Risk-based Integration of Strategic and Tactical Capacity Planning

Sunder Kekre[a], Uday Rao[b], Samarth Sarthi[c], Jinxin Yi[a]

[a] GSIA, Carnegie Mellon University, Pittsburgh, PA 15213

[b] CBA, University of Cincinnati, Cincinnati, OH 45221

[c] SCA Technologies, LLC, Pittsburgh, PA 15213

April, revised November 2002

Abstract

We consider capacity management with a long-term strategic choice, such as the number of production lines to install before demand is known, and short-term tactical decisions relating to production, inventory and subcontracting (recourse actions made after demand is known). We present an integrated, scenario-based, mathematical modeling and solution framework. For a single-product environment, we examine properties of total profit as a function of demand and the long-term capacity z. We investigate two measures of risk (profit variance and mean downside risk) and their corresponding profit-risk frontiers. Computational experiments are used to illustrate parameter sensitivity results obtained from the model.

1 Introduction and Literature Review

Over the last decade, purchased items have accounted for an increasing percentage of cost of goods sold by firms, which underscores the importance of supply management. This research is motivated by capacity planning problems we have seen at suppliers such as North Side Foods (a supplier to McDonald's which was acquired in 1998 by Smithfield Foods) and Delphi Packard Electric. For example, North Side Foods (NSF) provides hamburger patties to warehouses and franchises of McDonald's in accordance with a supply contract between the firms. At the beginning of the contract year, NSF must decide the production capacity level (e.g., number of production lines used), given market demand estimates from McDonald's. Subsequently, it must satisfy 100% of realized demand over the year using regular production, overtime, inventory and subcontracting, as necessary. The goal of capacity planning in this environment is to meet dynamic demand (with planned market promotions by McDonald's), at the lowest possible cost, while incorporating new suppliers/ technologies and integrating disparate entities in the supply chain. In the past, some sub-optimality has been observed in supplier selection (based on freight costs only instead of total delivered cost) or in capacity installation and utilization. This paper presents a model that the customer, McDonald's, and the supplier, NSF, can use in a collaborative manner to understand the effects of different market promotions or supplier selection rules or capacity level choices on the system's performance.

In a wide variety of industries, business managers make capacity decisions in response to requirements of annual supply contracts and estimates of future demand. *Corporate-level* executives make *long-term* capacity choices, relating to product or technology choice, to maximize expected profit under uncertainty in future demand. *Plant* managers determine *short-term* operational plans, using regular production capacity and auxiliary resources to minimize expected costs. In this context, there is a need for research on coordinated optimization under uncertainty, model formulation with multiple objective criteria, solution methodology and sensitivity analysis. In this paper we explore how strategic and tactical capacity decisions interact when demand is uncertain. The objective of our model is to maximize expected profit with constraints on risk (profit variance or mean downside risk). Risk and sensitivity analysis in capacity planning is the primary focus of this paper.

For a survey on deterministic capacity expansion problems, we refer to Luss (1982). Davis, Dempster, Sethi, and Vermes (1987) consider a stochastic capacity expansion model. More recently, there has been renewed interest in models motivated by industrial applications that integrate long-term and short-term decisions under uncertainty: Eppen, Martin, and Schrage (1989) develop a large scale, integer programming model with scenario-based risk analysis applied to four General Motors automobile lines. Karmarkar and Kekre (1987) and Fine and Freund (1990) study investment in dedicated vs. product-flexible manufacturing capacity. Newsvendor-type capacity models are considered in Harrison and Mieghem (1999) and Carr and Lovejoy (2000). Rao, Scheller-Wolf, and Tayur (2000) combine network optimization, inventory theory and simulation to determine the structure of a distribution network for Caterpillar products, while incorporating operational inventory replenishment decisions. The interaction between business risk and parameters of an inventory policy is considered in Singhal, Raturi, and Bryant (1994).

Several papers specifically address issues relating to risk management in capacity planning. Eppen, Martin, and Schrage (1989) propose "expected downside risk" as an easily computed risk measure associated with each capacity decision. They show that this measure generates "non-dominated solutions" on the risk-return frontier. Rinks, Ringuest, and Peters (1987) use a utility function approach to capacity planning where uncertainty in capacity installation cost is the major source of the risk, and the standard deviation of the "equivalent cost rate" is used to evaluate the risk of capacity expansions. Paraskevopoulos, Karakitsos, and Rustem (1991) develop a general model in which their expected-cost objective is augmented by a penalty on the cost sensitivity to various types of uncertainty. Their goal is to provide a European chemical industry with robust capacity plans that are insensitive to forecast errors. Different procurement risks in supply chains have been considered from a qualitative perspective in Zsidisin, Panelli, and Upton (2000) using a survey of practitioners.

We use a quantitative modeling framework similar to Eppen, Martin, and Schrage (1989). Our paper differs from theirs in that we primarily investigate theoretical properties of several key functions while they focused on insights from application of their model. The main questions addressed in this paper are the following:

1. How do long-term (strategic) capacity levels affect short-term (tactical) capacity costs? What long-term capacity level will maximize total expected profits? How does a long-term capacity level affect the profit variance and the mean downside risk?

2. Given a risk measure, what is the effect of specifying a limit on this risk measure? Alternatively, what is the shape of the trade-off curve represented by the profit-risk Pareto frontier?

3. How is mean profit and its risk affected by problem parameters such as uncertainty and seasonality in demand, as well as cost coefficients?

In this paper, long-term and short-term capacity decisions are integrated into one model (Section 2) when demand fluctuation is the source of risk. In Sections 3 and 4, we study properties of the profit function including its average value and two risk measures, and their corresponding risk-return frontiers. In Section 5, we present computational results based on parametric analysis. We conclude with a brief summary and areas for future work.

2 The Mathematical Model

In this section, we present a single-product multi-period production and capacity planning model with uncertain demand. This model may be used to determine the optimal long-term capacity level which maximizes expected profits subject to a risk constraint, such as a limit on profit variance. Although it is possible to develop a more detailed, multi-stage, dynamic programming (DP) formulation of the problem, we use a simpler, static, stochastic programming formulation. Our preliminary experimentation suggested that (i) the optimal objective function value from the DP is close to the optimal objective value from our stochastic programming formulation, (ii) the DP approach is computationally time-intensive, particularly if quick what-if analysis capabilities are required to assess the impact of changes in model parameters. Furthermore, any increased accuracy in profit estimation using the DP model would probably be voided by the inaccuracies in demand forecasts.

Consider a planning problem with one product and H time periods. Let z denote the long-term capacity level, which is a scalar corresponding to a single resource type such as a bottleneck machine or a production line. This capacity level is determined at the beginning of the planning horizon and remains fixed over time[1]. Subsequently, operating decisions are made to satisfy 100% of the demand in each period. The short-term decisions include determining the regular time production quantity subject to capacity z, and recourse actions such as inventory and subcontracting. Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_H)^T$, where $\mathbf{x}_t = (Q_t, O_t, S_t, I_t)^T$ is the vector of short-term capacity decisions in period $t \in [1, \ldots, H]$ with Q_t, O_t, S_t, and I_t denoting, respectively, the regular and overtime production quantities, the amount subcontracted, and the ending inventory at period t. We formulate the problem using a scenario-based approach similar to Eppen et al. (1989). Let $\xi \equiv (\xi_1, \xi_2, \ldots, \xi_H)^T$ denote demand over the planning horizon H, where ξ_t are independent[2] but not identically distributed. Demand is assumed to be finite and discrete; when demand is continuous, we approximate it by a set of discrete values. Let $\mathcal{L}$ denote the set of demand scenarios; scenario $l \in \mathcal{L}$ has probability P_l. Each demand scenario is a H-vector of ξ_t values.

The objective function in our model uses a profit function, $R(z, \xi)$, which depends on capacity level z and demand ξ. The profit is equal to sales revenue minus total costs. Since, in our environment, demand must be fully satisfied, the sales revenue is equal to $p\mathbf{e}^T\xi$, where p is the unit selling price, and $\mathbf{e}$ denotes a vector of ones. The total cost to satisfy demand ξ consists of short-term cost, $C_S(z, \xi)$, and capacity cost, $C_L(z)$. In most cases, capacity cost is a concave function of capacity. In this paper, we restrict it to the commonly-used form of a fixed cost plus a linear variable cost. The fixed cost is $K\delta(z)$ with indicator function $\delta(z)$ equal to 1 if $z > 0$ and 0 otherwise; the unit

[1]This model fits the annual supply contracting environment at North Side Foods (see www.scatech.com/casehistory_mcdonalds.html); z may be continuous or discrete.

[2]Our scenario-based model and solution approach do not require independence; we assume it to simplify analysis.

variable cost of capacity is v. The only restriction on K and v are that they be nonnegative. However, clearly, large values of K and z will affect profitability and likely result in lower installed capacity.

$C_S(z, \xi)$ is the short-term cost resulting from optimal short-term decisions given long-term capacity level z and demand realization ξ. Thus short-term choices are made contingent on realized demand. For any short-term decisions $\mathbf{X}$, let $g(\mathbf{X})$ denote the short-term cost comprised of regular and overtime production costs, inventory holding costs and subcontracting costs. Although we use a linear $g(\mathbf{X})$ for illustration, the results apply to any increasing convex $g(\mathbf{X})$. Further, in our computational testing, we simplify the problem by not considering the overtime recourse option. Preliminary analysis and experimentation suggested that the overtime production option was similar to the subcontracting option, including it did not change either the analysis or the computational results; on the other hand, keeping it in the model introduced some model intricacies which did not provide additional insights. To avoid trivial cases, we assume the unit subcontracting cost is greater than or equal to the unit production cost, i.e. $c_s \geq c_r$, and the unit selling price is greater than or equal to the unit production cost, i.e., $p \geq c_r$. If $c_r > c_s$, we would outsource the entire production. If $p < c_r$, we would not produce at all. Note that we do not consider the time value of money in our objective since the planning horizon is small (of the order of one year). For longer horizons, cash flows over time would need to be discounted, so that more recent cash flows are assigned a higher weight and value than more distant cash flows.

Given the above notation, our two-stage optimization model may be formulated as follows:

$$
\begin{array}{llll}
\textbf{(P)} & \max\limits_{z \geq 0} & E[R(z,\xi)] & \text{subject to risk } \rho(z) \ \leq \ \overline{\rho} \\
& \text{where,} & R(z,\xi) & \equiv \ r(z,\xi) - C_L(z) \\
& & r(z,\xi) & \equiv \ p\mathbf{e}^T\xi - C_S(z,\xi) \\
& & C_L(z) & \equiv \ K\delta(z) + vz \\
& & C_S(z,\xi) & \equiv \ \min\{g(\mathbf{X}) \ | \ A\mathbf{X} \geq \xi, B\mathbf{X} \leq z\mathbf{e}, \mathbf{X} \geq 0\ \} \\
& & g(\mathbf{X}) & \equiv \ \sum_{t=1}^{H}[c_r Q_t + c_o O_t + c_h I_t + c_s S_t]
\end{array}
$$

Here, $r(z, \xi)$ denotes the "short-term" profit function which excludes long-term capacity costs. Constraint $A\mathbf{X} \geq \xi$ states that demand must be fully satisfied, and $B\mathbf{X} \leq z\mathbf{e}$ ensures that production does not exceed capacity. Different input values for A, B, and X may be used to tailor the formulation to specific applications. For instance, if it is not necessary to immediately satisfy 100% of the demand, you would add a variable for the amount of demand backlogged or lost and penalize this variable in the objective by the appropriate shortage penalty cost. For completeness, we present an example illustrating the formulation for an environment that motivated this research. Note that, throughout this paper, we focus on the case where installed production capacity can be reused[3] over time.

Example 1 *Consider a product such as hamburger patties whose manufacturing consists of a two-stage process: production (bottleneck), followed by packing. Demand over the next four quarters is modeled using a single scenario, ξ. Suppose no overtime is permitted (as would be the case for a processing factory that operates continuously, 24 hours in a day, seven days a week). Let z = # of production lines; $x_t = (Q_t, S_t, I_t)^T$; $\mathbf{X} = (x_1, x_2, x_3, x_4)^T$; $g(\mathbf{X}) = c\mathbf{X}$, with c obtained from cost accounting data.*

$$A = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 \end{bmatrix} \qquad \xi = \begin{bmatrix} 50 \\ 150 \\ 75 \\ 200 \end{bmatrix}$$

For this data, the second row of $A\mathbf{X} \geq \xi$ yields the constraint $I_1 + Q_2 + S_2 - I_2 \geq \xi_2$, which corresponds to the standard inventory balance constraint for period $t = 2$. Each row of matrix B is similarly defined. For example, suppose the regular time production rate is 100 units per quarter per production line. Then, with $B_{i,3(t-1)+1} = 1/100$ and all other $B_{i,\cdot} = 0$, row i of $B\mathbf{X} \leq z\mathbf{e}$ yields the capacity constraint $Q_t/100 \leq z$.

In $\mathbf{P}$, the risk associated with capacity level z is denoted by $\rho(z)$. We introduce two risk measures in this paper: profit variance and mean downside risk (defined in Section 3.1). The objective of the model is to find a capacity level z which maximizes the expected profit subject to an upper bound on risk, i.e., $\rho(z) \leq \overline{\rho}$. By solving problem $\mathbf{P}$ for different values of $\overline{\rho}$, the entire risk-return frontier can be generated. When scenario probabilities P_l can not be easily estimated, the objective in $\mathbf{P}$ could be replaced by $\max_z \min_\xi R(z, \xi)$ (pessimistic view) or $\max_z \max_\xi R(z, \xi)$ (optimistic view) or other regret-based criteria.

Besides permitting risk assessment, the model is an effective means for collaborative supply chain planning. Using model $\mathbf{P}$, suppliers such as NSF can determine the number of production lines to install, given input data from the customer. The customer (McDonald's) can use the model to evaluate the effect of different input values on the system's performance. For instance, currently, market demand is allocated by McDonald's to specific suppliers based on geographical regions, and little or no subcontracting by suppliers is permitted. (McDonald's is clearly the more powerful entity in this supply chain who can dictate supplier behavior.) As a result, suppliers tend to carry higher long-term capacity and often resort to anticipation and safety stock inventory build-up, which translates into higher unit purchase costs for McDonald's. In our industrial experience,

[3]There are other supply contracting environments where a cumulative capacity is contracted for and this supply capacity cannot be reused.

by using the model, McDonald's was able to determine that significant system-wide savings (over 20% reduction in costs) were possible simply by relaxing their geographical demand allocation and subcontracting policies.

3 Structural Properties

In this section we develop analytical properties of the short-term cost and total profit functions. We explore how these functions behave when there is a small perturbation in a demand scenario. Let $\Delta\xi$ to denote such a perturbation. When all components of $\Delta\xi$ are nonnegative, we call it *nonnegative*. When the sum of all components of $\Delta\xi$ is nonnegative, we say it is *sum-nonnegative*. Clearly, if $\Delta\xi$ is nonnegative, it is also sum-nonnegative.

Since $\mathbf{P}$ is a two-stage optimization problem, to solve for the optimal z, we need to characterize the shapes of the functions defined in the model:

Proposition 3.1 *(i) The short-term cost function, $C_S(z,\xi)$, is jointly convex and $r(z,\xi)$ is jointly concave in z and ξ. (ii) For each ξ, $C_S(z,\xi)$ is decreasing and $r(z,\xi)$ is increasing in z. (iii) For long-term capacity cost, $C_L(z) \equiv K\delta(z) + vz$, $E[R(z,\xi)]$ is concave in $z \in (0,+\infty)$; it loses concavity only at the point $z = 0$.*

Proof. This proposition follows from the fact that z and ξ are only in the right hand side of the optimization problem used to calculate $C_S(z,\xi)$; this is a *convex* minimization problem with linear constraints. ∎

Given Proposition 3.1 (iii), the value of z that maximizes total expected profits may be easily determined as either $z = 0$ or the z^* value maximizing $E[R(z,\xi)]$ over $z \in (0,+\infty)$. The concavity of the profit function is illustrated[4] in Figure 1, for two demand scenarios with $\xi_t^H > \xi_t^L$ for all t and $p > c_s > c_r > v$. Observe that, as expected, the optimal capacities satisfy $z^{*H} \geq z^{*L}$. It is interesting to note that the gap between $r(z,\xi^L)$ and $r(z,\xi^H)$ is widening as z increases, as is the gap between $R(z,\xi^L)$ and $R(z,\xi^H)$. We state and prove this observation more rigorously as follows:

Proposition 3.2 *For any ξ and nonnegative $\Delta\xi$, $r(z+\Delta z,\xi) - r(z,\xi) \leq r(z+\Delta z,\xi+\Delta\xi) - r(z,\xi+\Delta\xi)$ for all z and all $\Delta z > 0$.*

Proof. Define $\mathbf{w} = (-z\mathbf{e},\xi)^T$ and re-write $r(z,\xi)$ as follows,

$$r(z,\xi) \equiv \tilde{r}(\mathbf{w}) = (\mathbf{0},\ \mathbf{p})\mathbf{w} - \min_{\mathbf{X}\geq 0}\left\{ g(\mathbf{X}) \;\middle|\; \begin{pmatrix} -B\mathbf{X} \\ A\mathbf{X} \end{pmatrix} \geq \mathbf{w}. \right\}$$

[4] Although we illustrate our results in this paper using specific examples, the observations hold true for all the problem instances in our computational testing.

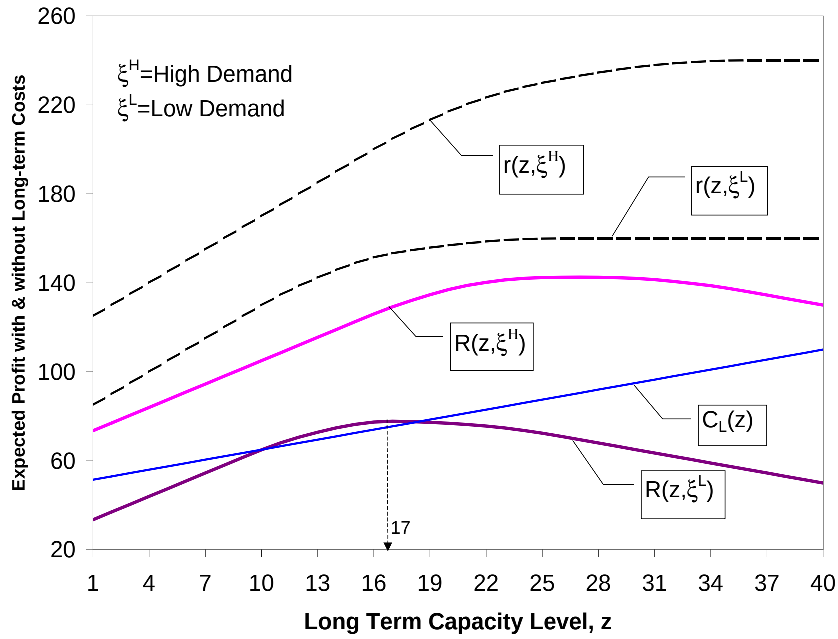

Figure 1: Change in Profit with z for High and Low Demand

Clearly, $\widetilde{r}(\mathbf{w})$ is concave in $\mathbf{w}$. Consider any three vectors: $\mathbf{w}^H$, $\mathbf{w}^L$ and $\Delta\mathbf{w}$, where every component of $\mathbf{w}^H$ is greater than or equal to the corresponding component of $\mathbf{w}^L$, and $\Delta\mathbf{w}$ is nonnegative. By concavity, we have

$$\widetilde{r}(\mathbf{w}^H - \Delta\mathbf{w}) - \widetilde{r}(\mathbf{w}^H) \geq \widetilde{r}(\mathbf{w}^L - \Delta\mathbf{w}) - \widetilde{r}(\mathbf{w}^L).$$

Using $\mathbf{w}^H = (-z\mathbf{e}, \xi + \Delta\xi)^T$, $\mathbf{w}^L = (-z\mathbf{e}, \xi)^T$, and $\Delta\mathbf{w} = (\Delta z\mathbf{e}, 0)^T$, yields the required result. ■

Corollary 3.1 *For any ξ and nonnegative $\Delta\xi$, $R(z + \Delta z, \xi) - R(z, \xi) \leq R(z + \Delta z, \xi + \Delta\xi) - R(z, \xi + \Delta\xi)$ for all z and all $\Delta z > 0$.*

Intuitively, Proposition 3.2 and and Corollary 3.1 state that the benefit of additional capacity is larger when demand is high than when demand is low. This is because, at any capacity level z, the cost of reactive short-term options will be larger when demand is higher. As z is increased, the long-term capacity costs increase, however, we save on short-term costs; these savings will be larger for the case of higher demand. Stated differently, the above results verify that the marginal profit generated from having extra demand will increase when we have a higher capacity level. This statement is not only true for fixed demand scenarios, but it also holds with random demands as follows.

Proposition 3.3 *Define $\Delta R(z, \xi^H, \xi^L) = E[R(z, \xi^H)] - E[R(z, \xi^L)]$. If ξ^H is stochastically greater than ξ^L, then $\Delta R(z, \xi^H, \xi^L)$ is increasing in z.*

Proof. By definition in Yao (1994), X is stochastically greater than Y iff $E[\phi(X)] \geq E[\phi(Y)]$ for any increasing function ϕ. Since ξ^H is stochastically greater than ξ^L and because $R(z, \xi + \Delta\xi) - R(z, \xi)$ is increasing in ξ (by Corollary 3.1), we see that $\Delta R(z, \xi^H, \xi^L)$ is increasing in z. ■

So far, we have seen that the expected profit is a concave function of capacity. Thus there exists a z^* which maximizes $E[R(z, \xi)]$. A company wanting to maximize expected profit should build its capacity at z^* (or get as close to z^* as the expansion budget permits). Proposition 3.3 shows that our model supports the intuition that, to maximize expected profit, a company that forecasts stochastically higher demand should install higher capacity (i.e., $z^{*H} \geq z^{*L}$). When capacity planners have other concerns such as the potential fluctuation in profit or the possible shortfall from a target profit, z^* may not necessarily be the *best* choice. Will this best choice be greater or smaller than z^*? In this context, the histograms in Figure 2 illustrate the spread in realized profits over 81 demand scenarios for two values of z. We see that the distribution of profit can change with z. Further, the profit spread at the higher capacity level ($z = 25$) is larger than that at the lower capacity level ($z = 10$), implying that the variation in profit increases as production capacity increases. The opposite can be seen to be true for short-term costs (whose variation decreases with z). We explore these properties analytically in the remainder of this section.

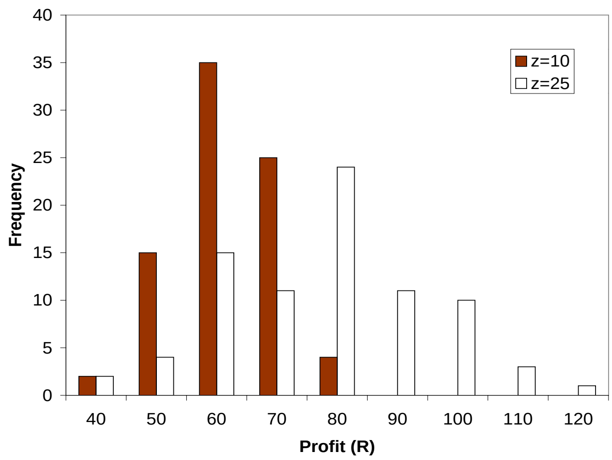

Figure 2: Profit spread under two different long-term capacity levels

3.1 Risk Measures

Often, when companies make production and capacity decisions, risks are as important as average return. There are different ways to measure these risks; we discuss two common measures in this paper. The first is the variance of total profit, which we show typically increases with z. The other measure is *mean downside risk* (Eppen et al. 1989) and may be defined as the expected shortfall of total profit below a specified target level. While this risk measure is not monotone in long-term capacity, it is convex and is a coherent risk measure (Artzner, Delbaen, Eber, and Heath 1999). These properties will be used in Section 4 to develop Pareto risk-return frontiers.

Risk-return frontiers have been widely studied in the finance literature on portfolio selection where the decision variables, X, correspond to the amount invested in each stock. In this case (Markowitz 2000), the mean return is typically linear in the decision variable (i.e., equal to $\vec{\mu}X$, given mean return vector $\vec{\mu}$) and the variance is of the quadratic form $X'\Sigma X$, where Σ denotes an input variance-covariance matrix. This does not apply to our optimization model since the mean and variance of our objective are non-linear functions of the decision variables. Moreover, we are optimizing the long-term capacity level (which maps to the investment budget in portfolio selection) under demand satisfaction constraints. These differences necessitate further analysis of risk in our operational setting.

The capacity level exhibits a different impact on the profit variance depending on the relation between the selling price p and the subcontracting cost c_s. The next two propositions describe the impact when $p \geq c_s$; the case with $p < c_s$ will be considered later. We use $V[x]$ to denote the variance of a random variable x, and $\mathsf{Cov}[x, y]$ to be the covariance between random variables x and y.

Proposition 3.4 *For any ξ and sum-nonnegative $\Delta\xi$, $r(z,\xi) \leq r(z,\xi+\Delta\xi)$ for all z when $p \geq c_s$.*

Proposition 3.5 *For any z and $\Delta z \geq 0$, (i) $V[R(z+\Delta z,\xi)] - V[R(z,\xi)] \geq 0$ when $p \geq c_s$, and (ii) $V[C_S(z+\Delta z,\xi)] - V[C_S(z,\xi)] \leq 0$.*

Proof. We prove (i) and (ii) can be shown analogously. Since $R(z,\xi) = r(z,\xi) - C_L(z)$, it suffices to show that $V[r(z+\Delta z,\xi)] - V[r(z,\xi)] \geq 0$. Clearly,

$$V[r(z+\Delta z,\xi)] - V[r(z,\xi)] = \mathsf{Cov}[r(z+\Delta z,\xi) + r(z,\xi), r(z+\Delta z,\xi) - r(z,\xi)]. \tag{1}$$

From Proposition 3.2 and 3.4, both $r(z+\Delta z,\xi) - r(z,\xi)$ and $r(z+\Delta z,\xi) + r(z,\xi)$ are increasing in ξ. Using the independence of demand and Theorem 1 in Lehmann (1966), $r(z+\Delta z,\xi) - r(z,\xi)$ and $r(z+\Delta z,\xi) + r(z,\xi)$ are concordant and hence *positively quadrant dependent*. Therefore, by Lemma 3 in Lehmann (1966), the covariance is nonnegative. ■

Proposition 3.5 shows that profit variance is increasing in the capacity level, which is consistent with our observation in Figure 2. Variance in short-term costs decreases with z. It is insightful to contrast this behavior of variance with the corresponding behavior of the expected value. Since $E[R(z,\xi)] \equiv p\mathbf{e}^T E[\xi] - E[C_S(z,\xi)] - C_L(z)$, profit and cost are equivalent in terms of an expected value objective. That is, maximizing expected profit will yield exactly the same solution (long-term and short-term decisions) as minimizing expected cost. In contrast, profit $R(z, \xi)$ and cost $C_s(z, \xi)$ behave very differently in terms of variance risk, since profit variance increases with z while cost variance decreases. Thus risk-averse decision makers focusing on cost variance would tend to install more capacity while risk-averse decision makers focusing on profit variance would tend to install less capacity. This may partially explain why plant operations managers responsible for cost control want more capacity installed, thereby avoiding expensive subcontracting options, while corporate managers responsible for revenue and profits prefer to install lower capacity. In our opinion, *it is important for operations managers to adopt an enterprise risk perspective and use a profit objective instead of cost, even though profit and cost may be equivalent in terms of expected values.* Also note that $V[r(z, \xi)] = V[p\mathbf{e}^T\xi] + V[C_S(z, \xi)] - 2\text{Cov}[p\mathbf{e}^T\xi, C_S(z, \xi)]$. $V[p\mathbf{e}^T\xi]$ is constant in z, and $V[C_S(z, \xi)]$ is decreasing in z. Consequently, for $V[r(z, \xi)]$ to increase in z, the covariance must decrease greater than *half* as fast as the variance in short-term costs. Furthermore, using Propositions 3.1 and 3.5, we can infer that, when $p \geq c_s$ and $z \in (0, z^*]$, both the mean profit and the profit variance are increasing in z, where z^* maximizes the mean profit. This is conceptually the same as in the finance literature: the higher mean profit we pursue, the more risk we are exposed to (see Markowitz 2000).

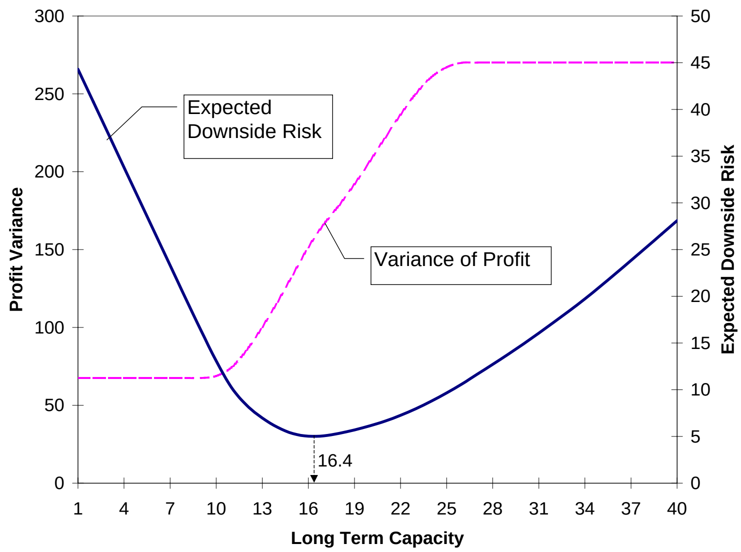

Figure 3: Two Risk Measures and Their Properties

Figure 3 provides an example of the profit variance as a function of z. We observe that the profit variance is fairly constant for low or high z values. With z large enough we do not need to use inventory and subcontracting, therefore the profit variance is close to $V[(p - c_r)\mathbf{e}^T\xi]$ which is independent of z. Similarly, at very low capacity levels, the expensive subcontracting is the dominant force for meeting demand, hence the profit variance is given by $V[(p - c_s)\mathbf{e}^T\xi]$ which is also independent of z and smaller than $V[(p - c_r)\mathbf{e}^T\xi]$. For intermediate values of z, the cost of meeting demand lies between c_r and c_s, and the profit variance increases from $V[(p - c_s)\mathbf{e}^T\xi]$ to $V[(p - c_r)\mathbf{e}^T\xi]$.

For the case of $p < c_s$, we see that the result in Proposition 3.5 does not apply. In fact, the profit variance no longer exhibits monotonicity when $p < c_s$. Instead, the relation appears to be unimodal according to our numerical experiments: the variance decreases when the capacity level is low, and increases when the capacity is high. This is illustrated in Figure 4 for a problem instance with the same data as that used to generate Figures 1 and 3, except for c_s being increased above p. Comparing Figures 1 and 4, we see that the capacity level z^* maximizing expected profits is larger ($z^* = 20$) for $p < c_s$ than the corresponding level ($z^* = 17$) for $p \geq c_s$. This is intuitive, since increasing the subcontracting cost makes extra installed capacity a more attractive option. What is less intuitive is that, while the expected profit remains concave, the profit variance in Figure 4 is not monotone, but unimodal in z. In this example, the mean demand is 20 per period, and $V[r(z,\xi)]$ is decreasing in z when $z \leq 20$ and is increasing when $z > 20$.

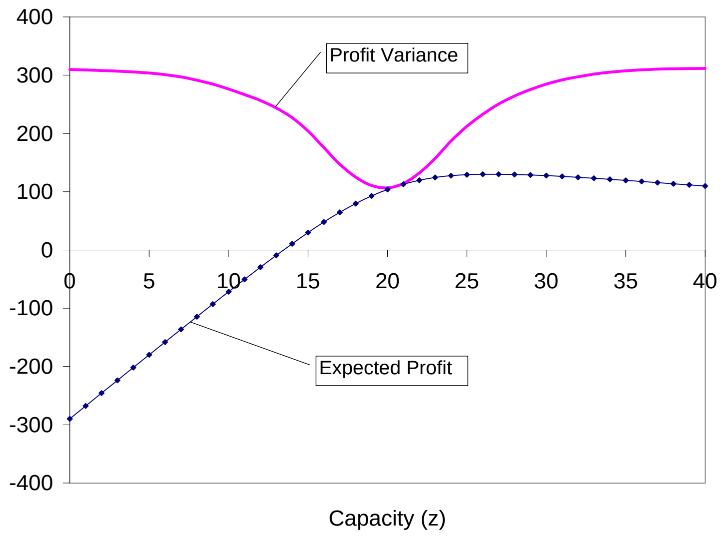

Figure 4: Profit Variance as a Function of z when $p < c_s$.

To better understand this behavior, recall that, by (1), $V[r(z + \Delta z, \xi)] - V[r(z,\xi)] = \mathsf{Cov}[r(z + \Delta z, \xi) + r(z, \xi), r(z + \Delta z, \xi) - r(z, \xi)]$. Hence, when the covariance in the above right hand side is nonnegative (nonpositive), the variance $V[r(z,\xi)]$ is non-decreasing (non-increasing) in z. Thus the monotonicity of profit variance requires an understanding of how the arguments of this covariance change with ξ. By Proposition 3.2, we know $r(z + \Delta z, \xi) - r(z, \xi)$ is always increasing in ξ. But this is not always the case with $r(z + \Delta z, \xi) + r(z, \xi)$. When $p < c_s$ and z is low, an increase in demand will require additional subcontracting which reduces the profit. Thus $r(z + \Delta z, \xi) + r(z, \xi)$

may decrease in ξ when the capacity level is below a certain level. If so, $r(z+\Delta z,\xi)+r(z,\xi)$ and $r(z+\Delta z,\xi)-r(z,\xi)$ are negatively quadrant dependent by Lehmann (1966), and so the covariance is nonpositive. When the capacity is large enough and the expensive subcontracting option can be avoided, $r(z+\Delta z,\xi)+r(z,\xi)$ will increase in ξ, and consequently, $r(z+\Delta z,\xi)+r(z,\xi)$ and $r(z+\Delta z,\xi)-r(z,\xi)$ will be positively quadrant dependent. In this case, the covariance will be nonnegative. Based on this behavior of the covariance term, we can find $\hat{z}$ and $\bar{z}$ such that the profit variance decreases in z when $z \leq \bar{z}$, and increases when $z \geq \hat{z}$. For example, when demand is finite, $\hat{z}$ could be the maximum demand, and $\bar{z}$ could be the minimum demand. However, analytically proving the unimodality of profit variance requires us to show that these $\hat{z}$ and $\bar{z}$ are equal, which appears to be a challenging area for future research.

Although the profit variance exhibits some nice properties, it may not be considered the appropriate risk measure in certain instances (Eppen, Martin, and Schrage 1989). An alternative measure is the following, for which the subsequent two properties are easily proven:

Definition 1 $E[(\overline{R}-R(z,\xi))^+]$ *is called the* mean downside risk *(*MDR*), where* $\overline{R}$ *is the target total profit and* R *is the total profit random variable (whose value depends on the realized demand).*

Proposition 3.6 $E[(\overline{R}-R(z,\xi))^+]$ *is convex in* z *and increasing in* $\overline{R}$.

Proposition 3.7 *Let* MDR(z) *denote* MDR *as a function of* z. *If* $z^* = argmax_z E[R(z,\xi)]$ *and* $z_r = argmin_{z\leq z^*} E[(\overline{R}-R(z,\xi))^+]$ *(where, in case of ties, the smallest value of* z_r *is chosen), then both* z_r *and* MDR(z_r) *are increasing in* $\overline{R}$.

Thus Proposition 3.7 states that increasing the target total profit increases both the capacity level that minimizes mean downside risk and the corresponding MDR value. We also note that although MDR is not monotonic in z, it is a convex and coherent (Artzner, Delbaen, Eber, and Heath 1999) risk measure. The form of MDR(z) is illustrated in Figure 3 for $\overline{R} = \max_z E[R(z,\xi)]$. We will return to this risk measure and further compare it to the profit variance after we develop risk-return Pareto frontiers.

4 Profit-Risk Frontiers

Similar to risk-return curves in finance, we use a variance vs. mean profit or a mean downside risk vs. mean profit curve to quantify the tradeoff between maximizing expected total profit and limiting risk. The Pareto optimal frontier corresponding to our risk-return curves are based on the following definition.

Definition 2 *Consider an expected profit function $E[R(z)]$ and a specified risk measure $\rho(z)$. For any two capacity choices, z_1 and z_2, we say that z_1 is more* <u>efficient</u> *than z_2 if $E[R(z_1)] \geq E[R(z_2)]$ and $\rho(z_1) \leq \rho(z_2)$ with at least one inequality being strict. The* <u>Risk Return Frontier</u> *is a curve of risk measure vs. expected total profit such that any point on the curve corresponds to a most efficient capacity decision for which it is impossible to increase the expected profit and reduce the risk.*

Illustrative risk return curves are shown in Figure 5. The rightmost point on the profit variance vs. expected profit curve corresponds to the maximum expected total profit over the capacity domain. This point splits the curve into upper and lower portions. By Definition 2, the lower portion of the curve, shown in bold, is *efficient*, since it has lower variance. By removing the upper half of the curve, we derive a risk profit frontier, named the *mean-variance* (MV) frontier. The corresponding long-term capacity on the frontier runs from z_a to z^*, where z^* is the long-term capacity maximizing the expected total profit and $z_a = 0$ ($z_a > 0$ when $p < c_s$). From the corresponding MDR efficient

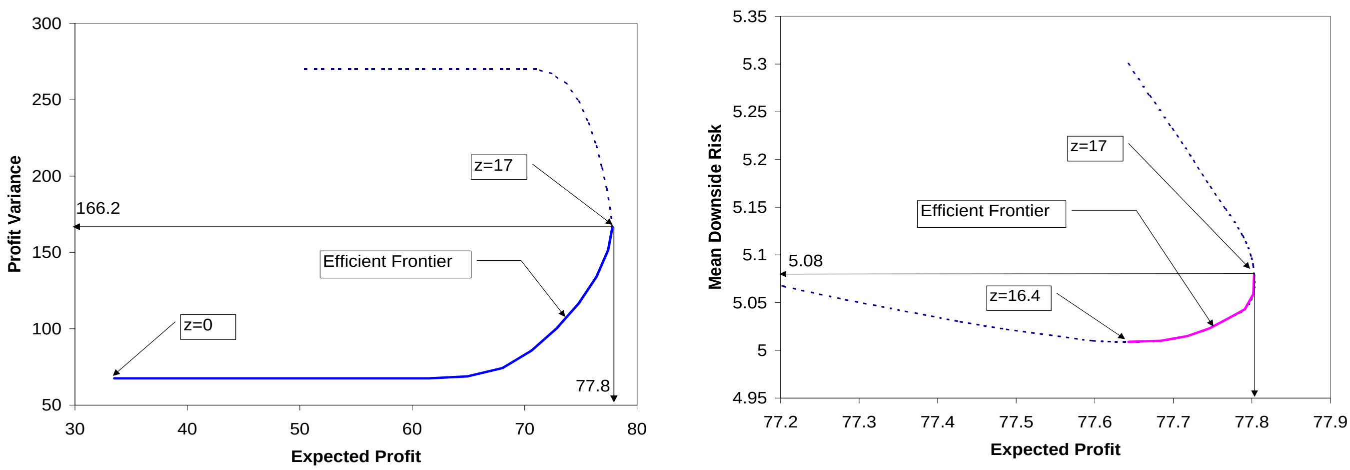

Figure 5: Profit Variance and Mean Downside Risk vs Expected Total Profit with $z^* = 17$, $z_a = 0$, $z_r = 16.4$. (The dotted line shows Profit-Risk combinations corresponding to inferior capacity choices; the solid line shows the efficient frontier.)

frontier in Figure 5 we see that the long-term capacity on the frontier (shown in bold) runs from z_r to z^*, where z_r is the long-term capacity level that minimizes the mean downside risk. Comparing the MV frontier and the MDR frontier, we note that

1. As we move from left to right along the two frontiers, the long-term capacity level, expected profit, and both risk measures increase.

2. The right-most point on both frontiers corresponds to the maximum expected total cost. The domain of the MDR frontier is smaller than that of the MV frontier.

3. Within the domain of MDR frontier, both risk measures are efficient. To the left of this domain, only the MV frontier is efficient. Hence, the domain of the MDR frontier may be a preferred region for choosing the long-term capacity level.

4. Table 1 illustrates how the range of the MDR frontier decreases with $\overline{R}$ (Proposition 3.7).

Table 1: Dependence of Frontier Range $z^* - z_r$ on $\overline{R}$

$\overline{R}$	30	40	50	60	70	$R^*(77.8)$	90	100	110	120
Range $(z^* - z_r)$	11.4	8.5	2.2	2.1	1.9	1	0.7	0	0	0

Although the MDR frontier exhibits the property described in item 3 above, numerical experiments show that it becomes very flat when $\overline{R}$ is very large (resulting in $z_r = z^*$) or very small (resulting in MDR $= 0$). In fact, the difference between the maximum and the minimum downside risk in Figure 5 is less than 0.05 as compared to a variance range of 100 or so in the MV frontier. The flatness of MDR frontier makes it an *unattractive* risk measure because if we want to reduce this risk by a small amount, the expected profit will have to drop significantly.

There are several other risk measures that could be computed, for instance, $P(C_S \geq \overline{C}_S)$, the probability that the short-term cost will exceed a target level $\overline{C}_S$, or $P(R \leq \overline{R})$. Such risk measures (including a utility function based approach) may be analyzed within our mathematical modeling framework. As illustrated above, the choice of which risk measure $\rho(z)$ to use and what risk limit $\overline{\rho}$ to specify is non-trivial and depends on the goals and risk aversion of the decision maker. In the next section, we conduct a computational experiment to explore how different problem parameters affect the profit mean and risk.

5 Computational Results

In this section, we summarize our computational study and provide insights obtained from the results. In particular, we examine the effects of demand variability, seasonality, and correlation on mean profit, its variance and mean downside risk. Unless otherwise stated, we restrict attention to the case $p \geq c_s$; results for tested cases with $p < c_s$ were qualitatively similar. For $p \geq c_s$, cost coefficients for regular production, subcontracting and inventory holding are given by $(c_r, c_s, c_h) = (2, 3, 0.5)$, smaller values of c_h were also tested. No overtime was permitted. The unit selling price p is 4. The fixed cost K and the variable cost of capacity v are 50 and 1 respectively. Details of the solution procedure used to find optimal solutions are provided in Appendix A.

Although many different problem instances were solved, we illustrate our computational results using the four demand instances in Table 2. The $\overline{\xi}_t$ in Table 2 denotes mean demand in period

Table 2: Four Demand Instances ($\overline{\xi}_t$ and V_t denote, respectively, the demand mean and variance in period t)

Instance	V_t	$\overline{\xi}_1$	$\overline{\xi}_2$	$\overline{\xi}_3$	$\overline{\xi}_4$	$\overline{\xi}_5$	$\overline{\xi}_6$	$\overline{\xi}_7$	$\overline{\xi}_8$	$\overline{\xi}_9$	$\overline{\xi}_{10}$	$\overline{\xi}_{11}$	$\overline{\xi}_{12}$
ξ^A	6.25	20	20	20	20	20	20	20	20	20	20	20	20
ξ^B	25	20	20	20	20	20	20	20	20	20	20	20	20
ξ^C	6.25	25	15	25	15	25	15	25	15	25	15	25	15
ξ^D	25	25	15	25	15	25	15	25	15	25	15	25	15

t. That is, if scenario l has probability P_l and corresponds to demand ξ_t^l in period t, then $\overline{\xi}_t = \sum_{l \in \mathcal{L}} \xi_t^l P_l$. The mean demand over the entire horizon is $\overline{\xi} = \frac{1}{H}\sum_{t=1}^{H} \overline{\xi}_t$. In multi-period problems, the variance of demand may be decomposed into two components, V_t and V: let $V_t = V[\xi_t]$ denote the variability of demand ξ_t in period t; let $V = V[\overline{\xi}_t]$ be the variance of mean demand, $\overline{\xi}_t$, over the planning horizon. That is, $V = \frac{1}{H}\sum_{t=1}^{H}(\overline{\xi}_t - \overline{\xi})^2$. Note that V_t represents the *variability of demand* in a period, whereas V corresponds to the *seasonality in mean demand* from period to period. In Table 2, each demand instance has identical V_t for all t, with either $V_t = 6.25$ for instances ξ^A and ξ^C or $V_t = 25$ otherwise. For instances ξ^A and ξ^B, $V = 0$, while ξ^C and ξ^D have $V = 25$. Before we show in Section 5.1 that the two demand parameters V_t and V have significantly different effects on the profit variance, we illustrate some basic what-if capabilities facilitated by the model.

Cost Coefficients and Expected Profit: Through computational experiments, we verified that the impact of cost coefficients on expected profits diminishes as the long-term capacity increases. More specifically, we note the following:

- When the long-term capacity is small (compared to the mean demand $\overline{\xi}$ or to z^*, the capacity level maximizing expected profit), it is likely that the production system will have to resort to subcontracting to meet demand since anticipation inventory can rarely be built due to the limited regular production capacity. Hence, the expected profit will be very sensitive to unit cost of subcontracting, and insensitive to the unit holding cost. Our computational experiments indicate that a ten percent increase in unit subcontracting cost could result in nearly a ten percent reduction in expected profit. However, for the same problem instances, a 500 percent increase in unit holding cost, decreases profit by at most two percent. This is a consequence of the fact that for our problem instances, typically, the inventory cost incurred over the planning horizon is much smaller than the corresponding regular production and subcontracting cost.

- When long-term capacity is large compared to $\overline{\xi}$, regular time production is often adequate to satisfy demand, and other production sources, such as inventory and subcontracting, will be seldom used. Hence their unit costs have little impact on expected profits.

When $p < c_s$, results are similar; however, the holding and subcontracting cost parameters have a much higher impact on expected profit (when compared to the case with $p \geq c_s$). The impact of holding cost is perceptible only at intermediate values of capacity (close to $\overline{\xi}$ or z^*). When installed production capacity is insufficient to meet demand, the impact of c_s on expected profit increases as c_s increases or z decreases. As expected, the z^* value maximizing expected profits increases as c_s increases.

5.1 Variability vs. Seasonality

To understand how the demand variability and seasonality affect profits, we calculated the expected profit, the profit variance, and the mean downside risk for the demand instances in Table 2. The results are summarized in Tables 3−6, from which we observe the following:

1. *Although the four demand instances have very different variability and seasonality, their expected profits are nearly equal at all capacity levels.* The demand instances with lower variability do generate a slightly higher mean profit, but the maximum difference is no more than 2.5%, which occurs at the medium capacity level. This may be explained by noting that our model provides enough flexibility in terms of recourse actions (contingent production and subcontracting) to keep expected profit relatively unchanged by second order moments of demand. (This insensitivity has been observed in a different global capacity planning context with a production switching recourse option under exchange rate fluctuations − see Cohen and Huchzermeier (1998) for a review.)

Table 3: Expected Profit

z	2	6	10	14	18	22	26	30	34	38
ξ^A	209.6	249.6	289.6	329.6	366.9	382.7	377.2	369.2	361.2	353.2
ξ^B	209.9	249.8	289.6	327.7	358.3	374.2	375.6	369.4	361.7	353.7
ξ^C	210.0	250.0	289.9	328.1	358.5	374.5	376.0	369.9	362.0	354.0
ξ^D	209.6	249.4	288.2	323.8	351.7	367.7	371.5	367.9	361.1	353.3

2. *Higher demand variability causes proportionally higher profit variance.* In Table 4, at any capacity level, the ratio of profit variance is approximately equal to the corresponding ratio of their V_t values.

3. *Demand seasonality has a complex, but much less significant impact on the variance in profit.* Demand ξ^A has $V_t = 6.25$ and $V = 0$, whereas ξ^C has the same V_t but $V = 25$. In other words, these two demand instances have the same demand variability but different seasonality. There appears to be no clear dominance relationship between the profit variances from

Table 4: Profit Variance

z	2	6	10	14	18	22	26	30	34	38
ξ^A	72.5	72.5	72.5	73.0	104.4	245.4	288.6	290.0	290.0	290.0
ξ^B	294.9	295.5	301.7	344.6	540.5	839.2	1087	1163	1178	1180
ξ^C	75.4	75.4	76.4	92.0	136.8	216.8	273.3	299.1	301.5	301.5
ξ^D	315.5	323.0	353.3	431.7	633.8	867.4	1086	1195	1236	1246

these two instances: ξ^C with higher V results in higher profit variance at lower and higher capacity levels, and lower variance at medium capacity levels. We observe a similar relationship between ξ^B and ξ^D. Also note that *the seasonality V has much smaller impact than the variability V_t on the profit variance.*

Observations 2 and 3 above may be explained as follows: The short-term cost model reacts to seasonality by carrying inventory when capacity is sufficiently high. Whenever little or no inventory is carried from one period to the next, the periods become independent of each other and, as result, profit variance is largely dependent upon the uncertainty within a period V_t, and less on seasonality across periods V. Low inventory is typically carried at low capacity levels (when no residual capacity is available to build inventory) or at high capacity levels (when inventory is unnecessary). Hence in these cases, we can expect instances ξ^A and ξ^C to have similar profit variance (since they have similar V_t), profit variance of ξ^C may be slightly higher due to the larger V. For medium capacity levels (near the z^* that maximizes expected profit), the interaction between demand variability and seasonality is complex, demand fluctuations can sometimes be effectively buffered by inventory, which may partially explain why the profit variance is lower.

4. Variability of demand affects mean downside risk in a manner that is opposite to that on the profit variance: *higher demand variability corresponds to lower MDR.* This is because demand with higher variability has a higher chance of generating profit exceeding the target profit thereby reducing MDR, which is a consequence of the fact that *MDR(z) is stochastically decreasing and convex* (Shaked and Shanthikumar 1994) in z. In our experiments, *the impact of demand seasonality on MDR had a scale similar to that of demand variability* (c.f. ξ^A and ξ^C). This is quite different from the effect on profit variance where seasonality has significantly smaller impact. Note that we used 95% of the maximum profit as the target profit for calculating the mean downside risks in Table 5.

   In order to assess the impact of the target profit on the mean downside risk, we computed MDR for different values of target profit $\overline{R}$ using demand instance ξ^A. Table 6 shows MDR(z) for low z values. Our computational results suggest that the mean downside risk has a near linear decreasing relationship with z when the capacity is very small (less than half the mean

Table 5: Mean Downside Risk (MDR)

z	2	6	10	14	18	22	26
ξ^A	99.96	51.96	5.73	0.0	0.0	0.0	0.0
ξ^B	93.68	45.78	5.92	0.06	0.0	0.0	0.0
ξ^C	93.89	46.01	6.00	0.06	0.0	0.0	0.0
ξ^D	89.26	41.60	5.21	0.15	0.04	0.03	0.02

demand $\overline{\xi}^A$), the MDR diminishes to zero as the capacity increases. As target profit increases the MDR(z) curves move upwards and rightwards. *MDR can be significantly affected by target profit* $\overline{R}$.

z	1	2	3	4	5	6	7
$\overline{R} = 100\%R^*$	130.65	118.65	106.65	94.65	82.65	70.65	58.65
$\overline{R} = 90\%R^*$	87.95	75.95	63.95	51.95	39.95	27.95	16.01
$\overline{R} = 80\%R^*$	47.95	35.95	23.95	12.16	3.11	0.23	0.00

Table 6: Mean Downside Risk (MDR) under Different Target Profits. (R^* denotes the maximum expected profit.)

Amongst our results we also noted that, for our problem instances, the short-term recourse option of carrying inventory was used much less than the subcontracting option, particularly at low or high capacity levels. This holds true, though to a lesser degree, even at intermediate capacity levels (possibly because the unit holding cost rate of 0.25 is quite high relative to the unit cost). In this case, the supplier needs to pick its subcontractors more carefully than it needs to plan for inventory storage. Thus, the model provides an effective means for evaluating the value of the subcontracting option, which can be a useful in negotiating with potential "fall-back" subcontractors. Other interesting research issues that could be considered include sensitivity analysis on factors such as the period length and the planning horizon, penalty costs for unfulfilled demand (which can be either lost or backlogged), and alternative cost functions for long-term and short-term costs. Our preliminary experiments suggest that these what-if questions raise new, interesting issues and are therefore promising avenues for future research. We conclude this section with a few observations from our experiments with autocorrelated demand.

Correlated Demand: To study the case when demand from period to period is not independent, we generated correlated demand using the three-step method in Erkip, Hausman, and Nahmias (1990): $\xi_t = \mu_t L_t + \epsilon_t$ where $L_t = \sigma_L Z_t + 1$ with $Z_t = \rho Z_{t-1} + \epsilon_{z_t}$ and ϵ is an appropriate normal variate. Using these formulas, the demand at period t, ξ_t, can be made to have mean μ_t and variance σ_t^2. The correlation of demand across periods is controlled by ρ. When $\rho = 0$, demands are independent. As in the instances considered earlier in Section 5, we consider a 12-period horizon with $\mu_t = 20$ and $V_t = \sigma_t^2 = 6.25$ for all t. All cost parameters were the same as the case considered earlier with $p \geq c_s$. Five cases of demand correlation were examined: $\rho = 0$, $\rho = \pm 0.5$ and $\rho = \pm 0.8$. For the generated demand scenarios at different values of ρ, the total demand variance over the 12-period horizons are respectively 62.12, 67.46, 72.49, 96.26, and 132.77, for ρ = -0.8, -0.5, 0, 0.5, and 0.8. Here, the total demand variance is defined to be the sum of demand variances in each period plus the covariance between demand in any two different periods Clearly, total demand variance is increasing in ρ. Note that, when $\rho = 0$, the theoretical total variance should be 75 (6.25×12). The discrepancy in the above table results from the limited sample size, which was 1000 demand scenarios. From these experiments, we observed the following:

- Expected profits and mean downside risks are not significantly affected by the demand correlation (under 0.5% change in expected profit with ρ).
- The demand correlations have a significant impact on profit variance (see Table 7). Negative correlation reduces the profit variance, whereas positive correlation increases the profit variance. When capacity is high (e.g., $z = 38$), the profit variance is close to $(p - c_r)^2 \times$ total demand variance. When the capacity is low (e.g. $z = 2$), the variance of profit is close to the total demand variance, since subcontracting is frequently used and $(p - c_s) = 1$.

Table 7: Profit Variance under Correlated Demands

z	2	6	10	14	18	22	26	30	34	38
$\rho = 0$	72.5	72.5	72.5	73.0	104.4	245.4	288.6	290.0	290.0	290.0
$\rho = 0.5$	96.3	96.3	96.3	97.3	137.5	316.7	383.2	385.0	385.0	385.0
$\rho = -0.5$	67.5	67.5	67.5	68.4	99.3	230.6	268.7	269.8	269.8	269.8
$\rho = 0.8$	132.8	132.8	132.8	134.1	191.0	439.0	528.8	531.1	531.1	531.1
$\rho = -0.8$	62.1	62.1	62.1	62.7	88.9	209.7	247.4	248.5	248.5	248.5

6 Summary and Future Research

The primary objective of this paper was to explore the interaction between long-term and short-term capacity decisions under demand uncertainty. By incorporating risk measures into our model, we are able to quantify the expected profit and its variability or downside risk at each long-term capacity level, thereby developing risk-return frontiers. Several analytical results are presented. For instance, if z^* maximizes expected profit, then $z < z^*$ would reduce profit variance (at the cost of reduced mean profit and increased cost variance). Our results help managers to better appreciate the tradeoff between profit and risk, and to make capacity decisions accordingly. Through computational experiments, we study the impact of cost coefficients, demand variability, seasonality and correlation, on profit mean and risk, which provides interesting insights into the annual capacity planning problem. For instance, the behavior of profit variance with capacity depends significantly on the relative values of unit selling price and unit subcontracting costs as well as the degree of demand correlation.

We view this research as a starting point in exploring profit and risk for the multi-product capacity planning model with shared capacity consisting of multiple resource types (e.g., dedicated vs. specialized), multiple sources of uncertainty, correlated demands, while incorporating other decisions such as selecting the subset of products for which capacity should be installed ($z > 0$) under fixed expansion budgets. A related area worthy of research is effective scenario sampling to balance the tradeoff between greater accuracy afforded by more scenarios and the increased solution time that would result with large sample sizes.

Acknowledgement: We thank Dr. Martin Levy at the University of Cincinnati for directing us to the Lehmann reference that simplified the proof of Proposition 3.5. We thank the editors and two annoynmous referees whose careful comments help improve the paper.

References

Artzner, P., F. Delbaen, J. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 9*, 203–228.

Carr, S. and W. Lovejoy (2000). The inverse newsvendor problem: Choosing an optimal portfolio for capacitated resources. *Management Science 46*(7), 912–927.

Cohen, M. and A. Huchzermeier (1998). Global supply chain management: A survey of research and applications. pp. 669–702. Chapter 21 in Quantitative Models for Supply Chain Management, edited by Tayur, Ganeshan, and Magazine.

Davis, M. ., M. Dempster, S. Sethi, and D. Vermes (1987). Optimal capacity expansion under uncertainty. *Advances in Applied Probability 19*, 156–176.

Eppen, G. D., K. R. Martin, and L. Schrage (1989). A scenario approach to capacity planning. *Operations Research 37*, 517–527.

Erkip, N., W. H. Hausman, and S. Nahmias (1990). Optimal centralized ordering policies in multi-echelon inventory systems with correlated demands. *Management Science 36*, 381–392.

Fine, C. and R. Freund (1990). Optimal investment in product-flexible manufacturing capacity. *Management Science 36*(4), 449–466.

Harrison, M. and J. V. Mieghem (1999). Multi-resource investment strategies: operational hedging under demand uncertainty. *European Journal of Operational Research 113*, 17–29.

Karmarkar, U. and S. Kekre (1987). Manufacturing configuration, capacity and mix decisions considering operational costs. *Journal of Manufacturing Systems 6*, 315–324.

Lehmann, E. (1966). Some concepts of dependence. *Annals of Mathematical Statistics 37*, 1137–1153.

Luss, H. (1982). Operations research and capacity expansion problems: A survey. *Operations Research 30*, 907–947.

Markowitz, H. (2000). *Mean-Variance Analysis in Portfolio Choice and Capital Markets.* Fabozzi Associates.

Paraskevopoulos, D., E. Karakitsos, and B. Rustem (1991). Robust capacity planning under uncertainty. *Management Science 37*, 787–800.

Rao, U., A. Scheller-Wolf, and S. Tayur (2000). Development of a rapid-response supply chain at caterpillar. *Operations Research 48*, 189–204.

Rinks, D., J. Ringuest, and M. Peters (1987). A multivariate utility function approach to stochastic capacity planning. *Engineering Cost and Production Economics 12n1-4*, 3–13.

Shaked, M. and G. Shanthikumar (1994). *Stochastic Orders and their Applications.* Academic Press, Inc.

Singhal, V., A. Raturi, and J. Bryant (1994). On incorporating business risk into continuous review inventory models. *European Journal of Operational Research 75*, 136–150.

Yao, D. (1994). *Stochastic Modeling and Analysis of Manufacturing Systems.* Springer.

Zsidisin, G., A. Panelli, and R. Upton (2000). Purchasing organization involvement in risk assessments, contingency plans, and risk management. *Supply Chain Management: An International Journal 5*, 187–197.

A Numerical Solution Procedure

In the computational experiments, for each problem instance, we compute the mean profit, its variance and mean downside risk for different capacity levels assuming that the optimal short-term decisions are implemented. A linear short-term function is used with regular production, subcontracting and inventory holding costs. In this section, for simplicity, we focus mainly on the variance risk measure with $p \geq c_s$ and zero demand correlation (the approach we present may be easily extended to consider mean downside risk, $p < c_s$, and demand correlation). The procedure entails the following two steps.

1. *Generate demand scenarios*: Since there are 12 time periods in Table 2, each scenario is a vector of 12 demand values. In our computational experiments, given mean and variance of demand, we use the finite normal distribution to generate demand scenarios. Number of scenarios generated is 1000, increasing this to 10,000 scenarios changed the profit mean estimate by under 0.3% in all tested instances (average change in profit variance was 0.36%), while running times increased by a factor of greater than 10.

2. *Solve the short-term LP model*: Initially we used Cplex to solve the short-term LP. Recall that we generate 1000 demand scenarios and solve the model for each scenario at an integer capacity level between 1 and 40. Therefore, to determine the profit-risk frontier, the total number of the problems

to be solved can be as large as 40,000. For such a large number of problems, Cplex 6.5 on a Sparc 20 workstation takes hours to solve all problems. Hence, we designed a new, efficient (greedy) optimal solution approach, which is described in detail in Section A.1 below. With this approach, we can solve all 40,000 problems in a few seconds on a PC with an 846MHz CPU and 256MB RAM.

A.1 Solution Approach for Model (P) with Variance Constraint

Model (**P**) on page 4 is solved as a two-stage optimization problem: In the first stage, we compute $r(z,\xi)$ for each z and ξ. In the second step, we calculate $E[R(z,\xi)]$, $V[R(z,\xi)]$, and MDR(z), for each z. A Pareto frontier such as the Mean-Variance frontier can then be drawn using $E[R(z,\xi)]$ and $V[R(z,\xi)]$. Given an upper bound value $\overline{\rho}$ on the profit variance (specified in model (P)), the optimal capacity level can be identified from the frontier.

The core part of this solution procedure is to efficiently solve $C_S(z,\xi)$ for each z and ξ. The LP for $C_S(z,\xi)$ and our greedy solution procedure are specified below.

$$
\begin{aligned}
C_S(z,\xi) \equiv \quad \min \quad & \sum_t c_r Q_t + c_h I_t + c_s S_t \\
s.t. \quad & I_{t-1} + Q_t + S_t - I_t = \xi_t, \ \forall t \\
& Q_t \le z, \ \forall t \\
& I_t \ge 0, Q_t \ge 0, S_t \ge 0, \ \forall t \text{ and } I_0 = 0.
\end{aligned}
$$

Preprocess – Determine Minimum Subcontract Amount:

When cumulative demand exceeds cumulative capacity over the first t periods, subcontracting will be necessary to satisfy demand. In this preprocess step, we determine the minimum amount of subcontracting required in each period by satisfying demand in any period with regular production and/or with inventory produced in a previous period. Let D_t be the portion of demand in period t being satisfied by regular production in period t, and E_t be the residual portion of demand in period t (which is satisfied by production in previous periods when $E_t \ge 0$ and represents residual capacity when $E_t < 0$).

1. Initialize $D_t = \xi_t$ for all t.
2. For $t = 1$ to T,
   - If $\sum_{\tau=1}^{t} D_\tau > z \times t$, set $S_t = \sum_{\tau=1}^{t} D_\tau - z \times t$ and set $D_t = D_t - S_t$.
   - Let $E_t = D_t - z$; when E_t is negative, capacity is higher than demand at period t, otherwise E_t units must be produced at some earlier period(s) (assuming minimal subcontracting). Note that $\sum_{\tau=1}^{t} E_\tau \le 0$ for t.

Construct a Feasible Production Plan: After the preprocess step, we have $\sum_{\tau=1}^{t} D_\tau \le z \times t$ for all t. We now can use regular production and inventory to satisfy the modified demand D_t. When $E_t \le 0$, we use regular production to meet demand. Otherwise, we use excess capacity in previous periods to produce some portion of the demand and carry it over to period t. Let $I_{i,j}$ be the amount of inventory produced in period i for consumption in period j.

1. Set $I_{i,j} = 0$ for $i, j \in \{1, T\}$.
2. For $t = T$ to 1,
   - When $E_t > 0$ do: For $\tau = t - 1$ to 1,
     - ⋆ If $E_\tau \geq 0$, do nothing.
     - ⋆ If $E_t + E_\tau \leq 0$, (i) $I_{\tau,t} = I_{\tau,t} + E_t$, (ii) $E_\tau = E_\tau + E_t$, (iii) $E_t = 0$.
     - ⋆ Else (i) $I_{\tau,t} = I_{\tau,t} - E_\tau$, (ii) $E_t = E_\tau + E_t$, (iii) $E_\tau = 0$.

In the above procedure, we produce inventory using excess capacity in earlier periods that are closest to the time when the inventory is needed. The resulting solution minimizes the total production and inventory holding cost for the modified demand. This solution is not optimal only because some of the inventory can be replaced by subcontracting at a lower cost.

Derive Optimal Solution: Since we assume $c_s \geq c_r$, we only need to consider the tradeoff between producing one unit early and holding it for a future period versus subcontracting it at the future period. If $c_s < c_r + m \times c_h$ for some integer m, then it is cheaper to subcontract a unit of demand instead of producing it m periods earlier and paying holding costs.

1. Find the minimum m such that $c_s < c_r + m \times c_h$. Let S_t be the minimum subcontracting amount derived in Preprocessing step.
2. For $\tau = 1$ to $T - m$,

   For $t = \tau + m$ to T,
     - ⋆ $S_t = S_t + I_{\tau,t}$ $E_\tau = E_\tau - I_{\tau,t}$.
     - ⋆ $I_{\tau,t} = 0$.
3. Compute total costs using the linear objective function with $Q_t = z + E_t$.