Faculty Research

University
of Michigan
Business
School

WORKING PAPER SERIES

Hospital Operating Room Capacity Expansion

William S. Lovejoy
University of Michigan Business School

Ying Li
University of Michigan Business School

Hospital Operating Room Capacity Expansion

William S. Lovejoy

Business School

University of Michigan

701 Tappan

Ann Arbor, MI 48109-1234

Ying Li

Business School

University of Michigan

701 Tappan

Ann Arbor, MI 48109-1234

1. Introduction

The analysis in this paper is motivated by a large midwestern hospital seeking alternative ways to expand capacity in their operating rooms. Operations are the financial engine that drives a significant portion of the costs and revenues in the hospital. Currently, the hospital has 25 operating rooms, and each has specified standard hours of operation. Hospital staff are scheduled to be available during these hours, and the available time is used by schedulers when determining how many cases to add to the schedule each day. There may, on some days, not be a sufficient number of cases to completely fill the regular working hours, and on other days a case may extend beyond working hours. Cases in process when regular time expires are not interrupted, but are completed on overtime as would be expected given the nature of the service.

The hospital is experiencing an increase in surgical caseload, which contributes to operating room congestion and delay, surgeon and patient frustration, and increasing waits to get on schedule. New operating room capacity can be had by building new OR's or extending the working hours in the current OR's. The choice between these options is complicated by the fact that patients, surgeons and surgical staff, and hospital administrators are all important stakeholders in the health service operation, and each has different priorities. Paying attention to each stakeholder population is important because each is mobile to some degree. Surgeons can change their hospital affiliations. Service levels to patients can have long term consequences when Health Maintenance Organizations periodically choose which hospitals will serve their patient populations.

This paper investigates the tradeoffs among three performance criteria (wait to get on schedule, scheduled procedure start time reliability, and hospital profits) which are of particular importance to three different constituencies (patients, surgeons and surgical staff, hospital administrators). The intent is to advise the hospital on how it can best invest in capacity to provide high quality service while protecting its profitability, acknowledging the key role that each constituency plays in that objective. We concentrate on these cost/service tradeoffs, and do not address the quality of the care itself (mortality rates, etc.), which is assumed constant in this study. Financial data has been altered to protect confidentiality, but in no instance does this influence the essential nature of the analysis.

Our model is a streamlined version of reality, but retains measures and influences important

to the capacity investment decision. For example, in reality there are weekly and seasonal fluctuations in case load and types of cases (due to the potential for accidents on weekends and the different weather/recreational environments in different seasons). Our analysis assumes Poisson arrivals at a constant rate λ, and so is better for longer-term financial tradeoffs than for day-to-day scheduling considerations. In reality procedures are of many types (hernias, hip replacements, heart surgery, etc.), and on any given day procedures of predictably different lengths may be performed in a specific room (one long one or several short ones, for example). In our model we assume that all procedures share a common distribution of procedure length, so that we essentially deal with the "average" procedure. Similarly, while the resources needed for individual procedures may vary the cost and revenue parameters used here reflect the averages over the appropriate case mix. In reality, not all of the operating rooms in our client hospital are suitable for all procedures, and some of the rooms are staffed for a different number of hours than others. In contrast, in our model all rooms are generic and share a single "regular time" working day length. We do not believe that this simplification greatly affects our conclusions, since in both cases the decision is how long to staff rooms (at cost) and how many procedures to schedule into the rooms given the risk of overtime. Each of these considerations is captured in our model. The actual scheduling system used differs from our model abstraction in ways that are discussed in section 6.1 below. Finally, the actual manner in which costs are incurred by the hospital is complicated by workforce policies and union rules, and includes different pay scales and workforce requirements for holidays, weekends, on-call duty, mixes of seniority, etc. In contrast, we use one average cost rate to staff an operating room for regular time operations, and a second (higher) rate for overtime operations. Again, this simplification may not be accurate on any specific day of operation, but will capture the general longer-range cost trends that the hospital will experience. Indeed, the costs and revenues used in our analysis were estimated from annual hospital financial data. In general, our model abstracts away from some tactical details but retains the major tradeoffs important to longer range capacity investment decisions.

Firm (hospital) profits and customer (patient) delays to get into service are natural metrics for service operations, and often studied (as they are here) using queueing relationships. In our setting it is appropriate to add start time reliability to this set of metrics. Surgeons

3

and hospital staff are very sensitive to the reliability of the scheduled start time for the procedure(s) that they will be attending. A surgeon who is scheduled to operate at 1:00 P.M. but does not actually get into the OR until 4:00 P.M. has essentially been robbed of 3 productive hours of his/her work time. That is, no large or substantive tasks can be started while waiting to get into the OR, leaving the surgeon or staff member with only administrative trivia that he/she can engage during the wait. The depths of this frustration, which can be fully appreciated only by talking directly to surgeons, is vividly apparent in our client facility and has also been documented elsewhere (c.f. Magerlein and Martin 1978, Gordon et al 1988). Naturally, when cases are scheduled close together (to make maximal use of capacity) uncertainty is compounded as variability in one transmits to all subsequent procedures. Hence, procedures scheduled later in the day are known to have less reliable start times than those scheduled earlier in the day. This is at the core of a well-known surgeon preference for early start times, and an equally common surgeon aversion to later start times.

We wish to advise our client hospital on potential capacity investments recognizing that there are three relevant constituencies with different preferences regarding system performance.

2. Relevant literature and relevant notation

Baily's (1952) early work on scheduling in health care stimulated research streams utilizing both queuing models and discrete-event simulation. In the first stream (c.f. Fandel and Hegemann 1989, Brahimi and Worthington 1991, Kim 1999) M/G/c queues or variants are used to explore the tradeoffs between throughput rate and patient wait times, and in some cases physician idle time. The latter research stream (c.f. Kuzdrall 1981, Charnetski 1984, Ho and Lau 1992) explores the same questions via simulation. A common theme in much of this literature is setting start times to minimize the sum of overage and underage costs. Weiss (1990) tackles this problem in a hospital setting and shows the optimality of a critical fractile policy. Similar themes arise in the due date scheduling literature for quoting lead times to customers in more standard product/service production settings (c.f. Wein 1991, Spearman and Zhang 1999, Hopp and Sturgis 2000 and references therein). Spearman and Zhang observe that expected tardiness, while a favorite with academic authors, has little

4

presence in practice where "serviceability" (the probability of being delayed) dominates as the relevant metric. They show, however, that focusing on serviceability will also do well by the alternative metric, and recommend a policy of constant serviceability for all jobs. This is accomplished by setting due dates equal to a critical fractile of the sojourn time distribution, making the link between serviceability and minimizing expected overage and underage costs apparent. We adopt this logic in part, in that we focus on "start time reliability" (probability of on-time start) as our metric of performance, and assume a policy that assigns one reliability to all surgical procedures. Babad et al (1996) analyze a service system in which customers arrive at regular intervals and are guaranteed service within a stated time after arrival. Van Ackere (1990) explicitly models surgeon behavior when facing a random start time, shows that the surgeon may rationally choose to arrive late and suggests how the scheduler can take this into account. Swanberg and Fahey (1983) note that a common response to case delays and long waits to get on schedule is to build more operating rooms, but that better use of existing assets may be a dominant strategy. Whether this is true at our client facility is the core question that we address. O'Keefe (1985) observes that hospital scheduling is fundamentally a political problem that requires consensus management among all of the affected parties. Doctors, nurses and administrators have different perspectives and hence differing objectives, a fact that we explicitly recognize in our analysis.

The following section introduces our model and presents some structural results. Throughout the paper, we will use F_Y to denote the cumulative distribution function of a random variable Y. $Y \sim Z$ means that random variable Y is distributed as random variable Z. $(a \vee b)$ denotes the maximum of a and b. We make use of several types of partial orders to order functions and random variables. For two real-valued functions f and g we say $f \leq g$ if $f(x) \leq g(x)$ for all x. We use $\preceq_d$ to denote first order stochastic dominance. Specifically, two random variables X and Y satisfy $X \preceq_d Y$ if and only if $E_X f \leq E_Y f$ for all nondecreasing functions f. It is well known that this is equivalent to $F_X \geq F_Y$. $\preceq_c$ refers to the convex stochastic partial order. $X \preceq_c Y$ if and only if $E_X f \leq E_Y f$ for all nondecreasing convex functions f. $\preceq_d$ and $\preceq_c$ survive convolutions. See Stoyan (1983) for a more complete discussion of these and other stochastic partial orders. For Markov transition matrices M and M' we will say that $M \preceq_d M'$ if each row of M is $\preceq_d$ the

corresponding row of M'. M is said to be "IFR" if its rows are $\preceq_d$ nondecreasing. More notation and definitions will be introduced as needed.

3. The model

The hospital operates m operating rooms. Patients needing procedures are assumed to arrive via a Poisson process with mean rate λ. Procedure lengths are assumed to be independent, identically distributed random variables distributed as a generic random variable X. The decision variables are the number of cases to schedule per OR per day (n), which is a daily capacity decision, and the probability that a scheduled procedure begins on time (π). These choices, along with the parameters of the "regular time" day length (T), the average margin per case (R) generated excluding OR costs, and the average costs of regular time and overtime staffing (C_r and C_{ot}, respectively) in the OR will determine the daily profits generated in the OR's (details below). We assume that T will be chosen optimally given n and π. Also, for a given arrival rate the caseload per day, n, will uniquely determine the capacity of the system and hence the wait (W) to get on schedule. So, a specific choice of n and π will determine the three performance metrics of interest to our stakeholders (π, W, and $Profit$). Note that π is both a decision variable and a performance metric. We wish to investigate the trade-offs among the three performance metrics.

Delay to get on schedule, starting and ending times and start time reliability

All procedures scheduled for completion on a given day are assumed to be completed that day, on overtime if necessary. Hence, for a specified n the capacity of the OR suite of rooms is nm cases/day with no ambiguity. An arriving customer sees an M/D/1 "bulk service" queue (c.f. Cohen 1982 or Chaudhry and Templeton 1983 and references there) with deterministic capacity of nm patients per day. At the beginning of each day, the scheduler inspects the queue of waiting cases and takes all waiting cases, or nm, whichever is less. Any remaining patients must wait until the following day for a chance to be placed on schedule. The wait time performance of this queue is worse than for the more standard M/D/m queue. This is because patients that arrive on a given day must wait until the following morning for a chance to get on schedule, even if there happens to be an idle room at the moment the patient arrives. This reflects reality in our client hospital for all cases except critical emergencies, which we do not consider in this paper.

The scheduler is assumed to load cases evenly (one case to each room before assigning two cases to any room, etc.) as this will minimize the hospital's overtime exposure. Hence, each room will be statistically identical to each other room. Some of the analysis below will focus on "per room" performance, which will then be aggregated for facility performance.

The decision variable π (the start time reliability) affects the distribution of procedure times within any one room as follows. Suppose that there are k procedures scheduled in a specific room. Let S_i denote the (random) actual start time for procedure i, $i = 1$ to k. Let X_i denote the random length of procedure i (by assumption, the X_i random variables have a common distribution), and let E_i denote the ending time for procedure i. The distribution of S_1 is assumed given and the same for all rooms. This distribution represents how well the hospital gets the first cases of the day into the OR's. From this, the choice of π, and the common distribution of procedure length (F_X) the statistics of all subsequent procedures (up through case k) are preordained. Specifically, $E_1 \sim S_1 + X_1$. Given the distribution of E_1, t_2 is set so that $F_{E_1}(t_2) = \pi$ to ensure the desired start time reliability for case 2. Then, the second procedure starts either at its scheduled time or whenever the first procedure ends, that is $S_2 \sim (E_1 \vee t_2)$. Working forward we have $E_i \sim S_i + X_i$ and $S_{i+1} \sim (E_i \vee t_{i+1})$ for each i.

Costs and revenues

We assume that each case processed generates average margins of R excluding OR costs. The hospital makes a policy decision to support n cases/day in each operating room and to ensure reliability π with its scheduling practices. The hospital also sets T, the regular time day length. Regular time staff is paid (at rate C_r/hour) up to time T regardless of whether or not there are a sufficient number of cases to fully utilize them. If cases extend beyond the regular time day, overtime premiums (C_{ot}/hour) are paid to the staff. Hence, the expected daily costs borne by the hospital in one room will be

$$Cost(T) = C_r T + (C_r + C_{ot}) \sum_{i=1}^{n} p_i \int_{T}^{\infty} (x - T) dF_{E_i}(x)$$

where p_i is the limiting and stationary probability that there are i cases in a generic room per day. We assume that the hospital sets the regular day length optimally, that is T is set equal to $T^* = \text{Min}_T \{Cost(T)\}$. Using standard newsvendor logic (c.f. Heyman and

7

Sobel 1984) it can be shown that $Cost(T)$ is strictly convex in T and that a necessary and sufficient condition for T^* to be optimal is $\sum_{i=1}^{n} p_i(1 - F_{E_i}(T^*)) = C_r/(C_r + C_{ot})$. The profit rate enjoyed by a hospital operating m OR's will be $Profit(T^*) = R\lambda - mCost(T^*)$. We will refer to $Profit(T^*)$ as $Profit^*$ when this is not ambiguous.

In summary, the decision variables are n (the number of cases that can be scheduled per room per day) and π (the start time reliability for scheduled procedures). The performance measures are W, $Profit^*$ and π. Each of these three performance measures is of primary importance to a different stakeholder constituency, and each constituency is important to satisfy.

It is useful in multi-criteria situations such as this to generate an "efficient frontier" of potential solutions. This is the set of triples $(W, \pi, Profit^*)$ such that no one of the three performance dimensions can be enhanced without undermining another. This frontier depends only on the physics of the problem and not on the preferences or utilities of the various stakeholders.

One way to approximate the efficient frontier (c.f Cohen 1978) is to solve a series of optimization problems:

Program EF: $\text{MAX}_{n,\pi} Profit^*(\pi, n)$

subject to $W \leq \tilde{W}$

                    $\pi \geq \tilde{\pi}$

for different points $\tilde{W}$ and $\tilde{\pi}$. Whenever this program is feasible and the solution unique, then the optimal triple $(W, \pi, Profit^*)$ is on the efficient frontier. In the case of nonunique optima, one of the set of optimal solutions is on the frontier. Therefore, the frontier can be approximated by solving the program at a spectrum of points $\tilde{W}$ and $\tilde{\pi}$ and graphing the solution surface. The next section provides some structural results that simplify the parameterized program EF and simultaneously provide insights into the dynamics of the trade-offs among the stakeholders.

4. Structural results

In this section we look at the tradeoffs among W, π, and $Profit^*$. First, we fix n, which results in a fixed W and hence constant patient utility, and ask what will happen as

8

we change π? Increasing π will make surgeons and surgical staff happier, but absent fundamental process improvements is accomplished by adding extra buffer time between procedures and it is natural to presume that this will increase costs. However, as we increase the reliability of the start times, we also increase the reliability of the ending times for each procedure. This decreases our uncertainty as to ending times and we can exploit this by adjusting the day length to decrease our overtime exposure. Consider, for example, the extreme case of unreliable S_1 but deterministic procedure lengths X. If we add enough buffer time after the first procedure, we can essentially drive out all remaining variability and reduce our overtime exposure to zero. It is interesting to ask whether it is possible to increase π *and* decrease total costs if overtime is sufficiently costly relative to regular time staffing costs. This, however, is impossible. Increasing π will generally decrease overtime costs, but these cannot decrease sufficiently to decrease total costs. We have the following proposition, which is proved in the appendix.

Proposition 1: For fixed n (and therefore W), if $\pi \geq \pi'$ then (a) $t_k \geq t'_k$ for all k, (b) $S_k \succeq_d S'_k$ for all k, (c) $E_k \succeq_d E'_k$ for all k, (d) $T^* \geq T'^*$, and (e) $Profit^* \leq Profit'^*$.

If we hold π constant, how do profits change with n? It is natural to assume that increasing capacity for the same arrival rate must increase costs. Proving this is complicated by the presence of the bulk service queue (for which closed form solutions do not exist) and the truncation statistics introduced by maintaining a given reliability π. Still, we can prove some structural results as follows. For any n define $Q_t(n)$ to be the number of patients waiting to get on schedule just before loading the operating rooms at the beginning of day t, and $I_t(n)$ the number still in queue after loading. Both $Q_t(n)$ and $I_t(n)$ are Markov chains. For example, $Q_t(n)$ has transition rule $Q_{t+1}(n) = (Q_t(n) - mn)^+ + A_t$ where A_t is the number of arrivals during period t. Let $M(n)$ be the transition matrix for this Markov chain. It can be shown that $M(n)$ is ergodic (when $\lambda < nm$), is IFR and if $n > n'$ then $M(n) \preceq_d M(n')$. Let $q(n)$ denote the limiting and stationary distribution for $Q_t(n)$. That is, $q_i(n) = \lim_{t \to \infty} \text{Prob}\{\{Q_t(n) = i\}$ and the probability vector $q(n)$ satisfies $q(n) = q(n)M(n)$. For any n and n', let W and W' denote the average wait to get on schedule experienced by an arriving customer. As before, let $p(n)$ denote the limiting and stationary distribution of the number of patients in one of the generic OR's each day. The following is proved in the appendix.

Proposition 2: For fixed π, if $n \geq n'$ then (a) $I_k \leq I_k'$ and $Q_k \leq Q_k'$ with probability one, (b) $W \leq W'$, (c) $q(n) \preceq_d q(n')$, (d) $E(p) \geq E(p')$, and (e) $p(n) \succeq_c p(n')$.

As expected, allowing a larger case load per day reduces the expected wait to get on schedule. The profit consequences are more complicated. One will intuitively expect that increasing n cannot decrease costs, but this is difficult to prove. Parts (c) and (e) of Proposition 2 show the two opposing forces affecting the statistics of the number of cases processed each day as the hospital increases the number of cases it allows on schedule. Increasing n allows more cases in on heavy days, but also stochastically decreases the queue to get on schedule (q, which will put a downward pressure on the number admitted p). The result is a simultaneous shifting of probability mass in the residency distribution p away from the middle and toward the upper and lower tails as n increases, hence the convex ordering.

This makes the profit consequences of increasing n difficult to identify analytically. The expected hospital profit at optimality is

$$Profit^* = R\lambda - m\Big[C_r T^* + (C_r + C_{ot}) \sum_{i=1}^{n} p_i(n) \int_{T^*}^{\infty} (x - T^*)dF_{E_i}(x)\Big].$$

Define $G_i(T) := \int_T^{\infty} (x - T)dF_{E_i}(x)$, the overtime exposure with i cases in a room and day length T. The following, which is proved in the appendix, connects hospital profits to decisions regarding n.

Proposition 3: (a) For fixed π, if $n \geq n'$ and if $G_i(T)$ is convex in i (that is, $G_{i+1} - G_i$ is nondecreasing) then $Profit(T) \leq Profit'(T)$ for all T, and $Profit^* \leq Profit'^*$; (b) For π so low (e.g. zero) that procedures run back-to-back, $G_i(T)$ is convex in i; (c) If the generic distribution of procedure length has finite support and $\pi = 1$, then $G_i(T)$ is convex in i.

The authors have run extensive simulation tests using normal, lognormal, beta, exponential, gamma and uniformly distributed procedure lengths. $G_{i+1}(T) - G_i(T)$ was in all cases nondecreasing or nearly so (within the random error limits of the simulation). Practically, we can expect $G_i(T)$ to be convex in i and that hospital profits will decline with n.

10

These structural results allow a reduction in the computational burden when approximating the efficient frontier via program EF. Specifically, $Profit^*$ is decreasing in π, so that there will always be an optimal solution to EF with $\pi^* = \tilde{\pi}$. W is decreasing in n and the only W values that are feasible are those associated with integer n. Further, $n > \lambda/m$ (equivalently, the utilization must be less than one) is required for the system to be stable. Hence, program EF can be reduced to

Program EF': $\qquad$ $MAX_n Profit^*(\tilde{\pi}, n)$

subject to $\qquad$ $n \geq \tilde{n}$

$\qquad\qquad\qquad\qquad\qquad$ n integer

where $\tilde{\pi}$ is allowed to range over $[0,1]$ and $\tilde{n}$ is allowed to range over integer values greater than λ/m. If n^* denotes the optimal n in EF' then $(\tilde{\pi}, n^*, Profit(\tilde{\pi}, n^*))$ will be on the desired trade-off surface. For practical purposes, Proposition 3 and the attending discussion suggest that $Profit^*$ will decline in n, as well. This implies that the trade-off surface can be generated very simply without optimization. Specifically, for $\tilde{\pi} \in [0,1]$ and integer $\tilde{n} > \lambda/m$, the point $(\tilde{\pi}, \tilde{n}, Profit^*(\tilde{\pi}, \tilde{n}))$ will be on the surface. Henceforth we will drop the "tilde" ($\tilde{n}$ and $\tilde{\pi}$) for notational convenience.

Actually constructing this surface requires more details regarding hospital finances, the procedure time distributions and the bulk service queuing model that predicts average wait. In the next section we develop a spreadsheet-compatible approximate model that allows the rapid generation of the desired surface. In later sections, we use this model to inform hospital policy.

5. Numerically approximating the efficient frontier

Analytical expressions are not available for the delay to get on schedule in a bulk service queue, or for the statistics of convolutions of truncated random variables. Here we suggest some simple approximations for these and report on their accuracy.

Bulk service queue approximation

Although there is no closed form expression for W in a bulk service queue, several authors have derived results using transform or generating function techniques (c.f. Cohen 1982, Chaudhry and Templeton 1983) for bulk service queues. Here we develop a simpler

approximation that derives from natural intuition for the light- and heavy-traffic cases. Since our approximation is within a few percent of simulated benchmarks, we feel that its simplicity and intuitive transparency recommends it in this application. We do not make this claim in general, pending to more extensive testing, but the approximation is well-suited for the tradeoffs appropriate in our capacity expansion problem and for data appropriate for our context.

The fundamental difference between our bulk service queue and a classical queue is that in the former a patient that arrives just after a batch of cases has been admitted to the hospital will have to wait a full day in queue, even if there are empty servers (rooms). However, for very high utilizations we might expect the delay in queue to be dominated by the long line seen by an arriving customer (which will be cleared at a rate of mn patients/day) rather than the details of what happens when the arrival finally reaches the front of the queue. That is, at high utilizations we should be able to approximate the expected delay with a classical $M/D/1$ queue with capacity nm per day. For very low utilizations, the queue will be cleared (with high probability) each day, in which case the day will "begin" (just after loading schedule) with an empty queue. We can exploit the feature of deterministic capacity in our bulk-service setting to derive a closed-form expression for W_1, the average wait in queue for patients that arrive on any day that begins with an empty queue. Then, an intuitive expression for the expected wait experienced by an arriving patient will be a weighted average (weighted by utilization) of what we think should happen in heavy traffic and what we know happens in light traffic. The following, which is proved in the appendix, provides a closed form expression for W_1.

Proposition 4: If an $M/D/1$ bulk service queue (with deterministic capacity mn each day of length τ) starts with an empty queue, then the average wait experienced by customers arriving on the first day is

$$\tau \Big[\frac{1}{2} + \sum_{i=1}^{\infty} \sum_{j=1}^{mn} \frac{2ij + i(i-1)mn}{2(imn+j)} Pr(A_1 = imn + j) \Big].$$

For stable systems ($\lambda < mn$), the probability of having $2mn$ patients arrive on any given day is small, and there will be little impact on this expression if we ignore the terms with

$i > 1$. With this simplification, the above intuition, and modeling the heavy traffic delay using the familiar Pollacek-Kinchine formula (c.f. Heyman and Sobel 1982), we found the following to perform very well as an approximate expression for W.

$$W \approx \rho^{2 \cdot \lceil \frac{nm}{10} \rceil + 1} \cdot \frac{1}{2} \cdot \left(\frac{\rho}{1 - \rho} \cdot \frac{1}{nm} \right) + \frac{1}{2} + \sum_{j=1}^{nm} \frac{j}{nm + j} Pr(A_1 = nm + j).$$

In simulation runs with 64 different capacity and arrival rate combinations the average error (($approximation - simulation$)/$simulation$) was .59% and the range of errors was from -1.93% to $+4.72\%$. This level of accuracy is sufficient for our purposes.

Truncated distributions

The statistics of convolutions of truncated random variables are not generally available in closed form. However, if X is normally distributed with mean μ and standard deviation σ, the first two moments of the truncated distribution can be computed as follows (c.f. Elandt 1961):

$$E(X \vee t) = t \, F_X(t) + \mu(1 - F_X(t)) + \sigma^2 \, f_X(t)$$

$$Var(X \vee t) = t^2 F_X(t) + (\sigma^2 + \mu^2)(1 - F_X(t)) + \sigma^2(t + \mu)f_X(t) - E^2(X \vee t).$$

Of course, the truncated distribution is not normally distributed. The essence of our approximation is to assume normalcy throughout, so that the above mean and variance are sufficient to completely determine the truncated distribution. While clearly not appropriate in all cases, this approximation has its merits. Recall that S_i and E_i are the starting and ending times for procedure i, and that X is the (now presumed normally distributed) generic procedure length. $E_i \sim S_i + X \sim (E_{i-1} \vee t_i) + X$ so that even if E_{i-1} and X are normal, E_i need not be. However, under these conditions E_i goes to normal in distribution as t_i gets very large or very small. In the former case, the start time is deterministic (in the limit) and in the latter case we essentially start operations back-to-back so E_i is the convolution of i procedure durations, which will be normal. Finally, the statistics of interest (start time reliability and overtime exposure) will usually depend on the upper tail of the E_i distributions while the nonsmooth effects of truncations in S_i tend to be manifest in the lower tail of $E_i \sim S_i + X_i$.

13

The distribution of the start time for the very first procedure of the day, S_1, is given. This reflects how good the hospital is at getting the first operation under way each morning. The desired reliability of scheduled starts (π) is also given. Assuming the E_i are normally distributed, the first two moments of S_i and E_i for all procedures on schedule can be computed recursively as follows. $E(E_i) = E(S_i) + \mu$ and $Var(E_i) = Var(S_i) + \sigma^2$ for all i. For $i \geq 2$, $E(S_i) = t_i \pi + E(E_{i-1})(1 - \pi) + Var(E_{i-1}) f_{E_{i-1}}(t_i)$, and $Var(S_i) = t_i^2 \pi + [Var(E_{i-1}) + E^2(E_{i-1})](1 - \pi) + Var(E_{i-1})[t_i + E(E_{i-1})] f_{E_{i-1}}(t_i) - E^2(S_i)$, where $t_i = F_{E_{i-1}}^{-1}(\pi)$ is the scheduled start time for the ith operation.

To test our approximation we simulated 1000 independent 7-procedure ($n = 7$) days for each of five different (μ, σ) pairs. For each of the seven procedures, our approximate mean and variance of the ending time is within the 99% confidence interval of the simulated value. For each of the (μ, σ) pairs we also simulated the actual probability of on-time starts achieved when our scheduled times are set using the approximation. That is, we computed the distributions of the E_i random variables using the above normal approximation and using that approximate distribution set the start times $t_{i+1} = F_{E_i}^{-1}(\pi)$. We then tested the actual (in simulation) start time reliability achieved. The average of absolute error relative to the simulated sample frequency was 1.09%, with a range from -5% to 3.6%. When $\pi > 0.5$, the absolute values of the errors were in all cases less than 3%. The absolute errors decreased as π increased, consistent with the above intuitive logic that normalcy obtains for large π.

Given the statistics of the ending time of each procedure, the computation of T^* and $Profit^*$ requires only the long-run average residency distribution p_i. This is computed from matrix multiplications with the appropriate transition matrix as is usual for Markov chains. The approximation was coded in Visual Basic using an Excel spreadsheet interface. It runs very quickly and is easily used for exploration of the efficient frontier.

6. The efficient frontier and process improvements

6.1 The efficient frontier

Figure 1a shows the efficient frontier in three dimensions ($W, \pi, Profit^*$) for our client hospital, computed using the results in sections 4 and 5. The first notable feature of the

14

efficient frontier is that at all reasonable values of n the delay to get on schedule is already near the minimum one-half day for a bulk service queue and it is financially very costly to reduce the wait even a fraction of a day. This argues strongly for operating on the highest profit profile in figure 1a, on which $W = .55$ days in queue and $Profit^*$ and π trade off as shown in figure 1b.

Data is not available for actual delays to get on schedule at our client hospital, but anecdotal evidence suggests that this is about 10 working days. Why is this, given our model's prediction? Scheduling at the hospital is actually done in two phases. Each surgical service (e.g. cardiac, neurosurgery, thoracic, urology, etc.) is "given" a room or set of rooms for certain days of the week to schedule as they see fit. This allocated time is called their "block time." If a service has more cases than it can schedule in its available block time, the surgeon can place the patient on a wait list. The power over the schedule for a given day (Wednesday, say) shifts from the surgical services to OR administrators in the morning of the previous day (Tuesday). The OR administrators attempt to bin-pack the cases on the wait list into any rooms that have not been fully utilized by the services. If this occurs as planned, the net effect should be very close to the bulk service queue model. Indeed, our model was crafted after speaking with hospital staff regarding scheduling practices. Naturally, we needed to explain the difference between the .5 days predicted by the model and the anecdotal evidence from schedulers.

We generated a detailed discrete-event simulation of hospital operations (which included different services and procedure lengths, block times, etc.) to compare with our simple model. This revealed the following. The detailed simulation generated a wait to get on schedule of 9.66 days without the wait room and .7 days with the wait room, providing good support for our simpler model and evidence that the difference was due to lower wait room use. We reported to the hospital that there is enough theoretical capacity in the system to greatly reduce wait times (with their current case load) if the wait room is used as designed. We also looked into reasons why the wait room may not be used and found differences among services. Oncology tended to be the most aggressive in the use of the wait room, probably because delays can be life-threatening for some of their patients. However, this indicates that even complex operations may be handled via the wait room. Surgeons suggested that the wait room was less desirable because they would not know for

15

sure until the day before whether they are on schedule, and that they preferred a familiar room (nurse/staff teams tend to stay with rooms and not surgeons) to an unfamiliar one. These frictions are real, and it is not clear how easy it will be to change behaviors. We reported to the hospital that if the wait room were more aggressively used they could reduce the wait to get on schedule significantly, with their current case load, with no other changes in the system.

The second notable feature of the efficient frontier is the dramatic tradeoff between schedule reliability and hospital profits. The decreased profits come from the addition of buffer time between procedures to increase the reliability of start times. The "optimal" reliability level is not obvious. The cost to the hospital for increasing π can be millions of dollars each year. Yet, the anecdotal evidence in the literature and gathered on-site by the authors suggests that lack of schedule reliability is a major source of frustration for surgeons and surgical staff, and decreases their overall productivity. We will return to this topic in section 7 below.

6.2 Process improvements

Process improvements (reduction in mean times or variability, for example) would shift the efficient frontier out to more preferred levels of $(W, \pi, Profit^*)$. The magnitude of the potential gains for each type of improvement can be estimated by assessing the benefits for process improvements along any one of these dimensions (e.g. profits), holding the other two constant. Also, how benefits are actually allocated to the various stakeholders is a policy decision with motivational consequences. The spreadsheet model provides the set of efficient allocations. Much could be reported here, but will not be due to space limitations. We stress, however, that the most likely individuals to engage a process improvement effort would be surgeons and surgical staff. Yet, these individuals would see little near-term personal benefit if attending to improvements did not relieve their own stresses, but rather flowed directly to the hospital's bottom line. For example, a reduction in the CV of procedure length from .4 to .3 can reduce costs by $360 thousand annually, or can be parlayed into a 11% increase in the reliability of on-time starts, or some combination of these.

To go further in recommending where on the efficient frontier the hospital may wish to

operate, we need some sense of what sorts of trade-offs are acceptable to the important stakeholders. That is, we need a sense of stakeholder utility. This is discussed in the following section.

7. Participation-inducing proposals for capacity enhancements

An increase in caseload is expected in the coming years that will necessitate some capacity expansion. Potential avenues to increased capacity include building new OR's or using the existing OR's more extensively. Specifically, the OR's currently operate about 8-10 hours each day but are then idle (except for emergency cases) in the evening. Rather than extend the operating day, however, the hospital has proposed spending $6 million to build two new operating rooms. When we asked why they would do this with a lot of expensive rooms lying idle for the evening hours, we were told that there is a lot of resistance from surgeons and staff to working during those hours. This is for a variety of reasons, including family (quality of life) considerations and the known fact that anything scheduled late in the day inherits all of the variability (in start and ending times) that has accumulated over the entire day. This resistance was real, but of unknown depth and ferocity. Could the staff be convinced to work late if the start time uncertainty is mitigated? Could they be enticed to work late with bonus pay? If so, what is the profit-maximizing combination of bonus pay and promised start time reliability sufficient to entice surgeons and staff to perform later procedures? Would the resulting contract save money for the hospital relative to the planned new construction? To answer these questions, we need to combine hospital financial considerations with the willingness of surgeons and surgical staff to work evening hours. The following section describes a method for doing this using estimated utility functions. The estimation of those functions can be difficult, but some suggestions and an illustrative example are presented in section 8.

7.1 Professional categories and utility functions

In this section we assume that we have fixed W, and that $Profit^*(\pi)$ is a nonincreasing, concave and continuously differentiable function of π. Such a function fits our client's data very well. Suppose there are K professional categories (henceforth PC's, for example nurses, surgeons, anesthesiologists, etc.) required to staff an operating room, and let N_k denote the number of members of each PC required to staff one OR. The hospital is

considering extending the workday for m_e operating rooms, but needs to do so in a way that appeals not just to hospital profits but also to the preferences of the hospital staff. For a given shift of work, we will assume that staff members will have preference orderings consistent with a linear utility function of pay, day versus late shift work, and the reliability of their start times. That is, we assume that members of PC k will have a utility function of the form

$$u_k(\delta(PM), \pi, b) = \gamma_k \delta(PM) + \alpha_k \pi + \beta_k b_k$$

where b_k is the bonus (\$/shift) paid to PC k, π is the start time reliability, and $\delta(PM)$ is an indicator function that equals one if the shift is in the evening hours and equals zero otherwise. α_k, β_k and γ_k are utility function coefficients to be estimated for each PC. In our application, described below, we employed conjoint analysis (c.f. Green and Srinivasan 1990) for this task.

Since the 1970's conjoint analysis has been the most popular market research technique for estimating multi-attribute utility functions, and there have been many commercial applications of this technique in an array of industries. Practical applications need to balance two influences. The first is a desire to estimate a complicated model, with many parameters, because many human choices are in fact complex. The second influence, however, is that the data collection effort increases in difficulty with the number of parameters to estimate, and overly complex tasks invite respondent fatigue and can decrease the validity of the procedure. The additive, linear form for the utility function assumed above responds to the need for efficient data collection efforts. This is a serious practical issue with health care professionals, who are often difficult or expensive to access. The separable form is well-supported in the marketing literature (c.f. Akaah and Korgaonkar 1983, Green 1984), where it has been found that including interaction effects can actually decrease the predictive power of the results for the reasons alluded to above. The linear form, however, need not be appropriate in all applications and should be verified with data. In section 7.2 below, we also investigate how robust the answer will be to misestimation of parameters in the utility model.

If the hospital builds new OR's and adds a shift of work during normal hours with normal operating procedures we assume that no bonus is paid and that the status quo start time reliability (which we will denote by π_0) will be realized. If the hospital does not build a

new OR but instead expands capacity by extending operations into the evening hours, it will have to do so in such a way that all necessary staff will willingly participate. That is, the hospital will have to offer a combination of bonus and start time reliability such that $u_k(1, \pi, b) \geq u_k(0, \pi_0, 0)$ for all PC's k (these are called "individual rationality" constraints in the principal-agent literature). Hence, the profit-maximizing participation-inducing combination of bonus pay and start time reliability can be found with the following math program, which we will refer to as PP (for Profit and Participation):

PP: $\qquad\qquad$ $\text{MAX}_{\pi, b_k} Profit^*(\pi) - m_e \sum_{k=1}^{K} N_k b_k$

Subject to $\qquad\qquad$ $u_k(1, \pi, b_k) \geq u_k(0, \pi_0, 0)$ for all k

$\qquad\qquad\qquad\quad$ $b_k \geq 0$ for all k

$\qquad\qquad\qquad\quad$ $0 \leq \pi \leq 1$

For any fixed π the optimal bonus is easily seen to be $b_k^*(\pi) = max\{0, -(\gamma_k/\beta_k) - (\alpha_k/\beta_k)(\pi - \pi_0)\}$. $-\gamma_k/\beta_k$ is the bonus amount that would be required to entice members of PC k into the evening shift if there were no change in start time reliability. If there is negative utility for the late shift γ_k will be negative and the needed bonus is positive. The required bonus descends linearly in π until it hits zero at $\hat{\pi}_k = \pi_0 - \gamma_k/\alpha_k$. Hence, $b_k^*(\pi)$ is piecewise linear and convex in π.

Substituting $b_k^*(\pi)$ into the objective function in PP reduces the problem to one of maximizing a concave (but not everywhere differentiable) function on the interval $\pi \in [0, 1]$. This can be solved using "subgradient" logic as in Rockafellar (1970, chapter 24), and despite some cumbersome notation is as conceptually straightforward as setting a derivative to zero to maximize a concave function.

Recall that $\hat{\pi}_k = \pi_0 - \gamma_k/\alpha_k$ is the minimal start time reliability π at which members of PC k will work the late shift with no additional bonus. Below $\hat{\pi}_k$ an additional bonus is required, and above $\hat{\pi}_k$ nothing additional is required. $\hat{\pi}_k$ is also the breakpoint of the derivative of $b_k^*(\pi)$. We begin by ordering the PC's by their $\hat{\pi}_k$ values. That is, let $k(1)$ be the PC with the lowest $\hat{\pi}_k$, $k(2)$ the next lowest, and so forth. Hence, we will have

$$0 \leq \hat{\pi}_{k(1)} \leq \hat{\pi}_{k(2)} \leq ... \leq \hat{\pi}_{k(K)}.$$

These are the breakpoints of the derivative of the objective function in PP. Define *jmax* to be the largest j such that $\hat{\pi}_{k(jmax)} \leq 1$. *jmax* may equal K or it may be less. If

jmax < *K* then there are some PC's that will require an additional bonus to work the late shift even if start time reliability is perfect. The only breakpoints we need worry about in PP are $0 \leq \hat{\pi}_{k(1)} \leq \hat{\pi}_{k(2)} \leq ... \leq \hat{\pi}_{k(jmax)} \leq 1$.

A subgradient is a natural generalization of a gradient for continuous but nondifferentiable functions. Denote the left-hand derivative of the objective function at a point π on $[0,1]$ by $D^-(\pi)$ and the right-hand derivative by $D^+(\pi)$. $D^-(\pi) = D^+(\pi)$ between breakpoints because $Profit^*$ is assumed to be differentiable, and $D^-(\pi) > D^+(\pi)$ at a breakpoint because we have dropped the bonus term for one of the PC's as we cross that point. We extend the objective function to the entire real line by defining it to be negative infinity for $\pi < 0$ or $\pi > 1$, so that the left-hand derivative at $\pi = 0$ is $+\infty$ and the right-hand derivative at $\pi = 1$ is $-\infty$. The subgradient of the objective function at any π is the set $\{x \in R : D^-(\pi) \leq x \leq D^+(\pi)\}$. A necessary and sufficient condition of π^* to solve PP is that zero be in the subgradient of the objective function at π^*. Such a point is guaranteed to exist because the graph of the subgradients descends (with some vertical segments) from $+\infty$ to $-\infty$ as π ranges from 0 to 1. Specifically, we have the one-sided derivatives shown in figure 2a.

Since the subgradient graph is descending, the optimal point is easily found by starting at $\pi = 0$ and finding the first breakpoint at which $D^+ \leq 0$. If at this point $D^- \geq 0$ that breakpoint is optimal. If $D^- < 0$ at that breakpoint then the optimal solution lies between it and the previous breakpoint, which is found by setting the derivative appropriate for that interval to zero. We make this formal in the following proposition. The proof follows directly from the above discussion and the results in Rockafellar and is omitted.

Proposition 5: If $Profit^*(\pi)$ is concave, continuously differentiable and nonincreasing then (a) For any fixed π, $b_k^*(\pi) = max\{0, (-\gamma_k/\beta_k) - (\alpha_k/\beta_k)(\pi - \pi_0)\}$ is the optimal bonus for PC k, and (b) an optimal π^* is any π such that zero is in the subgradient of the objective function, and can be found as described above.

Once the profit-maximizing manner in which to extend into the evening hours is determined, the results can be compared with the cost of building new OR's to indicate the most financially efficient way for the hospital to expand its capacity.

7.2 Sensitivity analysis

The method outlined in section 7.1 requires utility function estimates for each relevant PC. Getting marketing data from doctors and surgeons is often difficult and expensive, as any pharmaceutical or medical supply company can attest. There will typically be practical problems accessing all of the desired PC's in an efficient manner, and there are always potential estimation problems in real applications. Hence, we are interested in how robust the indications of program PP are to these real estimation problems.

First, note that start time reliability is a public good in that it is provided (and paid for) once and then enjoyed by all. Likewise, the social benefits of improved start time reliability are proportional to the population of affected hospital employees. The situation is different with bonus pay, which must be paid to each individual independently. Hence, the more employees that are involved with procedures the greater the incentive to use π instead of bonus in a participation-inducing contract, and the higher an optimal π^* will be. Consequently, a lower bound on the optimal π can be generated by dropping PC's from the analysis. We make this formal in the following, which is proved in the appendix.

Corollary 5.1: let J and $\tilde{J}$ be two sets of PC's and let π^* and $\tilde{\pi}^*$ be the optimal reliabilities in program PP for J and $\tilde{J}$, respectively. If $\tilde{J} \subseteq J$ then $\tilde{\pi}^* \leq \pi^*$. Further, if $\tilde{J} = \{k(K)\}$ and $\tilde{\pi}^* = \hat{\pi}_{k(K)}$ then $\pi^* = \tilde{\pi}^*$ for any J (e.g. even if all relevant PC's are considered).

Hence, if we analyze PP with just one PC, surgeons say, the resulting reliability level is always a lower bound on the optimal level considering all PC's. Also, the hospital will not want to increase π more than required to satisfy the most demanding (highest $\hat{\pi}$) PC. So, the second part of Corollary 5.1 claims that the entire problem is solved if the single (most demanding) PC problem solves at its no-bonus level. For example, suppose that surgeons (PC j, say) are the most demanding PC and the optimal solution considering only surgeons is to set $\pi^* = \hat{\pi}_j$. Then, $\pi^* = \hat{\pi}_j$ and paying no bonuses to any PC remains the optimal solution regardless of how many PC's we add to the analysis, provided none are more demanding than surgeons. In fact, all PC's other than surgeons will receive surplus utility beyond their point of indifference. Consequently, if one can identify or intuit the most demanding PC, a lot can be said about the general problem.

There are also, inevitably, estimation problems with the utility functions due to practical matters of correctly framing the questions and experimental design. Hence, we may be

interested in how our conclusions change, or not, as a result of changes in the estimated parameters. What is the effect on the overall desirability of expanding hours in the existing OR's relative to building new OR's? What is the effect on the optimal contract for expanding hours? The following answers to these questions are justified in the appendix. We continue to assume that $Profit^*(\pi)$ is concave, continuously differentiable and nonincreasing.

First, as long as $\pi^* > \hat{\pi}_k$ then PC k is not involved in any way in the optimality conditions for PP. Hence, the optimal profits and contract are robust to any changes in any of the coefficients of the utility function for PC k. In addition, we can say the following about the sensitivity of profits and contracting conditions to changes in utility function coefficients. We assume that $\gamma_k \leq 0$ for all k, indicating a disutility for evening shift work, because any PC with $\gamma_k > 0$ will volunteer for the late shift with no additional bonus or reliability enhancements and can be omitted from the analysis.

Effect of γ_k

Increasing γ_k (making it less negative) means that PC k is less resistant to working evening hours. This cannot increase the bonus paid to PC k at any specific π, and so cannot decrease optimal profits. Hence, the desirability of expanding into evening hours versus building new OR's is enhanced. Also, $\hat{\pi}_k$ (the no-bonus level of reliability for PC k) is decreasing in γ_k. The effect on the optimal contract can be either none at all (if the current and future solution has $\pi^* > \hat{\pi}_k$) or a potential decrease in the optimal π^*. Decreasing γ_k (more disutility for the evening shift) has the opposite effects, reducing the profitability and desirability of expanding hours but also increasing the use of reliability (increasing π^*) in the optimal contract.

Effect of α_k

Increasing α_k means that start time reliability increases in importance to PC k. This can never increase the bonus payments $b_k^*(\pi)$ (utility is enhanced for any level of π and so less bonus is required) or $\hat{\pi}_k$ (the level of π at which no bonus is required). Hence, profits cannot decrease and the desirability of expanding into evening hours versus building new is enhanced. The effect on the optimal contract is ambiguous. As long as $\hat{\pi}_k < \pi^*$ there

is no change, as before. If $\pi^* = \hat{\pi}_k$ then the optimal π^* may decline but will never fall below the new (lowered) "no bonus" level for class k. If $\hat{\pi}_k > \pi^*$ then π^* is nondecreasing in α_k. Decreasing α_k has the opposite effects, reducing the profitability and hence the desirability of expanding hours. However, the profit cannot drop below the lower bound of $Profit(\pi_0) - m_e \sum_k N_k(\gamma_k/\beta_k)$ which is independent of α_k and results from enticing all PC's into the evening shift with bonus pay only and no appeal to enhanced start times.

Effect of β_k

Increasing β_k means that PC k has a higher utility for money. In this case, $b_k^*(\pi)$ (the bonus paid to PC k) is nonincreasing so the optimal profits are nondecreasing. Hence, the desirability of expanding hours is enhanced. $\hat{\pi}_k$ (the no-bonus reliability level for PC k) is unaffected. This last observation may be surprising. The intuition is as follows. Changes in β for any reason will change our estimates for γ/β (the willingness to pay to avoid late shift work) and α/β (the willingness to pay for schedule reliability). However the ratio $(\gamma/\beta)/(\alpha/\beta) = \gamma/\alpha$ (willingness to trade off late shift work for reliability) is unaffected. Changing β is like changing the medium of exchange in an economy (e.g. from dollars to euros) which will not affect the ratio of the value of two goods (here, late shift work and reliability). As β_k increases, π^* is nonincreasing. As long as $\pi^* > \hat{\pi}_k$ then as before the optimal profits and contract are invariant to any changes in β_k. Otherwise, π^* may decrease as β_k increases. The opposite happens as β_k decreases, so that profits and the desirability of expanding into evening hours decline. As β_k decreases less money (and more π) is used to entice participation, and in particular if $\hat{\pi}_k$ is feasible (≤ 1) and $\beta_k \leq -\frac{m_e N_k \alpha_k}{Profit'(\hat{\pi}_k)}$ then an optimal solution to PP will have $\pi^* \geq \hat{\pi}_k$ and $b_k^* = 0$. In fact, if that condition holds for the most demanding PC ($k(K)$) then setting $\pi^* = \hat{\pi}_k$ and $b_i^* = 0$ for all PC's i is optimal in PP. In all cases, since the $\hat{\pi}$ values are unaffected by changes in β, it will always be true that $Profit(\hat{\pi}_{k(K)})$ (with no bonus paid to any PC) will be a lower bound on optimal profits, if it is feasible. This lower bound is independent of β.

In the next section, some of these sensitivity results are leveraged in our applied context.

8. Application

8a. Analysis and recommendations

The primary author teaches in an executive eduction program that introduces physicians, surgeons and hospital staff to some of the financial realities of business. He took the opportunity during several course offerings to ask the participants to execute a conjoint analysis experiment to reveal their preferences for money, evening work, and start time reliability. Respondents were told that there is the potential to reorganize a shift in the OR suite. The new time might be during the regular day, or in an afternoon/evening shift. Surgeons and staff that work in this time slot may receive a bonus in pay per shift worked, and will experience a given reliability of procedure start times. Respondents were asked to consider a series of concepts, each representing some combination of timing (day or evening), bonus, and reliability. Respondents were asked to rank the concepts from their most preferred to least preferred. From the manner in which each respondent orders the concepts, his/her individual utility function coefficients can be estimated (c.f. Srinivasan and Shocker 1973).

A total of 88 individuals sorted cards in two course offerings, but many of the respondents were physicians (not surgeons) whose preferences would not be directly affected by the capacity enhancement decision. The sample relevant for OR capacity management included 30 surgeons, 8 nurses and 6 nurse managers. The sample of surgeons is gratifyingly large, given typical difficulties accessing surgeons, and this is arguably the most important group to consider (from an economic perspective) because they are the most mobile and least easily replaced. The results for the 14 nurses or nurse managers are suggestive but potentially unrepresentative. Also, due to potential selection bias (do typical surgeons and nurses show up for such courses?) and some unavoidable contextual influences (the author's class presentation was on the effects of variability in OR's, which may temporarily influence respondents' attitudes) this experiment should be considered illustrative and suggestive, but not final. We investigate the robustness of our conclusions to misestimation errors below.

The first consistent impression to emerge from the data is that start time reliability is as important as bonus and shift timing in determining the overall utility for working conditions. This was consistent across all samples and PC's.

To illustrate the use of the results in section 7, consider three PC's: surgeons (PC 1), nurses (PC 2) and nurse managers (PC 3). The estimated linear utility functions from the

24

conjoint surveys for these three groups are as follows:

$$u_1(\delta(PM), \pi, b_1) = -37.1\delta(PM) + 83.5\pi + .023b_1$$

$$u_2(\delta(PM), \pi, b_2) = -13.1\delta(PM) + 93.8\pi + .026b_2$$

$$u_3(\delta(PM), \pi, b_3) = 8.07\delta(PM) + 119.6\pi + .033b_3$$

We can eliminate nurse managers from the PP problem because they prefer late shift work and need no bonus to convince them to sign on. The "no bonus" breakpoints for surgeons and nurses are $\hat{\pi}_1 = .94$ and $\hat{\pi}_2 = .64$, respectively.

The hospital is expecting increases in case load in the future and is considering ways to add capacity to accomodate that load. The current proposal is to build two new operating rooms at \$3 million each, and the caseload mix that is anticipated for that enhanced capacity has been determined. We would like to consider the altnernative of extending the hours in two existing rooms with a participation-inducing contract. Whether or not daytime frictions in the current block time structure can be reduced, new capacity available via new rooms or extended hours will probably not be assigned to any one service, at least initially, but rather shared by all. This means that any procedure can go to any new room, consistent with our model.

We assume that each case requires one surgeon and four nurses ($N_1 = 1$, $N_2 = 4$), and that the current start time reliability is $\pi_0 = .5$. The profit function for operating two rooms ($m_e = 2$) in the hospital can be found as in section 5, and results in the following regressed model ($R^2 = .97$) of profit (\$/shift) as a function of start time reliability:

$$Profit = 9,708 + 2,039\pi - 3,569\pi^2.$$

To maintain a nonincreasing function we will use $Profit'(\pi) = \{0 \wedge 2,039 - 7,138\pi^2\}$. Starting at $\pi = 0$ the directional derivatives are as shown in figure 2b.

The optimal solution is at the breakpoint $\pi^* = \hat{\pi}_1 = .94$ where the surgeons are just indifferent between late shift work and regular time work, without any bonus. In this case, the cost to the hospital to increase schedule reliability sufficiently to encourage the embrace of late shift work is less than the bonus required to entice these workers to work a late

shift at the old reliability $\pi_0 = .5$. Since no bonus need be paid, the profit in thousands of dollars per year that the hospital will enjoy for extending the hours in two operating rooms is $Profit(.94) = 9,708 + 2,039(.94) - 3,569(.94)^2 = 8,471$ \$/shift or about \$2.2 million per year. $Profit(\pi_0 = .5) = 9,835$ \$/shift (about \$2.56 million per year) but would require $2(1600) + 8(508) = 7,264$ \$/shift in bonus pay so clearly increasing π is preferable to paying bonuses to entice shift work.

If the hospital builds two new OR's and maintains its current schedule reliability, the profit is $Profit(.5) = \$2.56$ million per year. However, construction costs are about \$6 million for the two new OR's. The hospital uses a 7.7% cost of capital, at which rate it is better to extend hours and increase schedule reliability (in the extended hours) for any planning horizon. That is, the extra \$.36M each year with the new OR's will never equal in net present value the \$6M construction costs with a 7.7% cost of capital. In addition, the decision to extend hours can be reversed (by cancelling the late shift) at any time, whereas investments in construction are irreversible. Hence, extending hours maintains greater financial flexibility.

This convenience sample included few respondents from some PC's (nurses and nurse managers) and omitted others (e.g. anesthesiologists) entirely. How robust is this solution to missing PC's and potential distortions in utility estimates? The solution is invariant to the addition or deletion of any number of PC's, or misestimation of their utility parameters, providing they are less demanding (lower $\hat{\pi}$) than surgeons (Corollary 5.1 and following discussion). However, the convenience sample, contextual features, and other problems common to estimating utilities may inject some error in the coefficient estimates for surgeons. How robust is our conclusion to these?

First, re-estimating more complex utility models for surgeons revealed that the assumption of linear preferences for bonus payments is appropriate for the range of bonuses used in the experiment. Individual surgeon utilities for start time reliability showed some curvature, but with a few exceptions were close to linear within the relevant range.

As for the effect of estimation errors with the linear utility form, if γ_1 increases (becomes less negative) we reinforce the advantages of extending hours over building new. If γ_1 decreases we reduce this advantage, but increase the presence of start time reliability in any

26

optimal contract. If α_1 increases we also reinforce the advantages of extending hours over building new. If α_1 decreases we reduce this advantage, but can do no worse than the "pure bonus" solution. In this case, maintaining individual rationality with pure bonuses would cost the hospital $1.89 million per year, making extended hours preferable to building new OR's for time horizons of 3.78 years or less. Longer horizons could change our conclusion, reinforcing the importance of leveraging the opportunity to appeal to staff members with start time reliabilities in considering these expansion decisions. If β_1 increases we reinforce the advantages of extending hours over building new. If β_1 decreases profits may decline but cannot drop below the "no bonus" contract. This is the solution suggested by PP, which dominates new construction over all planning horizons. In summary, as long as we assume that surgeons are the most demanding job category, the recommendation to seek a participation-inducing contract to extend operating hours in existing rooms rather than building new rooms is robust to the omission or misestimation of PC's other than surgeons. This conclusion is also robust to misestimation of surgeon attitudes toward wealth, or low biases in the other utility coefficients. High biases in the coefficients for reliability and timing can change our conclusion, but these same biases would reinforce the need to leverage start time reliabilities in any optimal expansion contract.

8b. What happened?

The impact of our recommendations on the hospital is unfolding as this paper goes to press. We introduced the novel question of what manner of contract would get surgeons and surgical staff to voluntarily work in the evening, what the components of that contract may look like, and in particular the high social value and public good nature of start time reliability as part of the mix. The hospital administration supports our recommendations. There is support, but also resistance, among surgeons and staff. The most vocal opponents base their arguments on contract credibility. Our recommended contract is one in which the strong aversion to unreliable start times is leveraged by providing a public good substituting for a lot of bilateral payments. This implies that the administration begin an evening shift but also provide enough "white space" to guarantee reliable start times. Some staff are skeptical. Given the cost/profit pressures on the hospital they believe that the administration will begin down the right path but over time the temptation to fill that "white space" with revenue-producing procedures will be too strong. They believe that, in

the long run, the surgical staff will be right back where they are now, with unreliable start times but with longer days. That is, the hospital will gain and capture all the rents, and the surgeons and surgical staff will lose. They are arguing for building new ORs, which of course reverses this allocation (the hospital pays the cost and not the surgeons and staff). These are legitimate concerns and the administration is aware that how they handle this situation will either aggravate or relieve the reputation effect in future negotiations.

The hospital administration is considering a pilot program in which they seek a participation-inducing contract to extend hours in a few OR's. We expect that some such contract will eventually be adopted. To overcome historical mistrust, the first iteration of this contract may substitute more cash for reliability than the pure form recommended by our model, which assumes that both sides deliver on their contracted promises. However, the issue of start times and how to manage them in concert with other costs will remain on the table, because even a significant bonus will not get surgeons in into the evening OR without some assurances as to when they can get home again. The spreadsheet model can rapidly assess scenarios relevant to this discussion. For example, the model suggests that extending hours dominates building new for any reliability greater than 63%, which would require bonuses to surgeons of $1214 per shift and to nurses of $107 per shift. But, once we depart from the no-bonus contract for the most demanding PC, we need to worry about PC's not included in the analysis (e.g. anesthesiologists). In this example, a 75% start time reliability can gain the participation of surgeons (with an additional bonus of $705 per shift) and nurses (no additional bonus required), dominate building new OR's, and still leave a bonus pool of $1135 per shift for other job classes.

References

Akaah, I. and P. Korgaonkar. Empirical Comparison of the Predictive Validity of Compositional, Decompositional, and Hybrid Multiattribute Preference Models. Journal of Marketing Research, May 1983, 187-197.

Baily, N., "A Study of Queues and Appointment Systems in Hospital Outpatient Departments," *J. R. Statist. Soc.* 14 (1952), 185-199.

Babad, Y., M. Dada, A. Saharia, "An Appointment-based Service Center with Guaranteed Service," European Journal of Operational Research 89 (1996) 246-258.

Brahimi, M. and D. Worthington, "Queueing Models for Out-patient Appointment Systems – A Case Study," *J. Opl. Res Soc* 42 (1991), 733-746.

Charnetski, J., "Scheduling Operating Room Surgical Procedures with Early and Late Completion Penalty Costs," *J. Operations Management* 5 (1984), 91-102.

Cohen, J. Mulitobjective Programming and Planning, Acdemic Press, N.Y. 1978.

Elandt, R., "The Folded Normal Distribution: Two Methods of Estimating Parameters from Moments," *Technometrics* 3 (1961), 551-562.

Fandel, G. and H. Hagemann, "Capacity Planning of Diagnosis Systems in Hospitals," *Engineering Costs and Production Econ.* 17 (1989), 205-221.

Gordon, T, S. Paul, A. Lyles, and J. Fontain. "Surgical Unit Time Utilization Review: Resource Utilization and Management Implications," *J. of Med. Systems* 12 (1988), 169-179.

Green, P. Hybrid Models for Conjoint Analysis: An Expository Review. Journal of Marketing Research 21, 1984, 155-169.

Green and Srinivasan, "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," *J. of Marketing* 54 (1990), 3-19.

Heyman, D. and M. Sobel, Stochastic Models in Operations Research v. I, McGraw-Hill, N.Y. 1982.

Heyman, D. and M. Sobel, Stochastic Models in Operations Research v. II, McGraw-Hill, N.Y. 1984.

Kim, S., I. Horowitz, K. Young, and T. Buckley, "Analysis of Capacity Management of the Intensive Care Unit in a Hospital," *Eur. J. of Opl. Res.* 115 (1999), 36-46.

Kuzdrall, P., N. Kwak and H. Schmitz, "Simulating Space Requirements and Scheduling Policies in a Hospital Surgical Suite," *Simulation*, May 1981, 163-171.

Ho, C. and H. Lau, "Minimizing Total Cost in Scheduling Outpatient Appointments," *Management Science* 12 (1992), 1750-1764.

Hopp, W. and M. Sturgis, "Quoting Manufacturing Due Dates Subject to a Service Level Constraint," *IIE Transactions* 32 (2000), 771-784.

Magerlein, J. and J. Martin, "Surgical Demand Scheduling: A Review," *Health Services Research*, Winter 1978, 419-431.

O'Keefe, R., "Investigating Outpatient Departments: Implementable Policies and Quali-

tative Approaches," *J. Opl. Res. Soc.* 36 (1985), 705-712.

Rockafellar, R. Convex Analysis. Princeton University Press, N.J. 1970.

Spearman, M. and R. Zhang, "Optimal Lead Time Policies," *Management Sci* 45 (1999), 290-295.

Srinivasan, V. and A. Shocker. Linear Programming Techniques for Multidimensional Analysis of Preferences. Psychometrika 38, 1973, 337-369.

Stoyan, D. Comparison Methods for Queues and other Stochastic Models, edited by D. Daley, John Wiley and Sons, N.Y. 1983.

Swanberg and Fahey, "More Operating Rooms or Better Use of Resources?", *Nursing Management* 14 (1983), 16-19.

Van Ackere, A, "Conflicting Interests in the Timing of Jobs," *Management Science* 36 (1990), 970-984.

Wein, L. "Due-date Setting and Priority Sequencing in a Multiclass M/G/1 Queue," *Management Sci* 37 (1991), 834-850.

Weiss, E. "Models for Determining Estimated Start Times and Case Orderings in Hospital Operating Rooms," *IIE Transactions* 22 (1990), 143-150.

Appendix: Propositions and proofs

Proof of Proposition 1: The proof is facilitated by the following facts: (1) If $t \geq t'$ then for any random variable X, $(X \vee t) \succeq_d (X \vee t')$; (2) For any random variables X and Y and any constant t, $X \preceq_d Y$ will imply that $(X \vee t) \preceq_d (Y \vee t)$; (3) If random variables X and Y satisfy $X \succeq_d Y$, then for any z, $F_X(z) \leq F_Y(z)$ and for any $\pi \in [0,1]$, $F_X^{-1}(\pi) \geq F_Y^{-1}(\pi)$ where $F_X^{-1}(\pi) := \sup_z \{F_X(z) \leq \pi.\}$ We use these in various places to prove Proposition 1. Assume inductively that $E_{k-1} \succeq_d E'_{k-1}$, then $t_k = F_{E_{k-1}}^{-1}(\pi) \geq F_{E'_{k-1}}^{-1}(\pi) \geq F_{E'_{k-1}}^{-1}(\pi') = t'_k$, implying $S_k \sim (E_{k-1} \vee t_k) \succeq_d (E'_{k-1} \vee t_k) \succeq_d (E'_{k-1} \vee t'_k) \sim S'_k$, implying $E_k \sim S_k + X_k \succeq_d S'_k + X_k \sim E'_k$ because $\succeq_d$ survives convolution. This completes the induction. For any random start time S_1, we initiate the induction with $E_1 \sim S_1 + X_1 \sim E'_1$ so that $E_1 \succeq_d E'_1$ trivially. Hence we will have $t_k \geq t'_k$, $S_k \succeq_d S'_k$, and $E_k \succeq_d E'_k$ for all k, completing the proof for parts (a), (b) and (c). To prove (e) we have that $Cost(T) = C_r T + (C_r + C_{ot}) \sum_{i=1}^n p_i \int_T^\infty (x - T) dF_{E_i}(x) = C_r T + (C_r + C_{ot}) \sum_{i=1}^n p_i \int_{-\infty}^\infty (x - T)^+ dF_{E_i}(x)$. Since the function $(x - T)^+$ is nondecreasing in x,

$E_i \succeq_d E_i'$ will imply that $\int_{-\infty}^{\infty}(x-T)^+ dF_{E_i}(x) \geq \int_{-\infty}^{\infty}(x-T)^+ dF_{E_i'}(x)$ for all i and any fixed T. Hence, for any fixed T we have $Profit(T) = R\lambda - mCost(T) \leq R\lambda - mCost'(T) = Profit'(T)$. This implies that $Profit(T) \leq \max{}_T Profit'(T) = Profit'^*$ for all T, and hence $Profit^* = \max{}_T Profit(T) \leq Profit'^*$. This completes the proof for part (e). To prove part (d), note that $E_i \succeq_d E_i'$ for all i will imply that $F_{E_i}(T) \leq F_{E_i'}(T)$ for any i and T. Hence,

$$\sum_{i=1}^{n} p_i(1 - F_{E_i}(T^*)) = C_r/(C_r + C_{ot}) \geq \sum_{i=1}^{n} p_i(1 - F_{E_i'}(T^*))$$

and so either $T^* = T'^*$ or to recapture optimality (set the sum equal to $C_r/(C_r + C_{ot})$) we need to choose a $T'^* \leq T^*$.

Proof of Proposition 2: To prove (a), note that $I_{t+1} = (I_t + A_t - nm)^+$. Letting $n > n'$ an induction proof shows that for any sample path $\{A_t\}$, $I_t \leq I_t'$ for all t. A similar argument works for Q_t. The average wait experienced by an arriving patient cannot decrease as I_t increases, proving (b). To prove (c) we use the following facts: (1) For any n, $M(n)$ is IFR and if $n > n'$ then $M(n) \preceq_d M(n')$; (2) If $q \preceq_d q'$ then for any IFR transition matrix M, $qM \preceq_d q'M$; and (3) If $M \preceq_d M'$ then for any q, $qM \preceq_d qM'$. If $\lambda < nm$ the countable state Markov chain $\{Q_t(n)\}$ with transition matrix M (omitting the n for notational convenience) has a stationary distribution q that equals the limit of the sequence $\{q_t\}$ generated by $q_t = q_{t-1}M$, starting from any q_0. Assume inductively that $q_{t-1} \preceq_d q_{t-1}'$. Then $q_t = q_{t-1}M \preceq_d q_{t-1}'M \preceq_d q_{t-1}'M' = q_t'$. The induction can be initiated with any $q_0 = q_0'$ and taking limits completes the argument. An arrival has either been admitted into a room or is still in queue, so part (a) implies that along any sample path the cumulative number admitted (C_t say) into rooms up to any time t is greater with n than n'. Hence, the average number admitted each day ($lim_{t \to \infty} C_t/t$) is higher with n than n'. In the long run $(1/m)$ of these will be in any specific room, completing the proof of (d). To prove (e) we explicitly use the fact that the scheduler loads rooms evenly (e.g. rotating through the rooms) to "split" the Poisson arrivals into m streams, one for each room. The queue for each room is now an $M/D/1$ bulk service queue with capacity n and arrival rate λ/m per day. We reinterpret Q_t, q, etc. accordingly, and can reproduce parts (a) and (c) in this setting. The cdf for the number of patients in residence in one room satisfies $F_p(k) = 1$ if $k \geq n$ and equals $F_q(k)$ if $k < n$. Now, consider $n > n'$. Since $q \preceq_d q'$ we have $F_q \geq F_{q'}$. So, $F_{p'}(k) \leq F_p(k)$ for $k \leq n$ and $F_{p'}(k) = 1 \geq F_p(k)$ for $k > n$.

Hence, the cdf's "cross" exactly once and, together with the fact that the $E(p) \geq E(p')$ these distributions satisfy the "cut criterion" for the $\preceq_c$ partial order (c.f. Stoyan 1983). This completes the proof.

Proof of Proposition 3: $Profit(T) = R\lambda - m[C_r T + (C_r + C_{ot}) \sum_i p_i(n) G_i(T)]$ which will be decreasing in n if $G_i(T)$ is convex in i, because p is convexly stochastically increasing in n. This proves part (a). To prove (b), we have by hypothesis that $S_{i+1} \sim E_i$ and $E_{i+1} \sim S_{i+1} + X$ for all i. Let $f(z) = (z - T)^+$ so that $G_i(T) = \int_{-\infty}^{+\infty} f(z) dF_{E_i}(z)$. Then

$G_{i+1}(T) - G_i(T) = \int_{-\infty}^{+\infty} f(z) dF_{E_{i+1}}(z) - \int_{-\infty}^{+\infty} f(z) dF_{E_i}(z)$

$= \int_{-\infty}^{+\infty} \int_0^{+\infty} f(z+x) \, dF_X(x) \, dF_{E_i}(z) - \int_{-\infty}^{+\infty} f(z) \, dF_{E_i}(z)$

$= \int_{-\infty}^{+\infty} \int_0^{+\infty} [f(z+x) - f(z)] \, dF_X(x) \, dF_{E_i}(z)$. Define $g(x,z) = f(z+x) - f(z)$, which is continuous in (x,z) because the f functions are, and $g(x,z) \geq 0$ because $f(z)$ is nondecreasing and $x \geq 0$. Thus, we are allowed to change the order of the integrals by Fubini's Theorem, to reveal $G_{i+1}(T) - G_i(T) = \int_0^{+\infty} \int_{-\infty}^{+\infty} [f(z+x) - f(z)] \, dF_{E_i}(z) \, dF_X(x)$. $f(z+x) - f(z)$ is nondecreasing in z because $f(z) = (z-T)^+$ is convex in z. Thus, $\int_{-\infty}^{+\infty} [f(z+x) - f(z)] \, dF_{E_i}(z)$ is nondecreasing in i because $E_i \preceq_d E_{i+1}$. Hence, $G_{i+1}(T) - G_i(T)$ is nondecreasing in i. To prove part (c), let $S_1 = 0$ without loss of generality, and exploiting the finite support of X we will have $t_{i+1} - t_i = K$ (a constant) for all i. So, $t_i = S_i = (i-1)K$. Now, $G_{i+1}(T) - G_i(T) = \int g_i(x) dF_X(x)$ where $g_i(x)$ is defined equal to $(iK + x - T)^+ - ((i-1)K + x - T)^+$. If g_i can be shown to be nondecreasing in i the proof is complete. But, $g_i(x) = 0$ for $X < T - iK$; $g_i(x) = iK + x - T$ for $T - iK \leq x \leq T - (i-1)K$; and $g_i(x) = K$ for $x \geq T - (i-1)K$. Direct inspection reveals this is nondecreasing in i.

Proof of Proposition 4: Given that r patients arrive during a day of length τ, the expected arrival time of the patients is $\tau/2$. This is intuitive, but can be shown rigorously using the fact that the arrival times $\{a_i\}$ have the same distribution as order statistics corresponding to r independent random variables uniformly distributed on the interval $(0, \tau)$ (Ross 1996 Theorem 2.3.1). When there are nm or fewer patients arriving during the day, all the patients will be placed on schedule at the start of the second day (time τ). That is, given that there are $r \leq nm$ patients arriving during the first day, the waiting time in queue for the i_{th} patient to arrive is $\tau - a_i$. The next nm patients to arrive will be placed on schedule at the start of the second day hence, so that patient i in that cohort will experience a wait in queue of $2\tau - a_i$, etc. Using this,

$$E(W_1) = \sum_{r=1}^{nm} E\Big[\frac{\sum_{i=1}^{A_1} \tau - a_i}{A_1} \mid A_1 = r\Big] Pr(A_1 = r)$$

$$+ \sum_{i=1}^{\infty} \sum_{j=1}^{nm} E\Big\{\frac{\sum_{k=1}^{nm}(\tau - a_k) + \sum_{k=nm+1}^{2nm}(2\tau - a_k) + \ldots + \sum_{k=inm+1}^{A_1}((i+1)\tau - a_k)}{A_1} \mid A_1 = inm + j\Big\} \times$$

$$Pr(A_1 = inm + j)$$

$$= \frac{\tau}{2} Pr(A_1 \leq nm) + \sum_{i=1}^{\infty} \sum_{j=1}^{nm} E\Big[\frac{\sum_{k=1}^{A_1}(\tau - a_k)}{A_1} + \frac{\sum_{k=nm+1}^{2nm} \tau + \ldots + \sum_{k=inm+1}^{A_1} i\tau}{A_1} \mid A_1 = inm + j\Big] \times P(A_1 = inm + j)$$

From here arithmetic manipulations provide the result.

Proof of Corollary 5.1: The objective function (sub)gradient is nondecreasing as we add PC's.

Proofs of sensitivity analysis claims:

All of the claims follow from the definitions of the various variables and some simple observations. For example, we have defined $\hat{\pi}_k = \pi_0 - \gamma_k/\alpha_k$; $b_k^*(\pi) = (0 \vee -\gamma_k/\beta_k - (\alpha_k/\beta_k)(\pi - \pi_0))$ and the objective function is $Profit^*(\pi) - m_e \sum_{kactive} N_k b_k^*(\pi)$ where *kactive* denotes those PC's with $\hat{\pi} > \pi$. The directional derivative of the objective function as we increase π is $Profit'^*(\pi) + m_e \sum_{kactive} N_k(\alpha_k/\beta_k)$. The optimal π^* is nondecreasing in this directional derivative. Parameter changes can change the derivative directly and/or change $\hat{\pi}$ which changes the active set of PCs. All of the directional claims regarding π^* follow from some combination of these considerations. Naturally, if for any π the optimal bonus $b^*(\pi)$ increases or decreases the optimal overall hospital profits will vary accordingly. Also, it is apparent that any specific policy can provide a lower bound on optimal profits. For example, both the "no bonus" ($\pi = \hat{\pi}_{k(K)}$ and $b_k^*(\pi) = 0$ for all k) policy and the "no reliability enhancements" ($\pi = \pi_0$ and $b_k^*(\pi) = -\gamma_k/\beta_k$ for all k) policy provide such bounds.

Figure 2a: One-sided Derivatives

$$D^-(0) = +\infty$$

$$D^+(0) = Profit'(0) + m \sum_{k=1}^{K} N_k(\alpha_k/\beta_k)$$

$$D^-(\pi) = D^+(\pi) = Profit'(\pi) + m \sum_{k=1}^{K} N_k(\alpha_k/\beta_k) \text{ for } 0 < \pi < \hat{\pi}_{k(1)}$$

$$D^-(\hat{\pi}_{k(1)}) = Profit'(\hat{\pi}_{k(1)}) + m \sum_{k=1}^{K} N_k(\alpha_k/\beta_k)$$

$$D^+(\hat{\pi}_{k(1)}) = Profit'(\hat{\pi}_{k(1)}) + m \sum_{k \neq k(1)}^{K} N_k(\alpha_k/\beta_k)$$

$$D^-(\pi) = D^+(\pi) = Profit'(\pi) + m \sum_{k \neq k(1)}^{K} N_k(\alpha_k/\beta_k) \text{ for } \hat{\pi}_{k(2)} < \pi < \hat{\pi}_{k(1)}$$

$$D^-(\hat{\pi}_{k(2)}) = Profit'(\hat{\pi}_{k(2)}) + m \sum_{k \notin \{k(1),k(2)\}}^{K} N_k(\alpha_k/\beta_k)$$

$$D^+(\hat{\pi}_{k(2)}) = Profit'(\hat{\pi}_{k(2)}) + m \sum_{k \notin \{k(1),k(2),k(3)\}}^{K} N_k(\alpha_k/\beta_k)$$

etc.

$$D^-(\hat{\pi}_{k(jmax)}) = Profit'(\hat{\pi}_{jmax}) + m \sum_{k \notin \{k(i):i=1,jmax-1\}}^{K} N_k(\alpha_k/\beta_k)$$

$$D^+(\hat{\pi}_{k(jmax)}) = Profit'(\hat{\pi}_{jmax}) + m \sum_{k \notin \{k(i):i=1,jmax\}}^{K} N_k(\alpha_k/\beta_k)$$

$$D^-(\pi) = D^+(\pi) = Profit'(\pi) + m \sum_{k \notin \{k(i):i=1,jmax\}}^{K} N_k(\alpha_k/\beta_k) \text{ for } \hat{\pi}_{k(jmax)} < \pi < 1$$

$$D^-(1) = Profit'(1) + m \sum_{k \notin \{k(i):i=1,jmax\}}^{K} N_k(\alpha_k/\beta_k)$$

$$D^+(1) = -\infty$$

Figure 2b: Example application

$$Profit'(\pi) = \{0 \wedge 2,039 - 7,138\pi\}, \ \alpha_1/\beta_1 = 3,639 \text{ and } \alpha_2/\beta_2 = 3,620.$$

$$m_e = 2, \ N_1 = 1 \text{ and } N_2 = 4.$$

$$D^-(0) = +\infty$$

$$D+(0) = 0 + 2[3,639 + 4(3,620)] = 36,238$$

$$D^-(\hat{\pi}_{k(1)}) = D^-(.64) = -2,529 + 2[3,639 + 4(3,620)] = 33,709$$

$$D^+(\hat{\pi}_{k(1)}) = D^+(.64) = -2,529 + 2[3,639] = 4,749$$

$$D^-(\hat{\pi}_{k(2)}) = D^-(.94) = -4,671 + 2[3,639] = 2,607$$
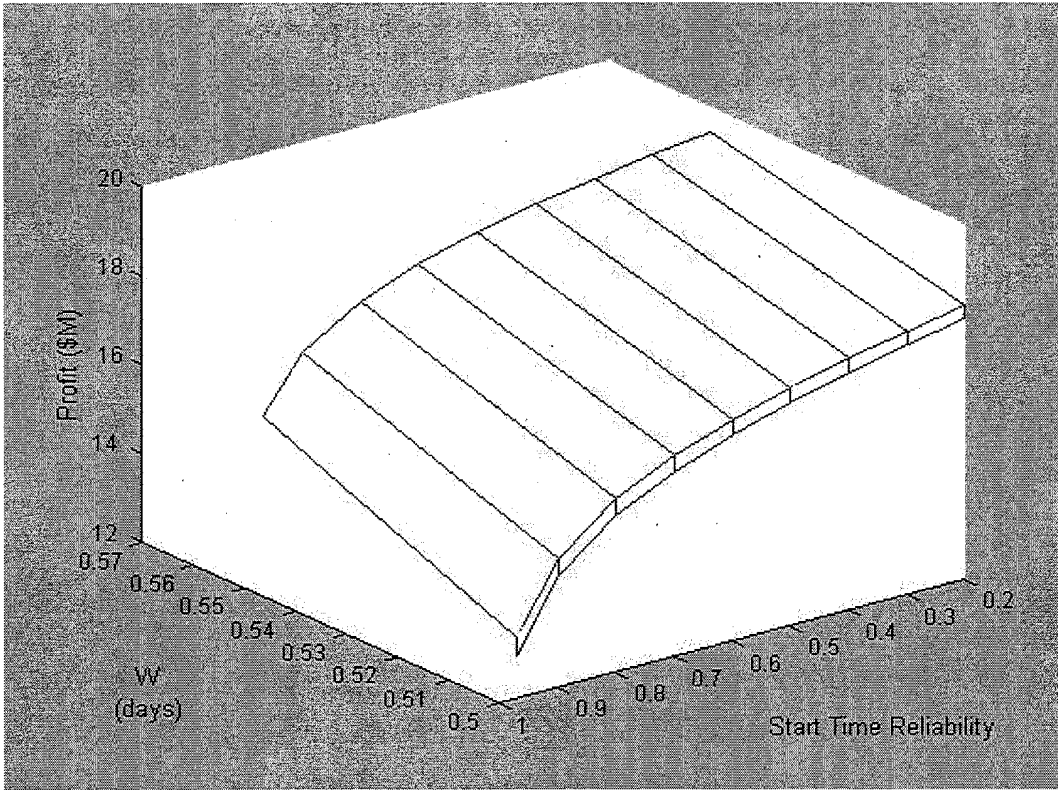
$$D^+(\hat{\pi}_{k(2)}) = D^+(.94) = -4,671 < 0$$
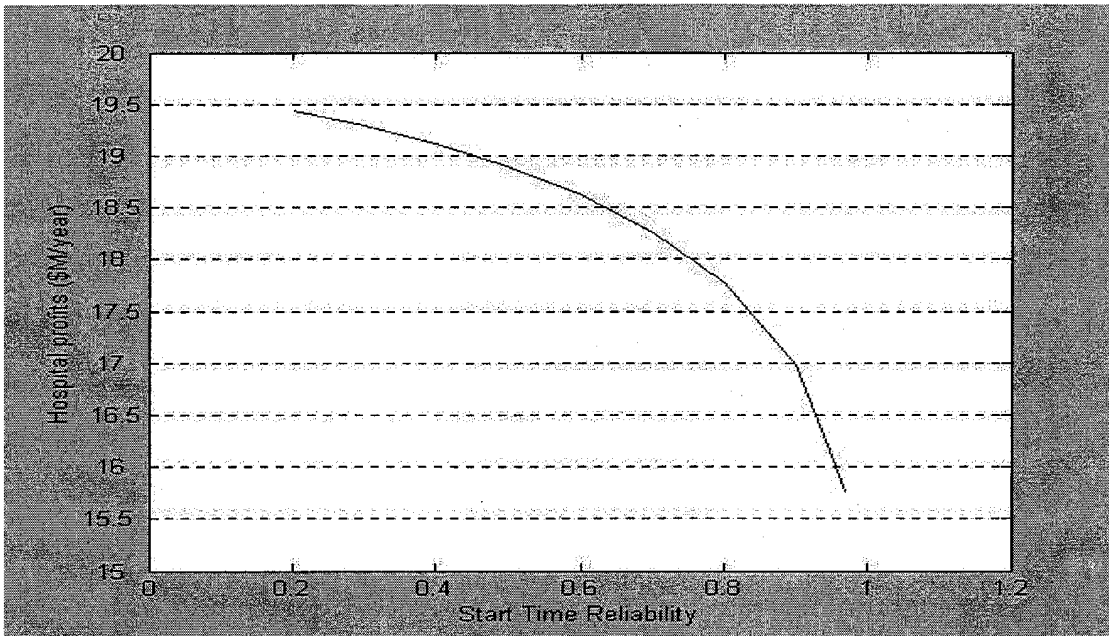
Figure 1a
The Efficient Frontier

Figure 1b
Profit vs Start Time Reliability
at W=.55 days