

## Optimal Decision Rules for Product Recalls

Ali Devin Sezer

Institute of Applied Mathematics, Middle East Technical University, 06800, Ankara, Turkey,  
[devin@metu.edu.tr](mailto:devin@metu.edu.tr), [http://www3.iam.metu.edu.tr/iam/index.php/devin\\_sezer](http://www3.iam.metu.edu.tr/iam/index.php/devin_sezer)

Çağrı Haksöz

Sabancı School of Management, Sabancı University Orhanlı, Tuzla 34956, İstanbul, Turkey,  
[cagrihaksoz@sabanciuniv.edu](mailto:cagrihaksoz@sabanciuniv.edu), <http://myweb.sabanciuniv.edu/cagrihaksoz>

We consider a hypothetical company that is assumed to have just manufactured and sold a number of copies of a product. It is known that, with a small probability, the company has committed a manufacturing fault that will require a recall. The company is able to observe the expiration times of the sold items whose distribution depends on whether the fault is present or absent. At the expiration of each item, a public inspection takes place that may reveal the fault, if it exists. Based on this information, the company can recall the product at any moment and pay back each customer the price of the product. If the company is not able to recall before an inspection reveals the fault, it pays a fine per item sold, which is assumed to be much larger than the price of the product. We compute the optimal recall time that minimizes the expected cost of recall of this company. We then derive and solve a stationary limit recall problem and show that the original problem converges to it as the number of items initially sold increases to  $\infty$ . Finally, we propose two extensions of the original model and compute the optimal recall times for these. In the first extension, the expired items are inspected only if they expire earlier than expected; in the second extension, the company is able to conduct internal/private inspections on the expired items. We provide numerical examples and simulation results for all three models.

**Key words:** product recalls; quality risk; supply chain risk; optimal stopping; stochastic optimal control; point processes; filtration shrinkage; sequential analysis; dynamic programming; Bayesian analysis

**MSC2000 subject classification:** Primary: 60G40, 60G55; secondary: 47H10, 49L20

**ORMS subject classification:** Primary: Decision Analysis, Reliability/Quality control; secondary: Manufacturing, Dynamic programming/optimal control

**History:** Received June 23, 2011; revised December 6, 2011, and March 2, 2012. Published online in *Articles in Advance* June 19, 2012.

**1. Introduction.** Product recalls are the main mechanism to handle manufacturing errors that are detected after production and sales. They are economically significant events; every year hundreds of recalls take place costing billions of dollars. A recall process consists of two stages: the detection of fault and the actual recall of items having the detected fault (Smith et al. [15], Teratanavat and Hooker [18], Tang [17]). A poor performance of either of these can do great harm to a company and its customers (Pinedo et al. [14], Jarrell and Peltzman [7], Tabuchi [16], *New York Times* [10], Vlasic [19], Maynard and Tabuchi [9]). This paper focuses on the first stage, i.e., on the detection of manufacturing faults that lead to recalls.

We think that a key component of an effective fault detection system is a good model of the *post-sales* environment and processes. These vary widely across countries, industries, and companies. Therefore, it is probably not possible to build an a priori universal recall model that will fit every company. However, the following observations seem to be fairly universal. First, a fault that may trigger a recall is (at least initially) not directly observable. The goal of a detection system is to dynamically understand the likelihood of fault from the available post-sales information flow. Second, the post-sales information is generated in parallel by the actions of many entities. In many cases, most of these entities are the customers who purchased instances of the product, but they can be other persons or organizations as well, such as companies and government agencies that regularly test products and publish the results. Third, there is an uncertainty as to whether an instance of a faulty product will lead to a problem that will reveal the fault. For example, in the case of the food industry, not every consumer gets sick from eating contaminated food. Fourth, there are two types of costs associated with a possible recall: the material cost of recall and a much greater cost arising from a late recall, which includes a tarnished reputation, costs arising from lawsuits against the company and fines associated with a late recall. This paper attempts to build the simplest possible model that incorporates the above characteristics. It then formulates the fault detection problem as the minimization of the expected cost of the recall decision.

In light of the forgoing discussion, we propose the following model. Imagine a hypothetical company, which we will refer to as “the seller,” that sells a single type of product,  $N$  copies of which are assumed to have been recently manufactured. Based on her current information and analysis, the seller knows that with probability  $\pi$ , the manufacturing was in accordance with the applicable standards and regulations and that with a small probability  $1 - \pi$ , despite seller’s best efforts, a regulation or a standard has been violated during manufacturing.

We will call this violation a “fault.”<sup>1</sup> Whether there is a fault is initially unobservable; i.e., at the time of sale, the seller knows not whether there is a fault, but only the probability of fault. Each item<sup>2</sup> is sold at a fixed price  $P$  and the sales of all items take place at the same time. This assumption is sensible in situations where buyers pre-order or when the seller is able to sell the product quickly.<sup>3</sup> Each sold item has a lifetime, which is conditionally exponentially distributed with rate  $\mu_1$  if there is a fault and with rate  $\mu_0 \neq \mu_1$  if there is no fault. When an item expires, i.e., when it reaches the end of its lifetime, the seller is informed about the expiration and the buyer inspects his/her unit to check if there was a manufacturing fault. If indeed the unit is faulty, s/he is able to detect it with probability  $1 - p$ . The inspection does not detect a fault if there is no fault; i.e., the inspection yields no false positives. If and when a fault is discovered in one of these inspections, everyone is alerted about it and the seller is forced to pay each customer  $K$  dollars per item sold, where  $K$  is a much larger amount than the price  $P$ ; the phrase “much larger” is made more precise in (14). Because everyone has access to the results of the inspections, we also refer to them as “public inspections.” For her part, at any moment before a fault is detected, the seller can recall. We assume a very simple recall process: it consists of collecting back the sold units and returning each buyer his/her  $P$  dollars. Finally, we assume that there is a constant interest rate  $r > 0$ .

By a recall decision rule we mean any measurable function that takes as input the parameters of this model and the post-sales information flow and tells the seller when to recall. The minimization of the seller’s expected cost of recall over all decision rules is the main problem of the present paper. Our first step toward its solution is to notice that this is an *optimal stopping problem*. The underlying processes of the problem are the inspection results and a point process with a state dependent intensity that keeps track of the number of expired items. We will use the filtration that these processes generate to model the post-sales information flow. The set of stopping times of this filtration represents the set of all recall decision rules. Our goal, then, is the characterization and computation of the stopping time that minimizes the expected cost of recall.

The model outlined above is developed in §2 and the analysis of the optimal stopping problem is in §§2, 3, and 4. Section 3 treats the special case when  $N = 1$ ; i.e., there is only a single item sold. The analysis of this section shows that there are two cases to consider: (1)  $\mu_1 < \mu_0$ : in this case, the fault somehow increases the expected lifetime of the item and there is an optimal static stopping time  $s^*$  until when the seller waits. If the item is still functioning at this time, she recalls. (2)  $\mu_1 > \mu_0$ : one ordinarily expects a fault to shorten the expected lifetime of a product; this holds when  $\mu_1 > \mu_0$ . The optimal recall decision rule for this case is: if the probability of fault  $1 - \pi$  is above  $(P/K)((\mu_1 + r)/\mu_1)(1/(1 - p))$  the seller recalls immediately (i.e., she doesn’t sell the item); otherwise, she sells and never recalls.

Section 4 treats the case when  $N > 1$ , i.e., when the seller sells more than one item, under the assumption  $\mu_1 > \mu_0$ . The optimal recall decision rule derived in this section is dynamic and is expressed in terms of a likelihood ratio process  $\Phi$ . The value of  $\Phi$  at any time is the ratio of the conditional probabilities of fault and no fault given all of the information up to that time. Right after manufacturing and before selling, this likelihood ratio equals  $(1 - \pi)/\pi$ . In time, as new information becomes available, it will evolve following the dynamics given in (7). The optimal recall rule (24) is of the following form: there is a threshold such that when the likelihood ratio goes above it, the seller recalls. The threshold depends on the number of products that are still functioning, and it needs to be updated every time an expiration occurs.

To the best of our knowledge, our work is the first to build a mathematical model of product recalls and formulate the recall decision problem as one of optimal stopping. As our analysis shows, the recall problem leads to an optimal stopping problem that is related to those that arise in the sequential hypothesis testing of point processes. The study of the sequential hypothesis testing of Poisson processes goes back to Wald (Wald and Wolfowitz [22], Wald [21]); more recent studies on it include Bayraktar et al. [1], Dayanik and Sezer [5], and Peskir and Shiryaev [13, §22] (see also the references in the latter). In the following paragraphs we give an outline of our solution to the recall problem that we have formulated above. The solution involves a number of ideas, some of which, to the best of our knowledge, appear here for the first time.

There are three main steps of the solution methods of §§2 and 4. Remember the assumption that the inspections never indicate a fault when there is no fault. The first step is to show that under this assumption there is no loss of generality to work with the smaller filtration generated by the expirations. This allows us to integrate out the public inspection results. Let us emphasize the stages of this argument: (a) show that the original filtration can

<sup>1</sup> Imagine that the seller has a list of possible problems that will require her to recall (such a list could be prepared based on past experience and data); we can also define a “fault” as the realization of one of the problems in this list.

<sup>2</sup> Or “unit;” we will use the words “item” and “unit” interchangeably.

<sup>3</sup> A recent example is the Apple iPad sales, which exceeded one million in less than a month (Patterson [12]).

be shrunk, i.e., can be replaced with a smaller filtration; (b) integrate out the processes (in this case there is only one) that are independent of the new filtration; and (c) study the resulting lower dimensional problem.

Second, we use a change of measure to replace the unobserved information whether a fault occurred or not with a likelihood ratio process. To our knowledge, this idea first appeared in Zakai [23]; it is used in the context of a hypothesis testing problem about a Poisson process in Bayraktar et al. [1] and Dayanik and Sezer [5]. The point process underlying the recall problem has an intensity that depends on the number of items currently functioning and the change of measure, and its Radon Nikodym derivative needs to be computed for this point process. The first step plays also a nontrivial role here; see Remark 2.1. The change of measure reduces the recall problem to an optimal stopping problem of a one dimensional likelihood ratio process whose jump intensity and dynamics depend on time but are independent of whether there is a fault or not; see (7) and (8). The likelihood ratio process acts like a continuous time price deflator in the expected cost (8). Each time this process jumps, a running cost determined by  $K$  and  $p$  is incurred, discounted at the rate indicated by the process. When the controller stops, a stopping cost of  $P$  (the price of the product) is incurred twice—once discounted at the rate of the actual interest rate  $r$  and once discounted by the likelihood ratio process and  $r$  combined.

The third step is to use dynamic programming (DP) to compute the value function of the control problem obtained from the second step. The value function depends on two variables:  $\phi$ , the initial position of the likelihood ratio process, and  $N$ , the number of items functioning (at time  $t = 0$ ,  $N$  is the number of items sold). We employ DP on the variable  $N$ , which yields an integral equation for the value function. This is an idea that goes at least back to Bertsekas and Shreve [3]; a recent paper that uses it in the context of the hypothesis testing of a Poisson process is Dayanik and Sezer [5]. The dependence of the dynamics of the likelihood process on  $N$  leads to a sequence of integral equations depending on  $N$ ; see (22). The successive application of the maps defined by these equations allows us to compute the value function of the optimal stopping problem and from it the optimal recall rule. An interesting feature of our analysis is that only calculus is used in proving that it is enough to consider the expiration times as the only candidate recall times; see the proof of Theorem 4.1.

Section 5 studies the asymptotics of the recall model of §2 when the number of items  $N$  sold increases to  $\infty$ . We show that, under proper scaling, a limit recall model emerges that is based on a Poisson process with fixed intensity  $\mu_0$  and in which the interest rate is zero. The optimal recall rule of the limit problem gives an almost optimal time independent recall rule for the original problem when  $N$  is large. The convergence problem treated in this section is of the following form: a sequence of optimal stopping problems that are based on a sequence of point processes with time/state dependent intensities converges to an optimal stopping problem based on a Poisson process with constant intensity. We haven't encountered a similar problem in the current point process literature. From a functional analysis point of view, this is a problem of showing the convergence of the composition of a sequence of operators to the fixed point of a limit operator. The nonstationarity of the prelimit processes and the presence of an interest rate preclude an argument based on monotonicity. The key idea of our analysis is the use of linear functions to express bounds; see Proposition 5.1 and Theorem 5.2. This seems to be a new argument and we hope to explain it and its generalizations in detail in a separate note. Our analysis also gives a precise rate of convergence (namely,  $O(1/N)$ ) to the limit. This rate of convergence is another novel feature of the problem (usually an exponential rate is obtained).

Section 6 introduces and solves two extensions. In the first extension, the public inspection of an expired item takes place only when the item expires before its initial expected expiration time  $1/\mu_0$ . In the second extension, the seller is able to conduct her own private inspection after each expiration. The first extension leads to the introduction of a new parameter (time to a deadline after which no inspections occur) and the value function begins to depend on two continuous parameters. A modification of the analysis of §4 that incorporates this new parameter gives the optimal recall rule for the first extension. In dealing with the second extension we show that its analysis can be reduced to a recursive application of the analysis of §4. The asymptotic analysis of the extensions is given in the same section and leads to simpler time independent recall decision rules for these problems.

Section 7 provides simulations and numerical examples running the recall models and optimal decision rules of §§4, 5, and 6. Further comments on our results are in §4.1 and in §8.

**2. The model.** The random variable  $d$  will represent whether a manufacturing fault occurred:  $d = 0$  indicates there was no manufacturing fault and  $d = 1$  otherwise.  $d$  is initially not observable but its distribution is known:  $\mathbb{P}(d = 0) = \pi$  and  $\mathbb{P}(d = 1) = 1 - \pi$ . Let  $L_i$  denote the lifetime of item  $i$ . Given  $d$ ,  $\{L_i\}$  is an exponentially distributed iid sequence with rate  $\mu_d$ . We will denote the sequence of expiration times by  $T_1 < T_2 < \dots < T_N$ . One obtains the sequence  $\{T_i\}$  from the sequence  $\{L_i\}$  by sorting the latter.  $C_t \doteq \sup\{i: T_i < t\}$  is the number of items that expired by time  $t$ . Let  $\tau_0 \doteq T_1$  and for  $i \geq 1$  let  $\tau_i \doteq T_{i+1} - T_i$  denote the time interval between two

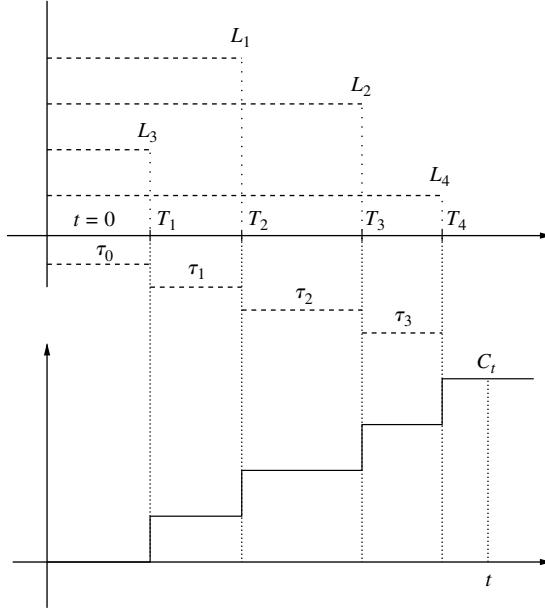


FIGURE 1. The times  $\{L_i\}$ ,  $\{T_i\}$ , and  $\{\tau_i\}$  and the process  $C_t$ .

consecutive expiration times. One obtains the distribution of  $\{\tau_i\}$  from the distribution of the lifetimes  $\{L_i\}$  as follows.  $\tau_0 = T_1$  is the time when the first expiration occurs; i.e.,  $\tau_0 = \min\{L_1, L_2, \dots, L_N\}$ . This implies that given  $d$ ,  $\tau_0$  is exponentially distributed with rate  $N\mu_d$ . After time  $T_1$  only  $N - 1$  items remain. The lifetimes are distributed exponentially and hence are memoryless. It follows that given  $d$ ,  $\tau_2$  is independent of  $\tau_1$  and is exponentially distributed with rate  $(N - 1)\mu_d$ . Repeating the same argument  $N$  times gives

**PROPOSITION 2.1.** *Conditioned on  $d$ ,  $\tau_i$  is exponentially distributed with rate  $(N - i)\mu_d$  and the  $\{\tau_i\}$  form an independent sequence.*

The random variables  $\{L_i\}$ ,  $\{T_i\}$ ,  $\{\tau_i\}$  and the process  $C$  are shown on an example in Figure 1;  $t = 0$  is when the sales of four items take place (i.e.,  $N = 4$ ).

A sequence of random variables  $\{i_k, k \in \mathbb{N}\}$  taking values in  $\{0, 1\}$  and with conditional distribution  $\mathbb{P}(i_k = 0 | d) = d(1 - p)$  models the inspections that take place after the expirations.  $i_k = 1$  if the  $k$ th inspection result is favorable for the company and  $i_k = 0$  if it (the  $k$ th inspection result) reveals the fault.  $\{i_k\}$  are assumed to be independent of each other given  $d$ . Note that  $i_k = 1$  for all  $k$  when  $d = 0$ , i.e., when there is no fault.

For the probability space that supports these random variables we take, for now, the canonical space  $\Omega \doteq \mathbb{Z}_2 \times \mathbb{R}^N \times \mathbb{Z}_2^N$ . For  $\omega = (e, \mathbf{x} = (x_0, x_2, x_3, \dots, x_{N-1}), \mathbf{i} = (j_1, j_2, \dots, j_N)) \in \Omega$  we realize our random variables as the coordinate projections  $d(\omega) \doteq e$ ,  $\tau_k(\omega) \doteq x_k$ ,  $i_k(\omega) \doteq j_k$ . The information that is available to the seller at time  $t$  is represented by the filtration  $\mathcal{F}'_t \doteq \sigma(C_s, s \leq t, i_k, k \leq T_t)$ ; i.e., the seller observes the lifetimes of the items that expired and the inspection results before time  $t$ . Let  $\mathcal{S}'$  denote the set of all stopping times of the filtration  $\{\mathcal{F}'_t\}$ .  $\mathcal{S}'$  represents the set of all recall decision rules based on the information flow  $\{\mathcal{F}'_t\}$ . Let us note the basic nature of the stopping times of the filtration  $\{\mathcal{F}'_t\}$ .

**LEMMA 2.1.** *For any stopping time  $\tau \in \mathcal{S}'$  and for any  $\omega_0 \in \Omega$ ,  $\tau = \tau(\omega_0)$  on the set  $E_{\tau, \omega_0} \doteq \{\omega: T_i(\omega) = T_i(\omega_0), \text{ if } T_i(\omega) < \tau(\omega_0)\}$ .*

Lemma 2.1 follows because  $C$  is deterministic between its jumps.

Define  $W \doteq \inf\{k: i_k = 0\}$ .  $W$  is the index of the inspection that reveals the fault and  $T_W$  is the time of the discovery of the fault ( $T_W = \infty$  if no fault exists). The minimum expected cost of recall is

$$\inf_{\tau \in \mathcal{S}'} \mathbb{E}[1_{\{d=0\}} e^{-r\tau} NP + 1_{\{d=1\}} (1_{\{\tau < T_W\}} e^{-r\tau} NP + 1_{\{\tau \geq T_W\}} e^{-rT_W} NK)]. \quad (1)$$

Dividing (1) by the number of items  $N$  yields

$$\inf_{\tau \in \mathcal{S}'} \mathbb{E}[1_{\{d=0\}} e^{-r\tau} P + 1_{\{d=1\}} (1_{\{\tau < T_W\}} e^{-r\tau} P + 1_{\{\tau \geq T_W\}} e^{-rT_W} K)], \quad (2)$$

which is the minimum expected cost per item. The  $1_{\{d=0\}}$  term represents the cost of an unnecessary recall, which can happen because  $d$  is not directly observable. The seller can be misled by an unlikely realization of the process  $C$  and issue a recall even though there is no fault. The  $1_{\{d=1\}}$  term represents the costs associated with a fault and is the sum of the cost of a timely recall ( $\tau < T_W$ ) and the cost of the fault being caught before the seller has a chance to recall. We will now simplify (2) in a series of steps. Define  $\mathcal{F}_t \doteq \sigma(C_s, s \leq t)$ . Let  $\mathcal{S}$  be the stopping times of this filtration.

**LEMMA 2.2.** *The optimal stopping problem (2) is equivalent to*

$$\inf_{\tau \in \mathcal{S}} \mathbb{E}[1_{\{d=0\}} e^{-r\tau} P + 1_{\{d=1\}} (1_{\{\tau < T_W\}} e^{-r\tau} P + 1_{\{\tau \geq T_W\}} e^{-rT_W} K)]. \quad (3)$$

**PROOF.** Let  $\mathbf{0}$  be the zero element of  $\mathbb{Z}_2^N$ . Take  $\tau' \in \mathcal{S}'$  and for any  $\omega = (e, \mathbf{x}, \mathbf{i}) \in \Omega$  define  $\tau(\omega) = \tau'(\omega')$  where  $\omega' = (e, \mathbf{x}, \mathbf{0})$ . Note that  $\tau \in \mathcal{S}$ . The expected cost of recall is the same for both of these stopping times because the inspections have no false positives and the recall problem ends as soon as an inspection reveals a fault. Thus, for every stopping time  $\tau' \in \mathcal{S}'$ , there is a stopping time  $\tau \in \mathcal{S}$  whose expected cost equals that of  $\tau'$ . This and  $\mathcal{S}' \supset \mathcal{S}$  imply that (2) and (3) are equivalent.  $\square$

In (3),  $\tau$  no longer depends on  $W$ . Furthermore, the value of  $W$  has no influence on the part of the expectation in (3) that is carried out over the set  $\{d = 0\}$ . Thus, there is no harm to change the conditional distribution of  $W$  to something arbitrary on this set, and in particular to the conditional distribution of  $W$  on  $\{d = 1\}$ . Thus, we may safely proceed as if  $W$  were completely independent of the rest of the random variables in (3) and integrate it out. This leads to

**PROPOSITION 2.2.** *The optimal stopping problem in (3) is equivalent to*

$$\inf_{\tau \in \mathcal{S}} \mathbb{E}\left[1_{\{d=0\}} e^{-r\tau} P + 1_{\{d=1\}} \left(p^{C_\tau} e^{-r\tau} P + \sum_{k=1}^{C_\tau} e^{-rT_k} p^k K'\right)\right], \quad (4)$$

where  $K' = (1 - p)K/p$ .

With this representation, the penalty  $K$  is translated into a fixed cost  $K'$  paid at each expiration time in the faulty case. The discount factor is  $p^k e^{-rT_k}$ ; the probability  $p$  of a failed inspection has become a discount factor accumulated at each jump. From this point on, we can reduce our probability space from  $\mathbb{Z}_2 \times \mathbb{R}^N \times \mathbb{Z}_2^N$  to  $\Omega = \mathbb{Z}_2 \times \mathbb{R}^N$  and discard the random variables  $i_k$ . Note that there is a stark difference between (3) and (4): in (3) the dynamics can stop before  $\tau$  (namely at time  $T_W$ ); in (4) the dynamics always continue all the way to time  $\tau$ . Nonetheless, these problems have the same expected cost for each  $\tau$ .

Integrating  $d$  out of (4) is less straightforward because, under  $\mathbb{P}$ , the distribution of  $\{L_i, i = 1, \dots, N\}$  depends on  $d$ . An idea that goes back to Zakai [23] suggests that one write  $\mathbb{P}$  in terms of another measure  $\mathbb{P}_0$  under which the distribution of  $\{L_i\}$  is independent of  $d$ . To this end, let measure  $\mathbb{P}_0$  be the measure on  $\Omega$  under which the sequence  $\{L_i\}$  is iid, exponentially distributed with rate  $\mu_0$ . Define

$$R_t \doteq R_{T_n} e^{-(N-n)(\mu_1 - \mu_0)(t-T_n)}, \quad t \in [T_n, T_{n+1}), \quad R_{T_n^-} \doteq \frac{\mu_1}{\mu_0} R_{T_n-}$$

for  $n \in \{1, 2, 3, \dots, N\}$  and  $R_0 = 1$ ; here  $R_{T_n-}$  denotes  $\lim_{t \nearrow T_n} R_t$ . Define  $\mathcal{G}_t \doteq \sigma(\mathcal{F}_t, d)$ .

**PROPOSITION 2.3.** *The measure  $\mathbb{P}$  is absolutely continuous with respect to  $\mathbb{P}_0$  and*

$$\mathbb{E}_0\left[\frac{d\mathbb{P}}{d\mathbb{P}_0} \mid \mathcal{G}_t\right] = 1_{\{d=0\}} + 1_{\{d=1\}} R_t. \quad (5)$$

See the appendix for a proof of this result. Using this proposition, one can write expectations with respect to  $\mathbb{P}$  as expectations with respect to  $\mathbb{P}_0$ . The independence of  $d$  and  $\{L_i\}$  (and hence of  $d$  and  $\{\tau_i, T_i\}$ ) under  $\mathbb{P}_0$  allows us to integrate out the  $d$  variable when (4) is written in terms of  $\mathbb{P}_0$ :

**PROPOSITION 2.4.** *The optimal stopping problem in (4) can be rewritten as*

$$\inf_{\tau \in \mathcal{S}} \mathbb{E}_0\left[\pi e^{-r\tau} P + (1 - \pi) \left(p^{C_\tau} e^{-r\tau} P R_\tau + \sum_{k=1}^{C_\tau} R_{T_k} e^{-rT_k} p^k K'\right)\right]. \quad (6)$$

See the appendix for a proof of this proposition.

**REMARK 2.1.**  $\mathbb{P}(\cdot \mid d = 1)$  is not absolutely continuous with respect to  $\mathbb{P}(\cdot \mid d = 0)$  because the inspection results cannot give false positives. Therefore, a change of measure similar to (5) would be less straightforward if the inspection results were not integrated out with Lemma 2.2 and Proposition 2.2.

Factor out  $\pi$  in (6) to get the equivalent problem

$$\inf_{\tau \in \mathcal{S}} \mathbb{E}_0 \left[ e^{-r\tau} P + \frac{1-\pi}{\pi} \left( p^{C_\tau} e^{-r\tau} PR_\tau + \sum_{k=1}^{C_\tau} R_{T_k} e^{-rT_k} p^k K' \right) \right].$$

Note that  $(1-\pi)/\pi$  is the likelihood ratio of fault at time zero. Incorporate this initial likelihood ratio and the  $p^k$  term in (6) into  $R_t$  and define

$$\begin{aligned} \Phi_t &\doteq \Phi_{T_n} e^{-(N-n)(\mu_1-\mu_0)(t-T_n)}, \quad t \in [T_n, T_{n+1}), \\ \Phi_{T_n} &\doteq \frac{\mu_1 p}{\mu_0} \Phi_{T_n-}, \quad \Phi_0 \doteq \frac{1-\pi}{\pi}. \end{aligned} \tag{7}$$

It follows from Proposition 2.4 that (4) and (6) are equivalent to

$$\inf_{\tau \in \mathcal{S}} \mathbb{E}_0 \left[ \sum_{k=1}^{C_\tau} e^{-rT_k} K' \Phi_{T_k} + e^{-r\tau} P(1 + \Phi_\tau) \right]. \tag{8}$$

In the rest of the paper we will use (8) as the primary representation of the recall problem. The initially unobservable variable  $d$  has disappeared from the final form (8) of our recall problem; in its place came its likelihood ratio as a price deflator.

We note that (5) and the abstract Bayes rule (Liptser and Shiryaev [8, §7.9]) give

$$\mathbb{P}(d = 1 \mid \mathcal{F}_t) = \frac{\pi R_t}{(1-\pi) + \pi R_t}, \quad \mathbb{P}(d = 0 \mid \mathcal{F}_t) = \frac{1-\pi}{(1-\pi) + \pi R_t}.$$

$\mathbb{P}(d = 1 \mid \mathcal{F}_t)[\mathbb{P}(d = 0 \mid \mathcal{F}_t)]$  is the Bayesian updated probability of (no) fault given the trajectory of  $C$  up to time  $t$ . The Bayesian updated likelihood ratio of fault is the ratio of these two probabilities and it equals  $((1-\pi)/\pi)R_t$ . A similar computation using  $\mathcal{F}'_t$  instead of  $\mathcal{F}_t$  gives

$$\mathbb{P}(d = 1 \mid \mathcal{F}'_t) = \begin{cases} \frac{\pi \Phi_t}{(1-\pi) + \pi \Phi_t}, & \text{if } i_k = 0 \text{ for all } i \text{ such that } T_i \leq t, \\ 1, & \text{otherwise.} \end{cases} \tag{9}$$

Therefore,  $\Phi$  is the likelihood ratio of fault with respect to  $\{\mathcal{F}'_t\}$ .

**REMARK 2.2.** We have chosen the likelihood ratio process  $\Phi$  as our state process and base our formulation on it; one can also use the process  $t \rightarrow \mathbb{P}(d = 1 \mid \mathcal{F}_t)$  for this purpose; see, for example, Peskir and Shiryaev [13]. The relation (9) connects these two approaches. The chief advantage of working with the likelihood ratio, at least in the context of the present problem, is the simple structure of its dynamics, which leads to simpler DP equations (DPEs). To the authors' knowledge, the idea of using the likelihood ratio instead of the conditional probability process is due to Zakai [23] and appeared first in the context of optimal nonlinear filtering. The main reason for its introduction seems to be, again, the simpler equations that it leads to.

**3. The static problem.** Lemma 2.1 implies that when there is only one buyer buying a single item, i.e., when  $N = 1$ , the seller can fix a deterministic recall time as soon as she sells the product. Suppose the seller is to recall the item at time  $s \in [0, \infty]$ , if it is still functioning. One can directly use (3) to compute the expected cost of the seller to be

$$v(p, \pi, P, K, r, s) \doteq \pi e^{-(\mu_0+r)s} P + (1-\pi)e^{-(\mu_1+r)s} P + K(1-\pi)(1-p) \frac{\mu_1}{\mu_1+r} (1 - e^{-(\mu_1+r)s}). \tag{10}$$

Let us refer to  $(p, \pi, P, K, r)$  as  $x$  and write  $v(x, s)$ . In terms of  $x$ , (8) is

$$\inf_{s \in [0, \infty]} \frac{1}{\pi} v(x, s). \tag{11}$$

Thus, for  $N = 1$ , the question of finding the optimal time to recall reduces to minimizing  $v$  with respect to the real parameter  $s$ .

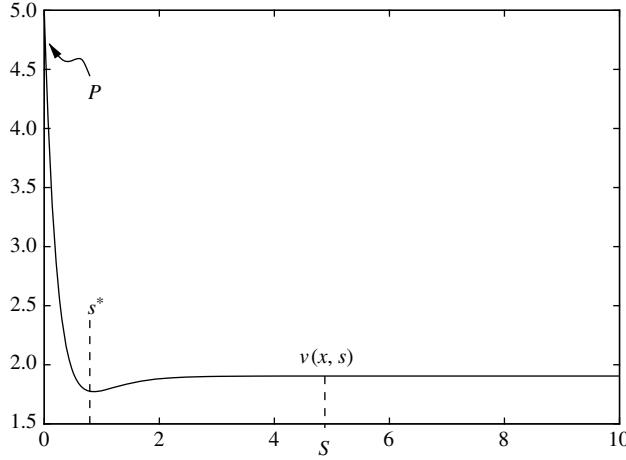


FIGURE 2. Expected cost of recall as a function of recall time when  $\mu_1 < \mu_0$ ,  $K = 100$ , and  $P = 5$ .

### 3.1. Solution.

There are two cases:

$\mu_1 < \mu_0$ . In this case, the manufacturing fault increases the expected lifetime of the item. A graph of  $v$  as a function of  $s$  is depicted in Figure 2; parameter values are  $P = 1.5$ ,  $K = 100$ ,  $\mu_0 = 4$ ,  $\mu_1 = 2$ ,  $r = 0.1$ ,  $\pi = 0.95$ , and  $p = 0.6$ . The figure shows that in this example there is a unique recall time  $s^*$ , and the optimal stopping rule for the seller is to wait until  $s^*$  and recall the item if it hasn't expired yet. Optimization of the value function  $v$  in (10) over all  $s$  using calculus proves that this picture is accurate in general; i.e., if  $\mu_1 < \mu_0$ ,  $(x, s) \rightarrow v(x, s)$  has at most one stationary point  $s^*$ , and if such a point exists, it must be a minimum. If  $\mu_1 < \mu_0$ , the seller expects the item to expire *sooner* in the case of no fault than in the case of fault. Thus, she takes the survival of the item to be longer than  $s^*$  to be evidence of fault.

$\mu_1 > \mu_0$ . In this case the fault shortens the expected lifetime of the item. The next lemma says that the graph in Figure 2, when turned upside down, shows the nature of the relation between  $s$  and  $v$  in this case as well.

**LEMMA 3.1.** *If  $\mu_1 > \mu_0$ ,  $v(x, s)$  has at most one stationary point, and if such a point exists, it must be a maximum.*

**PROOF.** The derivative of the expression in (10) with respect to  $s$  is

$$e^{-(r+\mu_1)s}(-(\mu_0+r)e^{(\mu_1-\mu_0)s}P - (\mu_1+r)(1-\pi)P + K(1-\pi)(1-p)\mu_1), \quad (12)$$

which has at most a single zero. If it exists, this zero corresponds to a maximum of  $v$  because (12) is negative as  $s \rightarrow \infty$ .  $\square$

Lemma 3.1 implies that when  $\mu_1 > \mu_0$  the minimizer of  $v$  is either  $s = 0$  or  $s = \infty$ . The value of  $v$  for these values of  $s$  are

$$v(x, 0) = P \quad \text{and} \quad v(x, \infty) = K(1-\pi)(1-p) \frac{\mu_1}{\mu_1 + r}. \quad (13)$$

The first of these is the price of the item and the second is the expected cost of never recalling. Then the optimal recall decision rule for the seller is to recall immediately (i.e., not sell) if the price of the product is less than the expected cost of never recalling and to sell and never recall if the price of the product is greater. The optimal decision rule can be rephrased in terms of a threshold for the probability of fault: don't sell if  $1 - \pi$  is above  $(P/K)((\mu_1 + r)/\mu_1)(1/(1-p))$ ; otherwise, sell and never recall.

Note that if  $K(1-p)(\mu_1/(\mu_1 + r)) \leq P$ , a recall is never optimal for the seller even when she is certain that there is a fault in the product. In the rest of the article we will assume that the opposite is true:

$$K(1-p) \frac{\mu_1}{\mu_1 + r} > P. \quad (14)$$

**4. Optimal recall of multiple items.** Our goal is now to solve (8) for general  $N$  under the assumption  $\mu_1 > \mu_0$ ; §4.1 comments on the case  $\mu_1 < \mu_0$ . For ease of notation, let

$$\delta \doteq \mu_0 - \mu_1, \quad c \doteq p \frac{\mu_1}{\mu_0}. \quad (15)$$

We will approach the problem recursively by writing the problem for  $N$  items in terms of the one for  $N - 1$ . The optimization problem (8) is an optimal stopping problem. The state of this control problem at time  $t$  is the pair  $(\Phi_t, N - C_t)$ .  $(\Phi, N - C)$  is a Markov process: this can be proved directly from the definitions of these processes or writing an stochastic differential equation (SDE) that they satisfy; for similar derivations we refer the reader to Peskir and Shiryaev [13, Chapter VI], Bayraktar et al. [1], and Liptser and Shiryaev [8]. Expand the problem so that the initial likelihood ratio  $\Phi$  can start from any point  $\phi \in \mathbb{R}_+$ :

$$V(\phi, N) \doteq \inf_{\tau \in \mathcal{S}} \mathbb{E}_0^\phi \left[ \sum_{k=1}^{C_\tau} e^{-rT_k} K' \Phi_{T_k} + e^{-r\tau} P(1 + \Phi_\tau) \right], \quad (16)$$

where  $\Phi_0 = \phi$ . The result of the minimization in (8) equals  $V((1 - \pi)/\pi, N)$ . The next lemma states the DPE that  $V$  satisfies in the variable  $N$ .

LEMMA 4.1.

$$V(\phi, N) = \inf_{s \in [0, \infty]} \mathbb{E}_0^\phi [e^{-rs} 1_{\{s < T_1\}} P(1 + \Phi_s) + 1_{\{s \geq T_1\}} e^{-rT_1} [K' \Phi_{T_1} + V(\Phi_{T_1}, N - 1)]]. \quad (17)$$

The proof of this result is given in the appendix. Define

$$C'(\mathbb{R}_+) \doteq \left\{ w: \mathbb{R}_+ \rightarrow \mathbb{R}_+, w \text{ concave}, w(0) \leq P, P \leq \frac{dw}{d\phi} \leq K \right\},$$

$C(\mathbb{R}_+) \doteq C'(\mathbb{R}_+) \cap \{w: w(0) = 0\}$  and

$$\begin{aligned} I_k: C'(\mathbb{R}_+) \times [0, \infty] &\rightarrow C'(\mathbb{R}_+), & I_k: (v, s) &\rightarrow w, \\ w(\phi) &\doteq \mathbb{E}_0^\phi [e^{-rs} 1_{\{s < T_1\}} P(1 + \Phi_s) + 1_{\{s \geq T_1\}} e^{-rT_1} [K' \Phi_{T_1} + v(\Phi_{T_1})]], \end{aligned} \quad (18)$$

where  $T_1$  is exponentially distributed with rate  $k\mu_0$ . The definition (7) of the process  $\Phi$  and the exponential distribution of  $T_1$  with rate  $k\mu_0$  imply that the  $w$  in (18) equals

$$w(\phi) = P(1 + \phi e^{k\delta s}) e^{-(k\mu_0 + r)s} + \frac{K' p k \mu_1}{k \mu_1 + r} (1 - e^{-(k\mu_1 + r)s}) \phi + \int_0^s v(\phi e^{k\delta u} c) e^{-(k\mu_0 + r)u} k \mu_0 du. \quad (19)$$

Note that 1) if  $f$  is concave and  $A$  is a linear operator on the domain of  $f$ , then  $f(A \cdot)$  is also concave and 2) the average of concave functions depending on a parameter with respect to a positive measure on the parameter space yields a concave function. These and the last expression for  $w$  imply that the range of  $I_k$  specified in (18) is indeed correct; i.e.,  $I_k(v, s) \in C'(\mathbb{R}_+)$  for all  $s \in \mathbb{R}_+$  and  $v \in C'(\mathbb{R}_+)$ . One can rewrite (17) using  $I_k$  as follows:

LEMMA 4.2.  $V(\cdot, N) \in C(\mathbb{R}_+)$  for all  $N$  and

$$V(\cdot, N) = \inf_{s \in [0, \infty]} I_N(V(\cdot, N - 1), s). \quad (20)$$

PROOF. It follows from (13) that

$$V(\phi, 1) = (1 + \phi) P \wedge (1 - p) K \frac{\mu_1}{\mu_1 + r} \phi. \quad (21)$$

$V(\cdot, 1)$  is concave, satisfies  $P \leq (dV/d\phi)(\cdot, 1) \leq K$  and  $V(0, 1) = 0$ , and is therefore a member of  $C(\mathbb{R}_+)$ . The expectation in (17) equals  $w$  of (18) with  $v = V(\cdot, N - 1)$ . This and (17) imply  $V(\phi, 2) = \inf_{s \in [0, \infty]} I_2(V(\cdot, 1), s)$ . For each  $s$ ,  $I_2(V(\cdot, 1), s)$  is a member of  $C'(\mathbb{R}_+)$ ; it follows that their infimum  $V(\phi, 2)$  is also in  $C'(\mathbb{R}_+)$ .  $V(0, 1) = 0$  implies  $V(0, 2) = 0$  and therefore  $V(\cdot, 2) \in C(\mathbb{R}_+)$ . The rest follows from induction and a repetition of the same argument.  $\square$

Lemma 4.2 is true independent of the the order of the  $\mu_i$ . The assumption  $\mu_1 > \mu_0$  comes into play with the next result, which says that the optimizer of the infimization in (20) must be either 0 or  $\infty$  if  $\mu_1 > \mu_0$ .

THEOREM 4.1. If  $\mu_1 > \mu_0$ ,

$$V(\phi, N) \doteq (1 + \phi) P \wedge [I_N(V(\cdot, N - 1), \infty)](\phi), \quad \phi \in \mathbb{R}_+. \quad (22)$$

**PROOF.** The proof employs only elementary calculus and the monotonicity of the value function. We will prove that

$$\inf_{s \in [0, \infty]} I_N(v, s) = (1 + \phi)P \wedge I_N(v, \infty) \quad (23)$$

for any  $v \in C(\mathbb{R}_+)$ . This and Lemma 4.2 will imply (22). The derivative of the expression for  $I_N(v, s)$  in (19) with respect to  $s$  is

$$w_s = -P(N\mu_0 + r)e^{-(N\mu_0+r)s} - \phi(N\mu_1 + r)Pe^{-(N\mu_1+r)s} + (1 - p)KN\mu_1 e^{-(N\mu_1+r)s}\phi + v(\phi e^{\delta s}c)e^{-(N\mu_0+r)s}N\mu_0.$$

We are interested in the number of times this expression is zero. Factor out the nonzero term  $e^{-(r+N\mu_0)s}$  from the above display:

$$= e^{-(r+N\mu_0)s}(-P(N\mu_0 + r) + \phi[KN\mu_1(1 - p) - P(N\mu_1 + r)]e^{\delta s} + v(\phi e^{\delta s}c)N\mu_0).$$

Let  $A \doteq [KN\mu_1(1 - p) - P(N\mu_1 + r)]$ . The question is the number of times the function  $f(s) \doteq \phi Ae^{\delta s} + v(\phi e^{\delta s}c)N\mu_0$  takes the value  $P(N\mu_0 + r)$ . This count can at most be one if  $f$  is monotone.  $A > 0$  by assumption (14) and  $\delta = \mu_0 - \mu_1 < 0$ ; therefore,  $\phi Ae^{\delta s}$  is decreasing in  $s$ .  $v$  is increasing by assumption and  $c > 0$ . Then  $v(\phi e^{\delta s}c)$  is decreasing in  $s$ . These imply that  $f$  is decreasing in  $s$ . Thus,  $f$  can take the value  $P(N\mu_0 + r)$  at most once. The rest of the argument is the same as the last part of the proof of Lemma 3.1.  $\square$

Theorem 4.1 gives the following optimal recall rule:

1. At each  $T_i$  compute the cost to stop  $(1 + \Phi_{T_i})P$  and the cost to continue

$$(1 - p)K \frac{(N - i)\mu_1}{(N - i)\mu_1 + r} \Phi_{T_i} + \int_0^\infty V(\phi e^{\delta u}c, N - (i + 1))e^{-(N-i)\mu_0 u}e^{-ru} du \quad (24)$$

2. Recall if the cost to stop is lower.

The process  $\Phi$  is completely observable, and its value is available to the seller at all times including the expiration times  $T_1, T_2, \dots, T_N$ .  $N - i = N - C_{T_i}$  is the number of items still functioning after the  $i$ th expiration. The value function  $V(\cdot, \cdot)$ , which is needed to run the optimal recall algorithm, can be computed recursively and numerically with (22) starting from (21). Subsection 7.1 has numerical examples and simulations using (24).

**LEMMA 4.3.** *Under assumption (14), the equation*

$$(1 + \phi)P = [I_k(V(\cdot, k - 1), \infty)](\phi) \quad (25)$$

*has a unique solution  $\phi_k^*$  and (24) can be written as*

$$\text{Recall as soon as } \Phi_{T_i} \geq \phi_{N-C_{T_i}}^*. \quad (26)$$

**PROOF.**  $\phi \rightarrow (1 + \phi)P$  is an affine function of  $\phi$ .  $\phi \rightarrow [I_k(V(\cdot, k - 1), \infty)](\phi)$ , on the other hand, is a concave function of  $\phi$  whose derivative is always greater than  $P$  (see the first term in (24)) and which maps 0 to 0. Then these functions must meet at a unique point  $\phi_k^* > 0$ , and the former is less [greater] than the latter for  $\phi > \phi_k^*$  [for  $\phi < \phi_k^*$ ]. These facts imply the results of this lemma.  $\square$

The inequality  $\delta = \mu_0 - \mu_1 < 0$  means that the likelihood ratio process  $\Phi$  is decreasing between expiration times and increases only when an expiration occurs. Therefore, (26) is equivalent to “Recall as soon as  $\Phi_t \geq \phi_{N-C_t}^*$ .”

**4.1. Discussion.** The process  $\Phi$  is multiplied by  $c$  at every expiration (see (7) and (15)). If  $c < 1$ , that is, if  $p < \mu_0/\mu_1$ ,  $\Phi$  becomes a decreasing process and the optimal recall decision rule reduces to the one obtained for the static case: if the initial probability of fault  $1 - \pi$  is below a threshold value, the seller sells the product and never recalls; otherwise, she doesn’t sell. Note that the condition  $p < \mu_0/\mu_1$  is equivalent to  $1 - p > (\mu_1 - \mu_0)/\mu_1$ .  $1 - p$  can be referred to as the “power” of the inspection, and  $(\mu_1 - \mu_0)/\mu_1$  can be thought of as a measure of the statistical effect of the fault on the post sales environment. The inequality  $1 - p > (\mu_1 - \mu_0)/\mu_1$  can then be interpreted to mean that the test applied to each item after expiration is more effective in catching faults than is studying the statistics of the post sales environment. If such a powerful test exists, the seller should make it a part of the quality control process before sales. Thus,  $p > \mu_0/\mu_1$  is not an unreasonable assumption, at least within the boundaries of our model. Under this assumption, a dynamic recall process is always optimal.

The case  $\mu_1 < \mu_0$  can be treated with the same tools as those used to handle  $\mu_0 < \mu_1$ . The optimal recall strategy will be a generalization of the one given in §3.1 for the same case. The value function will consist

of two pieces: as in the case of  $\mu_1 > \mu_0$ , it will equal  $P(1 + \phi)$  above a threshold  $\phi^*$ . This affine part will correspond to recalling immediately if the likelihood ratio goes above  $\phi^*$ . An algorithm based on (20) can be constructed to approximate  $V(\cdot, \cdot)$ .

An important statistic about the optimal recall rule is the probability that it will lead to an unnecessary recall, i.e., a recall when there is no fault. We haven't yet studied the computation of this probability. See §7 for comments on it in the context of the examples presented there.

**5. Convergence and the limit recall problem.** Our goal in the present section is to let  $N \rightarrow \infty$  and obtain a simpler limit recall problem and relate its solution to that of the original/prelimit recall problem. Remember that the recall threshold for the likelihood ratio function depends on  $N$ . The asymptotic analysis that follows implies that these thresholds converge to the fixed threshold of a limit problem. Thus, if  $N$  is large, this constant limit recall threshold can be used to decide when to recall, rather than recomputing a new threshold at the expiration of each item. This is one of the practical reasons to conduct an asymptotic analysis.

Define  $T: C(\mathbb{R}_+) \times \mathbb{N} \rightarrow C(\mathbb{R}_+)$  as

$$T(v, n) \doteq (1 + \phi P) \wedge \left[ \frac{n\mu_1}{r + n\mu_1} K(1 - p)\phi + \int_0^\infty v(\phi e^{n\delta u} c) e^{-n\mu_0 u} e^{-ru} n\mu_0 du \right]. \quad (27)$$

Note that (22) is

$$V(\cdot, N) = T(V(\cdot, N - 1), N). \quad (28)$$

Change the variable  $u$  to  $nu$  in (27) to rewrite it as

$$T(v, n) \doteq (1 + \phi P) \wedge \left[ \frac{n\mu_1}{r + n\mu_1} K(1 - p)\phi + \int_0^\infty v(\phi e^{\delta u} c) e^{-\mu_0 u} e^{-(r/n)u} \mu_0 du \right]. \quad (29)$$

As  $n \rightarrow \infty$  one expects the operator  $T(\cdot, n)$  to converge to  $T: C(\mathbb{R}_+) \rightarrow C(\mathbb{R}_+)$ ,

$$T(v) \doteq (1 + \phi P) \wedge \left[ K(1 - p)\phi + \int_0^\infty v(\phi e^{\delta u} c) e^{-\mu_0 u} \mu_0 du \right]. \quad (30)$$

The plan for the convergence analysis is as follows:

- (i) Write down a control problem corresponding to  $T$  and use it to identify a fixed point of  $T$ .
- (ii) Show  $T$  has a *unique* fixed point in  $C(\mathbb{R}_+)$  and that one can obtain it by repeatedly applying  $T$  to any element of  $C(\mathbb{R}_+)$ .
- (iii) Show that for any  $v \in C(\mathbb{R}_+)$ ,  $T(\cdots T(T(v, 1), 2), \dots, n)$  and  $T^n(v)$  get arbitrarily close to each other uniformly on compact sets.
- (iv) Conclude that  $V(\phi, N)$  converges to the unique fixed point of  $T$ .

The next subsection implements the first two steps and the one after it implements the rest.

**5.1. The limit recall problem.** Let  $M$  be a Poisson process with rate  $\mu_0$  and jump times  $S_k$  and

$$\begin{aligned} \Psi_t &\doteq \Psi_{S_n} e^{\delta(t-S_n)}, & t \in [S_n, S_{n+1}), \\ \Psi_{S_n} &\doteq c\Psi_{S_n-}, & \Psi_0 \doteq \phi. \end{aligned} \quad (31)$$

Let  $\mathcal{R}$  be the set of all stopping times of the filtration generated by  $M$  and define the optimal stopping problem

$$V(\phi) \doteq \inf_{\tau \in \mathcal{R}} \mathbb{E}_\phi \left[ \sum_{k=1}^{M_\tau} K' \Psi_{S_k} + 1_{\{\tau < \infty\}} P(1 + \Psi_\tau) \right]. \quad (32)$$

The underlying jump process of (32) is Poisson with a constant jump rate  $\mu_0$ . The cost structure of (32) is exactly the same as that of the prelimit problem except that it involves no interest rate and is determined by a limit likelihood ratio process  $\Psi$ , which decreases with the same rate at all times.

The change of variable that connects (27) to (29) implies that the events that happen in time interval  $[0, 1]$  for the limit problem correspond to the events that happen in the time interval  $[0, 1/N]$  in the original recall problem. Therefore, the control problem (32) is an approximation of what happens in the early stages of the original recall problem when the number of items  $N$  is large.

The cost of not stopping in (32) is

$$V_0(\phi) = \mathbb{E}_\phi \left[ \sum_{k=1}^{\infty} K' \Psi_{S_k} \right] = \frac{1-p}{p} K \phi \sum_{k=1}^{\infty} \mathbb{E}_\phi [\Psi_{S_k}] = \frac{1-p}{p} K \sum_{k=1}^{\infty} p^k = K\phi,$$

which also implies  $0 \leq V(\phi) \leq K\phi$ .

**REMARK 5.1.** The same result can also be inferred from the original recall problem.  $N = \infty$  has two implications: 1) if no recall is made, and unless  $p = 0$ , one of the inspections will eventually reveal the fault and 2) this revelation will take place very quickly, and therefore the interest rate will play no role.

Let  $V_n$  be the value function of (32) when stopping is allowed only before the  $n$ th jump of  $M$ . We have just seen that  $V_0 = K\phi$ . The first result of this section is the following:

**LEMMA 5.1.** Let  $V$  and  $V_n$  be as above.  $V_n \searrow V$  and  $V$  is a fixed point of the operator  $T$  of (30); i.e., it satisfies  $V = T(V)$ .

See the appendix for a proof of this lemma.

**REMARK 5.2.** In what follows we will often write  $|f - g|$  to denote the function mapping  $\phi$  to  $|f(\phi) - g(\phi)|$ . For example,  $|T_n(v_1) - T_n(v_2)|$  is the function mapping  $\phi$  to  $|w_1(\phi) - w_2(\phi)|$  where  $w_i = T_n(v_i)$ ,  $i = 1, 2$ .

The next step is to show that  $V$  is the unique fixed point of  $T$  and that iterations of  $T$  converge to  $V$ .

**THEOREM 5.1.**  $V$  of (32) is the unique fixed point of  $T$  in  $C(\mathbb{R}_+)$  and

$$|T^n(v) - V|(\phi) \leq (K - P)p^n\phi \quad (33)$$

for any  $v \in C(\mathbb{R}_+)$  and  $n \in \mathbb{N}$ .

**PROOF.** For any  $v_2, v_3 \in C(\mathbb{R}_+)$  with  $|v_2(\phi) - v_3(\phi)| \leq c_1\phi$  the following sequence of inequalities holds:

$$\begin{aligned} |[T(v_2) - T(v_3)](\phi)| &\leq \int |v_2 - v_3|(\phi ce^{\delta u})e^{-\mu_0 u}\mu_0 du \leq \int c_1\phi ce^{\delta u}e^{-\mu_0 u}\mu_0 du \\ &\leq pc_1 \int \phi e^{-\mu_1 u}\mu_1 du = pc_1\phi. \end{aligned} \quad (34)$$

Now suppose  $|T^{n-1}(v_2) - T^{n-1}(v_3)|(\phi) \leq p^{n-1}c_1\phi$ . Then

$$|T^n(v_2) - T^n(v_3)| = |T(T^{n-1}(v_2)) - T(T^{n-1}(v_3))| \leq pp^{n-1}c_1 = p^n c_1, \quad (35)$$

where the last inequality is implied by (34). Induction now implies that (35) is true for all  $n$ . By definition  $|v - v_1|(\phi) \leq (K - P)\phi$  for any  $v, v_1 \in C(\mathbb{R}_+)$ . This and (35) give  $|T^n(v) - T^n(v_1)| \leq p^n(K - P)\phi$  for all  $v, v_1 \in C(\mathbb{R}_+)$ . By Lemma 5.1,  $T(V)$  and therefore all  $T^n(V)$  equal  $V$ . Choosing  $v_1 = V$  in the last inequality gives (33).  $\square$

Lemma 5.1 gives the following optimal recall rule for the limit problem: at each expiration  $S_i$ , compute the cost to stop  $(1 + \Psi_{S_i})P$  and the cost to continue  $(1 - p)K\Psi_{S_i} + \int_0^\infty V(\Psi_{S_i} e^{\delta u} c) e^{-\mu_0 u} du$ ; recall if the cost to stop is lower. Let  $\phi^*$  be the unique solution of

$$(1 + \phi)P = \int V(\phi e^{\delta u} c) e^{-\mu_0 u} \mu_0 du. \quad (36)$$

An argument similar to the one given in the proof of Lemma 4.3 implies that the limit optimal recall decision rule is equivalent to “recall as soon as  $\Psi_{S_i} \geq \phi^*$ .” The results of the next section will show that when  $k$  is large, the threshold  $\phi^*$  is a good approximation of the threshold  $\phi_k^*$  used in the prelimit algorithm (26).

**5.2. Relating  $T(\cdot, n)$  to  $T$ .** The prelimit recall rules are determined by the recursion (28), which is in terms of the operators  $T(\cdot, n)$  defined in (27). The next two results relate these to  $T$ .

**PROPOSITION 5.1.**  $|T(v, n) - T(v_1)|(\phi) \leq p(c_1 + (K' + K)(r/(r + n\mu_1)))\phi$  for any  $v, v_1 \in C(\mathbb{R}_+)$  that satisfy  $|v - v_1|(\phi) \leq c_1\phi$ .

**PROOF.** The definitions (29) and (30), the assumption (14), and  $v, v_1 \in C(\mathbb{R}_+)$  imply

$$\begin{aligned} |T(v, n) - T(v_1)|(\phi) &\leq \frac{r}{r + n\mu_1} K(1 - p)\phi + \int |v - v_1|(\phi e^{\delta u} c) e^{-\mu_0 u} \mu_0 du \\ &\quad + \int (1 - e^{-(r/n)u})v(\phi e^{\delta u} c) e^{-\mu_0 u} \mu_0 du. \end{aligned}$$

$v \in C(\mathbb{R}_+)$  implies in particular that  $v(\phi) \leq K\phi$ . This and the assumption  $|v - v_1|(\phi) \leq c_1\phi$  imply that the last expression is bounded above by

$$\leq \frac{r}{r + n\mu_1} K(1 - p)\phi + pc_1\phi + K\phi p \frac{r}{r + n\mu_1} = p \left( c_1 + (K' + K) \frac{r}{r + n\mu_1} \right) \phi. \quad \square$$

Our main convergence result is

**THEOREM 5.2.**  $|V(\phi, N) - V(\phi)| \leq c_N \phi$ , where  $c_N \rightarrow 0$ .

**PROOF.** Let  $c_1 \doteq (K - P)$  and  $c_{n+1} \doteq p(c_n + (K' + K)(r/(r + n\mu_1)))$ . A representation of  $c_n$  is

$$c_n = p^{n-1} c_1 + (K' + K)r \left( \sum_{i=1}^{n-1} p^i \frac{1}{r + (n-i)\mu_1} \right). \quad (37)$$

The last sum is a weighted average of the terms of a sequence converging to 0. This implies that  $c_n$  converges to 0. Note that  $|V(\cdot, 1) - V|(\phi) \leq (K - P)\phi$  because both functions are members of  $C(\mathbb{R}_+)$ . This, Proposition 5.1, and (28) imply  $|V(\phi, N) - V(\phi)| \leq c_N \phi$ , which is what we wanted to prove.  $\square$

Theorem 5.2 implies that the prelimit recall likelihood ratio thresholds (25) converge to the limit recall likelihood ratio threshold (36). Thus, if the number of items  $N$  is large, one can use the limit recall threshold  $\phi^*$  of (36) in the prelimit recall rule (26) instead of  $\phi_k^*$  of (25). This approximation gives a stationary recall rule. The rate of convergence of  $V(\cdot, N)$  to  $V(\cdot)$  is determined by the rate at which the sequence of real numbers  $\{c_n\}$  converges to zero.

**LEMMA 5.2.**  $0 < \underline{\lim}_{n \rightarrow \infty} nc_n \leq \overline{\lim}_{n \rightarrow \infty} nc_n < \infty$ .

**PROOF.**  $nc_n$  and  $e_n \doteq (K' + K)r(\sum_{i=1}^{n-1} p^i(n/(r + (n-i)\mu_1)))$  have identical limits.  $e_n$  is bounded above by  $(K' + K)r(\sum_{i=1}^{n/2} p^i(n/(r + (n/2)\mu_1)) + \sum_{i=n/2+1}^n p^i(n/(r + \mu_1)))$ . The first sum here is bounded and the second sum converges to 0 with  $n$ , which implies  $\overline{\lim} nc_n < \infty$ . On the other hand, bounding  $e_n$  from below by its last term implies  $\underline{\lim} nc_n > (K' + K)rp(1/\mu_1)$ .  $\square$

**6. Two extensions.** This section extends the product recall model of §2 in two directions: 1) an expired item is inspected only when its lifetime turns out to be less than the expected lifetime  $1/\mu_0$  of the items when no fault exists and 2) after each inspection, the seller conducts her own private inspection; the inspection reveals the fault, if the fault is present, with probability  $1 - q$ .

**6.1. Buyer inspects only when the item dies before time  $1/\mu_0$ .** Let us now suppose that an inspection takes place only when the item dies before its initial expected expiration time. This condition appears natural: if a product lasts longer than expected, its owner may think that there is no reason to inspect it for a fault. Remember that in our model all of the items are sold at the same time. Thus, the new condition that we would like to introduce for inspections amounts to setting a single deadline for all items: an item is inspected if it cannot survive up to this fixed deadline. Thus, to deal with this new feature, it is enough to introduce a “time-to-deadline” parameter  $\eta \in \mathbb{R}_+$ ; no public inspections occur if an item survives  $\eta$  more units of time. We are especially interested in the case when  $\eta = 1/\mu_0$ . The optimal recall problem is now

$$\inf_{\tau \in \mathcal{T}} \mathbb{E} \left[ 1_{\{d=0\}} e^{-r\tau} P + 1_{\{d=1\}} 1_{\{T_W < \eta\}} (1_{\{\tau < T_W\}} e^{-r\tau} P + 1_{\{\tau \geq T_W\}} e^{-rT_W} K) \right]. \quad (38)$$

An argument that parallels the one that connects (3) and (8) implies that

$$\inf_{\tau \in \mathcal{T}} \mathbb{E}_0 \left[ \sum_{k=0}^{C_\tau} e^{-rT_k} 1_{\{T_k < \eta\}} K' \Phi_{T_k} + e^{-r\tau} P(1 + \Phi_\tau) \right]$$

is a representation of the optimal stopping problem (38). The value function  $V(\phi, \eta, N)$  of this problem satisfies

$$V(\phi, \eta, N) = \inf_{s \in [0, \infty]} \mathbb{E}_0^\phi [e^{-rs} 1_{\{s \leq \eta\}} 1_{\{s < T_1\}} P(1 + \Phi_s) + 1_{\{T_1 \leq \eta \wedge s\}} e^{-rT_1} [K' \Phi_{T_1} + V(\Phi_{T_1}, \eta - T_1, N - 1)]]. \quad (39)$$

This is a generalization of (17). Define  $J_k(v, s): (v(\cdot, \cdot), s) \rightarrow w$  with

$$w(\phi, \eta) \doteq 1_{\{s \leq \eta\}} Pe^{-(r+k\mu_0)s} (1 + \phi e^{k\delta s}) + (1 - e^{-(r+k\mu_1)s}) \frac{pK'k\mu_1\phi}{r + k\mu_1} + \int_0^{s \wedge \eta} v(\phi e^{k\delta t} c, \eta - t) e^{-(k\mu_0+r)t} k\mu_0 dt.$$

The DPE (39) in terms of the operator  $J_k$  is  $V(\phi, \eta, N) = \inf_{s \in [0, \infty]} J_N(V(\cdot, \cdot, N - 1), s)$ . The value function  $V(\phi, \eta, N)$  is an increasing function of  $\eta$ . This fact allows one to generalize Theorem 4.1 to the present setup as

$$\begin{aligned} V(\phi, \eta, N) = (1 + \phi)P \wedge & \left[ (1 - e^{-(r+N\mu_1)\eta}) \frac{(1-p)KN\mu_1}{r + N\mu_1} \phi \right. \\ & \left. + \int_0^\eta V(\phi e^{N\delta t} c, \eta - t, N - 1) e^{-(N\mu_0+r)t} N\mu_0 dt \right]. \end{aligned} \quad (40)$$

Define  $\phi_k(\eta)$  to be the unique solution of

$$(1 + \phi)P = \left[ (1 - e^{-(r+k\mu_1)\eta}) \frac{(1-p)Kk\mu_1}{r+k\mu_1} \phi + \int_0^\eta V(\phi e^{k\delta t} c, \eta-t, k-1) e^{-(k\mu_0+r)t} k\mu_0 dt \right], \quad (41)$$

if the solution exists; otherwise, set  $\phi_k(\eta)$  to  $\infty$ . The DPE (40) gives the following optimal recall rule:

$$\text{Recall when } \Phi_t \geq \phi_{N-C_t}(\eta - t). \quad (42)$$

It follows from its definition that  $\phi_n(\eta)$  is decreasing in  $\eta$  and  $\phi_n(\eta) \rightarrow \infty$  as  $\eta \rightarrow 0$ . Therefore, one can replace (42) with

$$\text{Recall when } \Phi_t \geq \phi_{N-C_t}(\eta - T_{C_t}). \quad (43)$$

In (42), the Equation (41) that defines  $\phi_n(\eta)$  needs to be solved continuously, whereas in (43), this needs to be done only at the expiration times.

In (40), change the variable  $\eta$  to  $\bar{\eta} = N\eta$  and  $t$  to  $Nt$ . After these, letting  $N \rightarrow \infty$  on both sides of (40) suggests  $\lim_N V(\cdot, \eta, N) = V(\cdot)$ , where  $V$  is the unique fixed point of  $T$  (see Lemma 5.1); the methods of §5 can be used to prove this. Thus, when  $N$  is large, the steady state recall rule will be an almost optimal recall rule for the present problem.

**6.2. The seller conducts her private inspection.** Let us go back to the original recall model of §2 and consider a new extension. Suppose now that after each expiration two inspections happen: a public one, as in the previous sections, and a private/internal inspection conducted by the seller. The seller's inspection reveals the fault, if it exists, with probability  $1 - q$ . As with public inspections, it is assumed that internal inspections yield no false positives. The seller has the right to keep private the information these inspections reveal to her. The problem is the same as before: what is the optimal time for the seller to recall?

One way to think about the present problem is to imagine it to consist of two phases: before and after a private inspection reveals a fault. The first phase proceeds exactly as before: there is a likelihood ratio process that represents all of the information up to present and all decisions are made based on it. If and when a private inspection detects a fault, the seller knows perfectly that there is a fault; this makes the likelihood ratio process immaterial (equivalently, the likelihood ratio process becomes  $\infty$  and stays there for the rest of the problem). The question for the seller then reduces to the following: given the number of remaining public inspections ahead, whether and when to recall? Let us first study the problem faced by the seller in the second phase.

Let  $n$  denote the number of items that continue to function and remain with their buyers at the moment when a private inspection reveals a fault. The memoryless property of the exponential distribution implies that whatever happens after this moment can be modeled with the model of §2, with the additional assumption that  $d = 1$  with probability 1; i.e.,  $\pi = 0$  and the seller knows with certainty that there is a fault in the sold products. Therefore, we can imagine that the whole problem starts afresh at this moment, i.e., that the time is  $t = 0$ , the seller has just sold  $n$  items, and the parameter  $\pi$  has the value 0.

Because the seller knows that  $d = 1$ , the only information she needs to decide when to recall is the number of items still functioning. Therefore, she will be optimizing over the set of stopping times  $\mathcal{S}''$  of the filtration  $\{\sigma(C_{T_i}, T_i \leq t)\}$ . The optimal recall problem of the seller in this setting is

$$U(n) \doteq \inf_{\tau \in \mathcal{S}''} \mathbb{E}[(1_{\{\tau < T_W\}} e^{-r\tau} P + 1_{\{\tau \geq T_W\}} e^{-rT_W} K)]; \quad (44)$$

one obtains (44) from (2) by setting  $\pi = 0$  and replacing  $\mathcal{S}'$  with  $\mathcal{S}''$ . The value function  $U$  of this problem is a sequence of real numbers. The DPE for (44) (which is a special case of (23)) gives the following recursion for  $U$ :

$$\begin{aligned} U(n) &= P \wedge \left[ \int e^{-rt} e^{-n\mu_1 t} [(1-p)K + pU(n-1)] n\mu_1 dt \right] \\ &= P \wedge [(1-p)K + pU(n-1)] \frac{n\mu_1}{n\mu_1 + r}. \end{aligned} \quad (45)$$

When  $n = 0$ , there is nothing sold and recalled, which gives the initial value  $U(0) = 0$ . This, (45), and (14) imply that  $U(n) = P$  for all  $n > 0$ . Thus we have the following:

**LEMMA 6.1.** *Suppose that (14) holds. Suppose further that (1) a private inspection has just revealed a fault and (2) at least one of the sold items is still functioning. Then it is optimal for the seller to recall immediately.*

Redefine the likelihood ratio process  $\Phi$  of (7) as follows:

$$\begin{aligned}\Phi_t &\doteq \Phi_{T_n} e^{-(N-n)(\mu_1-\mu_0)(t-T_n)}, \quad t \in [T_n, T_{n+1}), \\ \Phi_{T_n} &\doteq pq \frac{\mu_1}{\mu_0} \Phi_{T_n-}, \quad \Phi_0 \doteq \frac{1-\pi}{\pi}.\end{aligned}\tag{46}$$

The arguments that connect (2) to (8) and the idea of thinking of the problem in two phases as outlined above lead to the following representation of the extended recall problem:

$$\bar{V}(\phi, N) \doteq \inf_{\tau \in \mathcal{S}} \mathbb{E}_0 \left[ \sum_{k=0}^{C_\tau} e^{-rT_k} \Phi_{T_k} \left[ \frac{1-p}{pq} K + \frac{1-q}{q} U(N - C_{T_k}) \right] + e^{-r\tau} P(1 + \Phi_\tau) \right], \tag{47}$$

where  $N$  is the number of items sold at time 0 and  $\phi = (1-\pi)/\pi$  is the initial likelihood ratio. That  $U(n) = P$  for all  $n > 0$  implies that (47) is in almost the same form as (8) and hence that its solution is identical to that of the latter. Therefore, we simply state the final result. Define, for  $k > 1$ ,  $I_k(v, \infty)$ :  $v \rightarrow w$  with

$$w(\phi) \doteq \frac{k\mu_1}{r+k\mu_1} pq \left[ \frac{1-p}{pq} K + \frac{1-q}{q} P \right] \phi + \int_0^\infty v \left( \phi e^{k\delta t} \frac{\mu_1}{\mu_0} pq \right) e^{-(k\mu_0+r)t} k\mu_0 dt.$$

The value function  $\bar{V}$  of (47) satisfies

$$\bar{V}(\phi, 1) = V(\phi, 1), \quad \bar{V}(\phi, N) = (1+\phi)P \wedge [I_N(\bar{V}(\cdot, N-1), \infty)](\phi), \quad \phi \in \mathbb{R}_+ \tag{48}$$

for  $N > 1$ . The DPE (48) is the extension of (22) to the current setting. The first part of (48) handles the case  $N = 1$  separately because  $U(0) = 0$ . This corresponds to the following fact: if the seller sold only one item (or only one functioning item is remaining), the internal inspection is unnecessary and has no impact on the timing of the recall. Thus, for  $N = 1$ , the present model reduces to the original model.

Let us now combine (48) with Lemma 6.1 to obtain the optimal recall decision rule for the extended problem. Let  $\bar{\phi}_k^*$  be the unique solution of

$$(1+\phi)P = [I_k(\bar{V}(\cdot, k-1), \infty)](\phi). \tag{49}$$

Extend the dynamics of  $\Phi$  in (46) to include  $\Phi_{T_n} = \infty$  if the  $n$ th internal inspection reveals fault. The optimal recall rule is “recall as soon as  $\Phi_t \geq \bar{\phi}_{N-C_\tau}^*$ .”

**REMARK 6.1.** The dynamics (22) and (48) are the same. Therefore, the asymptotic analysis of §5 applies to (47) without modification.

**REMARK 6.2.** One can combine the models of the last two subsections. The solution of the resulting model will be analogous to the solutions we have presented above.

**7. Numerical examples.** The following parameter values will be used in all of the examples in this section:

$$\begin{aligned}N &= 15, & K &= 100, & P &= 4, & \pi &= 0.99, & p &= \frac{9}{10}, \\ \mu_0 &= \frac{1}{4}, & \mu_1 &= \frac{1}{2}, & r &= \frac{1}{10}, & d &= 1.\end{aligned}\tag{50}$$

The last of these means that all of the simulations are conditioned on  $d = 1$ , i.e., that there is a manufacturing fault, unobservable at time 0. All of the integral equations are discretized for numerical evaluation. Because the value function is affine above a bounded threshold, any level of precision can be attained by working with a fine enough grid over a compact interval for the variable  $\phi$ . We used the Octave computing environment (Eaton [6]) to carry out the numerical computations reported in this section.

**7.1. First model.** Let us begin with the model of §2. A simulation of the post sales environment with parameter values (50) is given in the left panel of Figure 3. The 15 simulated expiration times are (iid, exponentially distributed with rate  $\mu_1$ )

$$\begin{array}{cccccccc}0.097, & 0.131, & 0.220, & 0.319, & 0.674, & 0.772, & 0.834, & 0.866 \\ 0.996, & 1.163, & 1.179, & 1.709, & 1.729, & 1.831, & 5.198.\end{array}\tag{51}$$

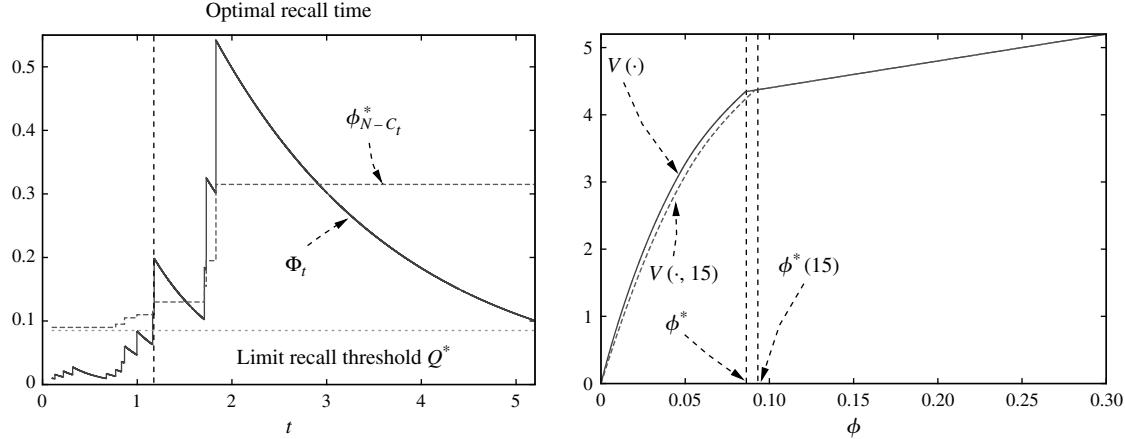


FIGURE 3. Simulated sample paths of  $\Phi$  and  $t \rightarrow \phi_{N-C_t}^*$  when there is a fault; the value function  $V(\cdot, 15)$  and  $V(\cdot)$ .

The jagged curve in this figure is the sample path of  $\Phi$ .  $C_t$ , the number of expirations at time  $t$ , is not depicted but can be observed by counting the jumps of  $\Phi$ . The increasing curve is the sample path of  $t \rightarrow \phi_{N-C_t}^*$ , the recall likelihood ratio thresholds. The straight line is the limit recall threshold  $\phi^*$ , which one obtains by solving (36) numerically. The times between the jumps of  $\Phi$  are the time intervals between consecutive expiration times, which are listed in (51).  $\Phi$  is computed from these using (7).  $t \rightarrow \phi_{N-C_t}^*$  ( $\phi^*$ , respectively) is computed by numerically solving (25) [(36)]; this requires the knowledge of  $V(\cdot, \cdot)$  [ $V(\cdot)$ ], which is computed by iterating (22) [(54)]. The right panel of Figure 3 depicts the value functions  $V(\cdot, 15)$  and the limit value function  $V(\cdot)$ .

Because  $\mu_1 > \mu_0$ , the likelihood ratio  $\Phi$  is decreasing between expiration times. At each expiration, the value of  $\Phi$  is multiplied by  $p\mu_1/\mu_0 = 1.8$ . In this example,  $\Phi$  goes above  $t \rightarrow \phi_{N-C_t}^*$  at the 11th expiration and therefore that is the optimal time to recall for this sample. None of the first 11 expirations will reveal the fault with probability  $0.9^{11} \approx 0.3$ . Then, conditioned on the expiration times listed in (51), the recall will be successful with the same probability. Intuitively, 0.3 sounds like a low probability; however, one must remember that the initial fault probability is only 1%.

**7.2. First extension.** In this model, an inspection occurs only if a purchased item expires earlier than its expected lifetime  $1/\mu_0$ . The value function now is a function of three variables  $(\phi, \eta, n)$ :  $\phi$  is the likelihood ratio of fault,  $\eta$  is the time that remains until  $1/\mu_0$ , and  $n$  is the number items still functioning. As before, the parameter values are those listed in (50). A simulation of the post sales environment under this model is depicted in Figure 4. The fifteen simulated expiration times are

$$0.133, \quad 0.177, \quad 0.205, \quad 0.225, \quad 0.346, \quad 0.357, \quad 0.531, \quad 0.549, \\ 0.916, \quad 1.082, \quad 3.075, \quad 3.210, \quad 3.799, \quad 4.784, \quad 9.546.$$

The increasing curve in Figure 4 that ends around  $t \approx 4 = 1/\mu_0$  is the likelihood ratio threshold  $t \rightarrow \phi_{N-C_t}(\eta - T_{C_t})$ ; see (43). That  $t \rightarrow \phi_{N-C_t}(\eta - T_{C_t})$  vanishes around  $1/\mu_0$  is a consequence of  $\lim_{\eta \rightarrow 0} \phi_n(\eta) = \infty$ . The second increasing curve in the same figure is the threshold for the original recall model; it is included so that the reader can see the effect of the  $\eta$  parameter on the recall thresholds. The straight line, as before, is the limit recall threshold.

The threshold process  $t \rightarrow \phi_{N-C_t}(\eta - T_{C_t})$  is computed by solving (41); this requires the knowledge of the value function  $(\phi, \eta, n) \rightarrow V(\phi, \eta, n)$  and this is computed by iterating (40). Two graphs of  $V(\cdot, \cdot, 15)$  are depicted in Figure 5.

In the present simulation, the expirations occur very rapidly, and at the fifth expiration it becomes optimal to issue a recall. This recall is successful if none of the public inspections reveals the fault earlier, which happens with probability  $0.9^5 \approx 0.6$ .

**7.3. Second extension.** In the model of §6.2, the seller privately inspects the expired product after each public inspection.  $1 - q$  is the probability that the seller's inspection reveals the fault, in case it exists. Let us assign 0.85 to  $q$  for the purposes of the simulation below. Note that  $q < p$ ; i.e., the seller's inspection is more

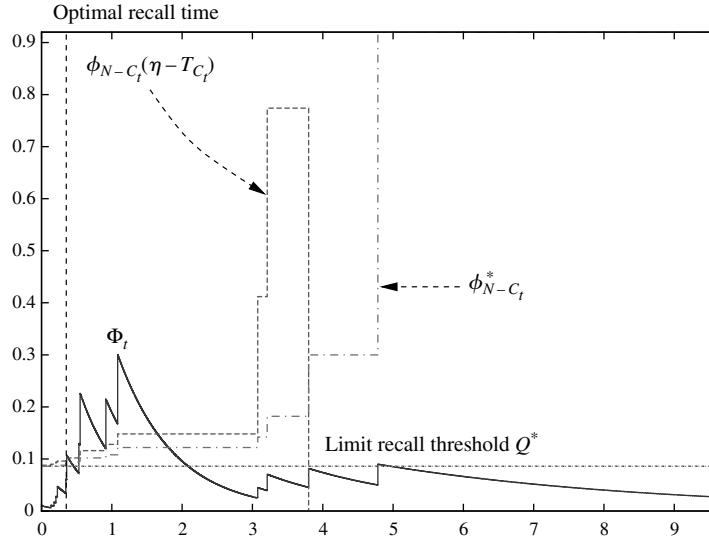


FIGURE 4. Simulated sample paths of  $\Phi$ ,  $t \rightarrow \phi_{N-C_t}(\eta - T_{C_t})$ .

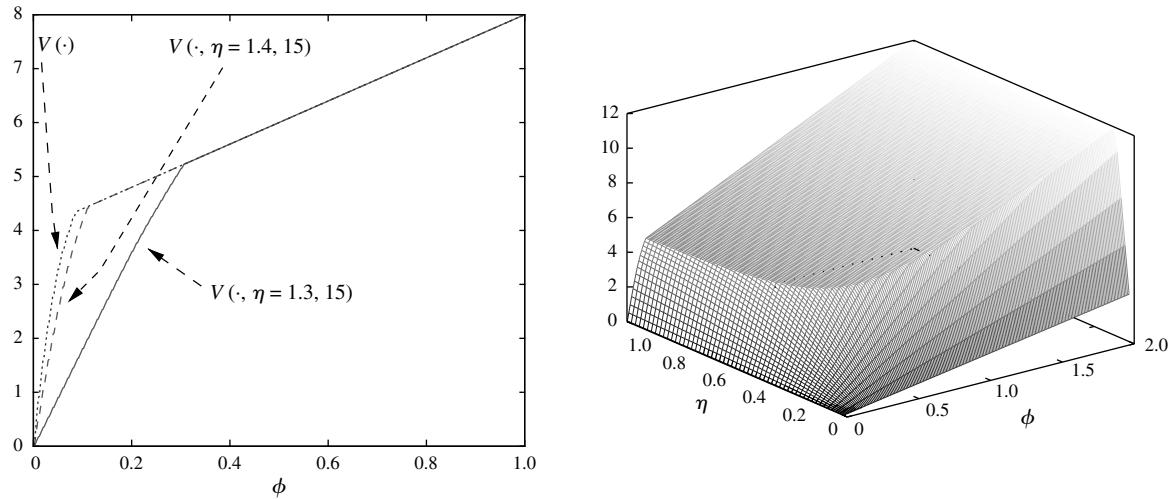


FIGURE 5. The value function  $V(\cdot, \cdot, 15)$ .

effective in catching faults than is the public inspection. The rest of the parameter values are listed in (50). The fifteen simulated expiration times are

$$\begin{aligned} 0.008, & \quad 0.030, & \quad 0.138, & \quad 0.152, & \quad 0.194, & \quad 0.197, & \quad 0.368 \\ 0.404, & \quad 0.604, & \quad 0.667, & \quad 0.707, & \quad 0.812, & \quad 1.368, & \quad 1.642, & \quad 3.041. \end{aligned} \tag{52}$$

In addition, we simulate the results of the 15 private inspections that the seller conducts:

$$1, \quad 1, \quad 0, \quad 1, \quad 1, \quad 1. \tag{53}$$

This is a sequence of iid Bernoulli trials with distribution  $(1 - q, q)$  (we remind the reader that 1 denotes an inspection result that indicates no fault). The likelihood ratio process  $\Phi$ , depicted in the left panel of Figure 6, is computed from the expiration times in (52) using (46). There are two threshold processes depicted in this figure: the threshold process  $t \rightarrow \bar{\phi}_{N-C_t}^*$  of the extended model and the threshold process  $t \rightarrow \phi_{N-C_t}^*$  of the original model. The first line of (48) implies that the threshold processes become equal to each other when only one item remains functioning.

The threshold process for the present model is computed by solving (49);  $\bar{V}(\cdot, \cdot)$  is computed by iterating (48). The value function  $\bar{V}(\cdot, 15)$  of the extended model and the value function  $V(\cdot, 15)$  of the original

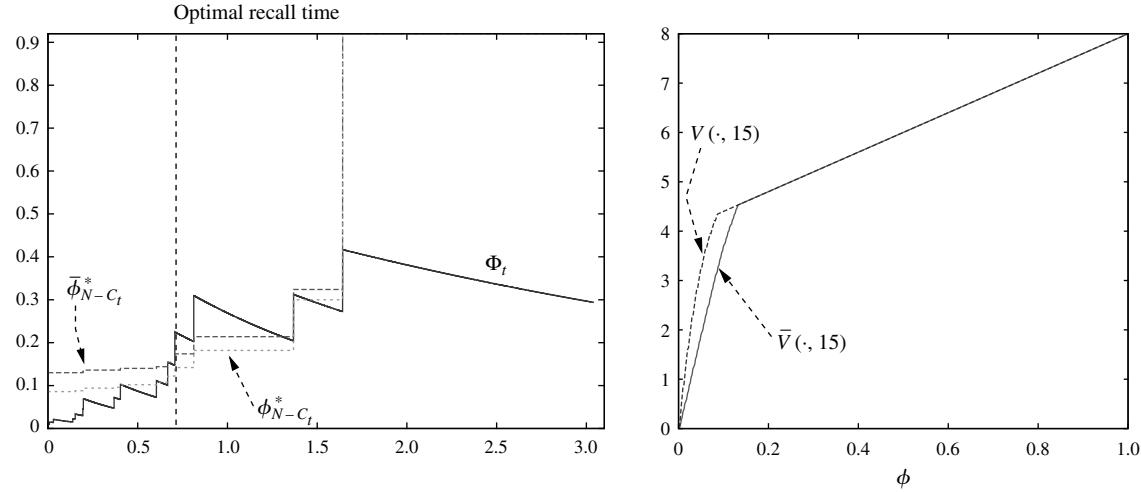


FIGURE 6. Sample paths of  $\Phi$  and  $t \rightarrow \bar{\phi}_{N-C_t}^*$ ; the value functions  $\bar{V}(\cdot, 15)$  and  $V(\cdot, 15)$ .

model are depicted in the right panel of Figure 6. Because the inspections have no cost,<sup>4</sup> the additional internal inspections allow the seller to reduce her expected cost of recall; i.e.,  $\bar{V}$  is always less than  $V$ , and this implies  $\bar{\phi}^*(n) \geq \phi^*(n)$ .

In the present simulation,  $\Phi$  goes above the recall threshold at the 11th expiration, and therefore right after this expiration is the optimal time to recall. Note that the first 11 simulated internal inspections whose results are listed in (53) are not able to catch the fault. Therefore, in this simulation, the expiration times prove to be more useful than the internal inspections in catching the fault.

**8. Conclusion.** Further thoughts on the study we presented in this article, including possible directions for further research, are as follows.

Our assumption that expiration times are exponentially distributed corresponds to a constant hazard rate, i.e., a product that doesn't age. If  $\mu_1$  and  $\mu_2$  are not very small, as long as there are many items functioning, the overall expiration rate will be high and many of the expirations will occur when the items are relatively young. Thus, under these conditions, assuming a constant hazard rate may be reasonable. Here, two questions come to mind: under what conditions does it become essential to take into account the aging process? And if this is to be done, what would be an appropriate model?

An important issue is the determination of  $\mu_1$ ,  $\mu_0$ ,  $\pi$ , and  $p$ . In practice, one will estimate these parameters from historical data. An idea that may be further explored is a recall model in which these parameters are assumed to be random as well; see Bayraktar et al. [2], in which a rate parameter is assumed to be random with a known prior distribution.

The effect of the interest rate  $r$  on the recall decision can be best seen when there is only one item for sale, i.e., in the static case treated in §3. It is clear from (10) and (13) that  $r$ 's effect on the recall decision depends on the ratios  $r/\mu_1$  and  $r/\mu_0$ . Typically, the interest rate is below 0.1 per year. When taken as the expiration rate of a product, this value corresponds to a lifetime of 10 years on average. Thus, unless the product lasts in the range of decades, the interest rate plays a minor role in the recall decision process.

We have formulated the recall problem as one of optimal stopping, which naturally led to a Bayesian likelihood ratio process. One could have directly used a Bayesian framework to tackle the problem. See, for example, Paté-Cornell [11], which takes this approach in the context of another problem. One advantage of the optimal stopping formulation is its ability to naturally combine *statistical data* with *financial data* in the decision process. The financial data in our model are  $r$ ,  $P$ , and  $K$ . Within the optimal stopping framework, these parameters have natural roles in the determination of optimal product recall thresholds.

Most products are sold over a time span rather than all at once. The framework in the present article can be used recursively as a building block to model continuing sales. We hope to do this in future work.

In our models the inspections yield no false positives. In real life, of course, tests will give false positives. A model that allows false positives will have to specify what happens when a false positive occurs. It seems possible to build such models and solve them with methods similar to those used and developed in this work.

<sup>4</sup> A model in which inspections do have a cost is an interesting direction for further research.

The most sophisticated model of the post sales environment suggested in this article contains four information flows: the results of the public inspections, the results of the internal inspections of the seller, the expiration times, and the amount of time remaining to a deadline after which no public inspections take place. We think that these minimalistically represent what information reaches a company about its products. For example, in the auto industry, every visit to an authorized mechanic can generate similar information flows. The information would be about the part of the car that was the reason for the visit. This type of information may be modeled with multidimensional point processes.

Smith et al. [15] suggest that a company can see the actual recall operation as marketing in reverse. In a similar way, a company can see the statistical study of its post-sales environment as a continuation of quality control. The methods (optimal stopping, sequential analysis) that we used in the solution of our proposed model were invented by Wald [20] for the purpose of quality control. All of the events that take place during post sales can be thought of as one big test of the product. The difference between this test and those conducted during production is that the latter are precisely designed and generate data that are easier to analyze. With careful modeling and data collection, quality control can be continued after sales.

Models of recalls can be useful in the regulation of recalls. If the goal of regulation is to keep manufacturing fault rates below a small level, models such as ours can be used to determine what fines imply a desired level.

**Acknowledgments.** The authors would like to thank Savaş Dayanik, Semih Sezer, Jerome Stein, Chris Tang, Ulrich Rieder, Bert Zwart, and the anonymous referees very much for their suggestions and comments.

### Appendix A. Proofs.

**PROOF OF PROPOSITION 2.3.** We adopt an argument parallel to the one given in the introduction of Bayraktar et al. [1]. The exponential distribution of the interarrival times under  $\mathbb{P}_0$  implies that under  $\mathbb{P}_0$ ,  $C$  has intensity  $\lambda_0(t) = (N - i)\mu_0$ , when  $T_i < t < T_{i+1}$ . This means that the process  $t \rightarrow C_t - \int_0^t \lambda_0(s) ds$  is a martingale under  $\mathbb{P}_0$ ; see Brémaud [4, T5 Theorem, p. 25]. Define  $\lambda_1(t) \doteq (N - i)\mu_1$ , when  $T_i < t < T_{i+1}$ . We would like  $C$  to have intensity  $\lambda_d$  under  $\mathbb{P}$ . Define  $Z_t = 1_{\{d=0\}} + 1_{\{d=1\}}R_t$ .  $Z$  is a martingale under  $\mathbb{P}_0$  with respect to the filtration  $\{\mathcal{G}_t\}$  and satisfies  $\mathbb{E}_0[Z_t] = 1$ . Define  $\mathbb{P}(A) \doteq \mathbb{E}_0[1_A Z_t]$  for  $A \in \mathcal{G}_t$ . By definition  $\mathbb{P}$  is absolutely continuous with respect to  $\mathbb{P}_0$  on  $\mathcal{G}_t$ ,  $t < \infty$ , and satisfies  $\mathbb{E}_0[(d\mathbb{P}/d\mathbb{P}_0) | \mathcal{G}_t] = Z_t$ . It can be shown by direct computation that  $C$  has intensity  $\lambda_d$  under  $\mathbb{P}$  (see also Brémaud [4], Bayraktar et al. [1]). This proves the proposition.  $\square$

**PROOF OF PROPOSITION 2.4.** For ease of notation denote by  $X_\tau$  the expression inside the brackets in (3). One can rewrite (3) as an expectation against  $\mathbb{P}_0$  by using the Radon Nikodym derivative  $d\mathbb{P}/d\mathbb{P}_0$  as follows:

$$\mathbb{E}[X_\tau] = \mathbb{E}_0\left[\frac{d\mathbb{P}}{d\mathbb{P}_0} X_\tau\right] = \mathbb{E}_0\left[\mathbb{E}_0\left[\frac{d\mathbb{P}}{d\mathbb{P}_0} X_\tau \mid \mathcal{G}_\tau\right]\right] = \mathbb{E}_0\left[X_\tau \mathbb{E}_0\left[\frac{d\mathbb{P}}{d\mathbb{P}_0} \mid \mathcal{G}_\tau\right]\right].$$

In the last equality  $X_\tau$  came out of the conditional expectation because all of the terms in  $X_\tau$  are measurable with respect to  $\mathcal{G}_\tau$ . If  $\tau$  is bounded, the last sequence of equalities, (5), and the optional sampling theorem imply

$$\mathbb{E}[X_\tau] = \mathbb{E}_0[X_\tau(1_{\{d=0\}} + 1_{\{d=1\}}R_\tau)].$$

Expanding the  $X_\tau$  on the right gives

$$= \mathbb{E}_0\left[\left(1_{\{d=0\}}e^{-r\tau}P + 1_{\{d=1\}}\left(p^{C_\tau}R_\tau e^{-r\tau}P + \sum_{k=0}^{C_\tau} R_\tau e^{-rT_k} p^{k-1}(1-p)K\right)\right)\right].$$

Under  $\mathbb{P}_0$ ,  $d$  and the rest of the variables in the above display are independent and therefore  $d$  can be integrated out:

$$= \mathbb{E}_0\left[\pi e^{-r\tau}P + (1-\pi)\left(p^{C_\tau}R_\tau e^{-r\tau}P + \sum_{k=0}^{C_\tau} R_\tau e^{-rT_k} p^{k-1}(1-p)K\right)\right].$$

One obtains (6) by replacing the  $R_\tau$  inside the last sum with  $R_{T_k}$ , which is possible because each of the summands in the last sum is measurable with respect to  $\mathcal{F}_{T_k}$  and  $R$  is a martingale; the details of the argument, which are omitted, involve conditioning on the values that  $C_\tau$  can take. If  $\tau$  is unbounded, one can approximate it by a bounded sequence and use the dominated convergence theorem, which is applicable because the costs  $P$  and  $K$  are fixed and  $C$  is bounded by  $N$ .  $\square$

**SKETCH OF PROOF OF LEMMA 4.1.** For the details of a proof of a similar result, we refer the reader to Bertsekas and Shreve [3, Chapter 8, in particular Corollary 8.1.1 and Lemma 8.1] and provide here a sketch that essentially has the same features as those of the cited proofs. Suppose the seller is to use the stopping time  $\tau$  as her recall plan. Lemma 2.1 implies that  $\tau$  will equal a constant  $s$  for sample paths for which  $\tau < T_1$ . Consider the random time  $\tau' = \tau - T_1$  on the set  $\{T_1 < s\}$ . One can think of  $\tau'$  as a recall rule for whatever happens after time  $T_1$  when  $T_1 < s$ . In general  $\tau'$  may depend on  $T_1$ . However, one can show that it is sufficient to restrict attention to stopping times  $\tau$  for which  $\tau'$  depends only on  $\tau_2, \tau_3, \dots, \tau_n$  and  $\Phi_{T_1}$ . For such  $\tau$ ,  $\tau'$  is completely a function of the shifted process  $t \rightarrow \Phi(t + T_1)$  and is a stopping time of the same process. This and the Markov property of  $(\Phi, N - C)$  allow one to break the optimization problem (16) into two pieces (before and after the first jump  $T_1$ ); the part after  $T_1$  is an optimization problem over  $\tau'$ . Because  $\tau'$  doesn't depend on  $T_1$ , this optimization can be handled separately and yields the  $V(\Phi_{T_1}, N - 1)$  term in (17). Once this part is computed, one chooses an optimum stopping time for the time interval between 0 and  $T_1$ ; this is the optimization over  $s$  in (17).  $\square$

**PROOF OF LEMMA 5.1.** The proof is based on monotonicity and is similar to those in Bertsekas and Shreve [3, Chapter 9] or to Dayanik and Sezer [5, Propositions 3.1, 3.7]; therefore, we only indicate its main steps. The function  $V_n$  is the value function of (32) when stopping is allowed only before the  $n$ th jump of  $M$ . The Markov property of  $\Psi$  and  $M$  and the explicit dynamics of  $\Psi$  given in (31) suggest

$$V_{n+1} = T(V_n). \quad (54)$$

This is a DPE and its proof is analogous to that of (17); we omit the details and refer the reader to the sketch of proof of Lemma 4.1 given above.  $V_n$  is decreasing in  $n$ , i.e.,  $V_n \geq V_{n+1} \geq V$ , because every recall decision rule that allows recalls before the  $n$ th jump of  $M$  is trivially a recall rule that allows recalls only before the  $n+1$ st jump and (32) is an infimization problem. This implies  $\bar{V} \doteq \lim_n V_n \geq V$ . On the other hand, let  $\tau_\epsilon^*$  be such that

$$V(\phi) + \epsilon \geq \mathbb{E}_\phi \left[ \sum_{k=1}^{M_{\tau_\epsilon^*}} K' \Psi_{S_k} + 1_{\{\tau_\epsilon^* < \infty\}} P(1 + \Psi_{\tau_\epsilon^*}) \right]. \quad (55)$$

Let us denote with  $\kappa$  the cost inside the expectation on the right of the above inequality. Define  $\tau_{n,\epsilon}^* \doteq \tau_\epsilon^* 1_{\{\tau_\epsilon^* \leq S_n\}} + \infty \cdot 1_{\{\tau_\epsilon^* > S_n\}}$ .  $\tau_{n,\epsilon}^*$  is the same as  $\tau_\epsilon^*$  except that it revokes the recall decisions after the  $n$ th expiration. Note that

$$\kappa_n \doteq \sum_{k=1}^{M_{\tau_{n,\epsilon}^*}} K' \Psi_{S_k} + 1_{\{\tau_{n,\epsilon}^* < \infty\}} P(1 + \Psi_{\tau_{n,\epsilon}^*}) \leq \sum_{k=1}^{\infty} K' \Psi_{S_k} + 1_{\{\tau_\epsilon^* < \infty\}} P(1 + \Psi_{\tau_\epsilon^*}), \quad (56)$$

where  $\kappa_n$  is the cost associated with the recall decision rule  $\tau_{n,\epsilon}^*$ .  $S_n$ , being an iid sum of random variables with positive means, converges to  $\infty$  almost surely. This implies  $\kappa_n \rightarrow \kappa$ .  $\tau_{n,\epsilon}^*$  is a recall decision rule that allows recalls only before the  $n$ th expiration. This and the definition of  $V_n(\phi)$  give  $V_n(\phi) \leq \mathbb{E}[\kappa_n]$ . These, (56), and the dominated convergence theorem imply  $\bar{V}(\phi) = \lim_n V_n(\phi) \leq \mathbb{E}[\kappa]$ . This and (55) imply  $V(\phi) + \epsilon \geq \bar{V}(\phi)$ . The last inequality is true for all  $\epsilon > 0$ , and therefore,  $V(\phi) \geq \bar{V}$ . We already had the reverse inequality; therefore  $\bar{V} = V$  holds and so does  $V_n \searrow V$ . Finally, the monotone convergence theorem,  $V_n \searrow V$  and the DPE  $V_{n+1} = T(V_n)$  imply  $T(V) = V$ .  $\square$

## References

- [1] Bayraktar E, Dayanik S, Karatzas I (2005) The standard Poisson disorder problem revisited. *Stochastic Processes and Their Appl.* 115(9):1437–1450.
- [2] Bayraktar E, Dayanik S, Karatzas I (2006) Adaptive Poisson disorder problem. *Ann. Appl. Probab.* 16(3):1190–1261.
- [3] Bertsekas DP, Shreve SE (1978) *Stochastic Optimal Control: The Discrete Time Case* (Academic Press, New York).
- [4] Brémaud P (1981) *Point Processes and Queues, Martingale Dynamics: Martingale Dynamics* (Springer, New York).
- [5] Dayanik S, Sezer S (2006) Sequential testing of simple hypotheses about compound Poisson processes. *Stochastic Processes Appl.* 116(12):1892–1919.
- [6] Eaton JW (2002) *Gnu Octave Manual* (Network Theory Limited, Bristol, UK).
- [7] Jarrell G, Peltzman S (1985) The impact of product recalls on the wealth of sellers. *J. Political Econom.* 93(3):512.
- [8] Liptser R, Shiryaev A (2001) *Statistics of Random Processes*, Vol. 1. (Springer Verlag, Berlin).
- [9] Maynard M, Tabuchi H (2010) Rapid growth has its perils, Toyota learns. *The New York Times*. Accessed January 28, 2010, <http://www.nytimes.com/2010/01/28/business/28toyota.html>.
- [10] *New York Times* (2010) Toyota's recall woes. Accessed February 9, 2010, <http://www.nytimes.com/2010/02/09/opinion/09iht-edtoyota.html>.
- [11] Paté-Cornell E (2002) Fusion of intelligence information: A Bayesian approach. *Risk Anal.* 22(3):445–454.

- [12] Patterson B (2010) iPad sales cross million mark twice as fast as original iPhone. Accessed May 5, 2010, [http://news.yahoo.com/s/ytech\\_gadg/20100503/tc\\_ytech\\_gadg/ytech\\_gadg\\_tc1901](http://news.yahoo.com/s/ytech_gadg/20100503/tc_ytech_gadg/ytech_gadg_tc1901).
- [13] Peskir G, Shiryaev A (2006) *Optimal Stopping and Free-Boundary Problems* (Birkhäuser, Basel, Switzerland).
- [14] Pinedo M, Seshadri S, Zemel E (2000) The Ford-Firestone Case. Stern School of Business, New York University, New York.
- [15] Smith NC, Thomas RJ, Quelch JA (1996) Strategic approach to managing product recalls. *Harvard Bus. Rev.* (September–October):102–112.
- [16] Tabuchi H (2010) Toyota is to recall 2010 Prius model cars for brakes. *The New York Times*. Accessed February 8, 2010, <http://www.nytimes.com/2010/02/08/business/08webtoyota.html>.
- [17] Tang CS (2008) Making products safe: Process and challenges. *Internat. Commerce Rev.* 8(1):49–55.
- [18] Teratanavat R, Hooker NH (2004) Understanding the characteristics of US meat and poultry recalls: 1994–2002. *Food Control* 15(5):359–367.
- [19] Vlasic B (2010) Lawsuit over a crash adds to Toyotas difficulties. *The New York Times*. Accessed February 5, 2010, <http://www.nytimes.com/2010/02/05/business/05recall.html?pagewanted=all>.
- [20] Wald A (1941) *Sequential Analysis* (Wiley, New York).
- [21] Wald A (1950) *Statistical Decision Functions* (Wiley, New York; Chapman & Hall, London).
- [22] Wald A, Wolfowitz J (1948) Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* 19(3):326–339.
- [23] Zakai M (1969) On the optimal filtering of diffusion processes. *Probab. Theory Related Fields* 11(3):230–243.