Reminder Systems for Reducing No-shows in General Practices

Leon Cui
Simon School of Business
University of Rochester
leon.cui@simon.rochester.edu

Vera Tilson
Simon School of Business
University of Rochester
vera.tilson@simon.rochester.edu

Gregory Dobson
Simon School of Business
University of Rochester
greg.dobson@simon.rochester.edu

ABSTRACT

Personal services businesses such as doctor’s offices operate based on an appointment schedule to maximize the utilization of the personnel and equipment. Some businesses deal with a no-show rate of customers that cuts into profits when it is not possible to charge the customer for the full impact of missed appointment. We develop a model to understand the investment and ongoing expense that the firm should put into such systems when customers act strategically. We also consider how reminder systems interact with overbooking policies.

1. Introduction

Missed appointments (no-shows) are a concern for many medical practices, since they lead to under-utilization of costly resources in the doctor’s office. While estimates of the frequency of no-shows vary, most sources put it between 7 and 12 percent [1]. A survey conducted in the U.K. in the year 2000 estimates that this translates into 17 million missed general practice appointments and 5.5 million missed practice nurse appointments, for a total estimated cost of $240 million [2].

Recent studies in medical journals have examined reasons cited by patients for missing their appointments [3]. The most common are forgetfulness of the patient, and mistakes or misunderstandings by the practice. Most doctors’ offices have tried to combat no-shows by using a reminder system to supplement the patients’ own efforts to attend the appointment. Yet, even with reminder systems, the no-show problem still persists. The obvious solution is to offer reminders more frequently, or to offer more elaborate reminders. Surprisingly though, a study by Neal, Hussain-Gambles, et al., found that phone call reminders from either nurses or doctors were less effective than postcard reminders, suggesting that more elaborate reminder systems do not necessarily reduce the no-show rate [3]. One possible explanation for a higher no show rate with more aggressive reminder systems is that patients start to expect reminders, and become lax in their own efforts to keep track of appointments. A similar effect was explored in Peltzman’s famous study on seat belts. Peltzman found that seat belts may actually decrease safety through an increase in reckless driving induced by moral hazard [4].

In this paper we examine a model that could be used to determine the cost-effectiveness of a reminder system. We consider a population of heterogeneous patients differentiated by how responsible they are in keeping their appointments. The introduction of a reminder system allows a practice to remind patients of their appointments. A reminder system is not guaranteed to reach a patient, either due to a technical flaw in the machinery of the system, or because the patient engages in irresponsible behavior that may cause the reminder system to be ineffective (for example, if a patient does not bother to check his phone messages). A patient does not show up for an appointment if his own efforts to keep the appointment fail, and if the reminder system fails to reach him. We examine the long-term effects of the reminder system, assuming that expecting a reminder causes a patient to put less effort into remembering his appointments. In our stylized model we use the Beta distribution to model heterogeneity in patients’ probability of showing up to an appointment without a reminder. Introduction of a reminder system shifts that distribution. We consider analytical models of such shifts and calculate the resulting probability distribution of the number of scheduled patients showing up for their appointments. We consider the decision of how many patients to schedule on a given day, and examine how the optimal decision is affected by the quality of the reminder system and patients’ levels of responsibility.

This paper is organized as follows. We first go through a literature review in Section 2, then present our model in Section 3. Next we consider a numerical example in Section 4 and point out several insights that we will prove rigorously in Section 5. We conclude the paper with two applications of our model. One is a method for clinics to predict how reminder systems will affect no-show rates, given past data. The other application examines how reminder systems can affect overbooking policies.
2. Literature Review

In this section, we will outline some of the techniques that have been used to model no-shows in the healthcare operations literature (a survey of the field is given in Cayirli et al. [5]). However, it is important to note that no-shows are not the focus of these papers. Rather, they are incorporated only as a part of an overarching model that optimizes patient scheduling in clinics. In the simplest case, authors have assumed a constant probability to show up (e.g. [6], [7], [8]). In an extension of this, some authors assign each patient to a group based on their characteristics, and then assign a constant probability of a no-show to each group, as in [9]. This is supported by empirical evidence [10], where the authors first divide patients by different traits (e.g., age, socioeconomic status) and then look at how each of these factors affect the no-show rate.

Some authors have taken a different approach of modelling no-shows as dependent on the waiting time of the patient from the time the patient schedules the appointment to the scheduled appointment time, which has been observed empirically (Gallucci [11]). Green and Savin [12] do this by explicitly assuming a functional form. They assume that no-shows are increasing in waiting time, so that patients who have a same-day appointment are less likely to skip it than patients who have scheduled their appointment well in advance. To obtain the specific functional form, the authors use best-fit methods on a data set obtained from a medical center. Another formulation is that of Liu, Ziya, and Kulkarni [13], who assign a type to each patient based on how many days the patient has until his appointment. Given the patient type, they then probabilistically assign a ‘cancelation date’ to the patient that may even be past the appointment date. If it is before the scheduled appointment, then the patient has canceled; otherwise, the patient has kept their appointment.

Our model differs from all of the ones mentioned above in that we take into account the effects of patient heterogeneity and learning effects.

3. Model

We first make some assumptions about the distribution of patients and how they behave in equilibrium. After outlining these assumptions, we will give our justifications that these assumptions are sensible. After explaining our assumptions, we will conclude this section by modeling the patient’s no-show behavior in equilibrium. In this model, we assume that showing up to an appointment is a function of both the patient’s own efforts and of the reminder system that has been implemented. The patient’s own efforts are modeled by \( \tau \), and the quality of the reminder system is modeled by \( q \). Here, the quality is solely a measure of how well the machinery of the system works. A higher quality reminder system is one that suffers from few technical or mechanical flaws that would somehow cause the reminder system to fail. For instance, a system where the office repeatedly calls the patient until he answers his phone will have a higher quality than a system where the office sends a reminder in the mail to the patient. In equilibrium, the patient will not show up if and only if both his own efforts and the reminder system fail.

1) Without a reminder system, each patient shows up to an appointment with a probability \( \tau \). Patients are heterogeneous and we assume \( \tau \) is distributed in the patient population according to a Beta distribution with parameters \( \alpha \) and \( \beta \). This captures the fact that some people are already extremely responsible on their own, while other people are less responsible. From the clinic’s point of view, this is a random variable since they cannot know a priori how responsible each patient is.

2) Now the doctor’s office faces a choice of which reminder system to employ. Specifically, they have to choose a system of quality \( q \). A more sophisticated reminder system that has a higher probability of reaching the customer will have a higher value of \( q \). With the implementation of the reminder system, in equilibrium the probability that the reminder successfully reaches the patient becomes \( h(\tau, q) \). We assume that this function \( h(\tau, q) \) is increasing in both \( \tau \) and \( q \).

3) Then, with the implementation of the reminder system, the patient now would show up to his appointment (due to his own efforts) with probability \( g(\tau, q) \) in equilibrium. It is important to note that this new function \( g \) is different from the parameter \( \tau \) that we have described earlier, since \( \tau \) describes how likely a patient shows up, given no reminder system. In contrast, \( g(\tau, q) \) describes how likely the patient is to show up in equilibrium, after the clinic has implemented a reminder system and the patient has grown accustomed to it. This function \( g(\tau, q) \) is decreasing in \( q \) and increasing in \( \tau \). Specifically, for a given person, we assume that he will put forth less of his own effort to show up to an appointment in response to a better reminder system. On the other hand, for a given reminder system, responsible people are still more likely to show up than irresponsible people.

Therefore in equilibrium the probability that a patient \( \tau \) does not show up for an appointment is given by

\[
(1 - g(\tau, q))(1 - h(\tau, q))
\]

which means that the patient’s own efforts failed and the reminder system failed. Here, we make the assumption that the effect of the patient’s own efforts is independent from the effect of the reminder system. This makes sense as, intuitively, the behavior of a patient should have no effect on how well the machinery of the reminder system works. Therefore the probability \( p \) that a \( \tau \) patient does show up for
an appointment is given by:

\[ p(\tau, q) = 1 - (1 - g(\tau, q))(1 - h(\tau, q)) \]

\[ = g(\tau, q) + h(\tau, q) - g(\tau, q)h(\tau, q) \quad (1) \]

Assuming that \( s \) appointments are scheduled, we are now interested in describing the distribution of \( S \), the number of patients that do show up. Keeping in mind that each patient’s type \( \tau \) is modeled as a separate draw from a random distribution, technically the distribution of \( S \) has become the sum of \( s \) Bernoulli trials, each with probability \( p(\tau, q) \) of success. However, as the lemma below shows, the distribution of \( S \) reduces to a binomial distribution with \( s \) trials and probability of success \( E[p(\tau, q)] \).

**Lemma:** Let \( S = \sum_{i=1}^{n} X_i \), where \( X_i \sim \text{Bernoulli}(p_i) \), and \( p_1, \ldots, p_n \) have the same density function \( f \) on \([0,1]\). Then \( S \) has the same distribution as a Binomial distribution with \( n \) trials and probability of success \( E[p] \).

**Proof:** Consider a single Bernoulli trial \( X_i \), with probability of success drawn as an instance of the random variable \( p \). Then \( P(X_i = 1|p_i) = p_i \). Conditioning on \( p_i \) gives us \( P(X_i = 1) = \int P(X_i = 1|p_i) f(p_i) dp = \int p f(p) dp = E[p] \).

Now we are interested in deriving the distribution of \( \sum X_i \). Now that we have established that each individual trial has the same probability of success \( E[p] \), it follows that the distribution of \( \sum X_i \) is binomially distributed with probability of success \( E[p] \) and number of trials \( n \).

In our model, \( X_i \sim \text{Bernoulli}(p_i(\tau, q)) \), where \( \tau \sim \text{Beta}(\alpha, \beta) \).

**Discussion of Modeling Assumptions:**

1) We have modeled the patient type as the probability of showing up without a reminder, so we assume types are distributed on the interval \([0,1]\). We use a beta distribution in particular, because it is flexible enough to incorporate many different kinds of skewness and shapes in the population. For example having \( \alpha = 0.5, \beta = 5 \) gives a left-skewed distribution where most patients are irresponsible, while having \( \alpha = 0.5, \beta = 0.5 \) gives a bimodal distribution, corresponding to the case where patients are either very responsible or very irresponsible, but unlikely to be anywhere between.

2) Essentially, \( h(\tau, q) \) models the new reliability of the system in equilibrium, factoring in the patient’s response to the system. Our first assumption is that this function \( h(\tau, q) \) is increasing in \( q \). We assume that a higher quality reminder systems has a greater chance of successfully contacting the patient and reminding him. Next, we assume that the function \( h(\tau, q) \) is also increasing in \( \tau \). That is, for a given reminder system, the equilibrium probability of a successful reminder increases for more responsible people. For example, if the clinic’s reminder policy is to call patients and leave messages, this reminder policy will be unsuccessful to the ‘irresponsible’ portion of the population that either does not own a telephone or does not regularly check messages on voicemail. As another example, we can consider the example of a clinic that sends out reminders by email in advance. Upon receiving the email, a responsible patient will check and record the appointment in his planner, while a less responsible patient may just do nothing. As an example of functional form, we can consider a class of functions \( h(\tau, q) = \tau^\gamma q \). Parameter \( \gamma \geq 0 \) models the dependency between the system and a patient as above. The smaller the value of \( \gamma \), the larger the influence of a patient’s type on the probability of a reminder reaching the patient. We can capture the special case that the patient type has no effect with \( \gamma = 0 \).

3) That \( g(\tau, q) \) is increasing in \( \tau \) should be intuitively clear — more responsible patients will show up more often due to their own efforts. However, it is less clear that \( g(\tau, q) \) is decreasing in \( q \). For a given patient, the better the system is in equilibrium, the more the patient will decrease his own efforts. For instance, if the doctor’s office is employing a very elaborate system (high \( q \)), then the patient will get used to the reminders after a sufficiently long period of time and rely completely on them to remember his appointments, rather than expending any of his own effort to remember the appointment (say, writing it down in his calendar). One possible objection is that it may be the case that the effort levels of the patients are independent of the quality of the reminder system, which is contrary to our above assumption that patients decrease their efforts instead. Yet, this assumption that the patients exhibit this compensating behavior is well justified by the moral hazard literature. Not only have
economists observed this phenomenon empirically, but also they have shown that this behavior is a rational decision that is simply motivated by utility maximization (e.g. [14], [15], [16],[17].)

As an example of the functional form of \( g \) we consider a class of functions with parameter \( \theta \) of the form \( g(\tau, q) = \frac{\tau}{1 + q^\theta} \). The parameter \( \theta \) models how strongly the system affects the patient’s effort.

4. A Numerical example

Using the model outlined above, we take the following as our functions: \( g(\tau, q) = \tau q^{-\alpha} \), \( h(\tau, q) = q \tau \), and we use the formula in (1) by numerically integrating to solve for \( E[p(\tau, q)] \). It is important to note that with no reminder system (\( q = 0 \)), \( g(\tau, q) = \tau \) and \( h(\tau, q) = 0 \). This is consistent with our definitions in Section 3. With no system, the probability that a patient shows up to the appointment due to his own effort is his own type \( \tau \), and the probability that the patient shows up to the appointment because of the reminder is 0. We give our results in Figures 3-6:

Figure 3: \( E[p(\tau, q)] \) over \( q, \alpha = 1, \beta = 3 \), mostly irresponsible

Figure 4: \( E[p(\tau, q)] \) over \( q, \alpha = 3, \beta = 1 \), mostly responsible

Figure 5: \( E[p(\tau, q)] \) over \( q, \alpha = 1, \beta = 1 \), uniform distribution

Figure 6: \( E[p(\tau, q)] \) over \( q, \alpha = 0.5, \beta = 0.5 \), bimodal distribution

Skewed toward the right, meaning that they are extremely responsible. On average, there is an extremely high probability for the patient to show up. We would intuitively expect the clinic dealing with the patient population in Figure 3 to invest heavily in a reminder system, whereas the clinic dealing with the patient population in Figure 4 not to invest in a reminder system.

From the numerical results, there are a few things to notice. First, it looks like \( E[p(\tau, q)] \) is non-monotone in \( q \). This implies that blindly improving the reminder system \( (q) \) does not guarantee an improvement in the percentage of customers showing up. Unless the doctor’s office can find a good enough reminder system, it is better not to even employ one at all.

Second, we notice that the inflection point of the distributions that are skewed more towards the left (more responsible patient populations) occur at higher values of \( q \). This makes sense intuitively - it is harder for the doctor’s office to positively influence the behavior of people that are already very responsible.

As a final note, we point out that the change within each figure is not too significant. This suggests that the reminder system seems to have a smaller effect than the distribution of patient heterogeneity. However, even an effect of a few percent is still nontrivial, considering that the estimated no-show rate is around 12%.

In the next section we will prove the insights that we
5. Theoretical Results

Proposition: For the functions \( g(\tau, q) = \tau^{1+\frac{1}{\tau}} \), \( h(\tau, q) = q\tau \), \( E[p(\tau, q)] \) is convex in \( q \).

Proof: We need to show that \( (E[p(\tau, q)])'' \geq 0 \), where the derivatives are taken with respect to \( q \). Since we can take the derivatives within the expectation, this is equivalent to showing that \( E[p''(\tau, q)] \geq 0 \). We will simply show that \( p''(\tau, q) \geq 0 \) since if this is the case, then its expectation will also be positive.

The second derivative is:

\[
\frac{2\tau(q-1)\left(\tau^{\frac{1}{\tau}} - \tau^{\frac{2}{\tau}}\right)\text{Log}[\tau]}{(-1+q)^4} + \frac{\tau\left(-q\tau^{\frac{1}{\tau}} + \tau^{\frac{3}{\tau}}\right)\text{Log}[\tau]^2}{(-1+q)^4}
\]

Keeping in mind that \( \tau \in (0, 1) \), it is immediately apparent that the term \( \frac{\tau\text{Log}[\tau]^2}{(-1+q)^4} \) is negative. It remains to show that the last term,

\[
2(1+q)(1 + \tau^{\frac{1}{\tau}} - \tau^{\frac{2}{\tau}}) + (-q\tau^{\frac{1}{\tau}} + \tau^{\frac{3}{\tau}})\text{Log}[\tau]
\]

is negative. We will show that the two separate terms are both negative.

Call \( y_1(q) = 2(1+q)(\tau^{\frac{1}{\tau}} - \tau^{\frac{2}{\tau}}) \) and call \( y_2(q) = (-q\tau^{\frac{1}{\tau}} + \tau^{\frac{3}{\tau}}) \). From basic algebra, we can see that \( y_1(q) < 0 \) if \( q < 1 \), which is always true since \( q \) is a probability.

Next, we notice that the derivative \( y_2(q) = \frac{-1+q}{q} + \text{Log}[q] + \text{Log}[\tau] \), which is negative. Furthermore, \( y_2(q) \) is continuous and \( y_2(0) \geq 0 \) and \( y_2(1) = 0 \). Using all these facts, we are able to conclude that on the interval \((0,1)\), \( y_2(q) \) is positive. Therefore \( (-q\tau^{\frac{1}{\tau}} + \tau^{\frac{3}{\tau}})\text{Log}[\tau] \) is negative. \( \square \)

Note that we can use the exact same analysis on a class of functions \( h(\tau, q) = \gamma q \) indexed by \( \gamma \in (0,1) \) as outlined in Section 3, since above we are taking the derivative over \( q \), treating \( \tau \) as a constant. Because of this, under the new \( h(\tau, q) \) function, we just have to replace every \( \tau \) with \( \tau^\gamma \), which is still a constant between 0 and 1, leading us to conclude that \( E[p(\tau, q)] \) is still convex.

5.1. Results for general functions

Proposition: If the functions \( g(\tau, q) \) and \( h(\tau, q) \) are convex in \( q \), then \( E[p(\tau, q)] \) is also convex in \( q \).

Proof: Using the expression for (1), we can derive the expression for the second derivative. Simple manipulation yields

\[
(1 - h(\tau, q))g''(\tau, q) + (1 - g(\tau, q))h''(\tau, q) - 2g'(\tau, q)h'(\tau, q)
\]

Since \( g(\tau, q) \) and \( h(\tau, q) \) are probabilities, \((1 - h(\tau, q))\) and \((1 - g(\tau, q))\) are both positive. Furthermore, from assumptions (2) and (3) in Section 2, we know that \( g''(\tau, q) \) is negative and \( h''(\tau, q) \) is positive. Therefore, the convexity of \( g(\tau, q) \) and \( h(\tau, q) \) gives a sufficient condition for the convexity of \( E[p(\tau, q)] \). However, this is by no means a necessary condition — in our numerical example of Section 3, the function \( g(\tau, q) = \tau^{\frac{1}{\tau}} \) is actually concave in \( q \).

5.2. Submodularity

As noted in Section 3, the point of inflection of \( E[p(\tau, q)] \) increases for distributions that tend to get ‘more responsible’. Since \( E[p(\tau, q)] \) is convex, the point of inflection is the level of \( q \) after which performance improves. Intuitively then, we would expect that the point of inflection for a responsible distribution of patients would be much higher than that of an irresponsible distribution of patients, since it would be harder for the clinic to positively influence the behavior of patients that are already extremely responsible. We may formalize this as follows. Fixing \( \beta \) in our distribution of \( \tau \), if we increase \( \alpha \), the distribution gets skewed more to the right (more responsible). We define the function \( f(\alpha, q) = E[p(\tau, q)] \), where we have fixed \( \beta \) and \( \tau \) is distributed according to a Beta distribution with parameters \( \alpha \) and \( \beta \). Then, we wish to conclude that for \( \alpha' \geq \alpha \), argmin\( f(\alpha', q) \geq \text{argmin}_q f(\alpha, q) \).

A necessary and sufficient condition for this property to hold is for the function \( f(\alpha, q) \) to be submodular. That is, \( \frac{\partial^2 f(\alpha, q)}{\partial \alpha \partial q} \leq 0 \). While this is easily checked if the \( g(\tau, q) \) and \( h(\tau, q) \) functions are known, it also makes sense intuitively. The condition requires in some sense that \( E[p(\tau, q)] \) exhibits diminishing marginal returns - as the patient population gets more responsible, an increase in the quality of the reminder system \( q \) will have less effect. This is quite a reasonable assumption to make.

6. An Application

Suppose that a certain practice is considering purchasing a new reminder system, but is unsure what benefits, if any, the new system will bring. The information that they have access to is the data of other practices that have also used this system — specifically the performance before and after the implementation of the system. Given this, a very reasonable question that can be asked is ‘Given the performance of a reminder system that another doctor’s office has employed, can I predict how the same reminder system will work for me?’ Essentially this question boils down to esti-
An application to overbooking

7. An application to overbooking

Now that we have examined exactly how reminder systems affect the show-up rate to the clinic, a natural extension is to ask what sort of impact it has on overbooking policies. Specifically, if we implement a better reminder system, do we expect to overbook more or less people, and to what degree? We will outline a model below to answer this question that is heavily based on the newsvendor problem.

We assume the clinic has a fixed capacity per day of $c$. That is, the clinic can attend to $c$ people on a normal day, with the standard number of people staffed and with all of them working normal hours. The clinic incurs a cost $C_o$ per appointment if the number of patients that actually show up, $S$, is greater than $c$. The clinic also has the cost $C_u$ for each unused unit of regular capacity. In this context, we can think of $C_u$ as being the opportunity cost of an appointment. On the other hand, if normal capacity is less than the number of patients that show up, the clinic needs to schedule more employees at a higher cost. So the clinic schedules appointments $s$ to minimize the expected cost:

$$
\min_s f(s) = C_oE[c-S]^+ + C_uE[S-c]^+,
$$

$$
S \sim \text{Binomial}(E[p(\tau,q)],s)
$$

(5)

This is not exactly a newsvendor problem. In the standard newsvendor setup, we minimize the cost by choosing our capacity. Here, we are trying to minimize the cost by picking $s$, the number of people scheduled - which only shows up as a parameter of random variable $S$.

Furthermore, $f$ is quasiconvex in $s$, if we notice that $E[c-S]^+$ is monotonically decreasing in $s$ and $E[S-c]^+$ is monotonically increasing in $s$. This guarantees that the minimum of $f$ occurs at a point of $s$ that satisfies the first order condition.

We approximate the distribution of $S$ with a normal distribution with parameters $(sE[p], sE[p](1-E[p]))$. Then, we have a simple form for $E[c-S]^+$ and $E[S-c]^+$ by using the standard normal loss and inverse normal loss functions (the details are derived below). Letting $\mu = sE[p]$ and $\sigma = sE[p](1-E[p])$ our objective function simplifies to:

$$
C_uE[c-S]^+ + C_oE[S-c]^+ =
$$

$$
\sigma C_u \left( \frac{c-\mu}{\sigma} \Phi \left( \frac{c-\mu}{\sigma} \right) + \phi \left( \frac{c-\mu}{\sigma} \right) \right)
$$

$$
+ \sigma C_o \left( \phi \left( \frac{c-\mu}{\sigma} \right) - \frac{c-\mu}{\sigma} (1 - \Phi \left( \frac{c-\mu}{\sigma} \right)) \right)
$$

While we were not able to find a closed form solution to this optimization problem, it can be easily evaluated by numerical calculations. An example follows in the next section. For further details on this newsvendor problem, the
7.1. A numerical example

In this section, we go through a numerical example to illustrate how to use the newsvendor model outlined above to decide whether or not to implement a new system. Assuming a patient population distributed according to a Beta distribution with parameters $\alpha = 1.3$, $\beta = 0.2$, and using the model outlined in Section 3, suppose the clinic is considering whether or not to upgrade their current system of quality $q = 0.7$ to a new system with $q = 0.95$. From the model, we know that the expected probability $(E[p(r, q)])$ goes from 0.85 to 0.86. Should the clinic upgrade their system?

To answer this question, we turn to our newsvendor model. We take our initial parameters in the newsvendor problem to be $C_u =$ $65$, $C_s =$ $55$, $E[p] = 0.85$, and $c = 200$ patients per day. We calculate the optimal policy numerically, yielding the solution to schedule 234 patients for the day at a total cost of $260. Now suppose that we have the opportunity to increase the average probability of showing up by 1% (that is, $E[p] = 0.86$) for a cost of $\$k$. Is it worth it to implement this system? By changing the parameter $E[p]$ of our newsvendor problem, we arrive at the new solution to schedule 232 patients a total expected daily cost of $\$250. So the daily costs of operating such a reminder system would have to be less than $\$10$ in costs, and the practice would have to consider whether the payback period on fixed costs is reasonable.

As a final note, it should be apparent that the model we have outlined does not take into account the costs associated with buying and operating different reminder systems. While this is an important factor in the decision making process, we have elected to leave it out of our model in order to avoid making any assumptions about the cost function which may be unrealistic. Instead, as the example above shows, our model evaluates the monetary benefit that implementing such a system would bring. With this information, it should be easy for the doctor’s office to decide whether or not the benefits outweigh the costs.

8. Summary and conclusion

Patients that do not show up for their appointments are a concern for medical practices. While these medical practices have attempted to solve the problem by simply giving the patients more reminders, empirical studies have shown that in many cases this strategy has had little or no effect. In this paper, we introduce a model that attempts to explain this phenomenon. We show that under some basic assumptions, the expected probability of showing up may actually decrease if the introduction of a reminder system can cause some patients to significantly decrease their own efforts for keeping the appointment. In addition, we introduce some applications of the model that we believe may be useful to clinics. One application gives clinics a way to predict the impact of a certain reminder system on their own patient population. Once clinics are able to predict the effectiveness of different reminder systems, we then introduce a way for them to weigh the costs and benefits of each system via a newsvendor model.

References

A. Derivations of the normal loss and inverse normal loss functions

If \( X \) is a standard normal distribution, we can calculate \( E[X - c]^+ \) (referred to as the standard normal loss function) as follows:

\[
E[X - c]^+ = \int_{-\infty}^{\infty} (t - c) \phi(t) dt
\]

\[
= \int_{-\infty}^{\infty} \phi(t) dt - \int_{c}^{\infty} c \phi(t) dt
\]

\[
= \int_{c}^{\infty} t \phi(t) dt - c (1 - \Phi(c)) = \phi(c) - c (1 - \Phi(c))
\]

Then if \( X \) is a normal distribution with parameters \( \mu \) and \( \sigma \), we can calculate \( E[X - c]^+ \) by normalizing \( X \):

\[
E[X - c]^+ = \sigma E\left[\frac{X - \mu}{\sigma} - \frac{c - \mu}{\sigma}\right]^+
\]

\[
= \sigma \left( \phi\left(\frac{c - \mu}{\sigma}\right) - \frac{c - \mu}{\sigma} \left(1 - \Phi\left(\frac{c - \mu}{\sigma}\right)\right)\right)
\]

In a similar way, if \( X \) is standard normal, we can calculate \( E[c - X]^+ \) (referred to as the inverse standard normal loss function) as follows:

\[
E[c - X]^+ = \int_{-\infty}^{c} (c - t) \phi(t) dt - \int_{-\infty}^{c} t \phi(t) dt
\]

\[
= c \Phi(c) + \phi(c)
\]

And if \( X \) is normally distributed with parameters \( \mu \) and \( \sigma \), we normalize \( X \) to calculate \( E[c - X]^+ \):

\[
E[c - X]^+ = \sigma E\left[\frac{c - \mu}{\sigma} - \frac{X - \mu}{\sigma}\right]^+
\]

\[
= \sigma \left( \phi\left(\frac{c - \mu}{\sigma}\right) - \frac{c - \mu}{\sigma} \left(1 - \Phi\left(\frac{c - \mu}{\sigma}\right)\right)\right)
\]

B. Comparative statics for the newsvendor problem

In this appendix we show that, as expected, our decision \( s \) of the number of people scheduled is increasing in \( C_u \), decreasing in \( C_o \), and increasing in \( c \). To prove this, it is sufficient to show that \( \frac{\partial^2 f}{\partial C_u \partial s} \leq 0 \), \( \frac{\partial^2 f}{\partial C_o \partial s} \geq 0 \), and that \( \frac{\partial^2 f}{\partial c \partial s} \leq 0 \), where \( f \) represents the expected cost to the firm. If, for example, \( \frac{\partial^2 f}{\partial C_u \partial s} \leq 0 \), then if we increase \( C_u \) a little bit, the marginal benefit \( \frac{\partial f}{\partial s} \leq 0 \). Thus it is optimal for us to increase \( s \) in order to achieve a lower objective value.

**Fact 1:** \( \frac{\partial^2 f}{\partial C_u \partial s} \leq 0 \): Taking the derivative of \( f \) over \( C_u \) first leaves us with \( E[c-S]^+ \). Since \( E[c-S]^+ \) is the expected number of unused capacity, \( \frac{\partial E[c-S]^+}{\partial s} \leq 0 \) since increasing the number of people scheduled \( s \) will lower unused capacity.

**Fact 2:** \( \frac{\partial^2 f}{\partial C_o \partial s} \geq 0 \): In the same way, we first take the derivative of \( f \) over \( C_o \), leaving us with \( E[S-c]^+ \). Taking the derivative again over \( s \) should leave us with a positive term. That is, increasing the number of scheduled patients \( s \) will increase \( E[S-c]^+ \), the expected number of units of demand over capacity.

**Fact 3:** \( \frac{\partial^2 f}{\partial c \partial s} \leq 0 \): We first take the derivative over \( c \), and interchange the derivative with the expectation.

\[
C_u \frac{\partial E[c-S]^+}{\partial s} + C_o \frac{\partial E[S-c]^+}{\partial s}
\]

\[
= C_u E[1_{c>S}] + C_o E[1_{S>c}]
\]

\[
= C_u P(c-S \geq 0) + C_o P(S-c \geq 0)
\]

\[
= C_u + (C_u + C_o) P(S \geq c)
\]

Next we take the derivative over \( s \). Note that \( \frac{\partial P(S \geq c)}{\partial s} \geq 0 \). Since \( S \) is binomially distributed with number of trials as \( s \), if we increase \( s \) there will be a higher chance that \( S \geq c \). Therefore \( \frac{\partial^2 f}{\partial c \partial s} \leq 0 \).