

Resource and Revenue Management in Nonprofit Operations

Francis de Véricourt

European School of Management and Technology, Berlin, Germany, devericourt@esmt.org

Miguel Sousa Lobo

INSEAD, Abu Dhabi, miguel.lobo@insead.edu

Nonprofit firms sometimes engage in for-profit activities for the purpose of generating revenue to subsidize their mission activities. The organization is then confronted with a consumption versus investment trade-off, where investment corresponds to providing capacity for revenue customers, and consumption corresponds to serving mission customers. Exemplary of this approach are the Aravind Eye Hospitals in India, where profitable paying hospitals are used to subsidize care at free hospitals. We model this problem as a multiperiod stochastic dynamic program. In each period, the organization must decide how much of the current assets should be invested in revenue-customer service capacity, and at what price the service should be sold. We provide sufficient conditions under which the optimal capacity and pricing decisions are of threshold type. Similar results are derived when the selling price is fixed, but the banking of assets from one period to the next is allowed. We compare the performance of the optimal threshold policy with heuristics that may be more appealing to managers of nonprofit organizations, and we assess the value of banking and of dynamic pricing through numerical experiments.

Subject classifications: capacity allocation; revenue management; dynamic pricing; nonprofit.

Area of review: Manufacturing, Service, and Supply Chain Operations.

History: Received May 2005; revisions received May 2006, May 2007, January 2008; accepted June 2008. Published online in *Articles in Advance* July 29, 2009.

1. Introduction

There are, today, 20 million blind eyes in India, and as lifespans increase and more people reach the age for high incidence of cataracts, 2 million new blind eyes are added each year. Currently, only 1 million eyes are being operated on in India each year, resulting in much unmet demand. To address an even more dire situation in 1976, Dr. G. Venkataswamy built a small paying hospital, mostly with the small initial capital of his family.¹ The paying hospital was expanded through 1977, and a first free hospital was built in 1978. In Dr. Venkataswamy's words, "from the revenue generated from operations [at the ground floor of the first paying hospital] we built the next floor, and so on until we had a nice five-story facility. And then with the money generated there, we built the Free Hospital" (Rangan 1994, p. 7). After much growth over 30 years, the Aravind hospitals throughout Tamil-Nadu now perform well over 100,000 operations each year (Shah and Murty 2004).

Although the Aravind story has become noted as an especially successful case, this pattern is not unique and is becoming more common. Nonprofit organizations often engage in a mix of activities, some of which are designed primarily to generate excess revenue to subsidize other activities that more directly serve the organization's charitable mission. This has led nonprofit researchers to address the strategic question of how to construct an optimal portfolio of activities (Gruber and Mohr 1982, Oster 1995).

However, few have paid attention to the practical decisions about resource and capacity allocation that must be made once a given portfolio of activities has been selected.² In this paper, we use a modeling approach to develop insights regarding the optimal allocation of resources between revenue-generating and mission-serving activities over time in an organization that seeks to maximize its mission impact.

To isolate the trade-offs involved in balancing revenue-generating and mission-serving activities across many time periods, we have constructed a simplified model of an organization with just two activities and two corresponding customer groups. Adapting a convention introduced by Weisbrod (1998), we refer to *R-activities* as those that generate more revenue than costs. They serve *R-customers*, which at Aravind are the patients at the paying hospitals. *R-activities* also may, but need not, generate mission impact. Although the mission impact of operating on a paying customer at Aravind is not zero, it is significantly smaller than for free customers given that alternative for-profit hospitals are available to paying customers. We refer to *M-activities* as those that generate positive mission impact, but require a financial subsidy.³ They serve *M-customers*, which at Aravind are the patients at the free hospitals, who come from India's poorest economic classes. The organization begins with an endowment of resources and is attempting to be self-sufficient, with its *R-activities* funding its

M-activities (we do, however, also consider the case where the organization may receive grants over time).

With increased competition for philanthropic and government funding over the past two decades, nonprofits are under increased pressure to generate more revenue from their own R-activities, even if they are relatively low on mission contribution, to allow for cross subsidy within the nonprofit. As a result of this pressure, the creation of revenue-generating ventures has become a major topic of discussion in the nonprofit field over the past two decades. See, for example, Skloot (1983, 1987), Dees (1998), Young (2002), Oster et al. (1995), Foster and Bradach (2005), and Weisbrod (2005). Numerous “how to” guides have been produced to help nonprofits develop profit-making ventures. See, for example, Steckel et al. (1989), Alter (2000), Boschee (2001), Anderson et al. (2002), Larson and Forth (2002), Robinson (2002), and Dees (2004).

This kind of nonprofit cross-subsidy can take different forms, and other examples can be cited in addition to Aravind. The Boston Symphony Orchestra runs its “Boston Pops” series using some of the same musicians and facilities as the regular orchestra, and relies on the revenues to subsidize other orchestra activities that are artistically important, but less lucrative (Oster 1995). This pattern is common among performing arts organizations that schedule more popular performances to subsidize more artistically important, but less popular ones. Consider the reliance of ballet companies on seasonal performances of “The Nutcracker” or theatre companies that rely on popular musicals or plays. Sometimes social service agencies will use key resources, such as their staff, facilities, or equipment, to deliver marketable products that subsidize more charitable operations. For example, CIPO Productions, a Brazilian organization, provides impoverished young people with training in photography, video production, and Web design, and helps subsidize these activities with revenue from selling the use of its production and computer equipment to customers who can pay (Elstrodt et al. 2004). Brinckerhoff (2000, p. 16) describes an Illinois school for behaviorally challenged adolescents that has created a business using its staff expertise to deliver workshops on “Dealing with Difficult Teenagers,” for which it charges admission. The net proceeds help to fund the school’s core mission activities. Any social service provider that includes a practice focused on paying clients to subsidize its mission-related services faces the kind of resource-allocation decision we are discussing.

Once the strategic decision is made to engage in activities that produce a net financial contribution to subsidize other activities that contribute to the mission, nonprofit managers are faced with budgeting and resource-allocation decisions. In a given period, how much of their organization’s resources and capacity do they devote to activities with profit potential, as opposed to activities that serve the mission? Our model focuses on this resource-allocation decision. Because social impact is the ultimate objective,

this decision involves a trade-off between achieving mission impact now versus generating additional financial resources to achieve mission impact in the future. For increased robustness of our results, we include in the model the option of banking funds to earn interest. Managers then have three options in the allocation of funds: to revenue service capacity, to mission service capacity, or banking them.

Under a set of plausible assumptions, we solve for the optimal policy for the allocation of resources. We find that when resources are scarce, all resources should either be allocated to revenue activities or banked. Only when resources are above a threshold should the organization engage in mission activities, and all resources above the threshold should be allocated to mission activities. The threshold increases when we allow R-activities to generate mission impact. We also find that banking is never optimal toward the end of the time horizon, and our numerical results suggest that relatively small expected gains are obtained from including the option of banking.

Because nonprofits may be able to set their prices strategically, we also analyze a version of the model that allows for dynamic pricing decisions. The threshold effect still holds. Under some assumptions for the probability distribution of R-customer demand (but not for all distributions), the optimal price is decreasing in the available resources. Finally, when the price is fixed and banking is not allowed we find, perhaps surprisingly, that the optimal service capacity decision is myopic. The optimal threshold above which resources should be allocated to mission activities is not time dependent and can be analytically derived.

No doubt many nonprofit leaders will want to allocate some resources to M-activities in all periods simply because that is the organization’s mission. Their instincts may be based on factors that fall outside our model. To drop all M-activities could affect the culture of the organization, the motivation of staff, board members, volunteers, and donors, and it could even jeopardize the nonprofit legal status of the organization, with the consequent tax implications. Also, for some organizations, especially for some types of medical services, if there are no alternative organizations that can provide a replacement for critical services, the impact of M-activities can be significantly nonlinear. The marginal mission value of the first few invested resources will be very high. These may be sufficient reasons not to allocate all resources solely to R-activities, even when resources are below the threshold. However, our numerical studies under a range of parameters in our model show that a proportional allocation heuristic performs poorly compared to the optimal threshold policies. According to our model, this prioritization of resources toward M-activities will result in reduced total mission impact when future periods are considered, even given that future mission impacts are discounted. Any allocation of resources to M-activities before the threshold is met must be justified for reasons that are external to our model. The

important point is for nonprofit leaders to understand that the gains from their decision to support M-activities must offset the potential loss of future social impact. Our analysis should help nonprofit managers understand the cost (in terms of future mission impact) of diverting resources away from revenue-generating activities when the resource level is below the optimal threshold.

Although this work has been motivated by an interest in the management of nonprofit organizations, a related investment versus consumption trade-off arises in the dividend-payment problem, where a (for-profit) firm must decide how much of the available cash to pay out as dividends, and how much to reinvest into the operation. As a result, our work can be related to a large literature in the for-profit sector. When firms operate in a complete financial market, Modigliani and Miller (1958) show that these decisions can be made independently. As has often been noted, many, if not most, firms do not operate in such conditions. This is especially true for small firms that do not have access to reasonably priced capital, such as in the equity or bond markets, either because of their small size or because they operate in an economy without a well-developed financial system. The literatures on cash-flow management and corporate finance have paid scant attention to the interaction between financial and operational decisions that consequently arise. These interactions have been explored in recent work that is more closely related to our model. Buzacott and Zhang (2001) consider a manufacturer with limited funds who needs an asset-based loan from a bank to finance future growth. The objective is to maximize the retained earnings at the end of the time horizon. Babich and Sobel (2004) study production, sales, and loan-size decisions to optimize the expected discounted proceeds from an initial public stock offer. The control of dividends problem studied by Li et al. (2003), which includes inventory-replenishment decisions, seems to be the most closely related to our model. In particular, they establish that the optimal policy is myopic for linear holding and backorder costs. Their model is similar to our problem with banking, but with the organization having unlimited access to borrowing. Demand is backordered in their model, whereas we assume lost sales (lost sales is better suited to the problem faced by many firms, especially in service industries). Also, we consider pricing decisions, which generally have not been addressed in the literature on joint operational and financial decisions. The prices set by a firm affect its revenue, which in turn affects the cash constraints and future revenue and growth. Our model is most closely linked to the literature on dynamic inventory models with pricing and stochastic demand. Elmaghraby and Keskinocak (2003) and Bitran and Caldentey (2003) present comprehensive reviews of this literature. For recent results, also see Chen et al. (2003), who consider the lost-sales case, and Chen and Simchi-Levi (2004) for the backorder case. Chen et al. (2004) model the system with a Brownian motion and provide a worst-case bound for the value of

dynamic pricing. In this stream of research, the pricing strategy and the inventory stock replenishment decisions determine the inventory levels available for future periods. This differs fundamentally from the cash constraints that follow from the capacity decisions in our model. For example, Huh and Janakiraman (2008) propose an elegant framework based on sample path analysis to retrieve and extend most of the existing results for multiperiod dynamic pricing and inventory problems (their approach cannot be applied to our model because in our setting the pricing decision directly affects the budget constraint of the following period).

Before formally introducing the model, we outline its key assumptions and discuss their practical justification.

- *Periodic Capacity Shifting and Resource Allocation.* Our model assumes that capacity can be reallocated only periodically at decision times and not between these periods. This fits with nonprofits that (1) are using common core assets in both R-activities and M-activities, and (2) have to schedule the use of these assets in advance. The first condition fits with the recommendations of strategists in this area. According to Oster (1995, p. 91), the most promising new activity extensions, including those created to generate revenue, are ones that can “best share in the core assets of the organization.” The second condition states that these core assets cannot be shifted between activities during the period.

Aravind Eye Hospital may be able to shift resources (doctors and nurses) frequently, but is unlikely to do so mid-shift. The paying and free hospitals are kept separate due to different levels of service (even staff uniforms are different), and planning will require a commitment to staff assignments. A requirement to commit to capacity decisions will also arise from the need to organize the camps that are run in poor villages to recruit and screen candidates for surgery.⁴ Consider also performing arts groups that schedule their seasons in advance. Shifting capacity midseason is extremely rare. They make their bets and live with them. Even social service agencies that attempt to balance paying clients (R-customers) with charitable clients (M-customers) may need to decide in advance how much staff time to make available for paid and charitable work. It is important to keep in mind that the decision period can be short.

Of course, some organizations are not constrained in this way because resources can sometimes be shared by R- and M-activities during the production period. We briefly explore this case, and find that the threshold effect still holds. However, further work is needed toward a complete study of this variant of the problem.

- *Random Demand Model.* Although the newsvendor model provides a reasonable approximation for many situations faced by nonprofits and is widely used in the field of operations, we show that the threshold structure is maintained for a very general relationship between capacity decisions and demand from revenue customers. We use

the Boston Symphony as a motivating example, where the newsvendor model assumption that the total demand in each period is independent of the capacity decision (the number of performances scheduled) is not appropriate.

- *Linearity of Social Return.* We assume that nonprofit organizations can expand their M-activities without experiencing diminishing social returns. This is a reasonable assumption for nonprofit organizations that serve only a small share of the need for their mission activities. Aravind can only address a small portion of the need for cataract surgery among the poor in India. It is often the case that a social service agency could serve many more mission clients, if it simply had the capacity, without any dilution of impact. The case is less clear with performing-arts organizations, if we think of customer demand for the artistic productions. However, thinking in terms of valuable artistic pieces (ballets, plays, symphonies, etc.) that could be performed, the number of productions with artistic merit that the company might like to perform far exceeds its capacity. Additional capacity for M-performances may not lead to diminishing returns on this artistic level.⁵

- *Discounting Future Social Impacts.* We assume that projected future social outcomes should be discounted. There are two rationales for this. The first is mathematical. If we fail to use a mission-impact discount rate and also allow for the possibility of earning real interest (above the inflation rate of operating costs) on banked funds, the mathematical model will result in no investment in mission service capacity in the current period, ever (Keeler and Cretin 1983). It will always be superior to bank the funds in order to purchase more mission impact in later periods. The second reason is that, in most cases, addressing social problems (global warming, poverty, spread of disease, etc.) or delivering social goods (health care, education, the arts) is better done sooner than later. A discount factor captures this degree of urgency, which can vary widely. It is also often used in practice to reflect the uncertainty surrounding the delivery and value of projected future benefits. Using a discount rate implies that a more certain immediate impact should be valued more highly than a less certain future impact of the same magnitude. We recognize that the use of discount rates with regard to social impact is controversial (Klausner 2003), and it certainly oversimplifies the complex considerations that should inform time preferences in the nonprofit sector, such as considerations of intergenerational equity or the potential value of intervening at strategic times. However, we believe that the use of discounting, that is, some preference for immediate impacts versus later impacts, is reasonable for our purposes.

- *Grants and Fundraising.* We model grants as randomly distributed inflows in each period. Although Aravind chooses not to rely on grants, many nonprofit organizations do. The Boston Symphony receives grants from the National Endowment for the Arts and charitable foundations, as well as private and corporate donations. However, only grants that the organization can use to run either side

of its operations (revenue or mission) are relevant to the model. Grants that impose the restriction that they may only be used for the mission side of operations can be treated as entirely separate from the revenue-mission trade-off model (they do, however, have an impact in that by increasing the funding for mission activities, they reduce other concerns such as retention and motivation of volunteer staff). We explicitly leave out of our resource-allocation model the diversion of resources to cover fundraising costs, and reserve it as a topic for future extensions. Note, however, that one option for including fundraising costs would be to treat it, as Weisbrod (1998) suggests, simply as an R-activity. Fundraising costs could be deducted from donations to indicate a net financial yield, which presumably would be positive, but with low or zero direct mission impact. This approach is acceptable provided that operating capacity and resources can be allocated to fundraising, just as well as they can be allocated to other R-activities that more resemble operational programs.

- *Sorting Customers.* Our model distinguishes R-customers from M-customers. In some cases, it is the nature of the customers (e.g., income level) that determines the degree of mission impact of an activity. We assume that when it matters (with regard to mission impact), these customers will sort themselves to the appropriate activities. In particular, R-customers will not simply switch to being M-customers if a comparable M-activity is available as a substitute. In most cases, this is not a problem because the M-activities are often not good substitutes for the R-activities. Even when the activities are similar, various barriers are likely to lead to appropriate sorting. Few of Aravind's wealthier patients would pretend to be poor to get the free version of the surgery. In some cases, R-customers can also be M-customers without disrupting the model because the mission impact is not tied to the nature of the customers served. This is the case with the performing-arts organizations. The same person can attend a Pops concert and a regular symphony without affecting the model results.

To make the problem tractable, we have created a simple model of a very complex world. Nonprofits are complex and diverse, with widely varying economic structures. They include, for example, homeless shelters, ballet companies, hospitals, environmental advocacy groups, social service agencies, museums, sports and athletic associations, schools and colleges, and grassroots political groups. Most of them are far more complex than our simple two-activity model, and some nonprofits are not aiming to be economically self-sufficient. Nonetheless, we believe our model, by isolating this trade-off, highlights a dynamic that is relevant to many nonprofit organizations as they search for the kind of balanced portfolio recommended by Oster (1995). The underlying dynamic in our model captures an increasingly common tension that nonprofit managers face. We offer it as, we hope, a meaningful first step in what we anticipate will be a larger, longer-term project of applying modeling

techniques to decision making in nonprofit settings. We hope that it stimulates more interest by the operations community in a large section of the economy (7% to 10% of the U.S. economy, depending on the metric) to which it has dedicated little attention.⁶ It should also inform the thinking of nonprofit managers regarding the potential implications of their resource-allocation decisions as they pursue a balance between financial and mission activities.

The rest of this paper is organized as follows. Section 2 addresses the problem of optimal capacity and banking decisions, with a fixed selling price to R-customers. We do this for a broad class of demand models, which includes newsvendor-type demand as a special case. (In §2 of the electronic companion, we briefly discuss conditions under which it is optimal not to bank.) Section 3 considers the problem without banking and with a newsvendor-type demand model, where the organization makes a pricing decision in each time period. Section 4 presents numerical studies. We first investigate the relative merit of the optimal capacity policy relative to a proportional-allocation policy. We then explore the expected value of including banking and pricing decisions in the problem specification. Section 5 discusses extensions, including the flexible-capacity case and directions for future research. We conclude in §6. An electronic companion to this paper is available as part of the online version that can be found at <http://or.journal.informs.org/>.

2. Resource Allocation with Banking

2.1. The Model

We consider discrete-time problems. In each time period t , given a capacity y_t allocated for R-customers, the organization serves $y_t \wedge \Theta_t$ (the minimum of y_t and Θ_t) R-customers. The Θ_t are random variables with finite mean, mutually independent and stationary (most of our results extend, however, to the case of nonstationary independent Θ_t). The newsvendor model, which assumes that the distribution of Θ_t is independent of the capacity decision y_t , is widely used in the literature and is adequate in many instances, including the Aravind example. In this section, we consider a more general model for the random demand, in which its distribution depends on the capacity decision (we will later return to the newsvendor model). As a motivation, consider again the Boston Symphony example. Demand is not independent of capacity because scheduling more performances provides customers with more options. For example, people who are not able to attend a performance on a given day due to other commitments may be able to attend if another performance date is scheduled. Increasing capacity by scheduling more performances is therefore likely to shift upwards the distribution of demand. We allow for this by defining, for each y , $f_y(\cdot): [\underline{\theta}, \bar{\theta}] \mapsto \mathbf{R}^+$ as the probability density function of Θ_t given a capacity decision $y_t = y$, with $\underline{\theta}$ nonnegative and $\bar{\theta}$

possibly equal to $+\infty$. We also consider the corresponding cumulative distribution $F_y(\cdot)$, the tail distribution $G_y(\cdot) = 1 - F_y(\cdot)$, and the inverse of the cumulative $F_y^{-1}(\cdot)$, all assumed differentiable in y . The mean of Θ_t is a function of y_t , which we assume to be bounded above by a finite value $\bar{\Theta}$. The newsvendor model is the special case where the demand distribution is independent of the capacity decision, that is, $F_y(\cdot) = F(\cdot)$ for all y . A natural assumption in the Boston Symphony example is that increased capacity cannot lead to reduced demand. If a new performance is added, some patrons may switch from another performance that they would have gone to instead. However, patrons will not decide not to see, say, a Schubert performance just because a Shostakovich performance is added to the schedule. This translates to assuming that the family of distributions $f_y(\cdot)$ is ordered in y according to stochastic dominance, which corresponds to $F_y^{-1}(u)$ increasing in y for each $u \in [0, 1]$. The results in this section are derived under a related assumption of progressive market saturation. The marginal gains in demand are assumed to become smaller as capacity increases so that for each $u \in [0, 1]$, $F_y^{-1}(u)$ is concave in y .⁷

Serving one M-customer increases the social impact of the organization by s (in social impact units) but does not generate revenue. When revenue-generating activities are also related to the mission of the organization, we denote by τ the social impact of serving one R-customer. The unit service-capacity costs for M- and R-customers are c_M and c_R . We assume lost sales, so that the costs for R-customers are incurred as a function of the capacity allocated even if that capacity is not fully utilized (the salaries of nurses and doctors assigned to Aravind's paying hospitals have to be paid in full, independently of how many surgeries are performed; the production costs of a music performance are not any lower if the auditorium is half-full). Without loss of generality, we assume that $c_R = 1$. This is simply a change of asset units, and we also assume that the selling price and the cost of serving M-customers have been modified so as to be specified in these units.

We assume ample demand of M-customers, that is, resources allocated to this group will always be consumed without decrease in the marginal social impact. In the Aravind case, this corresponds to noting that the number of poor people in need of eye surgery is far greater than the combined capacity of all charity hospitals.⁸

In this section, we assume the unit selling price p is fixed. This will be the case when the organization is a price-taker, with the price set by the market in which the organization competes for R-customers. When this is not the case, the organization may still decide to offer the service at the same price in all time periods due to marketing or organizational constraints.

Also for this section, we assume that the organization has an alternative financial application for its assets in a risk-free investment with return β . The organization needs to determine what portion of its assets should

be “banked,” what portion should be invested in capacity for R-customers, and what portion should be allocated to capacity for M-customers. We will also briefly discuss the relative importance of including banking in the problem formulation. In particular, in §2 of the electronic companion, we show that if the time horizon is short enough, banking is never optimal.

The system state is the assets a_t held by the organization at the beginning of period $t \in [1, \dots, T]$. At the beginning of each period, the organization decides how much service capacity y_t to provide for R-customers, and sets the amount of current assets that are banked, $z_t \geq 0$. Note that because the unit cost is normalized to one, y_t can interchangeably be used to denote the resources allocated to provide capacity and the capacity itself. The choices of y_t and z_t are limited by the current resources, according to the constraint $y_t + z_t \leq a_t$. The remaining resources, $x_t = a_t - y_t - z_t$, are allocated to serving M-customers. Demand Θ_t is then realized, and the number of R-customers served by the organization is the smallest of demand and capacity, which we write as $y_t \wedge \theta_t$. The service is perishable, and any unused capacity is lost. Finally, the organization may receive donations in each period, with the amount a random variable Δ_t , assumed stationary and independent, and with continuous density function and finite mean. The resources available to the organization at the beginning of the following period are then

$$a_{t+1} = p(y_t \wedge \theta_t) + \beta z_t + \Delta_t.$$

The organization contributes to its mission by serving $(a_t - y_t - z_t)/c_M$ M-customers and $y_t \wedge \theta_t$ R-customers, yielding a social return of $s/c_M(a_t - y_t - z_t) + \tau(y_t \wedge \theta_t)$. Without loss of generality, we set $s/c_M = 1$, which is simply a change in social return units. We assume that the units of τ , the social return from serving an R-customer, are also modified so as to be specified in these units, so that τ represents the social impact of a unit of assets spent on serving R-customers relative to the social impact of a unit spent on M-customers. The units of the value-to-go function are also changed accordingly. We assume that the social return of serving an M-customer is higher than that of serving an R-customer, i.e., $\tau \leq 1$.

We begin by considering problems with a finite horizon of T periods. This may arise, for example, if the initial assets of the organization are from a grant that specifies a duration of time in which to achieve its objectives. We later extend results to the infinite-horizon case. The social discount factor for delaying service to an M-customer to the next period is α , with $0 \leq \alpha < 1$. This discount factor measures the urgency of the social need that the organization addresses. The overall social return from the organization’s activities is the total discounted sum of the social return from the number of customers served in each period. The objective is to determine a banking and capacity-provision policy that maximizes the expected social return

over the time horizon T . We formulate the problem of jointly determining the best banking and capacity decisions as a finite-horizon Markov decision process.

Denote by $v_t(a)$ the maximum social-impact-to-go at period t given current resources $a \geq 0$. In the last period, all assets are allocated to serving M-customers, so that $v_T(a) = a$. For $t < T$, $v_t(a)$ can be shown to satisfy the optimality equations (see, for exam Heyman and Sobel 1984)

$$v_t(a) = \max_{\substack{0 \leq y, 0 \leq z \\ y+z \leq a}} (a - y - z) + \tau \mathbf{E}_y(y \wedge \Theta) + \alpha H^{v_{t+1}}(y, z) \quad (1)$$

$$= a + \max_{\substack{0 \leq y, 0 \leq z \\ y+z \leq a}} J^{v_{t+1}}(y, z), \quad (2)$$

where the operators J^v and H^v are defined for any real-valued function $v(\cdot)$ as

$$J^v(y, z) = -y - z + \tau \mathbf{E}_y(y \wedge \Theta) + \alpha H^v(y, z), \quad (3)$$

$$H^v(y, z) = \mathbf{E}_\Delta \mathbf{E}_y v(p(y \wedge \Theta) + \beta z + \Delta), \quad (4)$$

and \mathbf{E}_Δ and \mathbf{E}_y are the expectation over Δ and over Θ given y . On account of the stationary assumptions, we will generally omit the time index from the random variables.

We denote by $(y_t^*(a), z_t^*(a))$ the optimal resource-allocation and banking decisions at period t given current assets a . The optimal policy $(y_t^*(a), z_t^*(a))$ corresponds to the maximizer of $J^{v_{t+1}}$ subject to the constraints $0 \leq y$, $z \leq 0$, and $y + z \leq a$. (In case of multiple optima, $(y_t^*(a), z_t^*(a))$ designates the minimum optimal decision using, say, the lexicographic order.)

As a side result, note that if $\alpha\beta \geq 1$ the organization always prefers to bank rather than to serve M-customers (except in period T). Likewise, conditions on p and τ can also ensure that serving R-customers is never optimal.

PROPOSITION 1.

- If $\alpha\beta > 1$, it is never optimal to allocate resources to M-customers before the final time period.
- If $\alpha p + \tau < \alpha\beta$ or $\alpha p + \tau < 1$, serving R-customers is never optimal.
- If $\alpha\beta < 1$ and $\alpha p + \tau < 1$, it is optimal to use all assets to serve M-customers in the first period.

(The proof is in §1 of the electronic companion.)

2.2. Optimal Resource Allocation

The interesting case is then $\alpha\beta < 1$ and $\alpha p + \tau > 1$, which we assume for the rest of this section. This also implies $\alpha p + \tau > \alpha\beta$. The following result shows that the concavity of the social-impact-to-go can be iteratively propagated to all time periods, and establishes that the optimal policy for R-customer capacity plus banking is of threshold type. Capacity should only be allocated to serving M-customers when assets are above a threshold, and all assets above that threshold should be allocated for this purpose.

THEOREM 1. For each time period t , $v_t(\cdot)$ is differentiable nondecreasing concave and $J^{v_t}(y, z)$ is differentiable and jointly concave. Further, there exists a threshold a_t^* such that the optimal resource decision $y_t^*(a)$ and the optimal

banking decision $z_t^*(a)$ satisfy $y_t^*(a) + z_t^*(a) = a_t^* \wedge a$ for all a , and $y_t^*(a) = y_t^*(a_t^*)$, $z_t^*(a) = z_t^*(a_t^*)$ for $a > a_t^*$.

(The proof is in §1 of the electronic companion.)

Under the additional assumption that the profit margin is higher than the banking return rate, $p \geq \beta$, which should be a reasonable assumption for most problems, and when banking is not possible, we can further characterize the structure of the optimal policy in the newsvendor case.

COROLLARY 1. *If $F_y(\cdot) = F(\cdot)$ for all y and $p \geq \beta$, the optimal banking decision $z_t^*(a)$ is nondecreasing in a . Further, if $\beta = 0$, the myopic policy is optimal: At time t , the optimal resource allocation is $y_t^*(a) = a^* \wedge a$, where*

$$a^* = F^{-1}\left(1 - \frac{1}{\tau + \alpha p}\right). \tag{5}$$

(The proof is in §1 of the electronic companion.)

When the time horizon is short enough, it can be further deduced from Corollary 1 that the organization should never bank. In §2 of the electronic companion, we show that when $\tau = 0$ the optimal policy is of threshold type, myopic, and never involves banking for $t \geq T - n_0$ with $n_0 = \lfloor -\ln(\alpha\beta)/\ln(\alpha p) \rfloor$.

Finally, we extend the results to the infinite-horizon case.

COROLLARY 2. *The optimal policy for the infinite-horizon case is of threshold type. Further, if $F_y(\cdot) = F(\cdot)$ for all y , when $p \geq \beta$ the optimal banking decision is nonincreasing in the current assets, and when banking is not allowed the optimal policy is myopic.*

(The proof is in §1 of the electronic companion.)

3. Resource Allocation with Pricing

3.1. The Model

We have so far assumed that the organization proposes its service at a fixed price to the revenue customers. A nonprofit organization may, however, have market power in its revenue-generating activities, and the operational flexibility to dynamically change prices. The organization needs to develop a pricing strategy jointly with its capacity-provision and banking strategy. Characterizing the optimal policy for these joint decisions with the general model of §2 appears, however, to be far more challenging. To be able to provide some insight into the pricing problem, we focus on the newsvendor case and make two key simplifications in this section. First, we consider the case where the organization does not rely on grants ($\Delta = 0$), and R-customers have no social impact ($\tau = 0$). The latter is a reasonable assumption if alternative for-profit providers for paying customers are readily available. Second, we assume that the organization never banks. This allows us to explore the structure of the optimal pricing and capacity-provision decisions in more detail, while still maintaining the more interesting characteristics of the problem.

We consider a demand function with multiplicative uncertainty, that is $D_t = \gamma(p_t)\Theta_t$, where $\gamma(\cdot): [0, \bar{p}] \mapsto [0, +\infty[$ is the price response function and Θ_t is a random variable with finite mean. The response function $\gamma(\cdot)$ is differentiable and decreasing, with $\gamma(\bar{p}) = 0$. The price elasticity $e(\cdot)$ is defined as $e(p) = -p\gamma'(p)/\gamma(p)$. We will also refer to the revenue function, defined as $r(p) = p\gamma(p)$ which is assumed to be nonmonotone and strictly concave in p . The maximand $\underline{p} = \arg \max_p r(p)$ is then well defined. As before, $f(\cdot)$ denotes the probability density function of Θ_t , and $F(\cdot)$ and $G(\cdot)$ the corresponding cumulative and tail distributions, which in this section we assume twice differentiable.⁹ We also introduce the generalized failure rate (GFR), defined as $g(\cdot): [\underline{\theta}, \bar{\theta}] \mapsto \mathbf{R}^+$ such that

$$g(\theta) = \theta \frac{f(\theta)}{G(\theta)}. \tag{6}$$

At the beginning of each period, the organization sets the selling price p_t and decides how much service capacity y_t to provide for R-customers. The choice of y_t is limited by the current resources, which corresponds to the constraints $0 \leq y_t \leq a_t$. The remaining resources, $a_t - y_t$, are allocated to providing capacity to serve M-customers. Demand D_t is then realized, and the number of R-customers served by the organization is $y_t \wedge \gamma(p_t)\theta_t$. The resources available to the organization at the beginning of the following period are

$$a_{t+1} = p_t(y_t \wedge \gamma(p_t)\theta_t).$$

The goal is to determine a pricing and capacity-provision policy that maximizes the social return over the time horizon T .

For $t < T$, the social-impact-to-go at period t can be shown to satisfy the optimality equations

$$v_t(a) = \max_{\substack{0 \leq p \leq \bar{p} \\ 0 \leq y \leq a}} a - y + \alpha H^{v_{t+1}}(p, y), \tag{7}$$

where the operator $H^v(p, y)$ is defined for any real-valued function $v(\cdot)$ as

$$H^v(p, y) = \mathbf{E}_\Theta v(py \wedge r(p)\Theta). \tag{8}$$

We denote by $(p_t^*(a), y_t^*(a))$ the optimal pricing and capacity decisions at period t given current assets a . The optimal policy $(p_t^*(a), y_t^*(a))$ corresponds to the maximizer of $\alpha H^{v_{t+1}}(p, y) - y$, subject to the constraints $p \geq 0$ and $0 \leq y \leq a$.

Solving (7), however, can be challenging, even for the single-period case (in the following, we refer to the single-period problem as the case where $T = 2$ because no decision is made in the last period). When $T = 2$, the objective can be written as

$$v(a) = a + \max_{\substack{0 \leq p \leq \bar{p} \\ 0 \leq y \leq a}} \alpha p \mathbf{E}_\Theta (y \wedge \gamma(p)\Theta) - y, \tag{9}$$

and the optimization problem is equivalent to a newsvendor problem with pricing and with capacity constraint. The unconstrained newsvendor problem with pricing has been

the subject of much research (see Petruzzi and Dada 1999 for a survey). More recently, Wang et al. (2004) provide sufficient conditions for the uniqueness of the optimal decisions y^* and p^* . They show that if Θ has increasing generalized failure rate (i.e., $g(\cdot)$ is increasing over $[\underline{\theta}, \bar{\theta}]$), then y^* and p^* exist and are unique. With the change of variable $z = y/\gamma(p)$ (the “stocking factor”) in Equation (9), they show that, for any z , an optimal $p(z)$ exists, and that the function

$$H(z) = \alpha \mathbf{E}_{\Theta} r(p)(z \wedge \Theta) - \gamma(p)z$$

is unimodal in z . However, nothing is said about the concavity of the objective function, even along the optimal price $p^*(z)$. Bernstein and Federgruen (2005) propose a sufficient condition on $g(\cdot)$ that guarantees that $H(\cdot)$ is log-concave, and hence unimodal.

The analysis of multiperiod dynamic problems typically requires stronger properties than unimodality or log-concavity. Most of the previous approaches to address dynamic joint pricing-inventory problems (for general probability distributions) rely on the joint concavity of $\alpha H^v(p, y) - y$ in (p, y) for any concave value function $v(\cdot)$ (see, for example, Federgruen and Heching 1999). However, even for the single-period case, the py term in the definition of H^v in (8) is not jointly concave. Changes of variable such as $z = y/\gamma(p)$, or the conditions on the generalized failure rate $g(\cdot)$ proposed by Wang et al. (2004) or Bernstein and Federgruen (2005) do not circumvent this problem. Sample path approaches as developed by Huh and Janakiraman (2008) cannot be applied to our model.

In the following, we assume that the generalized failure rate at the stocking factor $y/\gamma(p)$ is not smaller than the inverse of the price elasticity, that is, $\forall p \in [\underline{p}, \bar{p}]$ and $\forall y, y/\gamma(p) \in [\underline{\theta}, \bar{\theta}]$,

$$g\left(\frac{y}{\gamma(p)}\right) \geq \frac{1}{e(p)}. \quad (10)$$

Note that, for $r(\cdot)$ concave, $e(p) \geq 1$ for $p \geq \underline{p}$, so that condition (10) holds when the generalized hazard rate is bounded below by one, i.e., $g(\theta) \geq 1$. Condition (10) is similar to the condition for the single-period case proposed by Bernstein and Federgruen (2005), where the generalized hazard rate is bounded below by one half. Condition (10) is also neither more general nor more restrictive than the increasing generalized failure rate condition proposed by Wang et al. (2004) because it allows for nonmonotonic generalized failure rates. If Θ is uniformly distributed over $[\underline{\theta}, \bar{\theta}]$, condition (10) holds for any concave revenue function and $\underline{\theta}/\bar{\theta} \geq 1/2$. Another example satisfying this condition for any concave revenue function is an exponential distribution with parameter λ truncated at $\theta = 1/\lambda$, that is, $f(\theta) = 0$ for $\theta < 1/\lambda$ and $f(\theta) = \lambda e^{-\lambda(\theta-1/\lambda)}$ for $\theta \geq 1/\lambda$. More generally, for any distribution defined over $[0, +\infty[$ with an increasing generalized failure rate, $g(\theta) > 1$ holds

if the distribution is truncated at $\underline{\theta}$ such that $g(\underline{\theta}) > 1$, and scaled accordingly.

Condition (10) can be shown to be equivalent to the bound from below on elasticity of sales introduced by Kocabiyikoğlu and Popescu (2005). They show that if this condition holds, an optimal solution exists for the news vendor problem with pricing and the corresponding optimal price is decreasing with the inventory level. In our context, the concavity and monotonicity properties can be propagated across time periods under this condition. We will show that if condition (10) holds, then $H^{v_t}(p, y)$ is concave along the optimal price, which implies that

$$\tilde{H}^{v_t}(y) = \max_{0 \leq p \leq \bar{p}} H^{v_t}(p, y) \quad (11)$$

is concave in y . For any continuous increasing function $v(\cdot)$ such that $v(0) = 0$, $\tilde{H}^v(\cdot)$ is well defined and $\tilde{p}(y)$, the maximizer of $H^v(\cdot, y)$, is interior ($\tilde{p}(y) \in]0, \bar{p}[$) because $H^v(0, y) = H^v(\bar{p}, y) = 0$ for any $y \in \mathbf{R}^+$. Concavity of $\tilde{H}^{v_{t+1}}(\cdot)$ guarantees in turn that the social-impact-to-go function $v_t(\cdot)$ is increasing and concave and that the optimal policy is of threshold type. Furthermore, because $r(\cdot)$ is nonmonotone strictly concave, $r(\cdot)$ and hence $H^{v_t}(\cdot, y)$ are increasing in p for $p < \underline{p}$ and the optimal price $\tilde{p}_t(y)$ is bounded from below by \underline{p} (the maximizer of $r(\cdot)$).

A threshold policy in our context is characterized by $T - 1$ thresholds (\hat{p}_t, \hat{a}_t) , $t \in [1, \dots, T - 1]$, such that the decisions $(p_t(a), y_t(a))$ given assets a at the beginning of period t are

$$y_t(a) = a \wedge \hat{a}_t, \\ p_t(a) = \hat{p}_t \quad \text{if } a \geq \hat{a}_t,$$

and by a price $p_t(a)$ for $a < \hat{a}_t$. Under a threshold policy, the organization tries to guarantee a particular service capacity for R-customers. Only when this threshold is assured are M-customers served.

Finding the optimal price, however, is not a trivial problem, and solutions can be counterintuitive. One could indeed reasonably expect the optimal price to be decreasing in the capacity allocated to R-customers. However, a simple counterexample will show that this is not always the case (see §3 of the electronic companion). We will see, however, that with condition (10) the optimal price is nonincreasing.

3.2. Optimal Capacity and Pricing Strategy

Our proof of the structure of the optimal capacity-provision and pricing strategy relies on the first- and second-order derivatives of $\tilde{H}^{v_t}(y)$. However, $\tilde{H}^{v_t}(y)$ may not be twice differentiable due to the constraint $y \leq a$ in its definition. We introduce a family of unconstrained dynamic problems parameterized by $\epsilon > 0$. We show that the corresponding operators $\tilde{H}^{v_t^\epsilon}(y)$ are concave, and likewise for the optimal value functions v_t^ϵ . We then show that, for all t , $v_t^\epsilon \rightarrow v_t$

pointwise when $\epsilon \rightarrow 0$, where the $v_t(\cdot)$ satisfy the optimality equations.

Consider $\mathbf{R}_{-\infty} = \mathbf{R} \cup \{-\infty\}$ and the extension of any function $\varphi(\cdot):]\underline{x}, \bar{x}[\rightarrow \mathbf{R}$ such that $\varphi(x) = -\infty$ when $x \notin]\underline{x}, \bar{x}[$. (In the following, we use the same notation for a given function and its extension.) For example, we will consider the extension of the logarithm function such that $\varphi(x) = -\infty$ when $x \leq 0$. The family of unconstrained problems is obtained by omitting the constraints $0 \leq y \leq a$, and introducing logarithmic barrier functions in the objective function.

More precisely, for any $\epsilon > 0$, we consider the dynamic problem,

$$v_t^\epsilon(a) = \max_y \{a - y + \epsilon \log((a - y)y) + \alpha \tilde{H}^{v_{t+1}^\epsilon}(y)\},$$

$$t = 1, \dots, T - 1, \quad (12)$$

$$v_T^\epsilon(a) = a. \quad (13)$$

The optimal capacity decision never equals the bounds ($0 < \tilde{y}^\epsilon(a) < a$). Backward iteration then shows that $v_t^\epsilon(a)$ is nondecreasing and twice differentiable.

Using condition (10) on the generalized failure rate, we show in §4 of the electronic companion that $\tilde{H}^{v_t^\epsilon}(\cdot)$ is concave if $v_t^\epsilon(\cdot)$ is also concave.

LEMMA 1. *Assume that condition (10) holds. For any $\epsilon > 0$, if $v_{t+1}^\epsilon(\cdot)$ is twice differentiable, nondecreasing, and concave, then $\tilde{H}^{v_{t+1}^\epsilon}(\cdot)$ is twice differentiable and concave. Furthermore, the pricing decision $\tilde{p}_t^\epsilon(y)$ maximizing $H^{v_{t+1}^\epsilon}(p, y)$ is nonincreasing in y .*

It follows that Equation (12) preserves concavity, which we establish in the next result; the proof is given in the electronic companion.

LEMMA 2. *Assume that condition (10) holds. For any $\epsilon > 0$, if $v_{t+1}^\epsilon(\cdot)$ is twice differentiable, nondecreasing, and concave, then $v_t^\epsilon(\cdot)$ is also twice differentiable, nondecreasing, and concave. Also, the optimal pricing decision $p_t^{*\epsilon}(a)$ is nonincreasing in the current assets a .*

Starting from $v_T^\epsilon(a) = a$ (which is concave), Lemma 2 ensures that $v_t^\epsilon(\cdot)$ is concave for all t . By letting $\epsilon \rightarrow 0$, we obtain the same result for $v_t(\cdot)$ in the original problem (also proved in §4 of the electronic companion).

LEMMA 3. *If (10) holds, then for $\epsilon \rightarrow 0$, the $v_t^\epsilon(\cdot)$ converge pointwise to the $v_t(\cdot)$ that solve the optimality Equation (7). Further, the $\tilde{H}^{v_t}(\cdot)$ and $v_t(\cdot)$ are concave and the optimal pricing decision $\tilde{p}_t(y)$ is nonincreasing in the capacity decision y .*

The proof of the following theorem, which characterizes the optimal capacity and pricing decisions, is then immediate.

THEOREM 2. *If condition (10) holds, the optimal policies are of threshold type and the optimal pricing decisions $p_t^*(\cdot)$ are nonincreasing in the current assets.*

In §5 of the electronic companion, we provide additional information regarding the shape of the optimal pricing policy by showing that $ap_t^*(a)$ (the total revenue if all assets are allocated to R-customers and demand exceeds capacity) is decreasing in a , and that $\lim_{a \rightarrow 0} ap_t^*(a) = 0$. We conclude this section by noting that our results extend to the infinite-horizon case.

COROLLARY 3. *If condition (10) holds, the optimal policy for the infinite-horizon case is of threshold type. Further, the optimal pricing decision $p^*(\cdot)$ is nonincreasing in the current assets.*

4. Numerical Studies

We consider an example with a concrete choice of parameters. This choice of parameters is suggested from Aravind to provide a sense for their order of magnitude, but are in no way intended to be exact in the actual parameter values. We use a time period of two months (“Our staff are rotated between the paying segment and the free segment every month or two months,” R. D. Thulasiraj, Executive Director, Aravind Eye Hospitals, in an interview to *IIMB Management Review*, September 2004; see Shah and Murty 2004, p. 36) and, somewhat arbitrarily, a time horizon of four years, so that $T = 24$. We assume a social discounting factor of 0.75 on a yearly basis, which leads to $\alpha = 0.953$ (the social discounting factor is a subjective choice that the organization’s leadership will have to settle on for planning purposes; in Aravind’s case, because the time period in the model is relatively short and cataracts progress relatively slowly, the subjective intertemporal social discounting will not be too high). The factor for the banking of assets is 1.10 on a yearly basis, so that $\beta = 1.016$.

The following are approximations consistent with values indicated in the cited literature on Aravind and its reports. For the fixed-price model, the price charged to paying customers for eye surgery is set at 2,000 rupees. The cost of providing surgery is 1,000 rupees in paying hospitals, and 500 rupees in free hospitals. The mean demand from paying customers over a two-month period is 6,000, with a uniform distribution between 4,000 and 8,000 (so that if all assets are dedicated to revenue customers, the mean demand can be fully served when operating assets of 6 million rupees are available for a two-month period).

With the assumption that $\tau = 0$ (which is justified if paying customers have other easily accessible alternatives), we normalize units as follows. We define one unit of customers to be 4,000 patients, and one unit of assets to be 4 million rupees. We then have $c_R = 1,000$ rupees/customer = 1 asset units/customer units. Likewise, after this normalization, $c_M = 0.5$, $p = 2$, and the demand distribution is uniform in $[1, 2]$. We can then arbitrarily choose the units of social impact so that $s = 0.5$ in order to have $s/c_M = 1$. We ran most simulations for initial assets ranging from near zero, $a = 0$, to over six times the mean demand, $a = 10$ (40 million rupees).

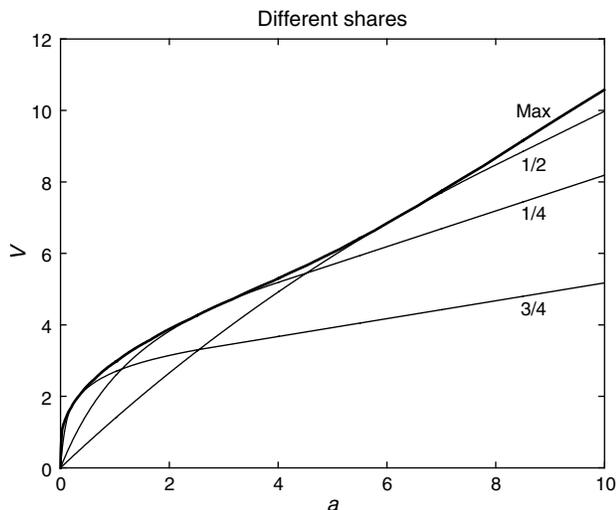
4.1. The Value of Optimal Capacity Decisions

We begin by considering the value of optimal capacity decisions by considering the case where the option of banking is not available and the price is fixed. Note that the distribution satisfies the conditions for Theorem 2. The optimal policy for this problem is of threshold type, as shown in Corollary 1.

A threshold policy may appear somewhat counterintuitive to many organizations. To our knowledge, no empirical studies exist on how resources are actually allocated in practice between mission and revenue customers in nonprofit organizations. However, strategies based on proportional allocations may be more appealing to managers. Such a policy, where the same fixed fraction of assets is allocated to mission customers in every period, is attractive in other respects. Under such policies, a fixed percentage of the current assets is allocated to the M-customers, so that the organization always contributes to its mission. In contrast, under a thresholds policy, no M-customers are served if $a_t \leq a^*$. This may, in the long run, jeopardize the culture of the organization as nonprofit oriented, which in turn can affect its ability to recruit volunteers and to provide high service quality to the M-customers, among other concerns (Dees 1998, Dees and Anderson 2003). However, if the gain is significant, the organization may benefit from obtaining buy-in from its stakeholders on the benefit of a threshold policy, while acting in other ways to maintain a nonprofit-oriented culture.

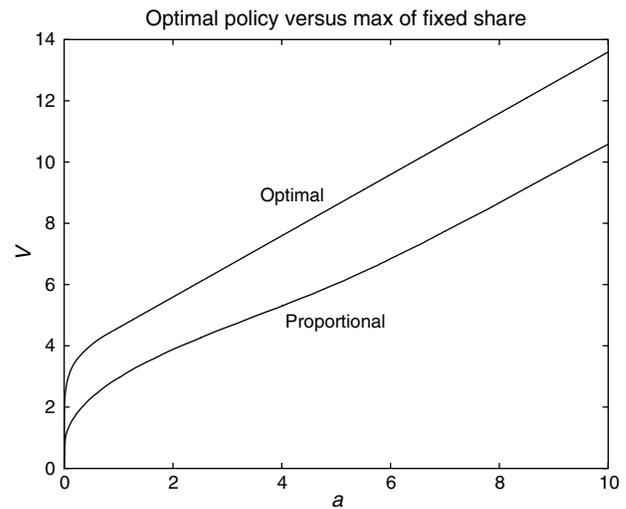
We take a look at the trade-offs involved by comparing the performance of the optimal policy with a proportional-allocation policy. Figure 1 plots the value function for three different y/a ratios. The best choice for y/a , as measured by the expected discounted social impact, depends on the initial endowment. We plot the maximum of the value

Figure 1. Capacity provided for each class of customers.



Note. Value functions for fixed-proportion policy with different ratios, and maximum over all ratios.

Figure 2. Capacity provided for each class of customers.



Note. Maximum of value functions for fixed-proportion policies, and value function for optimal policy.

functions associated with all possible choices for the fixed proportion. This corresponds to expected value when the organization optimally makes the decision at the beginning of the first period on the proportion based on the asset level at that time.

Figure 2 compares the optimal policy with the maximum over all proportional policies. The relative difference between the two curves is substantial, even for a large (at $a = 3$, or 12 million rupees, two times the assets required to provide capacity for twice the mean demand of R-customers, it is on the order of 30%). Near zero it becomes extremely large (higher than 100% for $a < 0.4$, or 1.6 million rupees). When the initial asset level is low, the optimal policy allows the organization to grow more aggressively and more quickly reach the asset level where it can fully exploit the available demand from revenue customers.

4.2. The Value of Banking

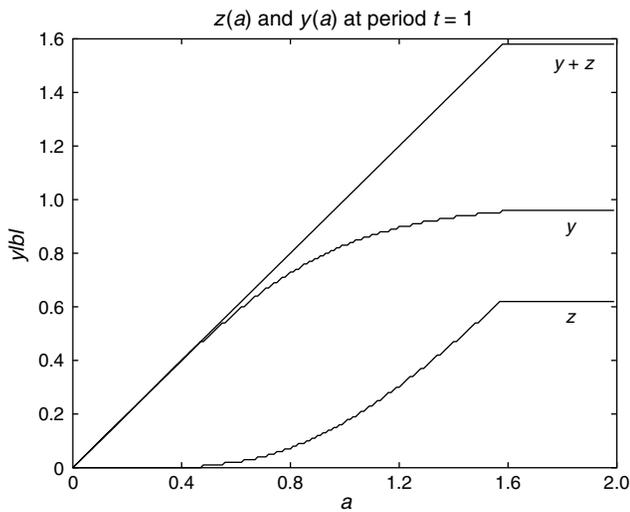
Figure 3 shows the optimal policy with a fixed price and banking decision. In agreement with Theorem 1, the optimal policy is of threshold type for $z + y$, and z is nondecreasing. In our numerical examples we have also always found y to be nondecreasing.

For low asset levels, the risk of unmet demand is low and the organization's focus is on growing the available cash, with all assets allocated to serving revenue customers. Banking becomes attractive when the organization has more assets and the expected marginal value of increasing service capacity for R-customers decreases.

The value of banking depends on the price being charged. To investigate this we compute the value function for different prices, holding constant the return on the banked assets. We also compute, for the same set of prices, the value function without banking. For each price, we

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at http://journals.informs.org/.

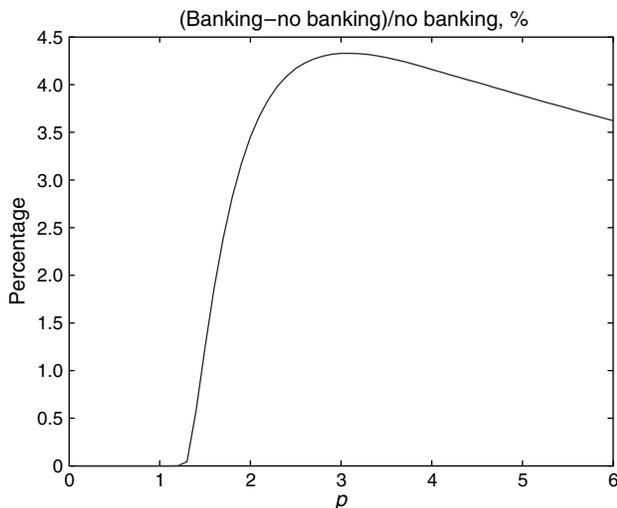
Figure 3. Optimal policy with banking.



compute the ratio of the two value functions, and the maximum of this ratio over all possible initial states. Figure 4 plots this maximum ratio as a function of the price. Banking is of relatively limited value, providing a maximum gain of about 4% or less.

The expected gain from having the banking decision available is not monotonic in the price. For small prices it is optimal to spend all the capital on M-customers in the first period. Therefore, the option of banking is of no value. For higher prices (and with the same demand distribution) the return rate from banking becomes, by comparison, less attractive. The organization has less of an incentive to hedge the outcome by banking because, given the high profit margin, it is easier to recover quickly from a period

Figure 4. Maximum percentage expected gain from banking over all possible initial asset levels.



Note. This maximum ratio of the two value functions is plotted over a range of prices (and same demand distributions).

with unusually low demand. For very large prices, the percentage gain was seen to converge to a nonzero value.

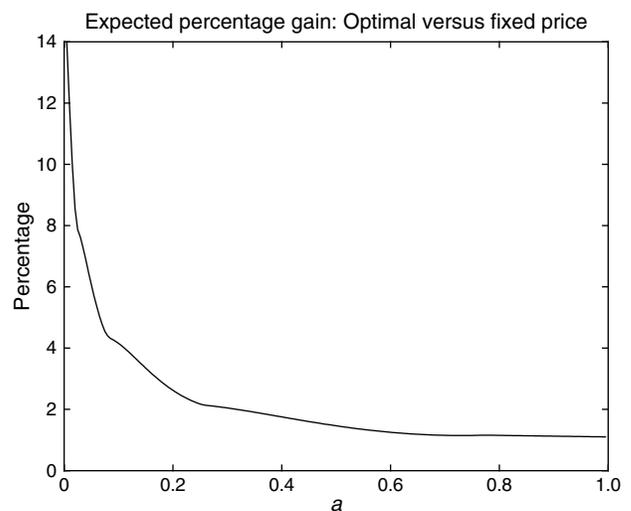
On the other hand, note that the value of the option of taking a loan, would be substantial for low asset levels. In particular, the value function would no longer be zero at zero net capital.

4.3. The Value of Pricing Decisions

Finally, we investigate the value of optimal pricing decisions, using the model of §3, which excludes the option of banking. The parameters used are otherwise as above, with the demand scaling as $4 - p$ (with this demand function the “monopoly price” is 2,000 rupees and the “shutdown price” is 4,000 rupees). The structure of the optimal policy is consistent with Theorem 2. The optimal capacity policy is of threshold type, and the optimal pricing policy is nonincreasing.

We compare the optimal policy with a policy which charges a fixed price of 2,000 rupees. Recent articles investigating different models in for-profit firms have found dynamic pricing to provide small gains in expected revenue (usually 5% or less; see, for instance, Gallego and van Ryzin 1994 and Chen et al. 2004). In this case, and as seen in Figure 5, the difference in expected value is high for a low asset level, when there will be a large amount of unmet demand if the organization charges a fixed price more adequate for periods in which the asset level is higher. When the assets are $a < 0.1$ (less than 400,000 rupees), the percent gain quickly rises above 5%. Although dynamic pricing in the sense of revising the price in each time period may be too complex and costly to be warranted, this points to the importance of considering capacity when setting the price, and therefore the importance of periodically revising the price as the available resources change. (If loans were available at a reasonable rate there would be little or no value in periodic price revisions, because the optimal policy when the

Figure 5. Ratio of the value functions with dynamic price and with fixed price.



INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at http://journals.informs.org/.

asset level is low would be to borrow capital in order to provide optimal capacity for the given demand distribution.) For mature organizations with an operation scaled to match the available demand, the asset level is likely to most often be above the threshold, where the expected gains from using dynamic pricing are small. The larger expected gains from using dynamic pricing when the asset level is small are of interest for organizations at the start-up stage, and for small organizations with a high variability of demand. Aravind, either in its early years or when setting up a new hospital in an underserved area, seems to fit this profile.

5. Extensions and Future Research

5.1. Demand Distribution and Demand Function

Condition (10) is used in the proofs only at the optimal price for each ϵ problem. If we assume that for all a and for ϵ sufficiently small

$$g\left(\frac{a}{\gamma(p^*(a))}\right) \geq \frac{1}{e(p^*(a))}, \quad (14)$$

then Theorem 2 still holds. This should be the case if (14) holds with strict inequality for $\epsilon = 0$. One interesting direction in which to extend our results is to find sufficient conditions on $f(\cdot)$ guaranteeing (14). Note, however, that establishing (14) for more general $f(\cdot)$ requires propagating properties stronger than concavity on the value function.

Also, we have restricted our analysis of the problem with pricing to the multiplicative demand case. For more general demand functions of the form $D = f(p, \Theta)$, the elasticity of sales, following Kocabiyikoğlu and Popescu (2005), is given by

$$\varepsilon(p, y) = p \frac{L'(p, y)}{1 - L(p, y)},$$

where $L(p, y) = P(D \geq y)$, with L' the derivative of L with respect to p . Condition (10) then becomes $\varepsilon(p, y) \geq 1$. This condition can be used to extend our results to the case of more general demand functions.

5.2. Objective Functions

We assumed a linear social return sx , where x is the capacity allocated to M-customers. This is a good model if there are no significant economies or diseconomies of scale, and if the size of the unmet need being addressed is large compared to the service capacity that the organization can potentially provide. A natural extension of our model is to the case of decreasing marginal social return, where the impact of serving x mission customers is described by a concave function $s(\cdot)$. A case of interest corresponds to random demand from M-customers. If this demand is given by a random variable X , the organization contributes to its mission in proportion to $s(x) = \mathbf{E}_X(x \wedge X)$. The optimal policy will no longer be of threshold type. It should, however, be possible to derive simple and effective heuristics for this case drawing from the results presented here.

5.3. Flexible Capacity

We assumed that the service capacity for R-customers needs to be committed at the beginning of each period, before demand is realized. In this section we briefly explore the case where the organization has the flexibility to allocate its service capacity between the two classes of customers after the actual demand in the current period is known. For instance, in a dual-hospital operation like Aravind it may in some cases be possible to share capacity. This would entail, among other things, moving staff across hospitals midway through their two-month assignments, and might in some cases require postponing the surgery of free patients already scheduled.

We show that the structure of the optimal policy is in many ways similar to our previous results. However, the idea of a threshold does not apply. The optimal policy is to, in each period, serve all R-customers first. This can be interpreted as essentially the same idea as before: Prioritizing R-customers will allow for greater mission impact. Note that to the extent that sharing of capacity is possible, the structure of the optimal policy implies that it is desirable to do so. When dynamic pricing is allowed we show the optimal price to be decreasing in the asset level, albeit under a somewhat more restrictive condition on the demand distribution. Both with and without dynamic pricing, we show decreasing marginal mission impact relative to the available resources.

Without loss of generality, we again assume that $c_R = c_M = s = 1$. At the beginning of each period, the organization makes an investment decision and chooses how much total service capacity $c_t \geq 0$ to reserve and sets the amount to be banked, $z_t \geq 0$. The choices of c_t and z_t are limited by the current resources, i.e., $z_t + c_t = a$. Demand is then realized and the organization must decide how to allocate the capacity c_t , subject to $c_t = x_t + y_t$, where x_t and y_t are the number of M-customers and R-customers served. The optimality equations are now

$$v_t(a) = \max_{\substack{0 \leq c, 0 \leq z, \\ c+z=a}} \mathbf{E}_\Theta \max_{\substack{0 \leq x, 0 \leq y, \\ x+y=c}} x + \tau(y \wedge \Theta) + \alpha v_{t+1}(p(y \wedge \Theta) + \beta z), \quad (15)$$

with $v_T(a) = a$. This equation can be simplified by writing $c = a - z$ and $x = c - y = a - y - z$, so that

$$v_t(a) = a + \max_{0 \leq z \leq a} \mathbf{E}_\Theta \max_{0 \leq y \leq a-z} J^{v_{t+1}}(y, z, \Theta), \quad (16)$$

where the operator J^v is defined as

$$J^v(y, z, \theta) = -z - y + \tau(y \wedge \theta) + \alpha v(p(y \wedge \theta) + \beta z). \quad (17)$$

Using a similar approach, Proposition 1 can be shown to hold when the service capacity is shared. Again, the interesting case is $\alpha\beta < 1$ and $\alpha p + \tau > 1$. The following result characterizes the optimal service-capacity decision $\tilde{y}_t(c_t, \theta_t)$ given the capacity choice c and the total number of R-customers θ_t .

PROPOSITION 2. *With a fixed price and shared capacity, and given the resources allocated to providing service capacity, it is optimal to serve as many R-customers as possible,*

$$\tilde{y}_i(c_i, \theta_i) = c_i \wedge \theta_i.$$

The proof is in §6 of the electronic companion. The optimal capacity decision is an infinite-threshold policy, and Proposition 2 allows us to simplify the optimality equations to

$$v_t(a) = a + \max_{0 \leq z \leq a} \mathbf{E}_\Theta J^{v_t+1}((a - z) \wedge \Theta, z, \Theta). \quad (18)$$

The following result states that the marginal social return is decreasing. The corresponding optimal banking decision $z_t^*(a)$, defined as the minimum optimal decision in case of multiple optima, is not necessarily of threshold type because the constraint a appears in the maximand in (18). The optimal decision $z_t^*(a)$ is monotone in the assets if $p \geq \beta$ and $\alpha(p - \beta) + \tau \geq 1$ (both hold when p is sufficiently large compared to β , i.e., the profit margin is sufficiently larger than the return rate from banking).

PROPOSITION 3. *For each time period t , $v_t(\cdot)$ is differentiable nondecreasing concave. Further, the optimal banking decision $z_t^*(a)$ is nondecreasing in a if $p \geq \beta$ and $\alpha(p - \beta) + \tau \geq 1$.*

The proof is in §6 of the electronic companion. We now consider the corresponding problem with dynamic pricing. Following the approach in §3, consider a demand function with multiplicative uncertainty, $D_t = \gamma(p_t)\Theta_t$. When banking is ignored ($\beta = 0$) and R-customers do not have mission impact ($\tau = 0$), the optimality equations are

$$\begin{aligned} v_t(a) &= a + \max_{0 \leq p \leq \bar{p}} \mathbf{E}_\Theta J^{v_t+1}(a \wedge \gamma(p)\Theta, 0, \gamma(p)\Theta) \\ &= a + \max_{0 \leq p \leq \bar{p}} -\mathbf{E}_\Theta[a \wedge \gamma(p)\Theta] + \alpha H^{v_t+1}(p, a), \end{aligned} \quad (19)$$

where $H^v(\cdot, \cdot)$ is defined as in (8). As before, we show that the optimal price is interior (in $]0, \bar{p}[$) and $\tilde{J}^v(a) = \max_{0 \leq p \leq \bar{p}} \mathbf{E}_\Theta J^v(p, a)$ is twice differentiable. Also, because $r(\cdot)$ is decreasing for $p \geq \underline{p}$, we have $r'(\cdot) \leq 0$ at the optimal price. As for Proposition 1, a sample-path argument shows that the optimal price satisfies $\alpha p > 1$. In order to characterize the optimal policy when capacity is not flexible, condition (10) was needed. When capacity is flexible, the corresponding condition is: $\forall p \in [\underline{p}, \bar{p}[$ and $\forall a, a/\gamma(p) \in [\underline{\theta}, \bar{\theta}]$,

$$g\left(\frac{a}{\gamma(p)}\right) \geq \frac{\alpha p}{\alpha p - 1} \cdot \frac{1}{e(p)}. \quad (20)$$

For instance, if $\alpha p = 2$, the condition is implied by $g(a/\gamma(p)) \geq 2/e(p)$. Condition (20) implies condition (10) of the inflexible capacity case and is therefore more restrictive.

THEOREM 3. *If (20) holds, then $v_t(\cdot)$ is increasing concave. Further, the optimal price is nonincreasing.*

See the electronic companion for a proof.

6. Conclusion

The success of Aravind, which has dramatically improved the quality of life of hundreds of thousands of people, demonstrates the potential of nonprofit organizations that generate resources by also running for-profit operations. The problem faced by such ventures differs in fundamental ways from the profit maximization problem of commercial ventures and, to our knowledge, we provide the first analytical approach for examining resource management issues in this context.

We have investigated how a nonprofit organization should dynamically allocate its assets over time between its revenue-generating activities and its mission, in order to maximize the organization's social impact. The optimal capacity allocation policy is of threshold type. Over all periods, this policy allows the organization to have a higher expected social impact by serving more mission customers. Although this needs to be balanced with other considerations, such as maintaining the organization's mission culture, which might be jeopardized if the organization only serves revenue customers for several consecutive periods, the analysis allows us to quantify the trade-off. Theoretical analysis and numerical studies suggest that there is limited value in the option of banking assets from one period to the next, but point to the importance of adjusting the price charged to revenue customers depending on whether assets are scarce or abundant.

Different nonprofit organizations face different problems and require different modeling approaches and, as a consequence, there is a need to add to the literature in several directions. We believe the approach proposed in this paper to be a productive initial framework to pursue such work.

7. Electronic Companion

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

Endnotes

1. "I mortgaged my house and raised enough money to start." (Dr. Venkataswamy in Rangan 1994, p. 7.)
2. One exception is Young (2004). He addresses a broad range of resource-allocation issues, within and across activities, using a marginal contribution framework. The main differences between his analysis and ours are (1) we capture the trade-off between spending on mission activities now versus in the future, (2) we allow for randomness in demand for the revenue-producing activities, and (3) our model does not assume decreasing marginal social return.
3. We are only concerned with net financial contribution. An M-activity may generate some revenue that covers part of its operational costs. As long as the net financial contribution of producing a unit of this good or service is negative, we can model it as a pure mission good with

costs equal to the net negative financial contribution. Similarly, an R-activity must produce a net positive financial contribution that can be used in future periods to subsidize M-activities.

4. In addition to the capacity decisions requiring significant precommitment, there is the ability to increase or decrease the total variable expenditures over both the paying and free hospitals. Although the flexibility in adjusting total staff costs from one period to the next is limited, it is not negligible. Because most staff work with long hours, there may be considerable flexibility to increase or decrease overtime. Other items are clearly variable costs and can be shifted to either revenue or mission customers, such as intraocular lenses, which represent 8% of operating costs (a share that is increasing as free hospitals also shift to this procedure).
5. For the nonlinear case, we expect that some form of threshold effect holds at least when the rate at which the marginal social return decreases is not too high relative to the “thickness” of the tail of the distribution or R-customer demand.
6. See McCardle (2005), who also suggests that the operations community, through its professional societies, set standards for pro bono work, similarly to what is done in the legal profession.
7. The assumption of F^{-1} increasing in y is not required for our results. However, the concavity condition (which is required) is more naturally understood in the context of increasing F^{-1} .
8. For the Boston Symphony, if the mission is understood as maintaining an active repertoire of skills and artistically important works, the demand for those performances is, in this sense, a secondary concern. Further, the organization can fill auditoriums to fulfill its mission of exposing a broader public to classical music by means such as offering discounts and student tickets.
9. The assumption of twice differentiability of F can be relaxed to differentiability by extending the limit argument in Lemma 3.

Acknowledgments

This work was supported in part by the Center for the Advancement of Social Entrepreneurship at the Fuqua School of Business, Duke University. Gregory Dees very generously provided much help in defining and framing the problem and in revising the paper. Bill Meehan of McKinsey greatly helped to improve a previous version of the paper with many suggestions and insightful comments. The authors are grateful to Dinah Vernik for her work on the numerical examples. We also thank, for useful comments, Beth Anderson, Paul Zipkin, James Smith, Gregory DeCroix, and the participants at the operations management seminars at the Fuqua School of Business, UCLA, INSEAD, LBS, University of Washington Business School, University of Chicago

Graduate School of Business, Yale School of Management, and Tepper School of Business at Carnegie Mellon University.

References

- Alter, S. K. 2000. *Managing the Double Bottom Line: A Business Planning Guide for Social Enterprises*. Pact Publications, Washington, DC.
- Anderson, B. B., J. G. Dees, J. Emerson. 2002. *Developing Viable Earned Income Strategies in Strategic Tools for Social Entrepreneurs: Enhancing the Performance of Your Enterprising Nonprofits*. John Wiley & Sons, New York.
- Babich, V., M. J. Sobel. 2004. Pre-IPO operational and financial decisions. *Management Sci.* **50**(7) 935–948.
- Bernstein, F., A. Federgruen. 2005. Decentralized supply chains with competing retailers under demand uncertainty. *Management Sci.* **51**(1) 18–29.
- Billingsley, P. 1995. *Probability and Measure*, 3rd ed. John Wiley & Sons, New York.
- Bitran, G., R. Caldentey. 2003. An overview of pricing models and revenue management. *Manufacturing Service Oper. Management* **5** 203–229.
- Bonnans, F., A. Shapiro. 2000. *Perturbation Analysis of Optimization Problems*. Springer, New York.
- Boschee, J. 2001. *The Social Enterprise Sourcebook: Profiles of Social Purpose Businesses Operated by Nonprofit Organizations*. Northland Institute, Minneapolis, MN.
- Brinckerhoff, P. C. 2000. *Social Entrepreneurship: The Art of Mission-Based Venture Development*. John Wiley & Sons, New York.
- Buzacott, J., R. Zhang. 2001. Inventory management with asset-based financing. *Management Sci.* **50**(9) 1274–1292.
- Chen, H., O. Wu, D. Yao. 2004. Optimal pricing and replenishment in a single-product inventory system. Technical report, Department of Industrial Engineering and Operations Research, Columbia University, New York.
- Chen, X., D. Simchi-Levi. 2004. Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The finite horizon case. *Oper. Res.* **52** 887–896.
- Chen, Y., S. Ray, Y. Song. 2003. Optimal pricing and inventory control policy in periodic-review systems with fixed ordering cost and lost sales. Technical report, The Chinese University of Hong Kong, Hong Kong.
- Dees, G. 1998. Enterprising nonprofits. *Harvard Bus. Rev.* **76**(1) 55–67.
- Dees, G. 2004. Putting nonprofit business ventures in perspective. S. M. Oster, C. Massarsky, S. Beinhacker, eds. *Generating and Sustaining Nonprofit Earned Income: A Guide to Successful Enterprise Strategies*. Jossey-Bass, San Francisco, 3–18.
- Dees, G., B. Anderson. 2003. Sector-bending: Blurring the lines between nonprofit and for-profit. *Society* **40**(4) 16–27.
- Elmaghraby, W., P. Keskinocak. 2003. Dynamic pricing in the presence of inventory considerations: Research overview, current practices and future directions. *Management Sci.* **49**(10) 1287–1309.
- Elstrodt, H.-P., A. Schindler, A. C. Waslander. 2004. Helping nonprofits do business. *McKinsey Quart., Web Exclusive* (December). http://www.mckinseyquarterly.com/Helping_nonprofits_do_business_1552.
- Federgruen, A., A. Heching. 1999. Combined pricing and inventory control under uncertainty. *Oper. Res.* **47**(3) 454–475.
- Foster, W., J. Bradach. 2005. Should nonprofits seek profits? *Harvard Bus. Rev.* **83**(2) 92–100.
- Gallego, G., G. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* **40** 999–1020.
- Gruber, R. E., M. Mohr. 1982. Strategic management for multiprogram nonprofit organizations. *California Management Rev.* **3**(Spring) 15–22.

- Heyman, D., M. Sobel. 1984. *Stochastic Models in Operations Research (II)*. Dover Publications, Mineola, NY.
- Huh, W., G. Janakiraman. 2008. Optimality results in inventory-pricing control: An alternative approach. *Oper. Res.* **56**(3) 783–790.
- Keeler, E., S. Cretin. 1983. Discounting of life-saving and other nonmonetary effects. *Management Sci.* **29**(3) 300–306.
- Klausner, M. 2003. When time isn't money: Foundation payouts and the time value of money. *California Management Rev.* **1**(1) 50–59.
- Kocabiyikoglu, A., I. Popescu. 2005. Joint pricing and revenue management with general stochastic demand. Technical report, INSEAD, Fontainebleau, France.
- Larson, R., V. Forth. 2002. *The Essential Guide to Starting a Money-making Business in Your Nonprofit Organization*. Amherst H. Wilder Foundation, St. Paul, MN.
- Li, L., M. Shubik, M. J. Sobel. 2003. Control of dividends, capital subscriptions, and physical inventories. Technical report, Yale University, New Haven, CT and Case Western Reserve University, Cleveland, OH.
- McCardle, K. 2005. Pro bono O. R.: O. R. for the public good. *OR/MS Today* **32**(6) 32–36.
- Modigliani, F., H. Miller. 1958. The cost of capital, corporation finance, and the theory of investment. *Amer. Econom. Rev.* **48** 261–267.
- Oster, S. M. 1995. *Strategic Management for Nonprofit Organizations: Theory and Cases*. Oxford University Press, New York.
- Oster, S. M., C. Massarsky, S. Beinhacker. 1995. *Generating and Sustaining Nonprofit Earned Income: A Guide to Successful Enterprise Strategies*. Jossey-Bass, San Francisco.
- Petruzzi, N. C., M. Dada. 1999. Pricing and the newsvendor problem: A review with extensions. *Oper. Res.* **47**(2) 183–194.
- Rangan, V. K. 1994. The Aravind Eye Hospital Madurai, India: In service for sight. Harvard Business School Case (9-593-098), Harvard University, Boston.
- Robinson, A. 2002. *Selling Social Change (Without Selling Out): Earned Income Strategies for Nonprofits*. Jossey-Bass, San Francisco.
- Shah, J., L. S. Murty. 2004. Compassionate, high quality health care at low cost: The Aravind model; In conversation with Dr. G. Venkataswamy and R. D. Thulasiraj. *IIMB Management Rev.* **16**(3) 31–43.
- Skloot, E. 1983. Should not-for-profits go into business? *Harvard Bus. Rev.* **61**(1) 20–25.
- Skloot, E. 1987. Enterprise and commerce in nonprofit organizations. *The Nonprofit Sector: A Research Handbook*. Yale University Press, New Haven, CT.
- Steckel, R., R. Simons, P. Lengsfelder. 1989. *Filthy Rich and Other Nonprofit Fantasies: Changing the Way Nonprofits Do Business in the 90s*. Ten Speed Press, Berkeley, CA.
- Wang, Y., L. Jiang, Z. Shen. 2004. Channel performance under consignment contract with revenue sharing. *Management Sci.* **50**(1) 34–47.
- Weisbrod, B. 1998. *To Profit or Not to Profit: The Commercial Transformation of the Nonprofit Sector*. Cambridge University Press, New York.
- Weisbrod, B. A. 2005. The pitfalls of profits. *Stanford Soc. Innovation Rev.* **2**(3) 40–47.
- Young, D. Y. 2002. Commercialization, social ventures, and for-profit competition. L. M. Salamon, ed. *The State of Nonprofit America*. Brookings Institution Press, Washington, DC, 423–446.
- Young, D. Y. 2004. Deploying resources effectively. S. M. Oster, C. Massarsky, S. Beinhacker, eds. *Generating and Sustaining Nonprofit Earned Income: A Guide to Successful Enterprise Strategies*. Jossey-Bass, San Francisco, 247–268.