{1}{2}(L + F^{-1}(1-\alpha) - I^0) + I^0) < F(F^{-1}(1-\alpha)) = (1-\alpha)$, as $\{\bar{D}/((\sum_{i=1}^{n_k} w_{i,n_k}X_i)/(\sum_{i=1}^{n_k} w_{i,n_k}p_i))\}_{k=1}^{\infty}$ converges to $\bar{D}$ in distribution (see Chow and Teicher 1997, Corollary 2, p. 272). Because the left-hand side of (22) equals $(1-\alpha)$, this leads to a contradiction. The proof for the second limit result is analogous to that of Theorem 3(b).

(c) and (d) The proofs are analogous to the discussion following Theorem 3. □

Thus, the CLT-based approximation of the probability of a shortfall continues to be asymptotically accurate for any set of orders, and the aggregate order determined by this approximation is asymptotically optimal for any given allocation scheme $\mathbf{w}$ that satisfies (13). When n is large, the aggregate order again approaches what would be ordered from a single supplier with a completely reliable yield factor $p = \sum_{i=1}^{n} w_{i,n}p_i$ given by the weighted average of the expected yields of the suppliers. It is noteworthy that part (c) applies for any allocation scheme $\mathbf{w}$, not just those satisfying (13), i.e., those under which the asymptotic properties of the CLT-based approximation can be guaranteed. However, if, for some allocation scheme $\mathbf{w}$, $R(n \mid \mathbf{w}) > z_\alpha^2$, so that a feasible solution exists under the CLT approximation, it also exists under $\mathbf{w}^*$, the scheme that allocates orders in proportion to the mean-to-variance ratio of the suppliers' yield distribution. Note that in view of the uniform lower bounds for $\{p_i\}$ and $\{\varsigma_i\}$, $\mathbf{w}^*$ satisfies (13). As in the case of identical suppliers, for a feasible set of orders to exist under any given allocation scheme $\mathbf{w}$, it suffices to show that $R(n \mid \mathbf{w}) > z_\alpha^2$. In particular, for a feasible set of orders to exist, it suffices that $z_\alpha^2 < \bar{R}(n) = \sum_{i=1}^{n}\gamma_i^{-2}$, the number of BSEs represented by the set of suppliers. Under this condition, feasibility of a set of suppliers can be assessed on the basis of a single measure characterizing this supplier set, i.e., its number of BSEs. Moreover, the condition is independent of the shape of the demand distribution, its mean and standard deviation included. It is, in general, considerably harder to identify a simple necessary condition for the existence of a feasible set of orders. However, when $I^* = 0$, i.e., in the absence of a starting inventory, the inequality $R(n \mid \mathbf{w}) \geqslant z_\alpha^2$, is a necessary condition for the existence of a feasible set of orders under the allocation scheme $\mathbf{w}$, whereas as mentioned, the strict inequality $R(n \mid \mathbf{w}) > z_\alpha^2$ is sufficient as well. The proof that $R(n \mid \mathbf{w}) \geqslant z_\alpha^2$ is necessary is analogous to that of Corollary 2(a), itself based on the discussion proceeding that corollary. For a given scheme $\mathbf{w}$ with $R(n \mid \mathbf{w}) > z_\alpha^2$, the optimal aggregate order is given by the unique root of (21). The optimal aggregate order under the best possible allocation scheme can thus be found, for a general demand distribution, by minimizing the root $\widetilde{Y}_E(n \mid \mathbf{w})$ over the feasible range $(\max(\gamma_n^{-2}, z_\alpha^2), \sum_{i=1}^{n}\gamma_i^{-2}]$ for $R(n \mid \mathbf{w})$. If the minimum is achieved for $R(n \mid \mathbf{w}) = R^*$, say, a corresponding allocation scheme can again be found; see the discussion following (14).

In some settings, we may be concerned about the likelihood that the shortfall is in excess of some critical threshold x, not necessarily $x = 0$. Let $P_n(Y, w_{1,n}, \ldots, w_{n,n} \mid x)$ denote the exact probability of the shortfall being larger than x, and $\widetilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n} \mid x)$ the approximate probability based on the CLT-approximation, i.e., $\widetilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n} \mid x) = \Pr[(\bar{D} + x)/Y > \sum_{i=1}^{n} p_i w_{i,n} - U\sqrt{\sum_{i=1}^{n} w_{i,n}^2\varsigma_i^2}]$. The following corollary is immediate from the proof of Theorem 4(a).

Corollary 3. *Consider an arbitrary allocation scheme* $\mathbf{w}$ *satisfying* (13). *Also, assume uniform positive lower bounds* $\underline{p}$ *and* $\underline{\varsigma}$ *such that* $p_i \geqslant \underline{p}$ *and* $\varsigma_i \geqslant \underline{\varsigma}$ *for all* $i = 1, \ldots, n$. *The difference between the exact and approximate probability of the shortfall being in excess of an arbitrary level converges to zero*:

$$\lim_{n\to\infty}\left[P_n(Y, w_{1,n}, \ldots, w_{n,n} \mid x) - \widetilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n} \mid x)\right] = 0 \quad \textit{for all } Y \textit{ and all } x.$$

Turning our attention to the case of Normal demands, we now show that a simple closed-form inequality captures the necessary and sufficient condition for the existence of a feasible set of orders. Also, in this case, the optimal order quantities are available in closed form, and various qualitative insights can be gleaned from these closed-form expressions. For all $i = 1, \ldots, N$, let $y_i = w_{i,n}Y$ denote the order for supplier i, under an aggregate order Y and the allocation scheme $\mathbf{w}$. Let $Y_E = \sum_{i=1}^{n} p_i y_i = \sum_{i=1}^{n} p_i w_{i,n} Y$ denote the expected effective supply resulting from this order. Any set of orders can be characterized by the value Y_E and the allocation scheme $\mathbf{w}$ $(Y = Y_E/(\sum_{i=1}^{n} p_i w_i))$.

Theorem 5 (Normal Demand Distribution). *Assume that D is Normal with mean* μ *and standard deviation* σ. *Under the CLT-approximation*:

(a) *A feasible solution exists if and only if it exists under the allocation scheme* $\mathbf{w}^*$.

(b) *A feasible solution exists if and only if condition* (F) *is satisfied*:

(F) (i) *If* $I^0 \leqslant \mu$, $\quad \bar{R}(n) = \sum_{i=1}^{n}\gamma_i^{-2} > z_\alpha^2$;

(ii) *If* $I^0 > \mu$, $\quad \bar{R}(n) = \sum_{i=1}^{n}\gamma_i^{-2} \geqslant z_\alpha^2 - (I^0 - \mu)^2/\sigma^2$.

(c) *Under* (F), *the optimal set of orders is given by the allocation scheme* $\mathbf{w}^*$ *and the expected effective supply level*:

$$Y_E^* = \begin{cases} \left(1-\dfrac{z_\alpha^2}{\bar{R}(n)}\right)^{-1} \\ \quad \cdot\left[(\mu-I^0)+z_\alpha\sqrt{\dfrac{(\mu-I^0)^2}{\bar{R}(n)}+\sigma^2\left(1-\dfrac{z_\alpha^2}{\bar{R}(n)}\right)}\right] \\ \qquad \text{if } \bar{R}(n)=\sum_{i=1}^n \gamma_i^{-2}\neq z_\alpha^2, \\ \left[z_\alpha^2\sigma^2-(I^0-\mu)^2\right]/\left[2(I^0-\mu)\right] \\ \qquad \text{if } \bar{R}(n)=\sum_{i=1}^n \gamma_i^{-2}=z_\alpha^2 \text{ and } I^0>\mu. \end{cases} \tag{24}$$

Proof. (a) It suffices to prove the "only if" part of the equivalency. By (19), under the CLT-approximation, the service-level constraint can be written as

$$\alpha \geqslant \Pr\left[D-I^0 \geqslant \sum_{i=1}^n p_i w_{i,n} Y - U\sqrt{\sum_{i=1}^n (w_{i,n}Y)^2 \varsigma_i^2}\right]$$
$$=\Pr\left[D-I^0 \geqslant Y_E - U\sqrt{\sum_{i=1}^n y_i^2\varsigma_i^2}\right]$$
$$=\Phi\left(\frac{\mu-I^0-Y_E}{\sqrt{\sigma^2+\sum_{i=1}^n y_i^2\varsigma_i^2}}\right)=1-\Phi\left(\frac{Y_E+I^0-\mu}{\sqrt{\sigma^2+\sum_{i=1}^n y_i^2\varsigma_i^2}}\right)$$
$$\Leftrightarrow\quad \frac{Y_E+I^0-\mu}{\sqrt{\sigma^2+\sum_{i=1}^n y_i^2\varsigma_i^2}} \geqslant z_\alpha=\Phi^{-1}(1-\alpha),$$

and this inequality is equivalent to the pair of inequalities

$$(Y_E-\mu+I^0)^2 - z_\alpha^2\left(\sum_{i=1}^n \varsigma_i^2 y_i^2\right) - z_\alpha^2\sigma^2 \geqslant 0, \tag{25}$$

$$Y_E \geqslant \mu - I^0 + z_\alpha\sigma. \tag{26}$$

(25) can be written as $(Y_E-\mu+I^0)^2 - z_\alpha^2 Y_E^2 \cdot (\sum_{i=1}^n \varsigma_i^2 w_{i,n}^2)/(\sum_{i=1}^n p_i w_{i,n})^2 - z_\alpha^2\sigma^2 \geqslant 0$, or

$$Y_E^2\left(1-\frac{z_\alpha^2}{R(n\mid\mathbf{w})}\right)-2Y_E(\mu-I^0) + (\mu-I^0)^2 - z_\alpha^2\sigma^2 \geqslant 0. \tag{27}$$

Thus, if the pair $(Y_E, \mathbf{w})$ results in a feasible solution, i.e., satisfies (25) and (27), so does $(Y_E, \mathbf{w}^*)$, because $\bar{R}(n)=R(n\mid\mathbf{w}^*)\geqslant R(n\mid\mathbf{w})$. This allows us to state the feasibility conditions as a linear or quadratic inequality in the single variable Y_E only, i.e., (26) in conjunction with

$$Y_E^2\left(1-\frac{z_\alpha^2}{\bar{R}(n)}\right)-2Y_E(\mu-I^0)+(\mu-I^0)^2-z_\alpha^2\sigma^2\geqslant 0. \tag{27'}$$

(b(i)) If $\bar{R}(n)>z_\alpha^2$, the coefficient of the quadratic term to the left of (27′) is positive, so that (26) and (27′) are satisfied for Y_E bigger than or equal to the largest root of (27′), i.e.,

$$Y_E \geqslant \frac{\mu-I^0}{1-z_\alpha^2/\bar{R}(n)} + z_\alpha\sqrt{\frac{(\mu-I^0)^2/\bar{R}(n)}{[1-z_\alpha^2/\bar{R}(n)]^2}+\frac{\sigma^2}{1-z_\alpha^2/\bar{R}(n)}}$$
$$\geqslant \frac{\mu-I^0}{1-z_\alpha^2/\bar{R}(n)} + \frac{z_\alpha\sigma}{\sqrt{1-z_\alpha^2/\bar{R}(n)}} \geqslant \mu - I^0 + z_\alpha\sigma.$$

(The other root equals

$$[1-z_\alpha^2/\bar{R}(n)]^{-1} \cdot\left[(\mu-I^0) - z_\alpha\sqrt{(\mu-I^0)^2/\bar{R}(n)+\sigma^2[1-z_\alpha^2/\bar{R}(n)]}\right]$$
$$=[(\mu-I^0)^2-z_\alpha^2\sigma^2] \cdot\left[z_\alpha\sqrt{(\mu-I^0)^2/\bar{R}(n)+\sigma^2[1-z_\alpha^2/\bar{R}(n)]}+(\mu-I^0)\right]^{-1}$$
$$\leqslant(\mu-I^0)^2\left[(\mu-I^0) + z_\alpha\sqrt{(\mu-I^0)^2/\bar{R}(n)+\sigma^2[1-z_\alpha^2/\bar{R}(n)]}\right]^{-1}$$
$$\leqslant (\mu-I^0)^2/(\mu-I^0) < \mu-I^0+z_\alpha\sigma.\Big)$$

Conversely, if $\bar{R}(n)=z_\alpha^2$ and $I^0<\mu$, (27′) is equivalent to $Y_E \leqslant [(\mu-I^0)^2-z_\alpha^2\sigma^2]/[2(\mu-I^0)] = (\mu-I^0)/2 - z_\alpha^2\sigma^2/[2(\mu-I^0)] < \mu - I^0 + z_\alpha\sigma$, which is inconsistent with (26). Similarly, if $\bar{R}(n)=z_\alpha^2$ and $I^0=\mu$, (27′) fails to hold for any value of Y_E. It remains to be shown that a feasible solution does not exist when $\bar{R}(n)<z_\alpha^2$. In this case, the quadratic function to the left of (27′) corresponds with a concave parabola, and a solution to (26) and (27′) exists only if the quadratic function has a largest root $\geqslant \mu - I^0 + z_\alpha\sigma$. However, the largest root, if it exists, is given by $[z_\alpha^2/\bar{R}(n)-1]^{-1}\cdot[z_\alpha\sqrt{(\mu-I^0)^2/\bar{R}(n)+\sigma^2[1-z_\alpha^2/\bar{R}(n)]} - (\mu-I^0)] \leqslant [z_\alpha^2/\bar{R}(n)-1]^{-1}[z_\alpha\sqrt{(\mu-I^0)^2/\bar{R}(n)} - (\mu-I^0)] = [z_\alpha/\sqrt{\bar{R}(n)}-1]^{-1}[z_\alpha/\sqrt{\bar{R}(n)}+1]^{-1}(\mu-I^0)[z_\alpha/\sqrt{\bar{R}(n)}-1] < \mu-I^0+z_\alpha\sigma$.

(b(ii)) If $\bar{R}(n)>z_\alpha^2$, (26) and (27′) are, again, satisfied if and only if Y_E is bigger than or equal to the largest root of (27′):

$$Y_E \geqslant [1-z_\alpha^2/\bar{R}(n)]^{-1} \cdot\left[(\mu-I^0)+z_\alpha\sqrt{(\mu-I^0)^2/\bar{R}(n)+\sigma^2(1-z_\alpha^2/\bar{R}(n))}\right]$$
$$=[z_\alpha^2\sigma^2-(I^0-\mu)^2] \Big/\left[z_\alpha\sqrt{[I^0-\mu]^2/\bar{R}(n)+\sigma^2[1-z_\alpha^2/\bar{R}(n)]}+(I^0-\mu)\right]$$
$$=[z_\alpha^2\sigma^2-(I^0-\mu)^2]$$

$$
\begin{aligned}
&\Big/\Big[(I^0-\mu)+z_\alpha\sqrt{[(I^0-\mu)^2-z_\alpha^2\sigma^2]/\bar{R}(n)+\sigma^2}\Big]\\
&\geqslant [z_\alpha^2\sigma^2-(I^0-\mu)^2]/[z_\alpha\sigma+(I^0-\mu)]\\
&= z_\alpha\sigma-(I^0-\mu)=\mu-I^0+z_\alpha\sigma.
\end{aligned}
$$

(The other root, $[1 - z_\alpha^2/\bar{R}(n)]^{-1}[(\mu - I^0) - z_\alpha\sqrt{(\mu-I^0)^2/\bar{R}(n)+\sigma^2[1-z_\alpha^2/\bar{R}(n)]}]$ is negative.) If $\bar{R}(n)=z_\alpha^2$, (27′) is equivalent to $Y_E \geqslant [z_\alpha^2\sigma^2-(I^0-\mu)^2]/[2(I^0-\mu)] = z_\alpha^2\sigma^2/[2(I^0-\mu)] - (I^0-\mu)/2 = z_\alpha^2\sigma^2/[2(I^0-\mu)] + (I^0-\mu)/2 + (\mu-I^0) \geqslant (\mu-I^0) + \frac{1}{2}\min_{x>0}\{(z_\alpha^2\sigma^2)x+1/x\} = (\mu-I^0)+z_\alpha\sigma$ so that (26) and (27′) are satisfied for Y_E larger than or equal to this root. Finally, if $\bar{R}(n) < z_\alpha^2$, the quadratic function to the left of (27′) corresponds, as mentioned, with a concave parabola achieving nonnegative values, only if the discriminant of the quadratic function is nonnegative. This verifies the necessity of the inequality in (F)(ii). To verify its sufficiency, note that under this condition, the smallest root of the quadratic function satisfies (26) as well because

$$
\begin{aligned}
&[z_\alpha^2/\bar{R}(n)-1]^{-1}\\
&\quad\cdot\Big[(I^0-\mu)-z_\alpha\sqrt{(\mu-I^0)^2/\bar{R}(n)+\sigma^2[1-z_\alpha^2/\bar{R}(n)]}\Big]\\
&=(I^0-\mu)[z_\alpha^2/\bar{R}(n)-1]^{-1}\\
&\quad\Big/\Big[1-z_\alpha\sqrt{1/\bar{R}(n)-(z_\alpha^2/\bar{R}(n)-1)\sigma^2/(I^0-\mu)^2}\Big]\\
&=(I^0-\mu)[z_\alpha^2\sigma^2/(I^0-\mu)^2-1]\\
&\quad\Big/\Big[1+z_\alpha\sqrt{(1/\bar{R}(n))(1-z_\alpha^2\sigma^2/(I^0-\mu)^2)+\sigma^2/(I^0-\mu)^2}\Big]\\
&\geqslant (I^0-\mu)[z_\alpha^2\sigma^2/(I^0-\mu)^2-1]/[1+z_\alpha\sigma/(I^0-\mu)]\\
&=(I^0-\mu)[z_\alpha\sigma/(I^0-\mu)-1]=\mu-I^0+z_\alpha\sigma.
\end{aligned}
$$

(The inequality follows from $I^0\leqslant \mu+z_\alpha\sigma$.)

(c) It is clearly optimal to choose the smallest feasible value of Y_E, and to allocate the aggregate order according to the allocation scheme $\mathbf{w}^*$. The following expressions for Y_E^* in (24) are immediate from the proof of part (b). (All allocation schemes are equally costly in this case, but as shown above, the scheme $\mathbf{w}^*$ has the best chance of achieving feasibility and is the only scheme when selecting the smallest feasible value of Y_E.) $\square$

Thus, under Normal demand, a minimum number of BSEs arises as the necessary and sufficient condition for feasibility, similar to the above sufficient condition for general demand distributions. Also, as long as $I^0 \leqslant \mu$, the minimum threshold is again given by z_α^2. When the initial inventory $I^0>\mu$ includes a safety stock of s^0 standard deviations of demand, whereas the permitted shortfall probability calls for z_α standard deviations, the minimum number of BSEs is given by $z_\alpha^2-(s^0)^2$.

Also, when the demand distribution is Normal and the starting inventory is less than the mean uncovered demand ($I^0\leqslant\mu$), we have, for any number of suppliers n, and similar to the case of identical suppliers, that the minimum expected available supply is the same as that required of a single fully reliable supplier, with a yield factor $p' = (\sum_{i=1}^n w_{i,n}^* p_i)(1-z_\alpha^2/\bar{R}(n)) < \sum_{i=1}^n w_{i,n}^* p_i$, the actual expected yield factor, and a larger c.v. of the uncertain uncovered demand $\sqrt{(1/\bar{R}(n))(1-z_\alpha^2/\bar{R}(n))^{-2}+\gamma_{\bar{D}}^2(1-z_\alpha^2/\bar{R}(n))^{-1}} > \gamma_{\bar{D}} \geqslant \gamma_D$ when $\gamma_D<1$. The required total order is again proportional to the mean uncovered demand and increases with its c.v. It is now optimal to split the aggregate order in proportion to the suppliers' mean-to-variance ratio, irrespective of the size of the aggregate order, the demand characteristics, or the permitted shortfall probability α. The same allocation scheme was shown to be optimal by Gerchak and Parlar (1990) in their EOQ model in which orders can be split among two suppliers, each with a random yield factor. This scheme also provides the best opportunity to obtain a feasible set of orders. Because the N potential suppliers are ranked in increasing order of the c.v. of their yield factors, the feasibility condition (F) translates into a simple equivalent lower bound $\underline{n}$ for the number of required suppliers:

$$
\begin{aligned}
&\underline{n}=\min\Big\{n:\ \sum_{i=1}^n \gamma_i^{-2} > z_\alpha^2\Big\} \quad \text{if } I^0\leqslant\mu \quad\text{and}\\
&\underline{n}=\min\Big\{n:\ \sum_{i=1}^n \gamma_i^{-2} > z_\alpha^2-(s^0)^2\Big\} \quad \text{if } I^0>\mu.
\end{aligned}
\tag{28}
$$

In the vaccine supply problem, it is striking that the United States, with a target population of 100 million, relies on two suppliers, whereas the United Kingdom, with a target population seven times smaller, employs six suppliers. Recall that each season starts without any inventory. The independence of the minimum number of suppliers in (28) with respect to the size of the target population implies that this counterintuitive situation could be optimal, but only if there is reason to believe that U.S. suppliers are considerably more reliable (or have considerably larger capacity) than their British counterparts. Finally, it is immediate from (24) that, at least when $I^0<\mu$, the optimal expected available supply, and hence the minimum (variable) costs, as determined by the CLT approximation, are increasing in the c.v. of the uncovered demand and decrease whenever the set of suppliers is expanded. It depends on the set of suppliers via a single measure, i.e., the number of BSEs represented by this set; Lemma 1 below shows that the optimal expected supply is convexly decreasing in this number of BSEs.

We have conducted a numerical study to assess the accuracy of the CLT-based approximation. We have, in particular, compared the optimal expected available supply, as determined by the CLT-based approximation, with that obtained by an exact analysis, employing the same allocation scheme $\mathbf{w}^*$ in (20), that was shown to be optimal under the CLT-based approximation. (Recall that it is prohibitively difficult to identify the allocation scheme that is optimal under the exact shortfall distribution.) Figure 2 displays the comparison for instances with a Normal demand distribution and $\mu = 100$, $\alpha = 0.025$, and two-point yield distributions with $p_i = \Pr(X_i = 1)$ and $1 - p_i = \Pr(X_i = 0)$. (All instances assume $I^0 = 0$.) For each supplier i, p_i is independently generated from a uniform distribution on $[0.9, 0.95]$. Figure 2(a) displays $\widetilde{Y}_E^*(n)$ and $Y_E^*(n)$ as functions of γ_D for $n = 4$ and $n = 10$. Employing the same parameter combination and $\gamma_D = 0.2$ and 0.4, Figure 2(b) compares the same values as n is varied from $n = 2$ to $n = 12$. The horizontal axis depicts the values of $\bar{R}(n)$; the values of $\widetilde{Y}_E^*(n)$ are displayed as a continuous function of the $\bar{R}$ value; see (24).

Our numerical study shows that the CLT-based approximation is quite accurate even for small or moderate values of n. The accuracy increases with α and the values of p_i. As illustrated by Figure 2, the accuracy is rather insensitive to the degree of demand uncertainty—i.e., the value of γ_D—whereas, not surprisingly, it increases with the number of suppliers, or the number of BSEs. Figure 2 also shows that Y_E^* is convexly decreasing in $\bar{R}$ (as will be shown in Lemma 1(a)) and convexly increasing in γ_D. (The latter property is easily verified analytically from (24) whenever $I^0 \leqslant \mu$.)

We now address the problem of selecting the best possible set of suppliers, considering the total of fixed and variable operating costs. Confining ourselves to the case of Normal demands and $I^0 = 0$, we design a heuristic, prove its remarkable worst-case optimality gap, and demonstrate its even more remarkable average performance on the basis of an extensive numerical study. The algorithm can be used for arbitrary demand distribution but the proof of the worst-case optimality gap is specific to the case of Normal demands. As mentioned, in terms of variable procurement cost, it is best to rank the potential suppliers in increasing order of the c.v.s of their yield distributions, and each additional supplier lowers the cost. These results, of course, may fail to hold when considering the fixed operating costs, particularly when these differ among the suppliers. Indeed, it may be expected that suppliers who have invested in more reliable production techniques and quality control processes incur significantly higher fixed costs K while benefiting from less-variable yield distributions.

In view of (24), the problem of selecting the optimal set of suppliers $S^* \subseteq \{1, \ldots, N\}$ may be formulated as

$$\min_S \{z(S): S \subseteq \{1, \ldots, N\}\}, \tag{29}$$

where $z(S) \stackrel{\text{def}}{=} \sum_{i \in S} K_i + C(\sum_{i \in S}(\gamma_i^{-2}))$ and

$$C(R) \stackrel{\text{def}}{=} \begin{cases} \dfrac{\mu}{(1 - z_\alpha^2/R)} \left(1 + z_\alpha \sqrt{\gamma_D^2 \left(1 - \dfrac{z_\alpha^2}{R}\right) + \dfrac{1}{R}}\right) \\ \qquad \text{if } R > z_\alpha^2, \\ \infty, \quad \text{if } 0 < R \leqslant z_\alpha^2. \end{cases} \tag{30}$$

More generally, we may wish to limit the number of suppliers to some maximum $\widehat{N} \leqslant N$, in which case the problem can be formulated as $\min_S\{z(S): S \subseteq \{1, \ldots, N\}$ and $|S| \leqslant \widehat{N}\}$. We now show that the selection problem (29) is a problem of minimizing a supermodular set function. (A set function h: $2^{\{1, \ldots, N\}} \to \mathbb{R}$ is supermodular if $h(T \cup \{j\}) - h(T) \geqslant h(S \cup \{j\}) - h(S)$ for all $S \subset T$ and $j \notin T$.)

Lemma 1. *Assume that $\gamma_D \leqslant 2\sqrt{3}/z_\alpha$.* (a) *The function $C(\cdot)$ is convexly decreasing.*

(b) *The set function $z(S)$ is supermodular.*

Proof. (a) The monotonicity of $C(\cdot)$ is immediate. To show that $C(\cdot)$ is convex on the positive half-line reduces to showing that it is convex on the half-line (z_α^2, ∞): because $C(R)$ has the same structural form as $\widetilde{Y}^*(n)$ in Corollary 2(b), this is immediate from Corollary 2(b). (b) Immediate from part (a). □

As demonstrated in §5, the condition $\gamma_D \leqslant 2\sqrt{3}/z_\alpha$ is without any loss of practical generality.

The class of combinatorial optimization problems that can be formulated as the minimization of a supermodular set function is broad and has been studied intensively; see, e.g., Nemhauser and Wolsey (1978, 1988), Nemhauser et al. (1978), and Cornuejols et al. (1977). The class includes, for example, the uncapacitated plant location problem and, more generally, the problem of finding a maximum weight independent set in a matroid. As such, the class is NP-complete. In our case, the set function $z(S)$ is of the special type

$$z(S) = f\left(\sum_{i \in S} a_i\right) + g\left(\sum_{i \in S} b_i\right) \tag{31}$$

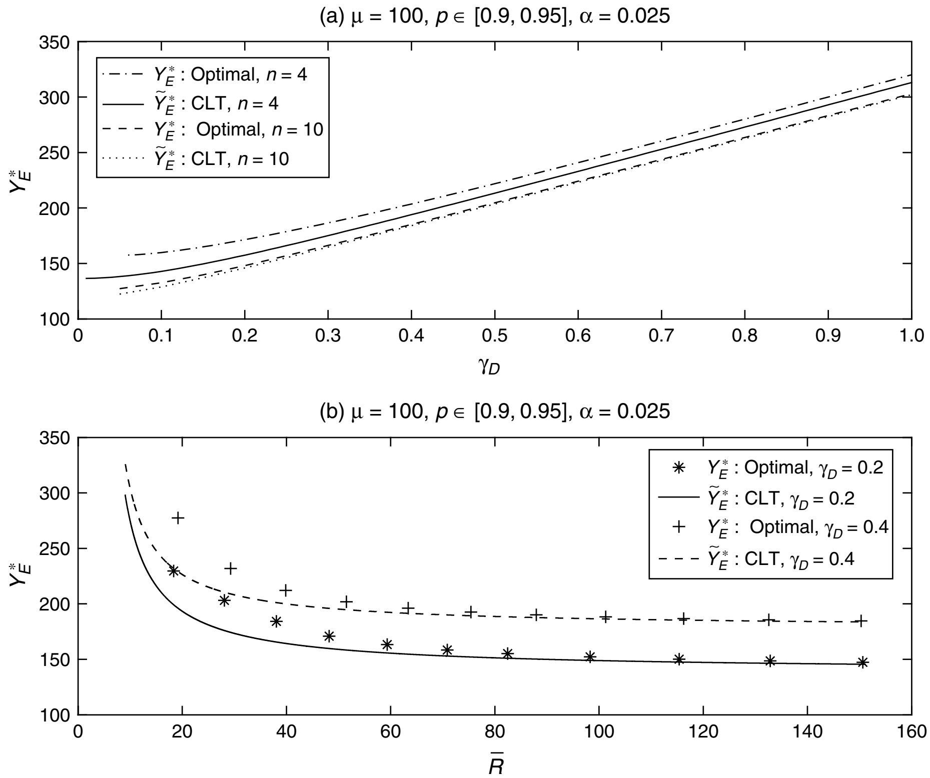
for a given sequence of positive pairs $\{(a_1, b_1), \ldots, (a_N, b_N)\}$ and with f: $\mathbb{R} \to \mathbb{R} \cup \{\infty\}$ and g: $\mathbb{R} \to \mathbb{R} \cup \{\infty\}$ convex. Federgruen and Groenevelt (1986) refer to this structure as "generalized symmetric" set functions. The following proposition shows that even this subclass is NP-complete.

Proposition 1. *Consider the class $\mathcal{P}$ of all set function minimization problems with set functions of the type* (31). *The class $\mathcal{P}$ is NP-complete.*

Proof. It suffices to show that for every instance of the knapsack problem: $\min\{\sum_{i=1}^N a_i x_i \mid \sum_{i=1}^N b_i x_i \leqslant B$ and $x_i = 0, 1,\ i = 1, \ldots, N\}$, there is an instance in $\mathcal{P}$ that reduces to it. Choose $f(x) = x$ and $g(x) = 0$ if $x \leqslant B$; whereas $g(x) = \infty$ for $x > B$. The problem of selecting the set S that minimizes (31) is clearly equivalent to the knapsack problem. □

Because the class $\mathcal{P}$ is NP-complete, no exact polynomial time procedure for the selection problem can be expected. However, Cornuejols et al. (1977) and Nemhauser et al. (1978) show that a simple greedy procedure has a low worst-case optimality gap; in practice, it comes within a few percentage points of optimality.

Figure 2. Total expected supply: Exact and approximate solutions.



Normally, the greedy procedure for a set selection problem operates as follows: starting with the empty set, in each iteration an element is added to the set whose addition results in the biggest cost reduction. The greedy procedure terminates after $\widehat{N}$ iterations (i.e., when the set reaches a cardinality of $\widehat{N}$) or in an earlier iteration if no cost saving can be achieved by adding a new element to the set. In our case, a slight modification of the procedure is required because $z(S) = \infty$ if $|S| < \underline{n}$; see (28). Thus, the greedy procedure cannot be started from the empty set, but must be initiated from a minimum set of suppliers enabling a feasible solution. We define a set S to be *minimally feasible* if $\sum_{i\in S} \gamma_i^{-2} > z_\alpha^2$, whereas $\sum_{i\in T} \gamma_i^{-2} \leqslant z_\alpha^2$ for all $T \subsetneq S$. Let $\bar{n} = \min\{n\colon \sum_{i=N-n+1}^{N} \gamma_i^{-2} > z_\alpha^2\}$. Clearly, all minimally feasible sets of suppliers have a cardinality between $\underline{n}$ and $\bar{n}$.

Algorithm 1. Supplier Selection Algorithm

INPUT: $p_1, \ldots, p_N$; $\gamma_1, \ldots, \gamma_N$; $K_1, \ldots, K_N$; α; μ; γ_D; $C(\cdot)$
STEP 0: $\bar{n} := \min_n\{\sum_{i=N-n+1}^{N} \gamma_i^{-2} > z_\alpha^2\}$
STEP 1: FOR $i_1 := 1$ TO N DO
  BEGIN
    $R := \gamma_{i_1}^{-2}$;
    IF $(R > z_\alpha^2)$ THEN
      add $\{i_1\}$ to $\mathscr{L}$
    ELSE
      FOR $i_2 := i_1 + 1$ TO N DO
        BEGIN
          $R := R + \gamma_{i_2}^{-2}$;
          IF $(R > z_\alpha^2)$ THEN
            add $\{i_1, i_2\}$ to $\mathscr{L}$
          ELSE
            FOR $i_3 := i_2 + 1$ TO N DO
              ⋮
                ELSE
                  FOR $i_{\bar{n}} := i_{\bar{n}-1} + 1$ TO N
                    add $\{i_1, i_2, \ldots, i_{\bar{n}}\}$ to $\mathscr{L}$
                ENDIF
              ⋮
          ENDIF
        END
    ENDIF
  END
STEP 2: For every set S in $\mathscr{L}$, expand the set S greedily until the set reaches cardinality $\widehat{N}$ or no cost improvement can be achieved by adding a new supplier to S.

Because $\gamma_i \leqslant 1$, $\bar{n} \leqslant \lceil z_\alpha^2 \rceil$. In practice, $\bar{n}$ is usually much smaller. For example, if all $p_i \geqslant 0.9$, $\bar{n} \leqslant \lceil z_\alpha^2/9 \rceil \leqslant 2$ with $\alpha \geqslant 1.1045 \times 10^{-5}$; if all $p_i \geqslant 0.8$, $\bar{n} \leqslant \lceil z_\alpha^2/4 \rceil \leqslant 2$ with $\alpha \geqslant 0.0023$. There are at most $\binom{N}{\bar{n}}$ minimally feasible sets because every set of cardinality $\bar{n}$ is feasible, but several such sets may contain the same minimally feasible set. The bound is tight in case all $\gamma_i = \gamma$.

Our proposed solution method consists of enumerating the list $\mathscr{L}$ of all minimally feasible supplier sets and applying the above described greedy procedure to each of them. The Supplier Selection Algorithm employs an efficient creation of the list $\mathscr{L}$. Starting from a given minimally feasible set $S \in \mathscr{L}$, the greedy procedure requires $O(N^2)$ elementary operations and square root calculations because in each of up to N iterations, up to N potential supplier sets have to be evaluated. Because there are at most $\binom{N}{\bar{n}}$ minimally feasible sets, the algorithm's complexity is $O(N^{\bar{n}+2})$. (As mentioned, usually $\bar{n} \leqslant 2$ and the algorithm is of complexity $O(N^3)$ or $O(N^4)$.) Let Z^* denote the total cost of the optimal set of suppliers and Z^G the cost associated with the set of suppliers generated by the greedy-type algorithm. Let ρ denote the maximum cost deterioration due to the addition of a single supplier to an existing set. By the convexity of the function $C(\cdot)$, $\rho = [\max_i\{K_i + (C(\sum_{j=1}^N \gamma_j^{-2}) - C(\sum_{j\neq i} \gamma_j^{-2}))\}]^+$. The following theorem is immediate from Theorem 4.2 in Nemhauser et al. (1978).

Theorem 6.

$$\frac{[Z^G - Z^*]}{\min_{S\in\mathscr{L}}[z(S) - Z^* + \widehat{N}\rho]} \leqslant \left(1 - \frac{1}{\widehat{N}}\right)^{\widehat{N}} \leqslant e^{-1}.$$

Disregarding the correction term $\widehat{N}\rho$, the optimality gap $(Z^G - Z^*)/(z(S) - Z^*)$ is somewhat unconventional, but, as argued in Cornuejols et al. (1977), it may actually be more descriptive of the quality of the heuristic: it relates the absolute gap between its cost value Z^G and the optimal cost value Z^* to the span between Z^* and the cost of an arbitrary starting solution.

We have evaluated the optimality gap of the Supplier Selection Algorithm for a collection of 5,637 instances, 4,800 with $N = 10$, 729 with $N = 15$, and 108 with $N = 20$ suppliers. Recall that the optimality gap assesses the difference in total costs, when the variable costs are approximated with the help of the CLT-based approximation, throughout. This CLT-approximation depends on the suppliers' yield distributions only via the means and standard deviations. All instances in our study employ a Normal demand distribution with mean $\mu = 100$ and two-point distributions for the suppliers' yields, i.e. $\Pr(X_i = 1) = p_i$ and $\Pr(X_i = 0) = 1 - p_i$. The values of p_i are selected randomly from a uniform distribution on an interval $[p_{\min}, p_{\max}]$. By selecting a large number of disparate $p_{\min}$, $p_{\max}$-values, we have evaluated a large spectrum of vector pairs $\{p, \gamma\}$, the only characteristics of the suppliers' yield factors, to affect the cost approximations. Similarly, the fixed-cost values K_i are generated randomly from a uniform distribution on an interval $(C(\bar{R}(N))/N)[\kappa_{\min}, \kappa_{\max}]$, i.e., the fixed operating costs are specified as a multiple of the average optimal variable cost per supplier. Table 1 exhibits the values of $p_{\min}$, $p_{\max}$, $\kappa_{\min}$, $\kappa_{\max}$, α, and γ_D chosen for each of these sets of instances with $N = 10$, $N = 15$, and $N = 20$ suppliers. (In each case, all parameter combinations have been evaluated.) The table displays for what percentage of instances the heuristic generates the optimal solution, as well as the average and the maximum optimality gap. We conclude that the heuristic is almost always generating the optimal solution and the average optimality gap is negligible.

7. Conclusions and Extensions

Traditional supply chain management has emphasized the benefits of consolidating supply sources. When there is a significant probability that all or some of a manufacturer's targeted supply does not become available because of disruptions or quality problems, this provides a powerful rationale to spread the supply over multiple suppliers. This raises the questions of which set of suppliers to patronize, how much to order in the aggregate, and how to allocate the aggregate order among the selected suppliers. Exact analysis reveals that, asymptotically as $n \to \infty$, the expected available supply under an optimal set of orders equals what would be ordered from a single completely reliable supplier. Moreover, asymptotically, all allocation schemes which assume that relative market shares remain bounded in the number of suppliers, are optimal. For small or moderate values of n, the allocation scheme is, however, of major importance. We have shown that even in the simplest case with identical suppliers, equal-size orders, the intuitively best allocation scheme, are guaranteed to be optimal only if the p.d.f. of the demand distribution is nonincreasing, and it is prohibitively difficult to determine the

Table 1. Performance: Supplier Selection Algorithm.

	$N = 10$	$N = 15$	$N = 20$
$p_{\min} \in$	$\{0.6, 0.7, 0.8\}$	$\{0.6, 0.7, 0.8\}$	$\{0.6\}$
$p_{\max} \in$	$\{0.85, 0.9, 0.95, 1\}$	$\{0.9, 0.95, 1\}$	$\{0.9, 1\}$
$\kappa_{\min} \in$	$\{0.25, 0.5, 0.75, 0.9\}$	$\{0.25, 0.75, 0.9\}$	$\{0.25, 0.9\}$
$\kappa_{\max} \in$	$\{1, 2, 3, 4\}$	$\{1, 2, 4\}$	$\{1, 2, 4\}$
$\alpha \in$	$\{0.01, 0.025, 0.05, 0.1, 0.2\}$	$\{0.01, 0.025, 0.05\}$	$\{0.01, 0.025, 0.05\}$
$\gamma_D \in$	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$	$\{0.1, 0.3, 0.5\}$	$\{0.1, 0.3, 0.5\}$
Number of instances	4,800	729	108
Percentage of optimal solutions (%)	98.62	97.52	98.15
Average optimality gap (%)	0.01	0.02	0.01
Maximum optimality gap (%)	4.55	2.20	1.27

optimal allocation scheme on the basis of the exact shortfall probabilities.

We have therefore developed an LDT- and a CLT-based approximation method for the shortfall probabilities. Both methods have been shown to be highly accurate. The former (LDT) has the advantage of generating a rigorous upper bound for the total order size and the variable procurement costs. The CLT-based approximation is *asymptotically accurate* and the order sizes determined by this method are *asymptotically optimal* as the number of suppliers grows. With Normal demands, the optimal total order size as well as the ones obtained by the two approximations, are proportional to the mean uncovered demand, with a proportionality factor that is an increasing function of σ or γ_D.

The CLT-based approximation method also permits many important managerial insights. For example, we have, for general demand distributions under the CLT approximation, derived a sufficient condition for a set of suppliers to be feasible, which is completely determined by a single measure, i.e., the number of BSEs represented by the set. A set of potential suppliers is feasible if the total number of BSEs is strictly larger than z_α^2, irrespective of the characteristics of the demand distribution. (When $I^0 = 0$, the number of BSEs being larger than or equal to z_α^2 is also necessary.) As far as the shape of the suppliers' yield distributions is concerned, only the c.v. values of these distributions matter. Based on this feasibility condition, the minimum number of suppliers required is thus given by $\min\{n \mid \sum_{i=1}^n \gamma_i^{-2} > z_\alpha^2\}$. The *suppliers' safety margin*, i.e., the additional BSEs beyond the minimum required number, allows for a reduction of the variable procurement costs. The above assumes that orders are split among the suppliers in proportion to the mean-to-variance ratios of their yield distributions. This allocation scheme is always optimal in terms of facilitating feasibility: if a feasible solution fails to exist under this allocation scheme, it fails to exist under any other scheme as well. When the demand distribution is Normal, the number of BSEs being strictly larger than z_α^2 is, in fact, not only sufficient but also necessary whenever the starting inventory is less than or equal to the mean demand. When it is larger than the mean demand, the same condition continues to be necessary and sufficient except that the minimum number of BSEs, z_α^2, needs to be reduced by $(s^0)^2$; with s^0 the number of standard deviations of demand, the starting inventory is in excess of the mean. We refer to §1 for a summary of many additional insights obtained on the basis of the CLT-approximation.

Next, we have addressed the problem of identifying a set of suppliers that minimizes all (fixed plus variable) costs, along with the associated set of orders. We have shown that under Normal demands, a greedy-type heuristic of low polynomial complexity for a fixed value of α has a worst-case optimality gap of 36.8%. An extensive numerical study shows that the heuristic finds the optimal selection in 98.5% of the 5,637 instances and has an average optimality gap of only 0.015%. The managerial implication here is that little, if anything, is lost when gradually expanding one's supplier base, as compared to establishing an "optimal" set of suppliers at once.

The analysis in this paper has assumed that the yield factors of the suppliers are independent. In some settings, supply risks may be correlated, for example, when natural disasters (storms, floods) or sabotage by terrorists are likely to hit multiple facilities in a given geographic region. (Recall the oil refineries example in §1.) To address these interdependencies, assume that the vector of yield factors $\{X_i\colon i = 1, \ldots, N\}$ has a general joint distribution, with correlation factors $\rho_{ij} = \operatorname{corr}(X_i, X_j)$: $1 \leqslant i \neq j \leqslant N$. The analysis based on the CLT-approximation is easily adapted to account for any interdependence of the yield factors. The service constraint (19) is now to be replaced by

$$\tilde{P}_n(Y, w_{1,n}, \ldots, w_{n,n}) = \Pr\left[\frac{\bar{D}}{Y} > \sum_{i=1}^n p_i w_{i,n} - U\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}\right] \leqslant \alpha, \tag{32}$$

where the $n \times n$-matrix $\boldsymbol{\Sigma}$ has $\Sigma_{ii} = \varsigma_i^2$ and $\Sigma_{ij} = \varsigma_i \varsigma_j \rho_{ij}$. For any allocation scheme $\mathbf{w}$, redefine $R(n \mid \mathbf{w}) = (\sum_{i=1}^n w_i p_i)^2 / \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$. Part (c) of Theorem 4 and Theorem 5 continue to apply with this adjusted specification of the function $R(n \mid \mathbf{w})$. Moreover, the allocation scheme $\mathbf{w}^*$ that maximizes $R(n \mid \mathbf{w})$ continues to simultaneously (i) provide the best chance for the existence of a feasible solution, and (ii) in the case of Normal demands, minimize variable procurement costs. This allocation scheme $\mathbf{w}^*$ continues to be independent of any of the demand characteristics as well as the permitted shortfall probability α. Moreover, it depends on the suppliers' yield distributions only via their vector of means $\mathbf{p}$ and their variance-covariance matrix $\boldsymbol{\Sigma}$. This optimal scheme $\mathbf{w}^*$ no longer allocates the aggregate order in proportion to the mean-to-variance ratios of the suppliers' yield distributions because the correlation pattern now impacts on the allocation. Although no longer available in closed form, $\mathbf{w}^*$ can still be computed efficiently: redefining $r^*(b) \stackrel{\text{def}}{=} \min\{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}/b^2 \mid \sum_{i=1}^n w_{i,n} = 1; \sum_{i=1}^n p_i w_{i,n} = b\}$, $\bar{R}(n) = \max\{R(n \mid \mathbf{w}) \mid \sum_{i=1}^n w_i = 1\}$ continues to satisfy the identity $\bar{R}(n) = 1/[\min_{p_{\min} \leqslant b \leqslant p_{\max}} r^*(b)]$. Note that for a given value of b, evaluation of $r^*(b)$ now amounts to solving a quadratic convex program, identical to those solved in the classical Markowitz mean-variance trade-off analysis of financial portfolios; see Markowitz (1952) or any elementary finance textbook like Brealey and Myers (1996). The computation of $\mathbf{w}^*$ thus amounts to the solution of a parametric quadratic program, with only two (linear) constraints. Most specifically, the closed-form expressions for the minimum expected effective supply in Theorem 5 continue to apply, under Normal demands, merely replacing $\bar{R}(n) = \sum_{i=1}^n \gamma_i^{-2}$ by the above expression. Similarly, one can continue to use Algorithm 1 to optimize total (fixed plus variable) costs and the worst-case optimality gap in Theorem 6 continues to apply.

Finally, the approximation $\widetilde{P}_n$ in (32) can still be justified on the basis of a central limit theorem, provided the dependence between the yield factors is sufficiently "weak." Here we need to be able to rank the suppliers in such a way that the "degree of dependence" between any pair of suppliers i and $i+m$ declines to zero sufficiently fast as m tends to infinity for any $i \geqslant 1$. The degree of dependence is best quantified by the so-called $\alpha(m)$ mixing coefficients; see, for example, Definition 3.41 in White (1999). In this case, the Wooldridge-White central limit theorem (see, e.g., Theorem 5.20 ibid) can be used to show that the shortfall distribution, properly scaled and centralized, converges to a Normal distribution. For example, if the interdependence between the suppliers' yield factors arises due to geographic proximity (as in the case of oil refineries), it is natural to rank the suppliers in accordance with their geographic locations. It is then indeed reasonable to assume that the degree of dependence between X_i and X_{i+m}, as measured by the α-mixing coefficient, declines to zero sufficiently fast. The sequence $\{X_n\}$ may, for example, be assumed to follow a general ARMA process, in which case the α-mixing coefficients decline geometrically; see, e.g., Mokkadem (1988). We defer the many details required for a formal proof to a future treatment.

Future work will address other various important generalizations of the model, allowing, for example, for supplier-dependent variable cost rates and capacity limitations, as well as multiple replenishment opportunities over a complete planning horizon.

8. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Endnotes

1. "In recent years there have been many significant disruptions of vaccine supplies. Between November 2000 and May 2003, there were shortages of 8 of the 11 vaccines for childhood diseases in the United States including those for tetanus, diphtheria, whooping cough, measles, mumps, and chicken pox. There have been flu vaccine shortages or miscues for four consecutive years." See *New York Times* (2004, p. A20).
2. This assumes $\mu - z_\alpha \sigma > 0$, i.e., the probability of the Normal demand distribution adopting a negative value is itself less than α.
3. Several procurement models with unreliable suppliers restrict themselves to this case, e.g., Anupindi and Akella (1993) and Swaminathan and Shanthikumar (1999). Bernoulli random yield factors represent settings where a supply disruption results in a complete shutdown of the facility, as in the case of hurricanes or contamination resulting in the closure of a vaccine supply plant.
4. The interchange of the differentiation and the expectation operators is justified by the fact that $E_{\widehat{X}}[e^{\theta \widehat{X}}] = \int_0^1 e^{\theta x} d\widehat{G}(x)$. Assume first that there are $m \geqslant 0$ loss factors where $\{x_i\colon 1, 2, \ldots, m\}$ such that $\Pr[\widehat{X} = x_i] = \pi_i$ and let $\hat{g}(\cdot)$ be the density of the continuous part of the $\widehat{G}(\cdot)$ distribution. Thus, $L(\theta) = \mathrm{E}[e^{\theta \widehat{X}}] = \sum_{i=1}^m \pi_i e^{\theta x_i} + \sum_{i=1}^{m+1} \int_{x_{i-1}}^{x_i} e^{\theta x} \hat{g}(x)\, dx$, where $x_0 = 0$ and $x_{m+1} = 1$. Therefore, $l'(\theta) = \sum_{i=1}^m \pi_i x_i e^{\theta x_i} + \sum_{i=1}^{m+1} \int_{x_{i-1}}^{x_i} x e^{\theta x} \hat{g}(x)\, dx = E_{\widehat{X}}[\widehat{X} e^{\theta \widehat{X}}]$, where the interchange of the integration and differentiation operators is justified by the fact that the two-variable function $(x, \theta) \to x e^{\theta x} \hat{g}(x)$ is continuous on each of the rectangles $(x_{i-1}, x_i) \times [0, \bar{\theta}]$, $i = 1, \ldots, m+1$. The remaining case with a countable number of values $\{x_i\}$ such that $\Pr[\widehat{X} = x_i] = \pi_i$ is positive can be handled in a similar way.
5. Theorem 3(a) does not imply that the exact and approximate shortfall probabilities are of the same order. The asymptotic optimality result in Theorem 3(b) follows nevertheless.

References

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

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, Cambridge, UK.

Brealey, R. A., S. C. Myers. 1996. *Principles of Corporate Finance*, 5th ed. McGraw-Hill, New York.

Chopra, S., G. Reinhardt, U. Mohan. 2005. The effect of supply chain disruption on sourcing strategy. Working paper, Department of Managerial Economics and Decision Sciences, Northwestern University, Evanston, IL.

Chow, Y., H. Teicher. 1997. *Probability Theory*: *Independence, Interchangeability, Martingales*, 3rd ed. Springer, New York.

Cornuejols, G., M. L. Fisher, G. L. Nemhauser. 1977. Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management Sci.* **23**(8) 789–810.

Department of Energy. 2005. Country analysis briefs. (January 12). Department of Energy, Washington, D.C.

Federgruen, A., H. Groenevelt. 1986. The greedy procedure for resource allocation problems: Necessary and sufficient conditions for optimality. *Oper. Res.* **34**(6) 909–918.

Gerchak, Y., M. Parlar. 1990. Yield variability, cost tradeoffs and diversification in the EOQ model. *Naval Res. Logist.* **37**(3) 341–354.

Gerchak, Y., Y. Wang, C. A. Yano. 1994. Lot sizing in assembly systems with random component yields. *IIE Trans.* **26**(2) 19–24.

Grosfeld-Nir, A., Y. Gerchak. 2004. Multiple lotsizing in production to order with random yields: Review of recent advances. *Ann. Oper. Res.* **126** 43–69.

Gurnani, H., R. Akella, J. Lekoczky. 1996. Optimal ordering policies in assembly systems with random demand and random supplier delivery. *IIE Trans.* **28**(11) 865–878.

Gurnani, H., R. Akella, J. Lekoczky. 2000. Supply management in assembly systems with random yield and random demand. *IIE Trans.* **32**(8) 701–714.

Harrison, J. M., J. A. Van Mieghem. 1999. Multi-resource investment strategies: Operational hedging under demand uncertainty. *Eur. J. Oper. Res.* **113** 17–29.

Heinrich, J. 2001a. Flu vaccine: Supply problems heighten need to ensure access for high-risk people. Report GAO-01-624, Washington, D.C.

Heinrich, J. 2001b. Flu vaccine: Steps are needed to better prepare for possible future shortages. Report GAO-01-786T, Washington, D.C.

Heinrich, J. 2004. Flu vaccine: Recent supply shortages underscore ongoing challenges. Report GAO-05-177T, Washington, D.C., November 18.

Henig, M., Y. Gerchak. 1990. The structure of periodic review policies in the presence of random yield. *Oper. Res.* **38**(4) 634–643.

Ilan, Y., M. Yadin. 1985. Second sourcing and transfer of technology—Tradeoffs under uncertainty. Working paper, Department of Operations Research, Statistics, and Economics, The Technion, Haifa, Israel.

Lee, H. L. 1996. Input control for serial production lines consisting of processing and assembly operations with random yields. *Oper. Res.* **44**(3) 464–468.

Lewis, B. M., A. L. Erera, C. C. White. 2005. An inventory control model with possible border disruptions. Working paper, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta.

Longitudes 04: Mapping the New World of Synchronized Commerce. 2004. Harvard Business School Publishing, Cambridge, MA.

Markowitz, H. M. 1952. Portfolio selection. *J. Finance* **7** 77–91.

Mokkadem, A. 1988. Mixing properties of ARMA processes. *Stochastic Processes and Their Appl.* **29** 309–315.

Müller, A., D. Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Ltd., West Sussex, UK.

Nemhauser, G. L., L. A. Wolsey. 1978. Best algorithms for approximating the maximum of a submodular set function. *Math. Oper. Res.* **3**(3) 177–188.

Nemhauser, G. L., L. A. Wolsey. 1988. *Integer and Combinatorial Optimization*. John Wiley & Sons, Inc., New York.

Nemhauser, G. L., L. A. Wolsey, M. L. Fisher. 1978. An analysis of approximations for maximizing submodular set functions—I. *Math. Programming* **14** 265–294.

New York Times. 2004. With few suppliers, shortage of flu vaccine was long in the making. (October 17) A20.

Parlar, M., D. Wang. 1993. Diversification under yield randomness in two simple inventory models. *Eur. J. Oper. Res.* **66** 52–64.

Rudi, N., Y. S. Zheng. 1997. Multi-item newsboy model with partial variety postponement. Presented at INFORMS Dallas Meeting, October.

Singh, M. R., C. T. Abraham, R. Akella. 1990. A wafer design problem in semiconductor manufacturing for reliable customer service. *IEEE Trans. Components, Hybrids, Manufacturing Tech.* **13**(1) 103–108.

Swaminathan, J. M., J. G. Shanthikumar. 1999. Supplier diversification: Effect of discrete demand. *Oper. Res. Lett.* **24**(5) 213–221.

Thorisson, H. 2000. *Coupling, Stationarity and Regeneration*. Springer, New York.

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

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

Van Mieghem, J. A., N. Rudi. 2002. Newsvendor networks: Inventory management and capacity investment with discretionary activities. *Manufacturing Service Oper. Management* **4**(4) 313–335.

White, H. 1999. *Asymptotic Theory for Econometricians*, revised ed. Academic Press, New York.

Yano, C. A. 1991. Impact of quality and pricing on the market shares of two competing suppliers in a simple procurement model. Technical report, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor.

Yano, C. A., H. L. Lee. 1995. Lot sizing with random yields: A review. *Oper. Res.* **43**(2) 311–334.

Yao, D. D. 1988. Optimal run quantities for an assembly system with random yields. *IIE Trans.* **20**(4) 399–403.

Zipkin, P. H. 1980. Simple ranking methods for allocation of one resource. *Management Sci.* **26**(1) 34–43.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.