ELSEVIER

European Journal of Operational Research 113 (1999) 17–29

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

Multi-resource investment strategies: Operational hedging under demand uncertainty

J. Michael Harrison [a], Jan A. Van Mieghem [b,*]

[a] *Graduate School of Business, Stanford University, Stanford, CA 94305-5015, USA*

[b] *J.L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2009, USA*

Received 1 May 1997; accepted 1 October 1997

Abstract

Consider a firm that markets multiple products, each manufactured using several resources representing various types of capital and labor, and a linear production technology. The firm faces uncertain product demand and has the option to dynamically readjust its resource investment levels, thereby changing the capacities of its linear manufacturing process. The cost to adjust a resource level either up or down is assumed to be linear. The model developed here explicitly incorporates both capacity investment decisions and production decisions, and is general enough to include reversible and irreversible investment. The product demand vectors for successive periods are assumed to be independent and identically distributed. The optimal investment strategy is determined with a multi-dimensional newsvendor model using demand distributions, a technology matrix, prices (product contribution margins), and marginal investment costs. Our analysis highlights an important conceptual distinction between deterministic and stochastic environments: the optimal investment strategy in our stochastic model typically involves some degree of capacity imbalance which can never be optimal when demand is known.

Keywords: Strategic planning; Capacity investment; Prices; Operational hedging; Multi-dimensional newsvendor model

1. Introduction

Consider a firm that markets n products that are jointly manufactured using m resources. Facing uncertain product demand, the firm has the option to change its investment in the m resources, thereby changing the capacity of its manufacturing process. Resources may be thought of as several types of labor and capital, and their increase usually comes at a certain cost, while a decrease may generate a revenue (e.g., in the case of capital resources with resale value) or an expense (if the factor is costly to retire, e.g., labor).

Although multi-resource investment problems may seem natural, they have received little attention in the literature. Economists traditionally identify two factors of production, capital and

* Corresponding author. Tel.: +1 847 491 5481; fax: +1 847 1220; e-mail: vanmieghem@nwu.edu

0377-2217/99/$ – see front matter
PII: S 0 3 7 7 - 2 2 1 7 (9 7) 0 0 4 2 5 - 6

labor, and focus on single resource investment, viewing capital investment as mostly fixed – and thus irreversible – and labor investments as variable and costlessly reversible. Operations researchers, inspired by the seminal work of Manne [1], have focused on capacity expansion and typically study irreversible investment in a single resource [2–4]. This emphasis on investment in a single capital good does not remain justified when employment changes are costly (the "Eurosclerosis" phenomenon). In addition, firms usually invest in a variety of resources with different financial and operational characteristics. Therefore, it is desirable to consider simultaneous investment in multiple resources, preferably under the uncertainty that is crucial to most investment problems.

The operations research literature has studied multi-resource models to solve production and capacity planning problems, typically by invoking mathematical programming techniques [5–7]. Recently an initial effort has been made to analyze true multi-resource investment dynamics: Dixit [8] studies the optimal investment dynamics in two resources, capital and labor. Eberly and Van Mieghem [9] present a general framework to study multi-resource investment under uncertainty in a dynamic, non-stationary environment. They show that under rather general conditions the optimal investment strategy follows a control limit policy, and they provide closed-form solutions for an investment model where uncertainty is modeled by a geometric Brownian motion in continuous time.

This article studies multi-resource investment under demand uncertainty, adding detail and special structure to the general framework developed in [9] in order to make connection with the control levers available to operations managers. While economics papers such as [8,9] take an operating profit function as a primitive, we increase the level of detail by modeling explicitly both manufacturing capacity and production decisions, which means that the operating profit function is endogenous in our model. The formulation remains general enough to include problems of reversible and irreversible investment. Having specialized the general structure proposed in [9], we are able to sharpen the characterization of optimal investment strategies developed in that earlier article. We present a graphical interpretation of the optimal strategy, defined directly in terms of the manufacturing process yet simple enough to be easily taught and remembered.

The decision problems we solve are known in the operations research literature as stochastic programs with recourse. Traditional solutions for such problems use discrete stochastic mathematical programming methods that can capture many practical details, but often at the expense of analytical tractability, so that one must resort to numerical methods. Our approach on the other hand, yields a parsimonious descriptive model that is a multi-dimensional generalization of the familiar "newsvendor model". Like the traditional newsvendor model, our model is amenable to analytic solution and graphical interpretation (but may be too stylized for practical decision support systems). Multi-dimensionality enriches the newsvendor model by incorporating product, demand and resource differentiation through price and cost *vectors*, a technology *matrix* and a *multivariate* demand distribution. In follow-up work, we apply the multi-dimensional newsvendor model that we introduce here to study flexible technology [10] and subcontracting and outsourcing [11]. Both models show how the capacity investment decision not only depends on correlation and risk in product demands, but also on price and cost differentials.

A final contribution of this article is to give qualitative insight into real-world capacity planning and capital budgeting practices, which typically involve two levels of decision making. Initially, lower-level production planners propose least-cost capacity adjustments necessary to enable execution of forecasted production and sales plans, which are treated as deterministic. Ultimately, however, senior-level managers approve an investment plan that is a "perturbed" version of the one recommended on the basis of such deterministic reasoning, seeking to minimize risk exposure due to demand uncertainties. Our model confirms and quantifies the optimal "operational hedge": capacity levels should be set so as to balance "underage costs" and "overage costs", properly defined. The optimal hedging strategy often leads to unbalanced capacities, in the sense that no single demand scenario may exist that implies full utilization of all resources. This observation, that optimal multi-resource investment decisions typically involve some degree of capacity imbalance when demand is uncertain, is not new. On the other hand, it does not seem to be sharply etched in the minds of either academics or practitioners, and our exposition of the unbalancing phenomenon in the context of a stylized model is intended to help remedy that situation. Under certain conditions common in high-tech industries, near-optimal hedging can be achieved by using a suitably defined one-dimensional newsvendor model to plan the capacity of the critical resource, and then investing in ample capacity for the remaining resources. The latter approximation may be helpful to guide outsourcing decisions.

The article is organized as follows. Section 2 briefly reviews the dynamics of multi-resource investment under uncertainty and specializes the general model developed in [9] to linear joint production of n products on m resources. Section 3 presents the multi-dimensional newsvendor solution for the optimal investment levels, which is illustrated with a comprehensive example in Section 4. Section 5 shows how the optimal investment can be interpreted as a hedging strategy against demand uncertainty and relates this to capital budgeting practice and outsourcing decisions. Section 6 summarizes the analysis.

We conclude this introduction with some notational conventions. We will not distinguish in notation between scalars and vectors. All vectors are assumed to be column vectors, and primes denote transposes, so $u'v$ is the inner product of u and v. Vector inequalities should be interpreted component-wise, as well as $\max(0,u)$ and $\max(0,-u)$ which are denoted by u^+ and u^-, respectively. As usual, $P(\cdot)$ denotes a probability measure and E is the associated expectation operator. Finally, $\nabla g(\cdot)$ denotes the gradient of a differentiable function $g(\cdot)$.

2. Linear manufacturing with demand uncertainty

Consider a firm that deploys m resources to manufacture n products in periods $t \in \{1,\dots,T\}$. At the beginning of each period t, the firm must decide on a non-negative m-vector of resource levels $K_t \in \mathbb{R}_+^m$ for period t production, before the product demand vector $D_t \in \mathbb{R}_+^n$ for that period is observed. After demand is observed, the firm chooses an n-vector $x_t \in \mathbb{R}_+^n$ of production quantities for the various products, constrained by its earlier resource investment. This multi-stage decision problem is characteristic of *real option* models [6,12]: first invest in capabilities, then receive some additional information, and finally exploit capabilities optimally, contingent on the revealed information. This decision problem also is known as a *stochastic program with recourse*. As in Dixit's work [8] on investment dynamics, successive periods are linked by resource investment decisions, but not by inventories (which would make the problem much more complicated). These assumptions are consistent with what is observed in practice where "each time period is of sufficient length (e.g. one year) so that production levels can be altered within the time period in order to satisfy as closely as possible the demand that is actually experienced" ([5], p. 519).

We model the firm's manufacturing process and production decisions as follows. In each period t, having chosen a resource vector K_t and observed a demand vector D_t, the firm chooses its production vector x_t so as to maximize profit in the following linear program (often called a *product mix problem*):

$$\max_{x_t \in \mathbb{R}_+^n} \quad p'x_t \tag{1}$$

$$\text{s.t.} \quad Ax_t \leqslant K_t \quad \text{(capacity constraints)}, \tag{2}$$

$$x_t \leqslant D_t \quad \text{(demand constraints)}, \tag{3}$$

where $p \in \mathbb{R}_+^n$ is an n-vector whose jth component represents the *unit contribution margin* for product j (that is, sales price minus variable cost of production), and A is an $m \times n$ *technology matrix* whose (i,j) component represents the amount of resource i required to produce one unit of product j. In accordance with standard terminology in linear programming, K_t often will be called a *capacity* vector. Note that the contribution margins and technology matrix are assumed for simplicity to be time invariant (not depending on t). Also, contribution margins do not depend on the production quantities chosen, so the firm implicitly is assumed to be a price taker in both the output and input market. The optimal objective value of the product mix problem (1)–(3) is the maximal *operating profit function*, and is denoted by $\pi(K_t, D_t) = p'x(K_t, D_t)$, where $x(K_t, D_t)$ is an associated optimal production vector. From linear programming theory we know that $\pi(\cdot, D_t)$ is concave, a fact that is important for the development to follow.

It is straightforward to incorporate demand shortage penalties into the model as follows. Assume product j carries a shortage penalty cost $c_{\mathrm{P},j} \geqslant 0$ for each unit of demand that is not satisfied so that the total shortage cost is $c_{\mathrm{P}}'(D_t - x_t)^+$. Using Eq. (3), it follows that all results presented in this article remain valid if we inflate unit contribution margins p to $p + c_{\mathrm{P}}$ and decrease the operating profit $\pi(K, D)$ by $c_{\mathrm{P}}'D$. Similarly, resource operating costs that depend on the installed capacity K_t, rather than on actual production x_t, can be included if they are convex as a function of K_t.

What connects successive periods in our model is that resource or capacity levels carry over without change unless explicit action is taken. Thus, resources are neither created, depleted, nor destroyed by production activity. If the firm decides to adjust resource vector K_{t-1} to K_t at the beginning of period t, it incurs an *adjustment cost* $C(K_t - K_{t-1})$, where $C(z) = c'z^+ - r'z^-$ and c and r are m-vectors. Thus, by assumption, the cost of adjusting any one resource level either up or down is linear in the size of the adjustment, and total adjustment cost is additive over resources. It is perhaps simplest to think of both c and r as positive, but disinvestment in some resources (for example, labor) may be costly, or even prohibitively expensive (which means that investment is irreversible), and investment in other resources may be so heavily subsidized as to generate a positive cash flow. In general, then, we allow components of both c and r to be either positive or negative, but we focus on resources that are costly to adjust, such that c and r are non-zero, subject to the additional requirement that $c \geqslant \delta r$, where $\delta \geqslant 0$ is the one-period discount factor (see below). If $c < \delta r$, then used capacity would be worth more than new capacity, and the firm could generate a cash stream with arbitrarily large present value by just making a large investment in one resource and reversing it in the following period. Readers should note that resource adjustment costs are assumed to be time invariant.

Because the firm's investment problem is assumed to end after T time periods, we must further specify the salvage value $f(K_T)$ associated with the final vector of resource levels K_T. In the interest of tractability, we assume that $f(K_T) = r'K_T$, which means that disinvestment in all resources is mandatory and the associated marginal revenues are exactly the same as in earlier periods. The firm seeks a multi-period strategy of investment and production that *maximizes* the *expected present value* of operating profits minus resource adjustment costs over T time periods, given an initial m-vector of resource levels K_0 and a single period discount factor $\delta \geqslant 0$. (One naturally thinks in terms of the case where $0 < \delta < 1$.)

The problem formulated above is a special case of the multi-period investment problem analyzed in [9]. The general model in [9] takes as primitive an operating profit function $\pi_t(K_t, \omega)$ for each period t, where ω is the "state of the world" and $\pi_t(K_t, \omega)$ is measurable with respect to information available in period t. In the current context, π_t is assumed to depend on ω only through the period t demand D_t. Furthermore, we have added detail to the model by explicitly recognizing production quantities x_t as decision variables, and then defining $\pi_t(K_t, D_t)$ as the maximal operating profit achievable in the linear programming problem (1)–(3). Because the vector p of contribution margins and the technology matrix A appearing in Eqs. (1)–(3) do not depend on t, the operating profit function $\pi(\cdot, \cdot)$ is itself independent of t, and we have further assumed that the marginal investment cost vector c and the marginal disinvestment revenue vector r are time invariant.

As a last and crucial element of the problem formulation in this article, let us further assume the following special structure: the demand vectors

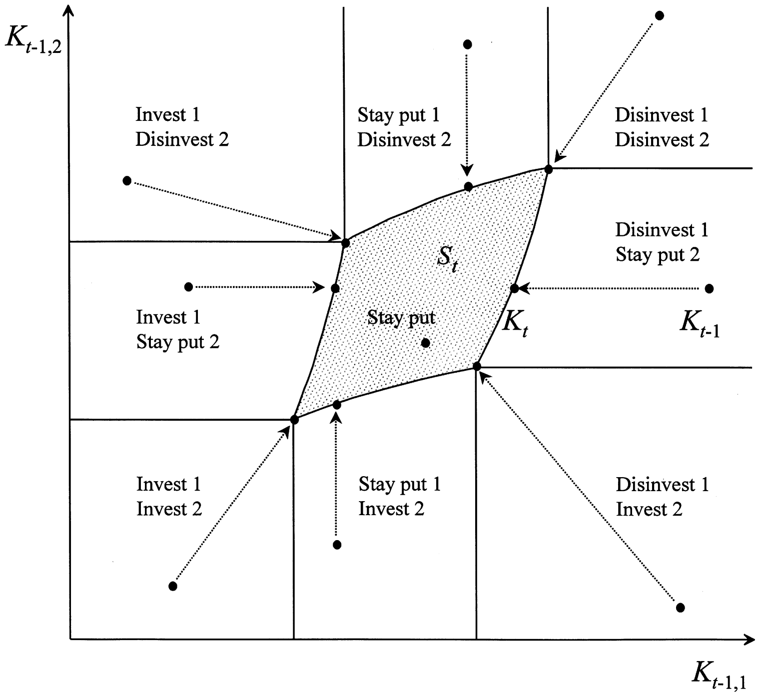

Fig. 1. Structure of the optimal ISD investment policy for $m = 2$ resources.

$D_1, D_2, \ldots, D_T$ are independent and identically distributed (IID), and there is no other relevant source of uncertainty in the problem environment. Under these assumptions, the "state" of our dynamic investment problem at the beginning of period t is fully summarized by the incoming capacity vector K_{t-1}. Because capacities are nonnegative by definition, our "state space" is the orthant $\mathbb{R}^m_+$. In [9] it was shown that there exists an optimal policy having an ISD (*Invest/Stay put/ Disinvest*) structure. Roughly speaking, this means that in each period t, the state space is partitioned into various domains, including a "continuation region" S_t (see Fig. 1) where no adjustment need be made in the vector of current resource levels. From any one of the domains outside this continuation region, the optimal investment action is to adjust the vector of resource levels to a new specified point on the boundary of the continuation region. In the special case treated here (problem data are time invariant, the salvage function is $f(K_T) = r'K_T$, demands are IID, and no other sources of relevant uncertainty exist), the optimal ISD policy is shown in [9] to have the following additional special structure. Denote the expected operating profit by Π, where

$$\Pi(K_t) = \mathrm{E}\pi(K_t, D_t)$$

for all $K_t \in \mathbb{R}_+$, which also is concave and is assumed to be smooth. [1] According to [7], Proposition 1, where the function Π was taken as a primitive, we have:

Proposition 1. *Starting from any initial capacity vector $K_0 \in \mathbb{R}^m_+$, the optimal investment strategy makes no capacity adjustments after the beginning of period 1 (that is, $K_1 = K_2 = \cdots = K_T$ under the optimal strategy). Moreover, the "stay-put region",*

[1] This is the case if D_t is a continuous random variable (see Proposition 2). In general, there may be a set of Lebesgue measure zero where the concave function Π is not differentiable. However, its left and right partial derivatives always exist and all following differential statements should be interpreted in subdifferential form.

or continuation region, characterizing the optimal ISD policy for period 1 is given by

$$S_1 = \left\{ K \in \mathbb{R}^m_+ :\ (1-\delta)r \leqslant \nabla\Pi(K) \leqslant \frac{1-\delta}{1-\delta^T}(c - \delta^T r) \right\}. \tag{4}$$

Effectively, the multi-period solution in our case can be found by solving a model with a single period of length T with equivalent discount rate $\delta^e = \delta^T$ and equivalent marginal adjustment costs:

$$r^e = \frac{1-\delta}{1-\delta^T} r \quad \text{and} \quad c^e = \frac{1-\delta}{1-\delta^T} c, \tag{5}$$

which represent the total marginal cost of (dis)investing in real terms. Because we have shown how to map this multi-period problem into an equivalent single period formulation, we will focus for the remainder of the article on the single period model and its continuation region S.

The fact that our multi-period problem effectively collapses to an equivalent single period problem is of course a result of our highly stylized model. In general, the form of an optimal investment strategy is not this simple (e.g., see the Brownian dynamic investment model in [9]). However, the qualitative insights regarding *operational hedging* that we will derive from our stylized model can be extended to more complex settings with non-stationary problem data, dependent demands, and additional sources of uncertainty, but no attempt will be made to prove that assertion in this article.

For ease of notation, we will suppress all time-subscripts and we will denote the components of a vector v by $v_x, v_y, \ldots$ to prevent confusion.

3. The multi-dimensional newsvendor solution

Continuing with the equivalent single period formulation, the continuation region S of the optimal policy for linear production according to Eqs. (1)–(3), can be expressed in terms of the *shadow values* or *dual prices* of the capacity constraints:

Proposition 2. *Let D be a continuous random variable that is finite with probability 1. Then, the expected operating profit function $\Pi(\cdot) = \mathrm{E}\pi(\cdot, D)$ is differentiable and for each capacity choice $K \in \mathbb{R}^m_+$ and realized demand $D \in \mathbb{R}^n_+$, the capacity constraints of the linear program (1)–(3) have optimal shadow values $\lambda(K,D) \in \mathbb{R}^m_+$ and $\nabla\Pi(\cdot) = \mathrm{E}\lambda(\cdot, D)$. The continuation region of the optimal ISD investment policy can be described in terms of these shadow values:*

$$S = \left\{ K \in \mathbb{R}^m_+ :\ (1-\delta^e)r^e \leqslant \mathrm{E}\lambda(K,D) \leqslant c^e - \delta^e r^e \right\}. \tag{6}$$

(The proof is relegated to the appendix.) The optimal investment solution is thus reduced to a fundamental quantity of a linear program, a shadow value vector of the capacity constraints (2), which can be calculated explicitly for sufficiently simple examples (see below). We interpret this result as the multi-dimensional generalization of the solution to the familiar newsvendor problem that considers a newsvendor who must purchase K newspapers at a cost c^e in anticipation of an uncertain demand D. Given that newspapers sell at $p + c^e$ (thus, p is the unit contribution margin) and have no resale value ($r^e = 0$), the newsvendor's problem is to determine how many newspapers should be purchased in advance. This is the single product, single resource, single period case of our linear manufacturing model where $A = 1$. The LP can be solved by inspection: $\lambda(K,D) = 0$ for $D < K$ and $\lambda(K,D) = p$ for $D \geqslant K$, and the optimal purchase quantity K^I is the well-known critical fractile solution:

$$\mathrm{E}\lambda(K^I, D) = pP(D \geqslant K^I) = c^e.$$

The critical fractile is found by balancing the expected "underage cost" (e.g., the weighted probability of the area $\{D \geqslant K^I\}$) with the "overage" or adjustment cost (c^e). Proposition 2 greatly enhances the intuitive content of the model by providing a similar solution technique with a graphical interpretation to the multidimensional case: calculate the shadow values $\lambda(K,D)$ by solving the linear program parametrically in terms of K and D. There are a finite number of different values for the shadow values. Thus, one can draw polygonal convex domains in the demand space in which $\lambda(K,D)$ is constant and the expected shadow values can easily be calculated by summing weighted probabilities of areas. The optimal investment level is found by adjusting the capacity vector K such that the sum of the areas equals the marginal adjustment costs. One can think of this result as saying that it is optimal to invest up to a critical "fractile" of the multidimensional demand distribution.

The following generalizations of the one-dimensional newsvendor problem are obtained. First, one will never invest so as to cover the maximum demand ("never go to the 100th percentile"). Indeed, if all resources are costly to invest in ($c^e > 0$), the expected shadow values $\mathrm{E}\lambda$ must be positive. Because a shadow value is positive only when the corresponding capacity constraint is binding, all constraints must be binding with positive probability at the optimal investment level. In addition, if marginal investment costs c^e decrease, expected dual prices $\mathrm{E}\lambda$ should decrease, lowering the probability that constraints are binding, and one will invest so as to cover more demand ("go to a higher fractile"). Second, the optimal investment level is increasing in the ratio of the scale of unit contribution margins to the scale of marginal adjustment costs. Indeed, the shadow values λ are linear in p and if *all* margins increase proportionally without changing the price-mix, i.e., there is a scalar $\theta > 1$ so that margins p become θp, the shadow values $\mathrm{E}\lambda$ increase likewise to $\theta \mathrm{E}\lambda$. Finally, we have the following economic interpretation of the ISD policy: adjust investment levels only if the expected marginal benefit of doing so, as measured by the shadow values, outweighs the marginal adjustment cost.

4. An example with two products and three resources

To illustrate the multi-dimensional newsvendor solution and to set the stage for the next section, consider a firm that produces two products on two dedicated assembly lines with a joint final test. This situation arises in the computer and disk drive industries and in the back-end of semiconductor manufacturing (packaging and final test). Many original equipment manufacturers in these high technology industries engage mainly in final assembly and test operations, purchasing most subassemblies and components from a network of suppliers. Their final assembly processes often are characterized by relatively inexpensive assembly steps followed by an expensive test step. Moreover, assembly capacity often is specific to a product model whereas test equipment is generic (test resources are typically computers with a changeable test bed) and therefore shared by multiple products.

Denoting the capacity of the two dedicated lines by K_x and K_y and the test capacity by K_z, the firm's linear manufacturing process can be modeled by a technology matrix:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \gamma & 1 \end{pmatrix},$$

where we assume that the ratio of test capacity requirement rates of product 1 to product 2 is $\gamma > 0$. The demand forecast for both products is modeled by a probability measure P on the demand (or sample) space $\mathbb{R}^2_+$, so formally, $P : \mathbb{R}^2_+ \to [0, 1]$. We also assume that the unit contribution margins p are sufficiently high to justify production of both products and that products are labeled such that $p_x \geqslant p_y$. It is clear that one should not have more capacity in either assembly line than total test capacity and that there should not be more test capacity than total assembly capacity. We therefore have that $\gamma K_x \leqslant K_z$, $K_y \leqslant K_z$ and $\gamma K_x + K_y \geqslant K_z$.

We want to determine the optimal investment strategy which follows an ISD policy whose continuation region S is calculated with our multi-dimensional newsvendor solution. The first step is to determine the optimal production decisions x contingent on the demand vector D. To actually write out the firm's optimal strategy in terms of the problem parameters, it is useful to partition the demand space $\mathbb{R}^2_+$ for a given a capacity vector $K \in \mathbb{R}^3_+$ as follows:

$$\Omega_0(K) = \{D \in \mathbb{R}^2_+ \colon \; AD \leqslant K\},$$
$$\Omega_1(K) = \{D \in \mathbb{R}^2_+ \colon \; \gamma D_x \leqslant K_z - K_y \text{ and } D_y > K_y\},$$
$$\Omega_2(K) = \{D \in \mathbb{R}^2_+ \colon \; \gamma^{-1}(K_z - K_y) < D_x \leqslant K_x \text{ and } \gamma D_x + D_y > K_z\},$$

$$\Omega_3(K) = \{D \in \mathbb{R}^2_+ \colon\ D_x > K_x \text{ and } D_y > K_z - \gamma K_x\},$$
$$\Omega_4(K) = \{D \in \mathbb{R}^2_+ \colon\ D_x > K_x \text{ and } D_y \leqslant K_z - \gamma K_x\}.$$

This partitioning is the direct result of the parametric analysis of the product mix problem (1)–(3). Indeed, by applying the Simplex method to that linear program where K and D are parameters, the linear inequalities that define the domains Ω_i are automatically manifested. Within each domain, there is an optimal basic solution with a corresponding optimal dual variable $\lambda(K, D)$. It directly follows that the dual variables $\lambda(K, D)$, representing shadow prices for the three kinds of capacity if the state of the world D obtains, are constant over each of the five domains identified above. Assuming $p_x > p_y$ to insure a unique solution, the optimal values of the primal and dual variables are:

	$x(K,D)$	$\lambda(K,D)$
if $D \in \Omega_0(K)$	D	$(0,0,0)$
if $D \in \Omega_1(K)$	(D_x, K_y)	$(0, p_y, 0)$
if $D \in \Omega_2(K)$	$(D_x, K_z - \gamma D_x)$	$(0, 0, p_y)$
if $D \in \Omega_3(K)$	$(K_x, K_z - \gamma K_x)$	$(p_x - \gamma p_y, 0, p_y)$
if $D \in \Omega_4(K)$	(K_x, D_y)	$(p_x, 0, 0)$.

The optimal production decisions are displayed in demand space in Fig. 2. The second step of the multi-dimensional newsvendor solution calls on the demand distribution to calculate the expected shadow value vector Eλ. Because the shadow value vector is constant in each of the five domains, its expectation is easily calculated and, using Eq. (6), the continuation region S of the optimal investment strategy becomes

$$S = \left\{ K \in \mathbb{R}^3_+ \colon (1-\delta^e) r^e \leqslant \begin{pmatrix} 0 \\ p_y \\ 0 \end{pmatrix} P(\Omega_1(K)) + \begin{pmatrix} 0 \\ 0 \\ p_y \end{pmatrix} P(\Omega_2(K)) + \begin{pmatrix} p_x - \gamma p_y \\ 0 \\ p_y \end{pmatrix} P(\Omega_3(K)) + \begin{pmatrix} p_x \\ 0 \\ 0 \end{pmatrix} P(\Omega_4(K)) \leqslant c^e - \delta^e r^e \right\}.$$

The continuation region S, a three-dimensional volume whose exact shape depends on the demand probability measure P, partitions the state space

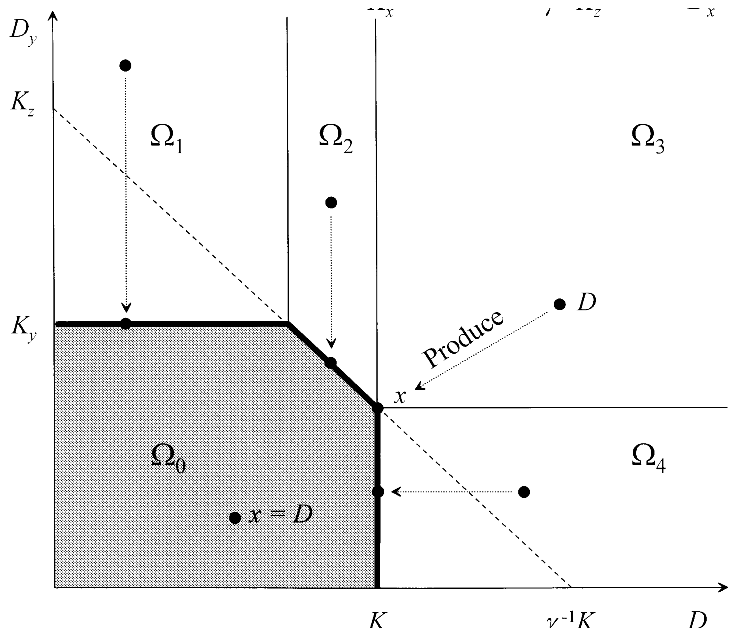

Fig. 2. Production decisions x and shadow values λ depend on the capacity K and demand D.

into nine domains. The optimal investment strategy makes no adjustment to an initial capacity vector that is contained in S. If the initial capacity vector is in any of the domains outside S, it is optimal to adjust capacity to a point on the boundary of S with a domain-specific action, similarly to the actions shown in the two-dimensional example in Fig. 1. If we focus on the investment decision starting with zero initial capacity the optimal investment level K^{I} is the lower left corner of the continuation region:

$$\begin{pmatrix} 0 \\ p_y \\ 0 \end{pmatrix} P(\Omega_1(K^{\mathrm{I}})) + \begin{pmatrix} 0 \\ 0 \\ p_y \end{pmatrix} P(\Omega_2(K^{\mathrm{I}})) + \begin{pmatrix} p_x - \gamma p_y \\ 0 \\ p_y \end{pmatrix} P(\Omega_3(K^{\mathrm{I}})) + \begin{pmatrix} p_x \\ 0 \\ 0 \end{pmatrix} P(\Omega_4(K^{\mathrm{I}})) = c^{\mathrm{e}} - \delta^{\mathrm{e}} r^{\mathrm{e}}. \tag{7}$$

Thus, the optimal investment level K^{I} can be interpreted graphically as in Fig. 3 where we have superimposed the joint correlated demand distribution (represented by the elliptical area that obtains as a level curve of the multivariate normal distribution) onto the state space partitioning of Fig. 2: one has to adjust the thick lines of the feasible region Ω_0 (by adjusting K) such that the probabilities of domains $\Omega_1, \ldots, \Omega_4$ offset the total marginal investment cost $c^{\mathrm{e}} - \delta^{\mathrm{e}} r^{\mathrm{e}}$ as in the optimality equation (7).

In general, the optimal solution requires solving the characteristic equations simultaneously using the multivariate demand distribution. However, if there exists a critical resource like the expensive test resource, a near optimal solution can be found with a one-dimensional newsvendor model as follows. If test capacity is much more expensive than assembly capacity ($c_z^{\mathrm{o}} \gg c_x^{\mathrm{o}}, c_y^{\mathrm{o}}$), optimality equation (7) yields that $P(\Omega_2) + P(\Omega_3) \gg P(\Omega_1), P(\Omega_4)$. Thus $P(\Omega_1)$ and $P(\Omega_4)$ should be very small in which case the graphical interpretation shown in Fig. 3 yields $P(\Omega_2) + P(\Omega_3) \simeq P(\gamma D_x + D_y \geqslant K_z)$. Then the third optimality equation in (7) becomes a traditional one-dimensional newsvendor equation

$$p_y P(\gamma D_x + D_y \geqslant K_z) \simeq c_z^{\mathrm{o}}. \tag{8}$$

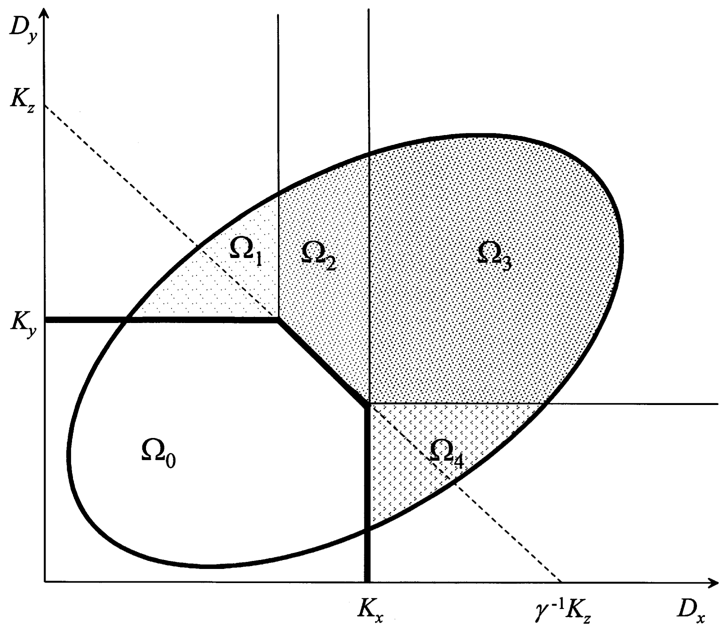

Fig. 3. Optimal investment K^{I} is found by balancing "underage costs" with investment costs.

Thus the optimal operational hedging strategy can be decomposed: the capacity K_z of the expensive resource (testers) is determined with a one-dimensional newsvendor model (8) using the (univariate) *total demand for test capacity* $\gamma D_x + D_y$ (effectively assuming infinite capacity at the upstream assemblies) and the *smaller* contribution margin p_y. Second, the optimal capacity of the inexpensive dedicated resources (assemblies) is the minimal capacity that does not decrease the expensive test capacity (i.e., make $P(\Omega_1)$ and $P(\Omega_4)$ positive, but small, solving Eq. (7)).

5. Operational hedging versus coordinated planning

In many firms, incremental capacity investments are decided upon through a two-stage decision process that can be roughly described as follows using the language and notation introduced in Sections 2–4. First, lower-level production planners propose least-cost required capacity adjustments based on a deterministic master production schedule. To arrive at these requirements, the firm's operations division calculates the technology matrix A and current capacity vector K using product routing, resource consumption and resource availability information, all of which (in the best case) reside in various engineering databases. Marketing and sales representatives develop forecasts of the demand vector D and contribution margins p for the coming period. Then, as discussed in [13,14], sales managers and production managers collaborate in the choice of a single vector x of production quantities for the coming period, a choice which they deem to be the best response to estimated market demand. The vector x of production quantities is usually known as *the production plan* or *Master Production Schedule* (MPS), and production planners then choose a least-cost capacity adjustment to enable execution of that plan. That is, using financial information captured in the capacity adjustment cost function C, they choose a new capacity vector K^c which minimizes adjustment costs subject to the requirement that the production plan x be feasible under K^c ($Ax \leqslant K^c$). We call this least-cost program of required capacity adjustments a *coordinated plan* because it typically implies full utilization of resources under the production plan x. (In the example of the previous section, a coordinated plan corresponds to a *rectangular* feasible region $\Omega_0(K^c)$.) By basing incremental capacity investments on a specific production plan x, coordinated planning effectively makes a bet on a particular future demand. Ultimately, however, senior-level management finalizes a capital budget and approves an investment plan which may be quite different from the coordinated plan. Acknowledging intrinsic demand uncertainty (or, similarly, price uncertainty), senior executives "perturb" the capacity requirements given by production planning in order to minimize risk exposure (thus improving expected profit). The perturbation is based on experience and individual appraisal of future uncertainties and risks and yields a *hedging* [2] position by a counterbalancing investment in various production factors. Following Huchzermeier and Cohen [16], we call this *operational hedging* to distinguish from *financial* hedging, which uses financial instruments and capital markets to minimize a firm's risk exposure. A major US disk drive manufacturer, facing an investment problem similar to the one described in the previous section almost every year, reduces the investment adjustment for expensive test capacity compared to the level specified by the coordinated plan while increasing less expensive final assembly capacities. In our stylized example, this "cuts off the corner" of the rectangular feasible region that results from coordinated planning and yields a feasible region as shown in Fig. 3. Similar practice is found in the automobile industry where car manufacturers often set the capacity of a body assembly line to less than the sum of the assembly line capacities of the V6 and V8 engines that will power the body.

Our model confirms the soundness of operational hedging: the optimal strategy minimizes the risk exposure of the investment by not committing to a single production scenario/plan. The model

[2] The Oxford Dictionary [15] describes hedging as "protecting oneself against loss or error by not committing oneself to a single course of action."

also quantifies the optimal hedge as the one that obtains from setting investment levels so as to balance the dual prices or "underage costs" of resources against the investment or "overage costs" of resource adjustment. (In our example of Section 4, the optimal balance is given by Eq. (7).) Two conclusions [3] follows from this interpretation: First, one should expect the required capacity adjustments as proposed in the coordinated plan to be different from the optimally hedged investments that senior management tries to attain, no matter how sophisticated the firm is in picking the production plan x. Second, it is optimal *not to balance* capacities: the optimal operational hedging position may be unbalanced in the sense that there does not exist a single demand scenario that will keep *all* resources fully [4] utilized. This implies that the performance of capacity planning software that starts with a single deterministic demand scenario may be improved.

As shown in Section 4, near-optimal hedging can be achieved when there exists a critical resource by using a suitably defined one-dimensional newsvendor model to plan the capacity of the critical resource, and then investing in ample capacity for the remaining resources. Then the optimal hedging solution can be applied successfully in a market mechanism where a firm can subcontract or outsource parts of its manufacturing process. Consider for example the high-tech example in the previous section where the firm has the option to outsource the assembly of product 2. The firm has to commit to (and pay for) subcontracting capacity K_y at a market [5] price of c_y^{e} per unit before product demand is realized. This is equivalent to capacity adjustment and this model gives the optimal quantity (or capacity) that the firm should outsource. Although optimally made simultaneously, this decision can be made sequentially after the firm has decided on its crucial (test) capacity according to the above decomposition scheme.

Finally, it may be interesting to estimate by how much senior management's operational hedging improves the net value of the firm as compared to the coordinated plan that derives from simple least-cost adjustment calculations. A general answer obviously depends on all parameters of the problem. To get a feeling for the magnitude of the improvement that may be obtained, we can compare operational hedging to coordinated planning in the simple example of the previous section when starting with no initial capacity. For simplicity, assume that both products are substitutes (e.g., laptop and desktop disk-drives or V6 and V8 engines) and their total demand of volume is reasonably well know while the product mix is uncertain. Thus we can approximate the demand distribution to be uniform on the line $D_x + D_y = 1$, and we will assume that both products require equal amounts of test capacity ($\gamma = 1$). Let $\bar{c}_z^{\mathrm{e}} = p_y(1 - (c_x^{\mathrm{e}}/p_x) - (c_y^{\mathrm{e}}/p_y))$. If $0 \leqslant c_z^{\mathrm{e}} \leqslant \bar{c}_z^{\mathrm{e}}$, the optimal hedging strategy [6] sets $K_x = 1 - (c_x^{\mathrm{e}}/p_x)$, $K_y = 1 - (c_y^{\mathrm{e}}/p_y)$, and $K_z = 1$ (see Fig. 4). The corresponding expected value of the firm under optimal hedging is

$$V = \frac{p_x}{2}\left(1 - \frac{c_x^{\mathrm{e}}}{p_x}\right)^2 + \frac{p_y}{2}\left(1 - \frac{c_y^{\mathrm{e}}}{p_y}\right)^2 - c_z^{\mathrm{e}}.$$

The best coordinated plan K^c has $K_z^c = K_x^c + K_y^c$ with constrained optimality equations $\mathrm{E}\lambda_i(K^c, D) + \mathrm{E}\lambda_z(K^c, D) = c_i^{\mathrm{e}} + c_z^{\mathrm{e}}$ for $i = x, y$ or, because $P(\Omega_2(K^c)) = 0$, $p_i P(D_i \geqslant K_i^c) = c_i^{\mathrm{e}} + c_z^{\mathrm{e}}$ and thus $K_i^c = 1 - (c_i^{\mathrm{e}} + c_z^{\mathrm{e}})/p_i$. The firm's expected value under the best coordinated plan is

$$V^c = \frac{p_x}{2}\left(1 - \frac{c_x^{\mathrm{e}} + c_z^{\mathrm{e}}}{p_x}\right)^2 + \frac{p_y}{2}\left(1 - \frac{c_y^{\mathrm{e}} + c_z^{\mathrm{e}}}{p_y}\right)^2.$$

[3] These conclusions hold almost always (i.e., whenever the solution K to the optimal investment equations yields capacity constraints that do not cross in a single point, that is there exist no x such that $Ax = K$).

[4] The optimal contingent production vector is either inside the feasible region, in which case no resource is fully utilized, or on the boundary of the feasible region, in which case not all resources are fully utilized because the capacity constraints do not intersect.

[5] If no efficient market for capacity exists, the capacity investment decision with subcontracting becomes "relationship specific" and will depend on the capacity decisions of the subcontractor as shown in [11].

[6] Under this assumption on c_z^{e}, the optimality equations (7) do not have a solution if $K_z > 1$ (and thus $P(\Omega_2) = P(\Omega_3) = 0$) or if $K_z < 1$ (and thus $P(\Omega_1) + P(\Omega_2) + P(\Omega_3) + P(\Omega_4) = 1$).

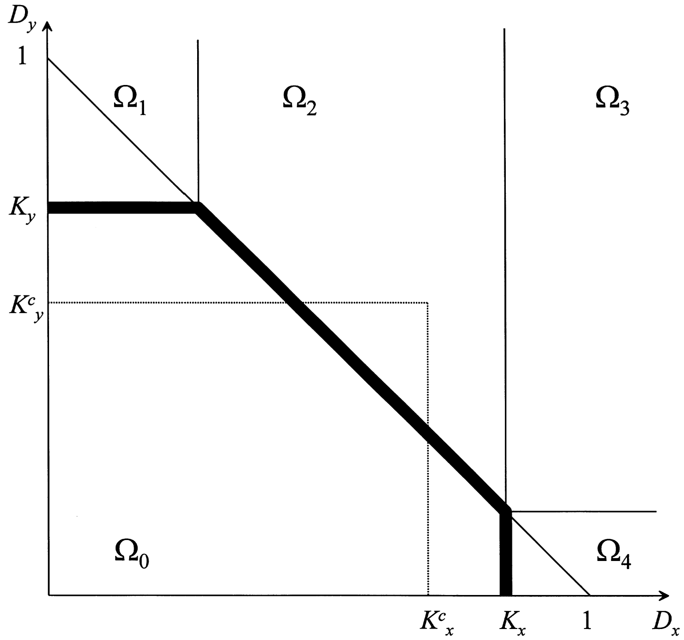

Fig. 4. Comparing the optimal hedge K with the best coordinated plan K^c.

Therefore, the improvement of operational hedging compared to all other coordinated plans is at least

$$V - V^c = c_z^e\left(1 - \frac{c_x^e + 0.5c_z^e}{p_x} - \frac{c_y^e + 0.5c_z^e}{p_y}\right) > 0,$$

and even compared to the best coordinated plan, the improvement may be significant:

$$0 \leqslant \frac{V - V^c}{V^c} \leqslant \frac{p_x p_y (p_x + p_y)\left(1 - \frac{c_x^e}{p_x} - \frac{c_y^e}{p_y}\right)^2}{p_x^3\left(1 - \frac{c_x^e - c_y^e}{p_x}\right)^2 + p_y^3\left(1 - \frac{c_y^e - c_x^e}{p_y}\right)^2}$$
$$\leqslant \frac{1}{2\left(1 + \frac{(p_x - p_y)^2}{p_x p_y}\right)} \leqslant 50\%,$$

where all bounds are tight (lower and upper bounds are attained when $c_z^e = 0$; $c_z^e = \bar{c}_z^e$; $c_x^e = c_y^e = 0, c_z^e = \bar{c}_z^e$; and $c_x^e = c_y^e = 0, c_z^e = \bar{c}_z^e$, $p_x = p_y$, respectively).

6. Conclusion

This paper has presented a model to determine the optimal investment strategies for a manufacturing firm that employs multiple resources to market several products to an uncertain demand. The model incorporates explicitly the production process and the production decisions and is able to consider reversible and irreversible investment. The optimal investment is determined with our multi-dimensional newsvendor solution using a demand distribution, a technology matrix, prices (product contribution margins), and marginal investment costs, and can be represented graphically.

The optimal investment position can be interpreted as a *hedge* against demand uncertainty. This has led us to four managerially relevant conclusions: First, the model explains current practice and quantifies the optimal operational hedge. Second, one should expect the required capacity adjustments as proposed in the coordinated plan by lower-level capacity planners to be different from the optimal hedged investments that senior management tries to attain. Third, it is optimal *not to balance* capacities. Finally, near-optimal hedging can be obtained by planning capacity of the critical resource with a suitably defined one-dimensional newsvendor model and by investing in ample capacity for the remaining resources.

Appendix A. Proof of Proposition 2

Consider the (single period) dual of the linear program (1)–(3):

$$\min_{\lambda \in \mathbb{R}^m_+, \mu \in \mathbb{R}^n_+} \quad K'\lambda + D'\mu,$$
$$\text{s.t.} \quad A'\lambda + \mu \geqslant p.$$

If D is finite, the primal program has a finite optimal solution and its corresponding objective function is equal to the optimal objective function of the dual according to the Duality Theorem of linear programming. Let $\lambda(K,D)$ and $\mu(K,D)$ denote an optimal solution of the dual. Fix a $K^o \in \mathbb{R}^m_+$. Then, for any $K \in \mathbb{R}^m_+$, it follows directly that $\pi(K,D) \leqslant K'\lambda(K^o,D) + D'\mu(K^o,D)$. Combining this with $\pi(K^o,D) = K^{o'}\lambda(K^o,D) + D'\mu(K^o,D)$, directly yields the familiar result that $\lambda(K^o,D)$ is a subgradient of $\pi(\cdot,D)$ at $K = K^o$:

$$\pi(K,D) \leqslant \pi(K^o,D) + \lambda(K^o,D)'(K - K^o), \tag{9}$$

or, interpreting ∇ as a subgradient, $\nabla\pi(\cdot,D) = \lambda(\cdot,D)$. In essence, we now must show that we can interchange differentiation and integration (a.k.a. Leibniz's rule) to yield $\nabla \mathrm{E}\pi(\cdot,D) = \mathrm{E}\lambda(\cdot,D)$. Use the arguments in [17], pp. 97–98 as follows: Because D is finite w.p. 1, taking expectations in Eq. (9) yields that $\mathrm{E}\lambda(K^o,D)$ is a subgradient of $\Pi(\cdot) = \mathrm{E}\pi(\cdot,D)$ at $K = K^o$. Also, because $\pi(\cdot,D)$ is concave, it is differentiable everywhere except on a possible set L of Lebesgue measure zero. Thus, λ is single valued except on L. If D is a continuous random variable, L has P-measure zero. Thus the subgradient $\mathrm{E}\lambda(K,D)$ is unique for all $K \in \mathbb{R}^m_+$ so that $\Pi(\cdot)$ is differentiable everywhere.

References

[1] A.S. Manne, Capacity expansion and probabilistic growth, Econometrica 29 (4) (1961) 632–649.

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

[3] M.H.A. Davis, M.A.H. Dempster, S.P. Sethi, D. Vermes, Optimal capacity expansion under uncertainty, Advances in Applied Probability 19 (1987) 156–176.

[4] J.C. Bean, J.L. Higle, R.L. Smith, Capacity expansion under stochastic demands, Operations Research 40 (Suppl. 2) (1992) S210–S216.

[5] G.D. Eppen, R.K. Martin, L. Schrage, A scenario approach to capacity planning, Operations Research 37 (4) (1989) 517–527.

[6] C.H. Fine, R.M. Freund, Optimal investment in product-flexible manufacturing capacity, Management Science 36 (4) (1990) 449–466.

[7] S. Li, D. Tirupati, Dynamic capacity expansion problem with multiple products: Technology selection and timing of capacity additions, Operations Research 42 (5) (1994) 958–976.

[8] A. Dixit, Investment and employment dynamics in the short run and the long run, Oxford Economic Papers 49 (1997) 1–20.

[9] J.C. Eberly, J.A. Van Mieghem, Multi-factor dynamic investment under uncertainty, Journal of Economic Theory 75 (2) (1997) 345–387.

[10] J.A. Van Mieghem, Investment strategies for flexible resources, Management Science, forthcoming.

[11] J.A. Van Mieghem, Capacity investment under demand uncertainty: The option value of subcontracting, Technical report 1208, Center for Mathematical Studies in Economics and Management Sciences, Northwestern University, Evanston, IL, 1998.

[12] A.K. Dixit, R.S. Pindyck, Investment Under Uncertainty, Princeton University Press, Princeton, NJ, 1994.

[13] N.A. Bakke, R. Hellberg, The challenges of capacity planning, International Journal of Production Economics 30 (31) (1993) 243–264.

[14] E.A. Veral, R.L. LaForge, The integration of cost and capacity considerations in material requirements planning systems, Decision Sciences 21 (3) (1990) 507–520.

[15] A.P. Cowie, Oxford Advanced Learner's Dictionary of Current English, 4th ed., Oxford University Press, Oxford, 1989.

[16] A. Huchzermeier, M.A. Cohen, Valuing operational flexibility under exchange rate risk, Operations Research 44 (1) (1996) 100–113.

[17] R.J.-B. Wets, Programming under uncertainty: The equivalent convex program, J. SIAM Appl. Math. 14 (1) (1966) 89–105.