Probability in the Engineering and Informational Sciences, **28**, 2014, 121–144.
doi:10.1017/S0269964813000363

PREVENTION OF CATASTROPHIC FAILURES WITH WEAK FOREWARNING SIGNALS

H. Dharma Kwon
Department of Business Administration, University of Illinois at Urbana-Champaign, Champaign, IL 61820
E-mail: dhkwon@illinois.edu.

We consider the problem of a firm facing failures with weak forewarning signals. In the base model that we study, the firm watches for signals of a random arrival of a disruptive innovation and continuously updates the posterior probability that a disruptive innovation has already happened. A disruptive innovation is marked by a rapid increase in the growth rate of the market for a new technology, and it is followed by a random arrival of catastrophic failure of the firm. The firm can invest capital to adopt the innovation to prevent failure. The optimal policy is to adopt it when the posterior probability exceeds an optimally chosen threshold. We investigate the probability of failure under the optimal policy when the cost of failure is large and the arrival rate of disruptive innovation is low. The probability of failure is close to one if the arrival rate is extremely low while it is close to zero if the arrival rate is moderate. We also consider an extension of the base model to incorporate recurrence of disruptive innovation; when the arrival rate is moderate, the optimal threshold and the failure probability can be significantly larger than those of the base model.

1. INTRODUCTION

Even successful firms may falter in the face of disruptive technological innovations [4]. *Eastman Kodak Co.*, once a dominant photographic film company, failed to respond fast enough to the emergence of digital imaging technology in the 1990s because it held a strong belief, based on a 100-year history, that ample profit opportunities still existed in its traditional chemical film market. When Kodak finally conceded that its traditional market was shrinking and restructured itself for the digital image market in 2003, it was too late; rival firms had already amassed experience and competitive advantage [7]. Between 1998 and 2010, Kodak's enterprise value declined by approximately 95%; it filed for Chapter 11 bankruptcy protection in January 2012. Kodak is not an isolated incident of a once-successful firm that failed to recognize sweeping technological changes; Polaroid Corp., Borders Group Inc., and Blockbuster Inc. are among the most well-known recent examples of firms that faltered spectacularly in the face of disruptive innovations [22].

There are two main difficulties with detecting and responding to a disruptive innovation. First, the signs of a disruptive innovation are often too weak for the incumbent firm to detect in the early stages of the innovation. A new disruptive innovation need not

 0269-9648/14 $25.00

compete head-to-head with the extant technology at the outset because it initially targets a completely different market, and hence, it might evade the incumbent's attention [5]. For example, in the 1990s, the digital imaging technology initially targeted a low-cost and low-performance camera market while Kodak was seeking a high-priced performance-competitive camera market. By the time the signs of disruption are conspicuous, the new technology will have encroached upon the incumbent firm's market, and it is already too late for the incumbent firm to respond due to, for example, first-mover advantage. Second, disruptive innovations are often very slow to happen, and hence, a successful incumbent firm may become complacent and believe that the current technology is unlikely to become obsolete for a long time to come, even in the face of a disruptive innovation. For instance, from the invention of the first digital camera in 1975, it took approximately 20 years for digital imaging technology to set foot in the consumer-level market. Other examples of innovations that took many decades to replace extant technologies include Watt's steam engine and the steamship [17].

In this paper, we investigate the incumbent firm's optimal policy of response to weak signals of a regime change which will propel enterprise-wide catastrophic failure. For example, the incumbent firm can invest in adopting the new technology to prevent failure. Our primary goal is to obtain insights into the efficacy of the optimal policy of adoption under extreme values of model parameters (weak signals, slow disruption, and high cost of failure) that often characterize disruptive innovations. Of particular interest to us is whether, even if the arrival of a disruptive innovation is extremely slow, an incumbent firm can effectively prevent failure by following an optimal (profit-maximizing) policy instead of succumbing to irrational complacency.

Our paper has two contributions. First, we extend the well-known change point detection theory to a model of catastrophic business failure with weak forewarning signals. Although our model has deep roots in the conventional change point detection models, it possesses mathematically and economically different salient features and applies to distinct settings. Second, we offer practical insights into the optimal policy, the expected return, and the probability of failure under the optimal policy.

In our theoretical model, we incorporate five salient features of an incumbent firm facing a potentially disruptive new technology. (1) The incumbent fails in two steps. In the first step, *disruption* takes place, that is, a new technology finds its footing in the current market, but its market share remains negligible for some time. In this stage, we say that the incumbent firm is in a *disrupted state*. The second step is a random arrival of catastrophic *failure* of the incumbent firm due to swift growth of the new technology. (2) On the alert for signals of disruption, the incumbent firm monitors the growth of the market for the new technology and updates its belief regarding whether disruption has already happened or not. However, even if a disruption happens, its signals are too weak and too noisy for the incumbent firm to detect it with certainty. (3) The incumbent firm can prevent failure if it recovers its original *undisrupted state* by investing a fixed amount of capital to adopt the new technology. (4) Once failure takes place, the firm incurs a colossal loss of its enterprise value due to the shrunk market demand for its extant technology. (5) The arrival rate of disruption is low, that is, it takes a long time for a new technology to become a disruptive innovation.

We model the situation as an optimal stopping problem for a Bayesian decision-maker. One possible sign of the arrival of disruption is a rapid increase in the market growth rate of a new technology [5]. In our model, the firm monitors the growth rate of the market for the new technology; if the rate of increase of the market growth rate is high, then it is a sign that disruption happened. We model the market growth rate as a Brownian motion with unknown drift. The drift can be interpreted as the rate of increase in the market growth rate. The value of the drift changes from zero to a higher value when disruption happens. Based on the history of the market growth rate observed, the firm can continuously update its belief characterized by the posterior probability that the firm is in a disrupted state. At any point in time, the firm can spend a fixed amount of capital to prevent failure. Utilizing optimal stopping theory, we find that the optimal policy is to adopt the new technology when the posterior probability exceeds a threshold value.

We obtain a number of qualitative insights into the optimal policy and the probability of eventual failure under the optimal policy. Throughout the paper, we assume that the cost of failure is much larger than the cost of recovery. We find that the optimal recovery threshold takes a small value, which is inversely proportional to the cost of failure. The optimal return and the probability of failure critically depend on the magnitude of the arrival rate of disruption, however. If the arrival rate of disruption is moderate (not very low), then the optimal policy is to recover the undisrupted state early. Consequently, the magnitude of the optimal expected return is approximately the cost of recovery, and the failure probability is very low. In contrast, if the arrival rate of disruption is extremely low, then the failure probability is very close to 1 even though the firm employs the optimal policy, and the optimal return is approximately the return from no recovery effort even though the optimal threshold is a very small number. This is because, due to the slow arrival of disruption, the firm updates its posterior belief very slowly; consequently, it takes a very long time for the posterior probability to reach the threshold of recovery, and failure is likely to happen before the firm invests in recovery.

Next we turn to the case of recurrent disruption. Under moderate arrival rates of disruption, we find qualitatively different results from those of the single-disruption model if the emergence of new technologies takes very little time. In particular, the recovery threshold is significantly higher, the optimal return is significantly lower, and the failure probability is significantly higher. In contrast, under extremely small arrival rates of disruption, there is very little difference from the results of the single-disruption model. Our results illustrate the danger of using a single-disruption assumption when disruption is a recurrent event.

The paper is organized as follows. We briefly discuss the related literature in Section 2. We investigate the base model of a single disruption in Section 3 and extend it to a model of recurrent disruption in Section 4. Conclusions are given in Section 5. All proofs appear in Appendix.

2. RELATED LITERATURE

Change point detection theory has a long history with rich literature. See Rapoport, Stein, and Burkheimer [16] and Poor and Hadjiliadis [15] for an overview and references therein. The most relevant to this paper is the literature on *Shiryaev's problem,* the objective of which is to minimize the cost of missing a sudden and random change of regime when the decision-maker observes a noisy signal of the change [21]. The signal process can be formulated as a Brownian motion, the drift of which suddenly changes at a random time, and the optimal detection problem can be solved by the well-established optimal stopping theory [2,20]. The optimal policy is to stop when the posterior probability of change exceeds a threshold value [14]. The problem can be also formulated in a discrete-time version or with an observable Poisson process whose arrival rate suddenly changes at a random time [14,15].

The basic framework of the change point detection theory has been applied to a wide range of problems. (See, e.g., p. 1 of Poor and Hadjiliadis [15]). In particular, it has been employed to analyze business decisions under a changing environment. Ryan and Lippman [19] studied the optimal policy of exit from a project that suddenly starts generating a negative profit stream. In their model, the cumulative profit stream is modeled as a Brownian motion with drift that suddenly drops in value at a random time. They employed Shiryaev's Bayesian formulation to obtain a threshold policy. However, their model did not incorporate exponential discounting. Beibel and Lerche [3] considered a problem of finding the optimal time to sell a stock whose growth rate suddenly changes at a random time. In this problem, they modeled the price of the stock as a geometric Brownian motion and incorporated exponential discounting. Gapeev [9] also incorporated exponential discounting in a problem of determining a sequence of stopping times that are as close as possible to multiple change points.

Our paper is also related to Bayesian sequential decisions of technology adoption. Jensen [10] studied the decision problem of a firm regarding adoption of a new technology when its profitability is uncertain: the decision-maker observes a stream of random (Bernoulli) signals that reflect the profitability of the new technology, and he sequentially updates his belief regarding its profitability. The optimal policy is to adopt it when the probability of high profit exceeds a threshold value. McCardle [12] and Ulu and Smith [23] extended this model by incorporating the cost of acquiring signals per unit time. They obtained an optimal policy characterized by two thresholds: adopt it if the probability exceeds the upper threshold, and stop acquiring signals if it hits the lower threshold. Kwon and Lippman [11] also investigated a Bayesian technology adoption and abandonment problem in which the signal of profitability emanates from an on-going pilot project, and they studied the comparative statics of time-to-decision with respect to uncertainty.

3. MODEL OF A SINGLE DISRUPTION

In this section, we first introduce the base model with a single disruption. Then we obtain the optimal policy in Section 3.1 and investigate the model parameter regimes of our interest in Section 3.2.

We consider an incumbent firm with an extant technology that earns a continuous stream of profits v per unit time. At time $t = 0$, the firm starts out in an *undisrupted state*, and a new technology that is potentially disruptive emerges. The new technology has not yet found its footing in any market in the undisrupted state. At a random time τ_1, which is an exponential random variable with mean $1/\lambda_1$, the new technology becomes a *disruptive* innovation by finding its way to a beach-head market (a strategically chosen initial target market), and the incumbent firm becomes *disrupted*. Upon disruption, the failure of the firm is not immediate. Instead, there is time to take remedial action. Specifically, failure takes place at a time τ_2 which has an exponential distribution with mean $1/\lambda_2$. We assume that τ_1 and τ_2 are mutually independent and that the arrival rates λ_1 and λ_2 are industry-specific values that are known to the firm. We interpret $\tau_1 + \tau_2$ as the last opportune time for the firm to respond and invest in the new technology. Once $\tau_1 + \tau_2$ has passed, entrants with the new technology will have accumulated competitive advantage so that the incumbent firm cannot avoid catastrophic failure. At the time failure happens, the net present value of the firm's future profit is reduced by w, which includes a partial or total loss of the future profit stream and the cost of shutdown. To prevent failure, the firm can invest a fixed amount k to adopt the new technology and *recover* the undisrupted state at any point in time before the failure happens. Note that the firm's investment in recovery can happen even before the disruption sets in. In this section, we assume that disruption can happen only once.

For all positive times $t > 0$, there is always a non-zero probability that the firm is disrupted, and the firm watches for signs of disruption at all times. The signs of disruption initially do not directly influence the incumbent firm's profit stream because the market for the disruptive innovation is separated from the incumbent's market in the early stages of the innovation. In this paper, we assume that the firm observes the process $X = \{X_t : t \geq 0\}$ of the instantaneous market growth rate of the new technology; Christensen, Anthony, and Roth [5] suggested that a substantially large instantaneous rate of increase of X is a sign of a disruptive innovation. We model the process X as the Brownian motion

$$X_t = \mu t + \sigma B_t,$$

where μ is interpreted as the average instantaneous increase of the growth rate of the market for the new technology, $B = \{B_t : t \geq 0\}$ is a Wiener process, and σ is a constant which quantifies the amount of noise in the market growth rate. Here, we set $\sigma = 1$ and take $\mu = 0$ in the undisrupted state and $\mu = 1$ in the disrupted state; the drift μ jumps from 0 to 1 at time τ_1 and stays constant thereafter. (We can take a general value of σ and $\mu = h$ for the disrupted state for some general value of $h > 0$, but we can always rescale X_t and the time t such that $h = 1$ and $\sigma = 1$ for convenience of notation. The specific values of h and σ do not affect our main results.) The firm knows the possible values that μ can take, but it does not know the current state of μ nor can it observe the process B. We assume that the process B is independent of τ_2 and τ_1. Finally, we assume that the firm is a maximizer of the expected discounted profit with a discount rate r.

Let X, μ, and B be defined on a probability space $(\Omega, \mathcal{G}, \mathbb{P})$. We also let $\mathcal{F} = \{\mathcal{F}_t : t \geq 0\}$ denote the filtration generated by the observable process X. We define the process of the posterior probability

$$P_t \equiv \mathbb{P}(\{\tau_1 < t\} | t < \tau_1 + \tau_2, \mathcal{F}_t),$$

that the firm is in a disrupted state at time t given that the failure has not happened by time t. Note that $\tau_1 < t$ implies that disruption happened before time t, and that $t < \tau_1 + \tau_2$ implies that the failure has not happened by time t. In order to apply the optimal stopping theory, we first need to obtain the stochastic differential equation (SDE) of the process $P = \{P_t : t \geq 0\}$.

Proposition 1: *Given the prior $P_0 = p$, the posterior process P_t is given by*

$$P_t = \frac{N(t)}{N(t) + (1-p)e^{-\lambda_1 t}},$$

where

$$N(t) \equiv \exp\left[\left(X_t - \frac{1}{2}t\right) - \lambda_2 t\right] \left\{p + (1-p)\int_0^t \exp\left[-\left(X_v - \frac{1}{2}v\right) + \lambda_2 v\right] \lambda_1 e^{-\lambda_1 v} dv\right\}.$$

The SDE of P_t is given by

$$dP_t = P_t(1-P_t)d\bar{B}_t + (1-P_t)(\lambda_1 - \lambda_2 P_t)dt, \tag{1}$$

where $\{\bar{B}_t, \mathcal{F}_t, t \geq 0\}$ is a Wiener process given by

$$\bar{B}_t = X_t - \int_0^t E[\mu | u < \tau_1 + \tau_2, \mathcal{F}_u] du.$$

The SDE of P_t is expressed in terms of an observable Wiener process $\bar{B} = \{\bar{B}_t : t \geq 0\}$ that can be constructed from the observable process X. In the SDE, the coefficient $P_t(1-P_t)$ of $d\bar{B}_t$ can be interpreted as the rate of diffusion of the process P_t. The drift term (the coefficient of dt) from Eq. (1) is $(1-P_t)(\lambda_1-\lambda_2 P_t)$, which represents the average direction of the evolution of P_t over time. The drift increases in λ_1 because λ_1 is the arrival rate of disruption, and the possibility of disruption tends to increase P_t. The other effect that influences the drift term is that the longer you survive without failure, the less you believe that the firm is in a disrupted state. This effect is enhanced if λ_2 is larger or if P_t is larger, and it explains why the drift decreases in $\lambda_2 P_t$.

3.1. Optimal Policy and Return Function

In this subsection, we define the firm's objective function and obtain the optimal policy. The firm's cumulative discounted profit Π is given by

$$\Pi = \int_0^{\min(T,\tau_1+\tau_2)} ve^{-rt}dt + \chi_{\{T<\tau_1+\tau_2\}}\left[e^{-rT}(-k) + \int_T^\infty ve^{-rt}dt\right] + \chi_{\{T\geq\tau_1+\tau_2\}}\left[e^{-r(\tau_1+\tau_2)}(-w) + \int_{\tau_1+\tau_2}^\infty ve^{-rt}dt\right],$$

where T is the stopping time at which the firm recovers the undisrupted state by adopting the new technology, and $\chi_{\{\cdot\}}$ is an indicator function. Note that Π is comprised of the profit stream v per unit time from the current technology, the cost k of recovery, and the cost w of failure. (In case failure happens, the term $\int_{\tau_1+\tau_2}^\infty ve^{-rt}dt$ is interpreted as the *projected discounted profit* that the firm would have earned if failure had not happened. The cost w of failure is interpreted as the amount of reduction in the discounted profit due to failure.) Recall that we assume that v is constant, so we can re-express Π as

$$\Pi = \frac{v}{r} + \chi_{\{T<\tau_1+\tau_2\}}e^{-rT}(-k) + \chi_{\{T\geq\tau_1+\tau_2\}}e^{-r(\tau_1+\tau_2)}(-w).$$

For convenience, we drop the constant term v/r from the following objective function:

$$V_T(t,p) = E^{(t,p)}\left[\chi_{\{T<\tau_1+\tau_2\}}e^{-rT}(-k) + \chi_{\{T\geq\tau_1+\tau_2\}}e^{-r(\tau_1+\tau_2)}(-w)\right], \tag{2}$$

where $E^{(t,p)}[\cdot] \equiv E[\cdot|t<\tau_1+\tau_2, P_t=p]$ is the conditional expectation operator given that $t<\tau_1+\tau_2$ and $P_t=p$. The firm's objective is to maximize $V_T(t,p)$ with respect to T. In particular, we consider the class of stopping times of the form

$$T = \inf\{t>0 : P_t \notin C\},$$

in which T is the exit time of the process P from an open interval $C\subset(0,1)$; in Proposition 2, we show that the optimal policy T^* indeed belongs to this class of stopping times via verification theorem.

From Proposition 1, the characteristic operator [13] for P is given as follows:

$$\mathcal{L} \equiv \frac{1}{2}p^2(1-p)^2\partial_p^2 + (1-p)(\lambda_1-\lambda_2 p)\partial_p. \tag{3}$$

Now we introduce a linear operator $\mathcal{A}$ for $V_T(\cdot,\cdot)$:

$$\mathcal{A}V_T(t,p) \equiv \mathcal{L}V_T(t,p) - (r+p\lambda_2)V_T(t,p), \tag{4}$$

which will be used below to characterize $V_T(\cdot,\cdot)$. The following Lemma is necessary for establishing the optimality conditions of $V_T(\cdot,\cdot)$:

LEMMA 1: *The return function* $V_T(\cdot,\cdot)$ *satisfies* $\mathcal{A}V_T(t,p) - p\lambda_2 w \cdot e^{-rt} = 0$ *for all* $p \in C$.

To obtain the optimal stopping time T^*, we consider a candidate stopping time $T = \inf\{t > 0 : P_t \notin [0, p^*)\}$, which is the exit time from an interval of the form $C = [0, p^*)$, and find $V(t,p)$ and a threshold $p^* \in (0,1)$ that satisfy the following conditions:

$$\mathcal{A}V(t,p) - p\lambda_2 w \cdot e^{-rt} = 0 \quad \text{for } p < p^* \tag{5}$$

$$V(t,p^*) = -ke^{-rt} \tag{6}$$

$$\partial_p V(t,p^*) = 0. \tag{7}$$

$$V(t,p) > -ke^{-rt} \quad \text{for } \ p < p^* \tag{8}$$

$$V(t,p) = -ke^{-rt} \quad \text{for } \ p \geq p^* \tag{9}$$

$$\mathcal{A}V(t,p) - p\lambda_2 w \cdot e^{-rt} < 0 \quad \text{for } p \geq p^* \tag{10}$$

By Theorem 10.4.1 of [13], once the solution $V(t,p)$ to the conditions above is found, the sufficient conditions for optimality are satisfied by the function $V(t,p)$, and we are assured that $V(t,p) = \sup_{T \in \mathcal{T}} V_T(t,p)$, where $\mathcal{T}$ is the set of all stopping times for X.

PROPOSITION 2: *(I) There exist a unique value of* $p^* \in (0,1)$ *and a unique function* $V(t,p)$ *that satisfy Eqs. (5)–(10) if the inequality*

$$w\frac{\lambda_2}{(\lambda_2 + r)} > k, \tag{11}$$

is satisfied.

(II) Under the condition (11), the optimal policy is to recover the undisrupted state as soon as the posterior P_t *exceeds* p^**, where* p^* *solves Eqs. (5)–(10). Moreover, the optimal return* $\sup_T V_T(t,p)$ *is given by* $V(t,p) = e^{-rt}V(p)$*, where*

$$\begin{aligned} V(p) &= c\phi\left(\frac{p}{1-p}\right) + V_1(p) \quad \textit{for } p < p^*, \\ &= -k \quad \textit{otherwise}, \end{aligned}$$

and

$$\begin{aligned} c &= \frac{-k - V_1(p^*)}{\phi(p^*/(1-p^*))}, \\ \phi(z) &= \frac{1}{\Gamma(a)\cdot(z+1)} \int_0^\infty e^{-u} u^{a-1} \left(1 + u\frac{z}{2\lambda_1}\right)^{b-a-1} du, \\ a &= \lambda_1 - \lambda_2 - \frac{1}{2} + \sqrt{\left(\lambda_1 - \lambda_2 - \frac{1}{2}\right)^2 + 2(\lambda_1 + r)}, \\ b &= 1 + 2\sqrt{\left(\lambda_1 - \lambda_2 - \frac{1}{2}\right)^2 + 2(\lambda_1 + r)}, \\ V_1(p) &= -w\frac{\lambda_2(\lambda_1 + rp)}{(\lambda_1 + r)(\lambda_2 + r)}. \end{aligned}$$

Here, $\Gamma(\cdot)$ is the Gamma function.

The optimal policy is a threshold rule because the firm must adopt the new technology only if the probability of disruption is sufficiently high. The condition given in Eq. (11) ensures that the cost k of recovery must be sufficiently smaller than the cost w of failure so that the firm has an incentive to invest in recovery.

Finally, in order to obtain the economic insight into the efficacy of the optimal recovery policy, we inspect the *probability of preventing the event of failure* given the optimal stopping time T^* and the current posterior $P_0 = p$:

$$F(p) = E^{(0,p)}[\chi_{\{T^* < \tau_1 + \tau_2\}}].$$

Its complement probability $1 - F(p)$ is the probability that the failure eventually occurs under the optimal policy. For future reference, we obtain the functional form of $F(p)$ in the following Lemma:

LEMMA 2: *The prevention probability $F(\cdot)$ is given by*

$$\begin{aligned} F(p) &= c' \cdot \psi\left(\frac{p}{1-p}\right) \quad \text{for } p < p^*, \\ &= 1 \quad \text{otherwise}, \end{aligned}$$

where

$$\begin{aligned} c' &= \left[\psi\left(\frac{p^*}{1-p^*}\right)\right]^{-1} \\ \psi(z) &= \frac{1}{\Gamma(a') \cdot (z+1)} \int_0^\infty e^{-u} u^{a'-1} \left(1 + u\frac{z}{2\lambda_1}\right)^{b'-a'-1} du, \\ a' &= \lambda_1 - \lambda_2 - \frac{1}{2} + \sqrt{\left(\lambda_1 - \lambda_2 - \frac{1}{2}\right)^2 + 2\lambda_1}, \\ b' &= 1 + 2\sqrt{\left(\lambda_1 - \lambda_2 - \frac{1}{2}\right)^2 + 2\lambda_1}. \end{aligned}$$

3.2. Large w and Small λ_1 Limits

In this subsection, we investigate large w and small λ_1 limits when the other parameters, λ_2, r, and k, are on the order of unity. For notational brevity, we define

$$\varepsilon \equiv \frac{k(\lambda_2 + r)}{w\lambda_2},$$

which is a small parameter in the large w limit. Because λ_1 is also small, we have two small parameters, and thus, we need to consider three different parameter regimes: $\varepsilon \ll \lambda_1$, $\varepsilon = O(\lambda_1)$, and $\varepsilon \gg \lambda_1$. In this paper, we obtain analytical results on the regimes $\varepsilon \ll \lambda_1 \ll 1$ and $\lambda_1 \ll \varepsilon \ll 1$ and illustrate a numerical example of the regime $\varepsilon = O(\lambda_1)$.

PROPOSITION 3: *(I) In the small λ_1 and small ε/λ_1 limit, the threshold probability is given by*

$$p^* = \frac{kr}{w\lambda_2}\left[1 + O\left(\frac{\varepsilon}{\lambda_1}\right)\right],$$

and the optimal return at $p = 0$ *is given by*

$$V(0) = -k\left[1 + O\left(\frac{\varepsilon}{\lambda_1}\right)\right].$$

Under the optimal policy, the probability of eventual failure when $p = 0$ *is given by*

$$1 - F(0) = \frac{\lambda_2}{2\lambda_1}(p^*)^2\left[1 + O\left(\frac{\varepsilon}{\lambda_1}\right)\right] = \frac{O(\varepsilon^2)}{\lambda_1}. \qquad \textbf{(12)}$$

(II) In the small ε *and small* λ_1/ε *limit, the threshold probability is given by*

$$p^* = \frac{b - a - 1}{b - a - 2}\varepsilon[1 + o(1)], \qquad \textbf{(13)}$$

where a *and* b *are given by Proposition 2(II). The optimal return at* $p = 0$ *is given by*

$$V(0) = -\frac{w\lambda_2\lambda_1}{r(\lambda_2 + r)}[1 + o(1)].$$

Under the optimal policy, the probability of eventual failure when $p = 0$ *is given by*

$$1 - F(0) = 1 - \frac{\Gamma(a')}{\Gamma(b' - 1)}\left(\frac{p^*}{2\lambda_1}\right)^{1-b'+a'}[1 + o(1)] = 1 - O\left(\frac{\lambda_1}{\varepsilon}\right)^{b'-a'-1}. \qquad \textbf{(14)}$$

The analytical results of Proposition 3 are obtained from the asymptotic properties of $\phi(\cdot)$ in the limits $\varepsilon/\lambda_1 \to 0$ and $\lambda_1/\varepsilon \to 0$, but it is difficult to obtain analytical insights into the case in which ε and λ_1 are comparable in size. Hence, we illustrate a numerical example in Figures 1 and 2 for p^* and $1 - F(0)$ in the regime $10^{-2} < \lambda_1/\varepsilon < 10^{1.8}$ when $\varepsilon = 10^{-2}$ and $\lambda_2 = r = k = 1$. Figure 1 (the solid curve) demonstrates that p^* is always of order ε irrespective of the size of λ_1. In contrast, $1 - F(0)$ strongly depends on the value of λ_1; $1 - F(0)$ is almost 1 for small λ_1/ε and almost zero for large λ_1/ε.

From Proposition 3 and Figure 1 (the solid curve), we observe that the optimal recovery threshold p^* is on the same order of magnitude as ε in the single-disruption model. This result is intuitively straightforward; if the potential cost of failure is large, then the decision-maker would take measures to prevent it at an early stage when the probability of disruption is still very small.

As p^* is a small number, one might navely expect that the probability of failure should be also very small. On the contrary, we find two very different results for the probability of failure in the limits $\varepsilon/\lambda_1 \to 0$ and $\lambda_1/\varepsilon \to 0$. In the limit $\varepsilon/\lambda_1 \to 0$, our intuitive expectation holds. The optimal policy is to recover the undisrupted state at p^*; because p^* is a very small number, the stopping time T^* for P_t to reach p^* is also very small. Hence, there is very little chance that failure occurs before T^*. Moreover, the discounting (the difference between e^{-rT^*} and 1) is negligible, and the expected return is very close to $-k$.

In the limit $\lambda_1/\varepsilon \to 0$, we obtain an opposite result in Proposition 3(II). If λ_1 is very small, then the firm updates its posterior probability very slowly; intuitively, it is difficult for the decision-maker to readily believe that an extremely rare event has happened. Mathematically, this effect of small λ_1 can be seen from Eq. (1), in which the drift term is a very small number or even a negative number when λ_1 and P_t are small so that the rate of increase of P_t is very low or even negative. Hence, even though p^* is small, it takes a very long time for P_t to reach p^* simply because λ_1 is small. As a result, there is a very high probability that failure happens before P_t reaches p^*. Thus, contrary to our nave expectation, we arrive at the conclusion that the efficacy of the optimal policy is critically sensitive to the value of the ratio ε/λ_1.

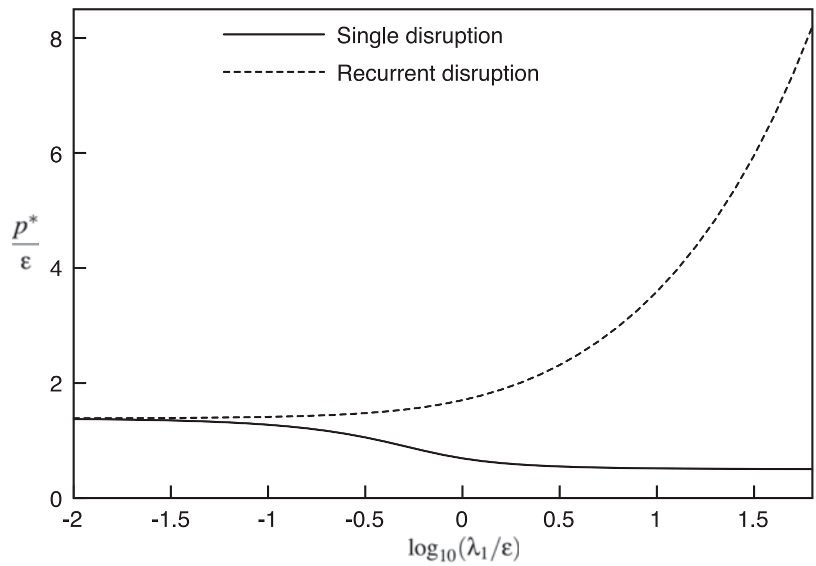

FIGURE 1. The threshold probability p^* as a function of $\log_{10}(\lambda_1/\varepsilon)$.

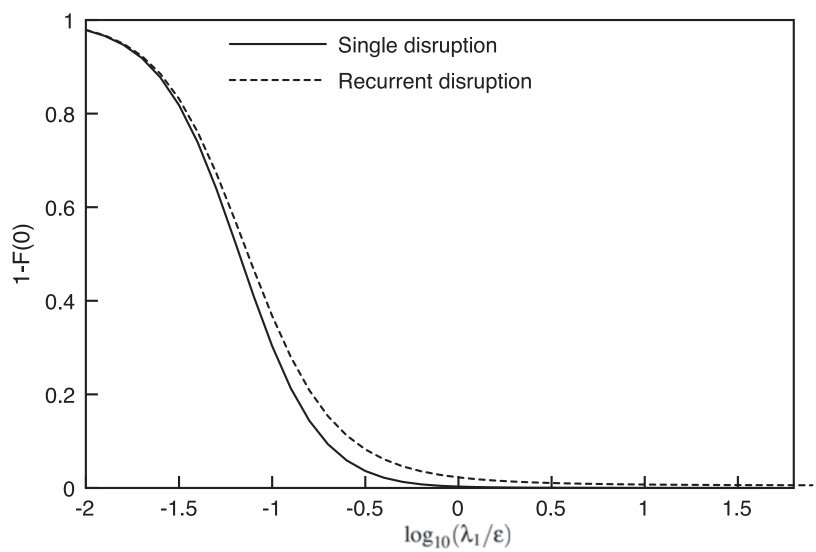

FIGURE 2. The failure probability $1 - F(0)$ as a function of $\log_{10}(\lambda_1/\varepsilon)$.

4. MODEL OF RECURRENT DISRUPTION

Disruptive innovations are often recurrent events within an industry. For example, the disk drive industry underwent a series of disruptive innovations [4]. Hence, we need to extend our model to one with recurrent disruption and explore the robustness of our results obtained in Section 3. In particular, the robustness of Proposition 3(I) is questionable if disruption is recurrent. If the policy is to invest k at almost every moment in time to prevent failure, then the cumulative discounted expected cost would amount to a very large number, possibly

https://doi.org/10.1017/S0269964813000363 Published online by Cambridge University Press

much larger than k. Thus, the frequency of recovery must be much lower than is prescribed by Proposition 3(I), and the probability of failure might turn out to be much higher.

In this section, we incorporate the recurrence of disruption in our model. We model the problem as a multi-epoch decision problem where each epoch begins in an undisrupted state and ends either when the firm recovers the undisrupted state or when failure happens. If failure happens, then the firm incurs a colossal loss w, terminates its operation, and exits the market. If the firm recovers the undisrupted state before failure happens, then the next epoch begins, and yet another new technology seeks a toehold in the economy to disrupt the current technology. We assume that there are an infinite number of epochs, each of which begins with a new technology that can be potentially disrupted by another technology. We also assume that the potentially disruptive technology of each epoch emerges only after the epoch commences. Hence, for each epoch, there is only one source of disruption.

Let t_i denote the time measured from the beginning of an epoch labeled by $i \in \{1, 2, \ldots\}$. At $t_i = 0$, the i-th epoch begins with a new technology that prevails in the market. At the same time, another technology that is potentially disruptive is being developed by a rival firm. The latter technology becomes disruptive at an exponential time $\tau_{1,i}$ at an arrival rate λ_1; the disruption is followed by failure, which arrives at an exponential time $\tau_{2,i}$ at a rate λ_2. Thus, the failure of epoch i happens at time $t_i = \tau_{1,i} + \tau_{2,i}$. We assume that $\{\tau_{1,i}\}$ and $\{\tau_{2,i}\}$ are mutually independent and i.i.d.

For each epoch i, the firm begins observing the process $X^i = \{X^i_t : t \geq 0\}$ of the market growth rate of the new technology at time $t_i = 0$. Similarly to the single-disruption model, we model the market growth rate as $X^i_{t_i} = \mu^i t_i + B^i_{t_i}$, where μ^i is the average increase rate of the market growth rate of the new technology for epoch i, and $B^i = \{B^i_t : t \geq 0\}$ is a Wiener process. At $t_i = 0$, the value of μ^i starts out as 0, but μ^i jumps from 0 to 1 at time $t_i = \tau_{1,i}$ and stays constant thereafter. For any positive t_i, the firm does not know the true value of μ^i, but it knows that μ^i is either 0 or 1. The firm continuously updates the posterior probability

$$P^i_{t_i} \equiv \mathbb{P}(\{\tau_{1,i} < t_i\} | t_i < \tau_{1,i} + \tau_{2,i}, \mathcal{F}_{t_i}),$$

of disruption by observing the process X^i. Assuming that $\{B^i\}_i$ are i.i.d., we can directly apply Proposition 1 and obtain the SDE of $P^i = \{P^i_t : t \geq 0\}$:

$$dP^i_t = P^i_t(1 - P^i_t)d\bar{B}^i_t + (1 - P^i_t)(\lambda_1 - \lambda_2 P^i_t)dt,$$

where $\{\bar{B}_t,^i \mathcal{F}_t, t \geq 0\}$ is a Wiener process given by

$$\bar{B}^i_t = X^i_t - \int_0^t E[\mu^i | u < \tau_{1,i} + \tau_{2,i}, \mathcal{F}_u] du.$$

Because P^i satisfies the same SDE as the posterior process P of the single-disruption model, they both share the same characteristic operator $\mathcal{L}$ given by Eq. (3).

Now we specify the objective function of the firm. The firm can recover the undisrupted state by investing k at any point in time during an epoch. Let T_i denote the stopping time of recovery for epoch i, and define

$$R_i(T_i) = \chi_{\{T_i < \tau_{1,i} + \tau_{2,i}\}} e^{-rT_i}(-k) + \chi_{\{T_i \geq \tau_{1,i} + \tau_{2,i}\}} e^{-r(\tau_{1,i} + \tau_{2,i})}(-w),$$

which is the discounted cashflow of epoch i measured at some time $t_i \geq 0$. (Just as in Section 3, we omit the profit stream v for brevity of notation). We let $\pi = (T_1, T_2, \ldots)$ denote the policy of recovery for all epochs. Let $\hat{T}_i = \min(T_i, \tau_{1,i} + \tau_{2,i})$ denote the stopping time that marks the end of the i-th epoch. Let $S_n \equiv \sum_{i=1}^n \hat{T}_i$ denote the time (measured

https://doi.org/10.1017/S0269964813000363 Published online by Cambridge University Press

from $t_1 = 0$) at which a new technology for the $(i+1)$-th epoch emerges, and let $K_n \equiv \prod_{i=1}^{n} \chi_{\{T_i<\tau_{1,i}+\tau_{2,i}\}}$ denote the indicator function of the event that the firm survives the n-th epoch without failure; $K_n = 1$ if the firm survives the first n disruptive innovations without failure, and $K_n = 0$ otherwise. Then we can express the total discounted return as follows:

$$R_\pi \equiv R_1(T_1) + \sum_{i=1}^{\infty} K_i \exp[-rS_i] R_{i+1}(T_{i+1}).$$

The objective of the firm is to maximize the expected value of R_π given a prior probability $P_0 = p$ that the first epoch is disrupted at time zero. Hence, the objective function is written as

$$V_\pi(p) \equiv E^p[R_\pi],$$

where $E^p[\cdot] = E[\cdot|P_0 = p]$. We let the optimal expected return be denoted by

$$V^*(p) = \sup_\pi V_\pi(p),$$

where the supremum is taken over all recovery policies of the form $\pi = (T_1, T_2, \ldots)$, in which each T_i is a stopping time for the process X^i. Note that we focus on stationary policies because $\{\tau_{1,i}\}$ and $\{\tau_{2,i}\}$ are i.i.d. In the following Proposition, we obtain the optimality condition for $V^*(p)$ and the optimal policy [18].

Proposition 4: *(I) The optimal expected return satisfies*

$$V^*(p) = \sup_{T\in\mathcal{T}} E^p[Y_T],$$

where $\mathcal{T}$ is the set of all stopping times for X, and

$$Y_T \equiv \chi_{\{T<\tau_1+\tau_2\}} e^{-rT}[-k + V^*(0)] + \chi_{\{T\geq\tau_1+\tau_2\}} e^{-r(\tau_1+\tau_2)}(-w).$$

(II) The optimal policy π^* such that $V_{\pi^*}(p) = V^*(p)$ is to recover the undisrupted state of each epoch at T^* that satisfies

$$E^p[Y_{T^*}] = \sup_{T\in\mathcal{T}} E^p[Y_T].$$

The difference between the objective function of the single-disruption model in Eq. (2) and the function $V^*(p)$ of the recurrent disruption model is that the cost k of recovery is replaced by the *net cost* $k - V^*(0)$ of recovery.

Next we investigate the existence of the optimal solution $V^*(p)$ in the parameter regime of our interest.

Proposition 5: *For sufficiently small λ_1, there exist a unique optimal threshold $p^* \in (0,1)$ and a unique optimal solution $V^*(p)$ if Eq. (11) is satisfied.*

In order for p^* to be less than 1, the cost w of failure must be significantly larger than the net cost $k - V^*(0)$ of recovery. If λ_1 is too large, then disruption happens much too often, and the expected future cost $|V^*(0)|$ also takes a large value. Consequently, the ratio of magnitudes of $|w|$ to $|-k + V^*(0)|$ is closer to unity. Thus, $p^* < 1$ only if λ_1 is sufficiently small.

Now that we established the existence and the uniqueness of the optimal solution for the recurrent disruption problem, we can compare the efficacy of the optimal policy for the recurrent disruption model to that of the single disruption model.

We first inspect a numerical example of the dependence of p^* and $1-F(0)$ on λ_1 in a single and recurrent disruption models in Figures 1 and 2, where $\lambda_2 = r = k = 1$. In Figure 1, we note dramatically different dependences of p^* on λ_1 between the two models. In the single disruption model, the optimal threshold probability p^*/ε is slightly smaller in the regime $\lambda_1 \gg \varepsilon$ than in the regime $\varepsilon \gg \lambda_1$; this result is consistent with the intuition that the firm should be more (less) proactive in preventing the failure when the arrival rate of disruption is higher (lower). In the recurrent disruption model, in contrast, p^*/ε appears to be increasing in λ_1; in particular, the value of p^*/ε is large (about 8) when $\lambda_1 \approx 70\varepsilon$. In the model of recurrent disruption, if the arrival of disruption is fast, the optimal recovery should not take place too frequently; otherwise, the cumulative discounted cost would be too high. Consequently, the optimal threshold p^* for the recurrent disruption model must be much higher (Figure 1). Because p^* is higher, the probability $1-F(0)$ of failure is also higher (Figure 2) than that of the single-disruption model. Our numerical example demonstrates that, if the arrival rate of disruption is not too low, the assumption of single-disruption can lead to qualitatively inaccurate results when disruption is recurrent.

Next we analytically explore the parameter regimes $\varepsilon = k(\lambda_2 + r)/(w\lambda_2) \ll \lambda_1 \ll 1$ and $\lambda_1 \ll \varepsilon \ll 1$.

Proposition 6: *(I) In the small λ_1 and small ε/λ_1 limit, the optimal threshold probability is given by*

$$p^* = \sqrt{\frac{k\lambda_1\lambda_2}{w}}\left[1 + O\left(\sqrt{\frac{\varepsilon}{\lambda_1}}\right)\right],$$

and the optimal return at $p = 0$ is given by

$$V^*(0) = -\frac{\sqrt{wk\lambda_1\lambda_2}}{r}\left[1 + O\left(\sqrt{\frac{\varepsilon}{\lambda_1}}\right)\right].$$

Under the optimal policy, the probability of eventual failure at $p = 0$ for each epoch is given by

$$1 - F(0) = \frac{k{\lambda_2}^2}{2w}\left[1 + O\left(\sqrt{\frac{\varepsilon}{\lambda_1}}\right)\right] = O(\varepsilon).$$

(II) In the small ε and small λ_1/ε limit, the optimal threshold probability is given by

$$p^* = \frac{b-a-1}{b-a-2}\varepsilon[1 + o(1))], \tag{15}$$

where a and b are given by Proposition 2(II). The optimal return at $p = 0$ is given by

$$V^*(0) = -\frac{w\lambda_2\lambda_1}{r(\lambda_2 + r)}[1 + o(1)]. \tag{16}$$

Under the optimal policy, the probability of eventual failure at $p = 0$ for each epoch is given by

$$1 - F(0) = 1 - \frac{\Gamma(a')}{\Gamma(b'-1)}\left(\frac{p^*}{2\lambda_1}\right)^{1-b'+a'}[1 + o(1)]. = 1 - O\left(\frac{\lambda_1}{\varepsilon}\right)^{b'-a'-1}. \tag{17}$$

Note the similarities and differences between Propositions 6 and 3. In particular, Propositions 6(II) and 3(II) are essentially identical to each other, and hence, the results of the

https://doi.org/10.1017/S0269964813000363 Published online by Cambridge University Press

single-disruption model is robust in the recurrent-disruption extension if $\lambda_1 \ll \varepsilon \ll 1$. When λ_1 is very small, the impact of the recurrent nature of disruption is negligible due to temporal discounting, and hence, the recurrence of disruption does not alter the qualitative features of the optimal policy or the optimal return.

In contrast, we note qualitative differences between Propositions 3(I) and 6(I) for the other regime $\varepsilon \ll \lambda_1 \ll 1$. First, the threshold probability p^* for the recurrent-disruption model is larger than that of the single-disruption model (roughly) by a factor of $(\sqrt{\varepsilon})^{-1}$. The intuitive reason for a high p^* is that, in the case of recurrent disruption, the cost of recovery is too high if p^* is too small because a small p^* entails frequent recovery. Second, the probability of failure for the recurrent-disruption model is larger (roughly) by a factor of ε^{-1}; this follows from the fact that the threshold p^* of recovery is much larger. Lastly, the optimal expected loss $|V^*(0)|$ for the recurrent-disruption model is larger than $|V(0)|$ for the single-disruption model (roughly) by a factor of $(\sqrt{\varepsilon})^{-1}$; this follows from the higher probability of failure for each epoch. We conclude that the efficacy of the optimal policy is significantly worse for the recurrent-disruption model if $\varepsilon \ll \lambda_1 \ll 1$.

5. CONCLUSIONS

History of technological advances shows that even successful firms may fail spectacularly in the face of disruptive innovations. On the other hand, many firms have successfully adapted to disruptive innovations by responding to the sign of technological change in a timely manner. For example, in contrast to Kodak, *Fujifilm* responded to the decline of the photographic film market by successfully diversifying its business in a timely manner [8]. This paper provides a theoretical model and analysis of a firm facing disruption accompanied by weak signals and followed by failure that causes a colossal monetary loss. In the model, the firm watches for the signals of disruption and updates its belief regarding whether the disruption has already happened or not. The firm can invest capital to recover the undisrupted state at any point in time. In particular, we have focused on the case in which disruption is slow to happen.

We find that the optimal policy is to recover the undisrupted state when the probability of disruption exceeds an optimally chosen threshold. In the single-disruption model, as long as the signals of disruption are sufficiently weak, the optimal threshold is small; it is inversely proportional to the cost of failure (Figure 1). However, the probability of eventual failure is strongly sensitive to the arrival rate of disruption; the failure probability is close to 1 when the arrival rate of disruption is very low while it is close to 0 when the arrival rate of disruption is very high (Figure 2). Thus, in an industry where disruptive innovations are very slow to happen, incumbent firms following the optimal policy still fail with very high likelihood *even without irrational complacency*.

If we incorporate the recurrent nature of disruption, qualitatively different features can emerge: the optimal threshold probability is much larger than that of the single-disruption model for moderate arrival rates of disruption (Figure 1), and, consequently, the probability of failure is significantly larger.

Our results have practical implications for enterprises facing disruptive innovations. In planning an adoption of a possibly disruptive technology, an incumbent firm must quantify the size of the projected cost of failure and the arrival rate of technological disruption because the qualitative features of the optimal return and the efficacy of the optimal policy strongly depend on the relative sizes of k/w and λ_1. Moreover, the adoption plan has to take account of the fact that disruption is recurrent, especially if $\lambda_1 \gg k/w$. Finally, although our model has been motivated by the problem of disruptive technological innovations, it is also applicable to any business or economic situations in which sudden catastrophic failure takes place with weak forewarning signals.

Our results also have an empirical implication regarding the statistical relationship between the rate of failure in the face of a disruptive innovation and the arrival rate of disruption. For instance, our results suggest that fewer firms survive disruptive innovations in an industry in which disruption is extremely slow to happen.

References

1. Abramowitz, M. & Stegun, I.A. (1965). *Handbook of mathematical functions*. New York: Dover Publications.
2. Alvarez, L.H.R. (2001). Reward functionals, salvage values and optimal stopping. *Mathematical Methods of Operations Research* 54: 315–337.
3. Beibel, M. & Lerche, H.R. (1997). A new look at optimal stopping problems related to mathematical finance. *Statistica Sinica* 7: 93–108.
4. Christensen, C.M. (2002). *The innovator's dilemma*. NY: Collins Business Essentials.
5. Christensen, C.M., Anthony, S.D. & Roth, E.A. (2004). *Seeing what's next*. Boston, MA: Harvard Business School Press.
6. Decamps, J.P., Mariotti, T. & Villeneuve, S. (2005). Investment timing under incomplete information. *Mathematics of Operations Research* 30(2) 472–500.
7. Economist (2005). Another Kodak moment (May 12).
8. Economist (2012). The last Kodak moment? (January 14).
9. Gapeev, P.V. (2010). Two switching multiple disorder problems for brownian motions. doi: arXiv:1011.0174.
10. Jensen, R. (1982). Adoption and diffusion of an innovation of uncertain profitability. *Journal of Economic Theory* 27(1) 182–193.
11. Kwon, H.D. & Lippman, S.A. (2011). Acquisition of project-specific assets with bayesian updating. *Operations Research* 59(5) 1119–1130.
12. McCardle, K.F. (1985). Information acquisition and the adoption of new technology. *Management Science* 31(11) 1372–1389.
13. Oksendal, B. (2003). *Stochastic differential equations: an introduction with applications*, 6th edn. Berlin, Germany: Springer.
14. Peskir, G. & Shiryaev, A. (2006). *Optimal stopping and free-boundary problems*. Berlin, Germany: Birkhauser Verlag.
15. Poor, H.V. & Hadjiliadis, O. (2009). *Quickest detection*. Cambridge: Cambridge University Press.
16. Rapoport, A., Stein, W.E. & Burkheimer, G.J. (1979). *Response models for detection of change*. Boston: D. Reidel Publishing Company.
17. Rosenberg, N. (1976). On technological expectations. *Economic Journal* 86(343) 523–535.
18. Ross, S.M. (1983). *Introduction to stochastic dynamic programming*. San Diego, CA: Academic Press.
19. Ryan, R. & Lippman, S.A. (2005). Optimal exit from a deteriorating project with noisy returns. *Probability in the Engineering and Informational Sciences* 19(03) 327–343.
20. Shiryaev, A. (1978). *Optimal stopping rules*. New York: Springer-Verlag.
21. Shiryaev, A.N. (1963). On optimum methods in quickest detection problems. *Theory of Probability and its Applications* 8 22–46.
22. Spector, M. & Mattioli, D. (2012). Kodak teeters on the brink. *The Wall Street Journal*. http://online.wsj.com/article/SB10001424052970203471004577140841495542810.html.
23. Ulu, C. & Smith, J.E. (2009). Uncertainty, information acquisition, and technology adoption. *Operations Research* 57(3) 740–752.

APPENDIX

Proof of Proposition 1: We closely follow the derivation of SDE in Section 22, [14]. Let p denote the prior probability that the firm is in the disrupted state at time 0. The probability measure conditional on the prior p can be expressed as follows:

$$\mathbb{P}_p = p\int_0^\infty \mathbb{P}^{0,u}\lambda_2 e^{-\lambda_2 u}du + (1-p)\int_0^\infty \left(\int_0^\infty \mathbb{P}^{v,u}\lambda_2 e^{-\lambda_2 u}du\right)\lambda_1 e^{-\lambda_1 v}dv,$$

where

$$\mathbb{P}^{v,u}(X \in \cdot) \equiv \mathbb{P}(X \in \cdot | \tau_1 = v, \tau_2 = u)$$

is a shorthand notation for conditional probability measure given that the disruption takes place at time v and the failure takes place at time $v+u$. Then from the Bayes rule,

$$P_t \equiv \mathbb{P}_p(\tau_1 \le t | t < \tau_1 + \tau_2, \mathcal{F}_t) = \frac{\mathbb{P}_p(\tau_1 \le t < \tau_1 + \tau_2 | \mathcal{F}_t)}{\mathbb{P}_p(t < \tau_1 + \tau_2 | \mathcal{F}_t)} = \frac{N}{D},$$

where

$$N = p \int_t^\infty \frac{d\mathbb{P}^{0,u}}{d\mathbb{P}^{\infty,u}}(t, X_t)\lambda_2 e^{-\lambda_2 u} du + (1-p) \int_0^t \left(\int_{t-v}^\infty \frac{d\mathbb{P}^{v,u}}{d\mathbb{P}^{\infty,u}}(t, X_t)\lambda_2 e^{-\lambda_2 u} du \right) \lambda_1 e^{-\lambda_1 v} dv$$

and

$$D = p \int_t^\infty \frac{d\mathbb{P}^{0,u}}{d\mathbb{P}^{\infty,u}}(t, X_t)\lambda_2 e^{-\lambda_2 u} du + (1-p) \int_0^t \left(\int_{t-v}^\infty \frac{d\mathbb{P}^{v,u}}{d\mathbb{P}^{\infty,u}}(t, X_t)\lambda_2 e^{-\lambda_2 u} du \right) \lambda_1 e^{-\lambda_1 v} dv$$
$$+ (1-p) \int_t^\infty \left(\int_0^\infty \frac{d\mathbb{P}^{v,u}}{d\mathbb{P}^{\infty,u}}(t, X_t)\lambda_2 e^{-\lambda_2 u} du \right) \lambda_1 e^{-\lambda_1 v} dv.$$

We define

$$Z_t \equiv \exp\left[X_t - \frac{1}{2} t \right],$$

for $t > 0$. Following the procedure from p. 309 of [14], we identify that

$$\frac{d\mathbb{P}^{v,u}}{d\mathbb{P}^{\infty,u}}(t, X_t) = \begin{cases} Z_t / Z_v & \text{if } v \le t \\ 1 & \text{if } v > t \end{cases},$$

and we obtain

$$N = p \int_t^\infty Z_t \lambda_2 e^{-\lambda_2 u} du + (1-p) \int_0^t \left(\int_{t-v}^\infty \frac{Z_t}{Z_v} \lambda_2 e^{-\lambda_2 u} du \right) \lambda_1 e^{-\lambda_1 v} dv,$$
$$D = N + (1-p) \int_t^\infty \left(\int_0^\infty \lambda_2 e^{-\lambda_2 u} du \right) \lambda_1 e^{-\lambda_1 v} dv = N + (1-p) e^{-\lambda_1 t}.$$

From the expression $P_t = N/D$, we obtain the posterior P_t as a functional of X_t, and we can derive the SDE in Eq. (1) using Ito's formula. ■

Proof of Lemma 1: We introduce a random variable Y given by

$$Y \equiv \chi_{\{T < \tau_1 + \tau_2\}} e^{-rT}(-k) + \chi_{\{T \ge \tau_1 + \tau_2\}} e^{-r(\tau_1 + \tau_2)}(-w) \quad \textbf{(A.1)}$$

so that $V_T(t,p) = E^{(t,p)}[Y]$. Now we consider a stopping time δ such that $t + \delta < T$, that is, $t + \delta$ is an exit time from an interval contained within C. Then we can write

$$Y = \chi_{\{\tau_1 + \tau_2 < t + \delta\}} Y + \chi_{\{\tau_1 + \tau_2 \ge t + \delta\}} Y.$$

Our goal is to express $V_T(t,p)$ in terms of $E^{(t,p)}[V_T(t+\delta, P_{t+\delta})]$ to derive the differential equation $\mathcal{A} V_T(t,p) - p\lambda_2 w \cdot e^{-rt} = 0$.

https://doi.org/10.1017/S0269964813000363 Published online by Cambridge University Press

One of the terms of $E^{(t,p)}[Y]$ is

$$\begin{aligned} &E^{(t,p)}[\chi_{\{\tau_1+\tau_2<t+\delta\}}Y] \\ &\quad = E^{(t,p)}\left\{\chi_{\{\tau_1+\tau_2<t+\delta\}}\left[\chi_{\{T<\tau_1+\tau_2\}}e^{-rT}(-k)+\chi_{\{T\geq\tau_1+\tau_2\}}e^{-r(\tau_1+\tau_2)}(-w)\right]\right\}. \end{aligned}$$

If $\tau_1+\tau_2<t+\delta$, then $T>\tau_1+\tau_2$ is automatically satisfied, so we have

$$\begin{aligned} &E^{(t,p)}[\chi_{\{\tau_1+\tau_2<t+\delta\}}Y] \\ &\quad = E^{(t,p)}[\chi_{\{\tau_1+\tau_2<t+\delta\}}e^{-r(\tau_1+\tau_2)}(-w)] \\ &\quad = -we^{-rt}E^{(t,p)}\left[p\int_0^\delta e^{-ru}\lambda_2 e^{-\lambda_2 u}du+(1-p)\int_0^\delta e^{-rv}\lambda_1 e^{-\lambda_1 v}\int_0^{\delta-v}e^{-ru}\lambda_2 e^{-\lambda_2 u}du\,dv\right] \end{aligned}$$

We note that

$$\frac{E^{(t,p)}[\chi_{\{\tau_1+\tau_2<t+\delta\}}Y]}{E^{(t,p)}(\delta)}\to -wpe^{-rt}\lambda_2 \quad \text{as} \quad \delta\to 0.$$

The other term of $E^{(t,p)}[Y]$ is

$$\begin{aligned} E^{(t,p)}[\chi_{\{\tau_1+\tau_2\geq t+\delta\}}Y] &= E^{(t,p)}[E[\chi_{\{\tau_1+\tau_2\geq t+\delta\}}Y|\mathcal{F}_{t+\delta}]] \\ &= E^{(t,p)}[\chi_{\{\tau_1+\tau_2\geq t+\delta\}}E[Y|\mathcal{F}_{t+\delta};\tau_1+\tau_2\geq t+\delta]] \\ &= E^{(t,p)}[\chi_{\{\tau_1+\tau_2\geq t+\delta\}}V_T(t+\delta,P_{t+\delta})]. \end{aligned}$$

In the last line of the equation above, the only dependence on τ_1 and τ_2 is in the factor $\chi_{\{\tau_1+\tau_2\geq t+\delta\}}$; hence, we can take its expectation with respect to τ_1 and τ_2:

$$E[\chi_{\{\tau_1+\tau_2\geq t+\delta\}}|P_t=p]=E[(\chi_{\{\tau_1\leq t\}}+\chi_{\{\tau_1>t\}})\chi_{\{\tau_1+\tau_2\geq t+\delta\}}|P_t=p].$$

The first term is

$$E[\chi_{\{\tau_1\leq t\}}\chi_{\{\tau_1+\tau_2\geq t+\delta\}}|P_t=p]=p\int_\delta^\infty \lambda_2 e^{-\lambda_2 u}du=pe^{-\lambda_2\delta},$$

and the second term is

$$\begin{aligned} &E[\chi_{\{\tau_1>t\}}\chi_{\{\tau_1+\tau_2\geq t+\delta\}}|P_t=p] \\ &\quad =(1-p)\left(\int_0^\delta \lambda_1 e^{-\lambda_1 v}\int_{\delta-v}^\infty \lambda_2 e^{-\lambda_2 u}du\,dv+\int_\delta^\infty \lambda_1 e^{-\lambda_1 v}\int_0^\infty \lambda_2 e^{-\lambda_2 u}du\,dv\right) \\ &\quad =(1-p)-(1-p)\int_0^\delta \lambda_1 e^{-\lambda_1 v}\int_0^{\delta-v}\lambda_2 e^{-\lambda_2 u}du\,dv. \end{aligned}$$

Thus, we have

$$\begin{aligned} &E^{(t,p)}[\chi_{\{\tau_1+\tau_2\geq t+\delta\}}Y] \\ &\quad = E^{(t,p)}[V_T(t+\delta,P_{t+\delta})] \\ &\qquad + E^{(t,p)}\left\{\left[p(e^{-\lambda_2\delta}-1)-(1-p)\int_0^\delta \lambda_1 e^{-\lambda_1 v}\int_0^{\delta-v}\lambda_2 e^{-\lambda_2 u}du\,dv\right]V_T(t+\delta,P_{t+\delta})\right\}. \end{aligned}$$

https://doi.org/10.1017/S0269964813000363 Published online by Cambridge University Press

We note that the second term above divided by $E^{(t,p)}[\delta]$ converges to

$$-p\lambda_2 V_T(t,p)$$

as $\delta \to 0$ because $V_T(t+\delta, P_{t+\delta})$ is a bounded function. Finally, collecting all the terms that constitute $V_T(t,p)$ and subtracting $E^{(t,p)}[V_T(t+\delta, P_{t+\delta})]$, we obtain

$$-(\mathcal{L}-r)V_T(t,p) = \lim_{\delta\to 0} \frac{V_T(t,p) - E^{(t,p)}[V_T(t+\delta, P_{t+\delta})]}{E^{(t,p)}[\delta]} = -p\lambda_2[w\cdot e^{-rt} + V_T(t,p)].$$

Thus, $\mathcal{A}V_T(t,p) - p\lambda_2 w\cdot e^{-rt} = 0$ is satisfied. ■

Proof of Proposition 2: (I) We first find a solution to Eq. (5) of the form $V(t,p) = e^{-rt}V(p)$ where $V(p)$ satisfies $\mathcal{A}V(p) - p\lambda_2 w = 0$ for $p < p^*$. We separate $V(\cdot)$ into two parts: $V_0(\cdot)$ and $V_1(\cdot)$, where $V_0(\cdot)$ satisfies $\mathcal{A}V_0(p) = 0$ and $V_1(\cdot)$ satisfies $\mathcal{A}V_1(p) - p\lambda_2 w = 0$. It is straightforward to verify that

$$V_1(p) = -w\frac{\lambda_2(\lambda_1 + rp)}{(\lambda_1 + r)(\lambda_2 + r)}.$$

The equation $\mathcal{A}V_0(p) = 0$ can be re-expressed as follows:

$$\frac{1}{2}p^2(1-p)^2 V_0''(p) + [\lambda_1(1-p) - \lambda_2 p(1-p)]V_0'(p) - (r + \lambda_2 p)V_0(p) = 0.$$

For convenience, we introduce a variable $z \equiv p/(1-p)$ and re-express the differential equation above in terms of z and derivatives with respect to z as follows:

$$V_0''(z) + 2\left[\frac{1}{z+1} + \frac{(\lambda_1 - \lambda_2)}{z} + \frac{\lambda_1}{z^2}\right]V_0'(z) - 2\left[\frac{r}{z^2} + \frac{\lambda_2}{z(z+1)}\right]V_0(z) = 0,$$

where $V_0''(z) \equiv d^2V_0(z)/dz^2$ and $V_0'(z) \equiv dV_0(z)/dz$. To solve the differential equation, we use 13.1.35 and 13.1.37 of [1] by substituting

$$A = a = \lambda_1 - \lambda_2 - \frac{1}{2} + \sqrt{\left(\lambda_1 - \lambda_2 - \frac{1}{2}\right)^2 + 2(\lambda_1 + r)},$$

$$b = 1 + 2\sqrt{\left(\lambda_1 - \lambda_2 - \frac{1}{2}\right)^2 + 2(\lambda_1 + r)},$$

$$f(z) = \ln(z+1),$$

$$h(z) = \frac{2\lambda_1}{z},$$

and the solution is given by $V_0(z) = c\cdot\phi(z)$ for some constant c, where

$$\phi(z) = \frac{z^{-a}(2\lambda_1)^a}{z+1}U(a, b, \frac{2\lambda_1}{z}) = \frac{1}{\Gamma(a)\cdot(z+1)}\int_0^\infty e^{-u}u^{a-1}\left(1 + u\frac{z}{2\lambda_1}\right)^{b-a-1} du.$$

Here $U(\cdot,\cdot,\cdot)$ is the Kummer's function (13.1 of [1]) that takes finite values as $z \to 0$, and its integral representation is given by 13.2.5 of [1].

https://doi.org/10.1017/S0269964813000363 Published online by Cambridge University Press

We re-expressing $V_1(\cdot)$ as a function of z as

$$V_1(p) = \tilde{V}_1(z) = -w\frac{\lambda_2[\lambda_1 + rz/(z+1)]}{(\lambda_1 + r)(\lambda_2 + r)},$$

so that we can express the candidate solution $V(\cdot)$ in terms of z as

$$V(p) = c \cdot \phi(z) + \tilde{V}_1(z).$$

Eqs. (6) and (7) combined are equivalent to the following pair of equations:

$$c \cdot \phi(z^*) + \tilde{V}_1(z^*) = -k,$$
$$c \cdot \phi'(z^*) + \tilde{V}_1'(z^*) = 0,$$

where $z^* \equiv p^*/(1-p^*)$. These two conditions lead to the following equations:

$$c = \frac{-k - \tilde{V}_1(z^*)}{\phi(z^*)},$$
$$[-k - \tilde{V}_1(z^*)]\phi'(z^*) - \phi(z^*)[-k - \tilde{V}_1(z^*)]' = 0. \qquad \textbf{(A.2)}$$

Because $\phi(\cdot)$ is positive definite, Eq. (A.2) is equivalent to the following equation:

$$\left(\frac{-k - \tilde{V}_1(z^*)}{\phi(z^*)}\right)' = 0. \qquad \textbf{(A.3)}$$

For notational convenience, we introduce the following functions:

$$\begin{aligned} n(z) &= (z+1)(-k - \tilde{V}_1(z)) \\ d(z) &= \frac{1}{\Gamma(a)}\int_0^\infty e^{-u}u^{a-1}\left(1 + u\frac{z}{2\lambda_1}\right)^{b-a-1} du \\ g(z) &= \frac{n(z)}{d(z)}, \end{aligned} \qquad \textbf{(A.4)}$$

so that we can succinctly express Eq. (A.3) as

$$g'(z^*) = 0.$$

We first assume that $c > 0$, which is possible for sufficiently large values of z^* as long as Eq. (11) is satisfied. (We will justify the condition $c > 0$ later). We prove that there exists a unique solution to $g'(z) = 0$ in the set $G \equiv \{z : g(z) > 0\}$. (Because we have $n(z^*) > 0$ from the assumption that $c > 0$, we focus on the region of z in which $n(z) > 0$, or equivalently, $g(z) > 0$.) In particular, we show that $g(z)$ achieves its positive maximum at $z^* \in (0, \infty)$ if the inequality (11) is satisfied. To do so, it is sufficient to prove the following statements: (i) $g(z) > 0$ for sufficiently large z; (ii) $g'(z)$ changes its sign from positive to negative at most once in the set G; (iii) $g'(z) > 0$ for some $z \in G$; and (iv) $g(z) \to 0$ as $z \to \infty$.

(i) Because of the inequality (11), $n(z)$ is an increasing linear function, and hence, $n(z) > 0$ is satisfied for sufficiently large z. The statement (i) follows because $d(z) > 0$ for all $z \in [0, \infty)$.

(ii) We first note that $d''(z) > 0$ from the integral representation of Eq. (A.7) because $b - a - 2 > 0$ for all model parameters, which is straightforward to prove. Next we use the relation

$$[d(z)^2 g'(z)]' = -n(z)d''(z) < 0,$$

which implies that $d(z)^2 g'(z)$ is a strictly decreasing function in the set G. Thus, $g'(z)$ changes its sign from positive to negative at most once in the set G.

(iii) Suppose $g(0) < 0$. From the fact that $n(z)$ is an increasing linear function of z, there is some $y > 0$ such that $g(z) < 0$ for $z < y$ and $g(z) > 0$ for $z > 0$. Hence, $g'(z) > 0$ for some $z > y$ by the mean value theorem.

If $g(0) > 0$, it is sufficient to prove that $g'(0) > 0$. From the definition of $g(\cdot)$, it is straightforward to show that

$$g'(0) = \frac{d(0)n'(0) - n(0)d'(0)}{d(0)^2} = \frac{rk}{\lambda_1} > 0.$$

(iv) In the large-z limits,

$$d(z) = \frac{\Gamma(b-1)}{\Gamma(a)} \left(\frac{z}{2\lambda_1}\right)^{b-1-a} [1 + o(1)],$$

so $g(z)$ approaches zero from above as z^{2+a-b} in the limit $z \to \infty$ because $2 + a - b < 0$.

Now we come back to the question regarding the sign of c and show that Eq. (8) is satisfied if and only if $c > 0$. We note that Eqs. (6), (7), and (8) lead to two conditions $c \cdot d'(z^*) = n'(z^*)$ and $c \cdot d(z) > n(z)$ for $z < z^*$. Because $d(\cdot)$ is strictly convex and $n(\cdot)$ is linear, the two conditions can be satisfied if and only if $c > 0$.

Finally, Eq. (10) translates into

$$\mathcal{A}(-k) - p\lambda_2 w = k(r + p\lambda_2) - p\lambda_2 w = p\lambda_2(k - w) + kr < 0 \quad \text{for all } p \geq p^*. \qquad \textbf{(A.5)}$$

From Eq. (11), we know that $k < w$, so the condition (A.5) is equivalent to $\mathcal{A}(-k) - p\lambda_2 w < 0$ at $p = p^*$. From conditions (8), (6), and (7), we have $V(p) > -k$ for all $p < p^*$, $V(p^*) = -k$, and $V'(p^*) = 0$. Using the expression $V(p) = c\phi(p/(1-p)) + V_1(p)$ and the fact that $d^2V_1(p)/dp^2 = 0$, we obtain the following:

$$\lim_{p \uparrow p^*} [\mathcal{A}V(p) - p\lambda_2 w] = \lim_{p \uparrow p^*} \left\{ \frac{1}{2} p^2 (1-p)^2 \frac{d^2}{dp^2} \left[c\phi\left(\frac{p}{1-p}\right)\right] + \mathcal{A}(-k) - p\lambda_2 w \right\}. \qquad \textbf{(A.6)}$$

From the expression of $\phi(\cdot)$, we obtain its second derivative with respect to p as follows:

$$\frac{d^2}{dp^2}\phi(z) = \frac{d^2}{dp^2}\left[\frac{1}{z+1} d(z)\right] = (z+1)d''(z) > 0.$$

By Eq. (A.6) and the fact that $\mathcal{A}V(p) - p\lambda_2 w = 0$ for all $p < p^*$, we obtain

$$\mathcal{A}(-k) - p\lambda_2 w = -\frac{1}{2} p^2 (1-p)^2 \frac{d^2}{dp^2} \left[c\phi\left(\frac{p}{1-p}\right)\right] < 0 \quad \text{for } p = p^*.$$

(II) Let $V^*(t,p) = \sup_T V_T(t,p)$ be the supremum return function and let $C^* = [0, p^*)$, where p^* is given by (I). We let $T^* = \inf\{t > 0 : P_t \notin C^*\}$ denote the exit time of P_t from C^*. Our goal is to prove that $V^*(t,p) = V(t,p) = V_{T^*}(t,p)$. Because $V^*(t,p)$ is a bounded function, we can use the same argument as in the proof of Proposition 7.1 in [6] to conclude that $V_{T^*}(t,p) = V(t,p)$. By Theorem 10.4.1 (variational inequalities) of [13], statement (I) implies that $V(t,p) \geq V^*(t,p)$. From the fact that $V^*(t,p) \geq V_{T^*}(t,p)$, it follows that $V^*(t,p) = V(t,p)$. ■

Proof of Lemma 2: We note that $F(p) = E^{(0,p)}[Y]$, where Y is the random variable given by Eq. (A.1) from the proof of Lemma 1 if we set $r = 0$, $k = -1$, and $w = 0$. Thus, by Lemma 1, $F(p)$ satisfies

$$\mathcal{L}F(p) - p\lambda_2 F(p) = 0 \quad \text{for } p < p^*,$$

along with the condition that $F(p) \leq 1$ for $p < p^*$ and the boundary condition $F(p^*) = 1$. It follows that $F(p)$ has the same functional form as $V_0(p)$ defined in the proof of Proposition 2 except that $r = 0$, and the statement of the Lemma follows. ■

https://doi.org/10.1017/S0269964813000363 Published online by Cambridge University Press

Proof of Proposition 3: (I) We search for the optimal threshold p^* of order $O(\varepsilon)$. The search is complete once we find such a solution because of the uniqueness of the optimal solution by Proposition 2.

Using the notation adopted by the proof of Proposition 2, we obtain the following expressions in the limit $z/\lambda_1 \to 0$:

$$d(z) = 1 + \frac{z}{\lambda_1}(\lambda_1 + r) + \frac{1}{2}\left(\frac{z}{\lambda_1}\right)^2 (\lambda_1 + r)(\lambda_2 + r) + O\left(\frac{z}{\lambda_1}\right)^3, \tag{A.7}$$

from 13.5.2 of [1]. The equation for p^* (or z^*) can be written as $d(z^*)/d'(z^*) = n(z^*)/n'(z^*)$ from the proof of Proposition 2, and its asymptotic form is given by

$$\frac{\lambda_1}{\lambda_1 + r}\left[1 + \frac{z^*}{\lambda_1}(\lambda_1 - \lambda_2) + O\left(\frac{z^*}{\lambda_1}\right)^2\right] = \frac{\lambda_1/(\lambda_1 + r) - \varepsilon}{1 - \varepsilon} + z^*,$$

from which we obtain

$$z^* = \varepsilon \frac{r}{\lambda_2 + r}\left[1 + O\left(\frac{\varepsilon}{\lambda_1}\right)\right],$$

and $p^* = z^*/(z^* + 1)$ follows. By Proposition 2,

$$\begin{aligned} V(0) &= c\phi(0) + V_1(0) \\ &= \frac{w\lambda_2\lambda_1}{\lambda_2 + r}\left[\frac{1}{\lambda_1 + r} - \frac{\varepsilon}{\lambda_1} + O\left(\frac{\varepsilon}{\lambda_1}\right)^2\right] - \frac{w\lambda_2\lambda_1}{(\lambda_2 + r)(\lambda_1 + r)} = -k\left[1 + O\left(\frac{\varepsilon}{\lambda_1}\right)\right], \end{aligned}$$

Here the asymptotic expression of $\phi(\cdot)$ is derived from $\phi(z) = d(z)/(z + 1)$ and Eq. (A.7).

Finally, by Lemma 2,

$$F(0) = \frac{\psi(0)}{\psi(z^*)} = 1 - \frac{1}{2}\lambda_1\lambda_2\left(\frac{z^*}{\lambda_1}\right)^2 + O\left(\frac{z^*}{\lambda_1}\right)^3,$$

from which Eq. (12) follows.

(II) Similarly to the proof of (I), we search for the optimal threshold p^* of order $O(\varepsilon)$. If p^* (or z^*) is of order $O(\varepsilon)$, then z^*/λ_1 is a large number in the limit that we consider. Thus, we inspect the asymptotic behavior of $d(z)$ in the limit $z/\lambda_1 \to \infty$:

$$d(z) = \frac{\Gamma(b-1)}{\Gamma(a)}\left(\frac{z}{2\lambda_1}\right)^{b-a-1}[1 + o(1)],$$

from 13.5.6-8 of [1]. The equation for p^* (or z^*) can be written as $d(z^*)/d'(z^*) = n(z^*)/n'(z^*)$ from the proof of Proposition 2, and its asymptotic form is given by

$$\frac{z^*}{b - a - 1}[1 + o(1)] = \frac{\lambda_1/(\lambda_1 + r) - \varepsilon}{1 - \varepsilon} + z^*, \tag{A.8}$$

from which we obtain Eq. (13) by $p^* = z^*/(z^* + 1)$.

By Proposition 2,

$$c\phi(0) = \frac{w\lambda_2\lambda_1}{\lambda_2 + r}\left(\frac{1}{\lambda_1 + r} - \frac{\varepsilon}{\lambda_1} + (1 - \varepsilon)\frac{z^*}{\lambda_1}\right)\frac{\Gamma(a)}{\Gamma(b-1)}\left(\frac{z^*}{\lambda_1}\right)^{a+1-b}[1 + o(1)].$$

Because $a + 2 - b < 0$, we obtain the limiting behaviors $(\varepsilon/\lambda_1)^{a+2-b} \to 0$ as $\varepsilon/\lambda_1 \to \infty$, and hence, we arrive at the asymptotic behavior of $V(0) = c\phi(0) + V_1(0) = V_1(0)[1 + o(1)]$.

Finally, by Lemma 2,

$$F(0) = \frac{\psi(0)}{\psi(z^*)} = \frac{\Gamma(a')}{\Gamma(b'-1)} \left(\frac{z^*}{\lambda_1}\right)^{a'+1-b'} [1 + o(1)],$$

from which Eq. (14) follows. ■

Proof of Proposition 4: We closely follow the proofs of Theorems 2.1 and 2.2 from Chapter II of [18]. (I) Let π be an arbitrary policy that chooses to recover the undisrupted state at a series of stopping times $T_1, T_2, T_3, \ldots$, where T_n is the recovery time for the n-th epoch. Let W_π be the return from the policy of recovery at stopping times $T_2, T_3, \ldots$, measured at the time the new technology for the second epoch emerges, that is, it is the return from the policy π starting from time $t_2 = 0$ in the second epoch. Then we have

$$V_\pi(p) = E^p \left\{ \chi_{\{T_1 < \tau_1+\tau_2\}} e^{-rT_1} [-k + W_\pi(0)] + \chi_{\{T_1 \geq \tau_1+\tau_2\}} e^{-r(\tau_1+\tau_2)} (-w) \right\}.$$

From the fact that $W_\pi(0) \leq V^*(0)$, we have the following inequalities:

$$\begin{aligned} V_\pi(p) &\leq E^p \left\{ \chi_{\{T_1 < \tau_1+\tau_2\}} e^{-rT_1} [-k + V^*(0)] + \chi_{\{T_1 \geq \tau_1+\tau_2\}} e^{-r(\tau_1+\tau_2)} (-w) \right\} \\ &= E^p[Y_{T_1}] \leq \sup_{T \in \mathcal{T}} E^p[Y_T]. \end{aligned}$$

The inequality above implies that $V^*(p) \leq \sup_{T \in \mathcal{T}} E^p[Y_T]$. It remains to prove the opposite inequality.

From Proposition 2, there exists $T^* \in \mathcal{T}$ such that $E^p[Y_{T^*}] = \sup_{T \in \mathcal{T}} E^p[Y_T]$ because $V^*(0)$ is simply a deterministic constant. Let π be a particular policy such that the recovery time for the first epoch is T^* and the subsequent recovery times are such that $W_\pi(0) \geq V^*(0) - \delta$ for some arbitrarily small number $\delta > 0$. Then we obtain

$$\begin{aligned} V_\pi(p) \geq\, & E^p\{\chi_{\{T^* < \tau_1+\tau_2\}} e^{-rT^*} [-k + V^*(0)] + \chi_{\{T^* \geq \tau_1+\tau_2\}} e^{-r(\tau_1+\tau_2)} (-w)\} \\ & - \delta E^p\{\chi_{\{T^* < \tau_1+\tau_2\}} \exp[-rT^*]\}. \end{aligned}$$

From the fact that $V^*(p) \geq V_\pi(p)$ and that δ is arbitrary, we obtain $V^*(p) \geq \sup_{T \in \mathcal{T}} E^p[Y_T]$.

(II) We can re-write the optimality condition for $V^*(p)$ as follows:

$$\begin{aligned} V^*(p) = E^p&[\chi_{\{T^* < \tau_1+\tau_2\}} e^{-rT^*} (-k) + \chi_{\{T^* \geq \tau_1+\tau_2\}} e^{-r(\tau_1+\tau_2)} (-w)] \\ &+ E^p\{\chi_{\{T^* < \tau_1+\tau_2\}} \exp[-r(T^* + e)]\} V^*(0). \end{aligned} \tag{A.9}$$

By the argument of the proof of Proposition 2, the stopping time T^*, if it exists, is characterized by an upper threshold p^* with respect to the posterior P_t; if $p < p^*$, then we have $E^p\{\chi_{\{T^* < \tau_1+\tau_2\}} \exp[-rT^*]\} \equiv \alpha_p \in (0,1)$. The term $V^*(0)$ in Eq. (A.9) can be replaced by the following:

$$V^*(0) = E^0[\chi_{\{T^* < \tau_1+\tau_2\}} e^{-rT^*} (-k) + \chi_{\{T^* \geq \tau_1+\tau_2\}} e^{-r(\tau_1+\tau_2)} (-w)] + \alpha_0 V^*(0).$$

Repeating the substitution n times, we can identify $V^*(p)$ as the return from the policy of recovering each epoch at T^* for n times and getting a final reward of $V^*(0)$ at the end, which is discounted by the factor of $\alpha_p \alpha_0^{n-1}$. Thus, in the limit $n \to \infty$, the policy of recovery at T^* yields $V^*(p)$. ■

Proof of Proposition 5: Let x represent the candidate for the value of $V^*(0)$ that satisfies

$$V^*(0) = E^0 \left\{ \chi_{\{T^* < \tau_1 + \tau_2\}} e^{-rT^*} \left[-k + V^*(0) \right] + \chi_{\{T^* \geq \tau_1 + \tau_2\}} e^{-r(\tau_1 + \tau_2)} (-w) \right\}, \qquad \textbf{(A.10)}$$

where T^* satisfies $E^p[Y_{T^*}] = \sup_{T \in \mathcal{T}} E^p[Y_T]$ as in Proposition 4 (II). Define

$$f(x) \equiv \sup_{T \in \mathcal{T}} E^0 [\chi_{\{T < \tau_1 + \tau_2\}} e^{-rT^*} (-k + x) + \chi_{\{T \geq \tau_1 + \tau_2\}} e^{-r(\tau_1 + \tau_2)} (-w)],$$

then the equation for $V^*(0)$ reduces to $x = f(x)$. (It is straightforward to verify that $f(\cdot)$ is continuous by Proposition 2). It is obvious that $f(0) < 0$, so $x > f(x)$ for $x = 0$. Moreover, by the envelope theorem, $df(x)/dx = E^0 [\chi_{\{T < \tau_1 + \tau_2\}} e^{-rT^*}] < 1$, and hence, the solution to $x = f(x)$ is unique if it exists.

For the optimal threshold $p^* \in (0, 1)$ to exist such that $E^p[Y_{T^*}] = \sup_{T \in \mathcal{T}} E^p[Y_T]$ for $T^* = \inf\{t > 0 : P_t \notin [0, p^*)\}$, the following condition must be satisfied:

$$w \frac{\lambda_2}{(\lambda_2 + r)} > k - V^*(0),$$

which is derived from Eq. (11). It implies that $p^* = 1$ (the optimal policy is to never recover the undisrupted state) if $-V^*(0) = -k + w\lambda_2/(\lambda_2 + r)$. Hence, if $x = k - w\lambda_2/(\lambda_2 + r)$, which is negative by the assumption Eq. (11), we obtain

$$f(x) = E^0 [e^{-r(\tau_1 + \tau_2)} (-w)] = (-w) \frac{\lambda_2 \lambda_1}{(\lambda_1 + r)(\lambda_2 + r)}.$$

For sufficiently small values of λ_1, the above value is greater than x because $x < 0$. Thus, we obtain $x > f(x)$ for $x = 0$ and $x < f(x)$ for $x = k - w\lambda_2/(\lambda_2 + r) < 0$, and it follows that there exists a unique solution to $x = f(x)$ for some x between $k - w\lambda_2/(\lambda_2 + r)$ and 0. ■

Proof of Proposition 6: For notational convenience, we introduce a new parameter:

$$\zeta = \frac{|V^*(0) - k|(\lambda_2 + r)}{|w|\lambda_2}.$$

(I) We first note that Proposition 2 exactly applies to the optimal solution $V^*(p)$ except that k is replaced by $k - V^*(0)$. Our goal is to find the optimal threshold p^* or $z^* = p^*/(1 - p^*)$. In order to find the solution, it is useful to make an assumptions about the solution and confirm that these assumptions are satisfied by the solution obtained. Hence, we assume that $z^*/\lambda_1 \ll 1$ and check to see if there exists a solution $V^*(0)$ of Eq. (A.10) that satisfies this condition. Once we find one solution $V^*(0)$, because of the uniqueness of the solution to Eq. (A.10), there would be no other solution.

The equation for the optimal threshold z^* can be obtained by a similar procedure to that of the proof of Proposition 2. Define

$$m(z) = \left[-k + V^*(0) \right] (z + 1) + w \frac{\lambda_2 [\lambda_1 (z + 1) + rz]}{(\lambda_1 + r)(\lambda_2 + r)},$$

which is the same as $n(z)$ in proof of Proposition 2 except that k is replaced by $k - V^*(0)$. By the same argument used in the proof of Proposition 2, the optimal threshold is given as the solution to $d(z^*)/d'(z^*) = m(z^*)/m'(z^*)$, where $d(z)$ is defined in the proof of Proposition 2. The small-z/λ_1 expansion of $d(z)$ is exactly given by Eq. (A.7). From the equation $d(z^*)/d'(z^*) = m(z^*)/m'(z^*)$,

https://doi.org/10.1017/S0269964813000363 Published online by Cambridge University Press

we obtain

$$\frac{z^*}{\lambda_1} = \frac{\zeta}{\lambda_1}\frac{r}{(r+\lambda_2)(1-\zeta)} + O\left(\frac{\zeta}{\lambda_1}\right)^2.$$

From the equation for $V^*(0)$ in (A.10), we have

$$V^*(0) = \frac{m(z^*)}{d(z^*)}d(0) - w\frac{\lambda_2\lambda_1}{(\lambda_1+r)(\lambda_2+r)} = [-k+V^*(0)]\left[1 - z^*\frac{r}{\lambda_1} + O\left(\frac{\zeta}{\lambda_1}\right)^2\right],$$

from which we obtain

$$-V^*(0)r\left[\frac{z^*}{\lambda_1} + O\left(\frac{\zeta}{\lambda_1}\right)^2\right] = k. \tag{A.11}$$

Hence, $k/|V^*(0)|$ is of order $O(\zeta/\lambda_1)$. From the definition of ζ and the fact that $k = V^*(0)O(\zeta/\lambda_1)$, Eq. (A.11) is re-expressed as

$$\left\{[V^*(0)]^2 - k\frac{\lambda_1\lambda_2}{r^2}w\right\}\left[1 + O\left(\frac{\zeta}{\lambda_1}\right)\right] = 0,$$

and we obtain

$$V^*(0) = -\frac{\sqrt{wk\lambda_1\lambda_2}}{r}\cdot\left[1 + O\left(\frac{\zeta}{\lambda_1}\right)\right].$$

We can confirm that

$$\frac{\zeta}{\lambda_1} = \frac{1}{r}\sqrt{\frac{k\lambda_2}{w\lambda_1}}\frac{(\lambda_2+r)}{\lambda_2}\left[1 + O\left(\sqrt{\frac{\varepsilon}{\lambda_1}}\right)\right],$$

and hence, $\zeta/\lambda_1 = O(\sqrt{\varepsilon/\lambda_1})$, which is a small parameter. Because $z^*/\lambda_1 = O(\zeta/\lambda_1)$, our assumption $z^*/\lambda_1 \ll 1$ is consistent with the solution of $V^*(0)$ that we obtained.

Finally, the probability of eventual failure can be obtained from Lemma 2 as follows:

$$1 - F(0) = \frac{\lambda_2}{2\lambda_1}(z^*)^2\left[1 + O\left(\frac{\zeta}{\lambda_1}\right)\right].$$

(II) Following the same argument as in the proof (I), we first make an assumption about the solution $V^*(0)$ and justify it later. Assume that $z^*/\lambda_1 \gg 1$ for the moment. By the definition of ζ and from the fact that $\lambda_1 \ll \varepsilon$, we also have $\lambda_1 \ll \zeta$ irrespective of the magnitude of $V^*(0)$. Then we can exactly follow the proof of Proposition 3(II) except that ε is replaced by ζ. From Eq. (A.8), we obtain

$$z^* = \frac{b-a-1}{b-a-2}\cdot\frac{\zeta - \lambda_1/(\lambda_1+r)}{1-\zeta}[1+o(1)] = \frac{b-a-1}{b-a-2}\cdot\frac{\zeta}{1-\zeta}[1+o(1)].$$

Thus, z^* is at least as large as $O(\zeta)$, and the assumption $z^*/\lambda_1 \gg 1$ is justified.

Now we inspect the magnitude of $V^*(0)$. Using the asymptotic property of $\phi(\cdot)$ and the same argument as in the proof of Proposition 3(II), we obtain Eq. (16). Because $V^*(0)/k = O(\lambda_1/\varepsilon) \ll 1$, we have $\zeta = \varepsilon\cdot(1+o(1))$.

Finally, from the fact that both ζ and z^* are of order $O(\varepsilon)$ and using Lemma 2, we obtain Eq. (17). ■