Managing Customer Churn via Service Mode Control*

Yash Kanoria[†], Ilan Lobel[‡], Jiaqi Lu[§¶]

December 1, 2022

Abstract

We introduce a novel stochastic control model for the problem of a service firm interacting over time with one of its customers who probabilistically churns depending on his satisfaction. The firm has two service modes available, which determines the drift and volatility of the Brownian reward process. The firm's objective is to maximize the rewards generated over the customer's lifetime. Meanwhile, the customer might churn probabilistically if his satisfaction, modeled as an Orstein-Uhlenbeck process controlled by the firm's service mode, is below a certain threshold. We build upon Markov processes with spatial delay to solve this problem, and we explicitly characterize the firm's optimal policy, which is either myopic or a sandwich policy. A sandwich policy is one where the firm deploys the service mode with inferior reward rate when the customer satisfaction level is in a specific interval near the satisfaction threshold, and uses the myopically optimal service mode for all other satisfaction levels. Specifically, we find that the firm should use the safe service mode when the customer is marginally satisfied and the risky service mode when the customer is marginally unsatisfied. We find numerically that the customer lifetime value under the optimal policy is large relative to that under the myopic policy. Our results are robust to a variety of alternative model specifications.

Keywords: Ornstein-Uhlenbeck process; spatial delay; stochastic control; risk-seeking; risk-averse; L'Hospital-type rules for monotonicity; goodwill model; customer lifetime value.

1 Introduction

In this paper, we study a stochastic control problem motivated by customer management applications. We consider a service firm repeatedly interacting with a customer whose churn risk is a function of his recent experiences, and explore how such a firm should adapt its service mode as a function of the customer's experiences. Specifically, suppose the firm have access to two service modes, one always produces stable outcomes, but in the long run is not sufficient to satisfy the customer, and another generates volatile outcomes. Should the firm prefer the latter? For example, an investment manager that recommends an aggressive portfolio may lose a customer when a market crash occurs. Meanwhile, a conservative portfolio, while offering protection against market turbulence, might generate insufficient returns over time. A similar tradeoff between a risky service mode and a safe service mode also exists in subscription-based businesses such as recommendation systems, where safe prescriptions may over time bore the customer, while risky recommendations are likely to surprise the customer in either a good or a bad way.

*An online appendix is available at https://tinyurl.com/czt4bcfr. It has also been submitted as supplementary material.

[†]Columbia Business School, Columbia University — ykanoria@gsb.columbia.edu

[‡]Stern School of Business, New York University — ilobel@stern.nyu.edu

[§]The Chinese University of Hong Kong, Shenzhen — lujiaqi@cuhk.edu.cn

[¶]We thank Rene Caldentey and Iosif Pinelis for pointing out relevant research and techniques.

Through solving a stylized stochastic control problem, we show when the firm should deviate from the myopically superior (in terms of expected immediate reward) service mode in order to maximize customer lifetime value. In particular, the firm should use the service mode that generates a higher expected return, with two exceptions. When the customer is currently not at risk of churn but may soon become so following a few bad experiences, the firm should be risk-averse, i.e., use the safe mode. Also, when the customer is currently at risk of churn but may quickly not be one after a few good experiences, the firm should be risk-seeking, i.e., use the risky mode.

Our model focuses on the interaction between a firm and a customer. This modeling assumption does not imply that the firm has a single customer, but instead it assumes the firm can personalize its response to each customer, and that we can decouple the firm's payoffs across customers. At each point in time, the firm chooses a service mode: Risky or Safe. The Safe mode generates rewards at a constant rate, while the Risky mode produces Brownian rewards with a drift that can be either higher or lower than the Safe mode. We assume the rewards accrue to both the firm and the customer according to some fixed proportion, so that interests are aligned. The firm is patient and rational and wants to maximize the long-term rewards produced by its interaction with the customer. Meanwhile, the customer is assumed to be subject to recency bias. More precisely, the customer's satisfaction level follows a goodwill model that weighs recent experiences more heavily, which we model as an exponentially-weighted moving average of recent experiences. As a result, the customer's satisfaction evolves according to an Ornstein-Uhlenbeck process, with parameters determined by the service mode currently being used by the firm. The customer departs stochastically, and has a hazard rate of leaving that is a function of his satisfaction level.[1] The hazard rate is assumed to be a step function where the customer quits with positive hazard rate if the satisfaction level is below a given satisfaction threshold and with zero hazard rate if the satisfaction level is at or above the threshold.[2]

An optimal policy we find (Theorem 1) is of one of two kinds: it is either a *myopic policy*, which always chooses the superior mode (superior in terms of instantaneous reward rate) regardless of customer satisfaction, or a *sandwich policy* that chooses the inferior mode in an intermediate satisfaction interval and the superior mode elsewhere. The emergence of the sandwich policy as the optimal one in a large part of the parameter space is interesting. In particular, when the Safe service mode is myopically superior, the optimal policy is always of sandwich type, where the firm utilizes the Risky mode when the customer satisfaction level is slightly below the customer satisfaction threshold. This is in contrast with the customary wisdom that a high volatility and low return option is dominated by a low volatility and high return option. When the Risky mode is superior, the optimal policy is either myopic (Risky always) or of sandwich type, where the firm utilizes the Safe mode when the customer satisfaction is at or slightly above the customer satisfaction threshold. The deviation from the myopic policy is more substantial in cases where the alternative service mode has a similar instantaneous reward rate, and in cases where the difference in riskiness is larger. Also, numerical investigations reveal a very substantial customer lifetime value (CLV) improvement compared with a myopic policy that doesn't internalize the service mode's impact on the customer's probabilistic churn, in a large part of the parameter space. For example, we see CLV increases of 100% or more in many cases (see Figure 5 in Section 4.2).

This paper is novel in several ways. We are the first to study service mode control using the framework of stochastic control. This framework allows us to capture the customer's satisfaction state evolution under recency bias by a well-understood Ornstein-Uhlenbeck process, and thus obtain closed-form solutions that offer clear structural insights[3] regarding when should the firm

[1]Throughout the text, we use male pronouns to refer to the customer.

[2]We provide robustness checks regarding several model assumptions in the online appendix. In particular, in online appendices B–E, we consider mixed strategies, geometric Brownian reward process, switching cost, and alternative hazard rate functions, respectively. In all of these extensions, we show that our main insights are preserved.

[3]We expect the optimal policy to be messier and more challenging to interpret under alternate modeling approaches

use the risky or the safe service mode depending on the customer's satisfaction state. The optimal interval policy we obtain differs from previously studied policies in that we establish the optimality of sandwich policies that are non-monotone.

The stochastic control problem we study is also novel and challenging as a mathematical problem. To reduce algebraic hardness when solving the Hamilton-Jacobi-Bellman (HJB) equation, we purposely let the safe mode's volatility be zero, and choose the hazard rate function to be a step function with a discontinuous jump. The cost of doing this is losing the C^2 property of the value function, and invalidating the usual sufficient conditions for the controlled diffusion process to be well-defined. In particular, the drift and volatility terms are discontinuous in time and may switch infinitely often between two possible values. We build upon recent advances about Markov processes with spatial delay (Salins and Spiliopoulos [29]) to be able to modify and apply classic stochastic control methodologies to our setting. Moreover, to solve the problem algebraically (while verifying optimality of the derived policy by showing nonpositivity in Condition 4 of Proposition 2; see Section 5.1), we crucially draw upon L'Hospital-type rules for monotonicity (Pinelis [24]) and Chernoff-type bounds for the error function (Chang [8]) to establish properties of functions involving the error function (see Lemma 5 in Section 5.2). In general, our analysis showcases a way to balance algebraic hardness with stochastic control tractability. This method can potentially be applied to other stochastic control applications where the verification step is equally challenging.

The paper is organized as follows. Section 2 discusses related literature. Section 3 sets up the model, with main results presented and discussed in Section 4. The key steps to proving the main results, including the optimality conditions and the verification steps, are in Section 5. Proofs of lemmas and propositions are in Section 6.

2 Related Literature

The most closely related paper to ours is Aflaki and Popescu [2], a paper that studies the relationship between a service firm and a customer who might abandon its service. The service firm aims to maximize the value it obtains from the relationship, and its control is the service effort level at each point in time. Increasing the service level is costly to the firm but increases the customer's likelihood of staying in the system. As in our method, they use a goodwill model to model satisfaction (they also study a habituation model we do not consider). They find that an optimal effort policy leads to stationary effort and satisfaction levels. Two key differences between our model and theirs is that our model is stochastic and that our control is the firm's service mode rather than its effort level.

Goodwill models were first introduced in a seminal paper by Nerlove and Arrow [20], a paper that focused on the dynamic impact of advertising on demand. Nerlove and Arrow's goodwill model captured the fact that the effect of advertising on demand decays exponentially fast. We make the same assumption regarding how consumer satisfaction responds to past quality of service. Goodwill models have been validated empirically (Zeithaml [39]) and have the support of celebrated behavioral economics experiments that show that individuals overweigh recent experiences (Kahneman et al. [15]). Goodwill models (and their variants) have been used in operations and marketing to model how customers respond to past fill rates (Gaur and Park [11], Adelman and Mersereau [1], Liu and van Ryzin [16]), how customers recall service experiences (Ho et. al. [13], Das Gupta et al. [10]), and how customers learn about prices (Ovchinnikov and Milner [23]).

In terms of methodology, we build on techniques from the stochastic control literature. We assume that consumer satisfaction is a continuous-time stochastic process that is affected by the firm's choice of service mode at each point in time. Using a continuous-time model increases the technical complexity of our paper, but it improves the paper in that it allows us to obtain crisper theorems. To solve our stochastic control model, we use a "smooth pasting" technique that matches

where time, customer satisfaction and/or rewards are discretized.

the value function and its derivatives at satisfaction values where the optimal service mode changes. To prove the optimality of the policy we obtain, we use the verification technique that is typical of the stochastic control literature (Borkar [7], Mirică [18], Radner and Shepp [25], Touzi [36], and Ata et al. [4]). We also build on ideas and results from recent papers on stochastic calculus, such Strulovici and Szydlowski [32]'s sufficient conditions for a value function to be twice continuously differentiable (which our models do not quite satisfy because the Safe service mode has zero volatility), and Salins and Spiliopoulos [29]'s characterization of Markov processes with spatial delays.

A concurrent work by Johari and Schmit [14] studies the problem of learning about a customer who may abandon. A customer leaves the first time the firm does not meet her expectations, and the firm tries to balance the risk of customer departure with rewards earned. The focus of that paper, including the model and results, are quite different from the present paper.

Our paper assumes the firm has full information about the customer's satisfaction state and his churn propensity. In the marketing literature, latent customer relationship dynamics are approximated using hidden Markov models (Netzer et al. [21] and Ascarza et al. [3]). The firm can then optimize decisions by solving a partially-observable Markov decision process (Montoya et al. [19]).

It is worthwhile to contrast our results with those of Radner and Shepp [25], who use somewhat similar mathematics to study optimal dividend control given bankruptcy risk. They find that a monotone interval policy is optimal and that firms should use increasingly risky policies (higher volatility-to-drift ratio with increasing drift) if they have more cash on hand. In contrast, we find that the optimal policy in many cases is a sandwich policy where risk-taking is a non-monotonic function of the state. A key driver of this difference is that unsatisfied customers in our model quit according to a hazard rate while, in Radner and Shepp, a firm without cash goes bankrupt immediately. Based on our findings, we believe that if [25] allowed bankruptcy to happen according to some hazard rate when cash drops to zero and below, their optimal policy would adopt a similar risky action below the bankruptcy threshold. Wang and Zenios [37] apply stochastic control methodology to the problem of new venture creation, and show supremacy of monotone interval policies in their setting. The entrepreneur chooses between two controls with different cost, drift and volatility profiles, and either succeeds or fails when the firm value hits an upper or lower bound (exogenous or endogenous). Since there are only two control actions, their monotone policy has at most two intervals (one for each control).

Atar and Lev-Ari [5] study a scheduling control problem for multiclass single server queue with abandonment. They identify an asymptotically-optimal workload-dependent dynamic interval policy which determines the priority ranking between customer classes based on the current rescaled workload. The HJB equation in the heavy traffic limit is similar to ours, but with two important distinctions. In [5], the coefficient of the value function's second derivative, i.e., the volatility term, is a constant. In our paper, the volatility depends on the control being employed. Also, the coefficient of the value function in [5] is a constant, whereas in our paper this term is a discontinuous function of the state. As a result, unlike in [5], classical solutions (C^2) to our HJB equation do not exist. We employ recent advances in stochastic processes [29] and L'Hospital-type rules for monotonicity [24] for verification of our HJB equation. The optimal policy we find has a very different structure from the interval policy found to be optimal in [5].

There are several other recent papers in operations research that study drift control problems in various contexts. For example, Matoglu and Vate [22] and Matoglu et al. [17] study a Brownian drift control problem for managing capacity in a build-to-order environment. The firm's objective is to minimize the long-run average cost of operation and switching. Harrison and Sunar [12] study a firm's optimal Bayesian learning policy with respect to a project's unknown value. Formulated as a Brownian control problem, the firm decides on the learning mode which affects the Brownian motion's drift, as well as a stopping rule that leads to either investment or abandonment of the project. Sunar et al. [33] study a joint optimal dynamic development, launch, and post-launch strategies when the product diffuses over a network of customers. They model the technology of the product via a stochastic process whose drift is determined by the firm's effort level, and are able to explicitly characterize the optimal policies. Finally, Sunar et al. [34] study a continuous-time Bayesian learning problem in a game-theoretic setup where two firms make investment decisions about an unknown project. They show that the leader firm's equilibrium expected discounted profit and investment can strictly decrease when the chance of favorable market increases. In contrast to this stream of literature, where the controlled diffusion process is either restricted to some interval or stops when it hits a given boundary, our paper considers a region of probabilistic stopping, which adds a second layer of stochasticity to the problem.

3 The Model

We study a continuous-time model of one firm repeatedly interacting with one customer. The firm has two alternative service modes, a Risky mode R and a Safe mode S. At any given time $t \geq 0$, the firm chooses service mode $u_t \in \{R, S\}$ which determines the drift μ_{u_t} and volatility σ_{u_t} of the reward process. The firm is able to switch between the two service modes over time. The reward Y_t accrues according to the following stochastic differential equation:

$$dY_t = \mu_{u_t} dt + \sigma_{u_t} dB_t, \tag{1}$$

where B_t is a standard Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ with $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ and $Y_0 = 0$. We are interested in cases where the expected rewards are positive. Therefore, we assume $\mu_R > 0$ and $\mu_S > 0$. We call the service mode with higher drift the (myopically) *superior* mode. Namely, the Risky mode is the superior mode if $\mu_R > \mu_S$ and the Safe mode is the superior mode if $\mu_R < \mu_S$.[4] Likewise, we call the service mode with lower drift the (myopically) *inferior* mode. We assume the Safe mode has no volatility, i.e., $\sigma_S = 0$.

We think of the rewards as accruing to both the customer and the firm, in the sense that both players seek higher rewards. The firm is rational and aims to maximize the total reward generated from its interaction with the customer over time. We give a precise description of the firm's objective in Eq. (8) below. The customer, meanwhile, is assumed to adaptively learn from past experience while being subject to recency bias, i.e., he weighs recent experiences more heavily. We describe the customer at each point in time via H_t, a one-dimensional state that we term *satisfaction*. To model recency bias, we model customer satisfaction as an exponentially weighted moving average of recent rewards. That is, the customer satisfaction at time t is

$$H_t = x \cdot e^{-t} + \int_0^t e^{-(t-s)} dY_s, \tag{2}$$

where $H_0 = x$ is the initial satisfaction level. Based on standard stochastic calculus arguments, the expression in Eq. (2) can also be represented as the following stochastic differential equation:

$$dH_t = dY_t - H_t dt, \tag{3}$$

with $H_0 = x$. Thus, customer satisfaction would follow an Ornstein–Uhlenbeck (O-U) process if the firm were to select the same service mode over the entire time horizon (for the Safe mode, the O-U process is degenerate since $\sigma_S = 0$). The value H_t can be interpreted as the customer's unsophisticated prediction at time t of the reward rate he expects to get in the future, and Eq. (3) captures that the customer iteratively modifies this satisfaction by the difference between the realized reward dY_t and the predicted reward $H_t dt$. There is no private information in our model, so the firm is able to observe the history of customer satisfaction and rewards. Note that $\mathcal{F}_t$ contains all information of the reward process Y_t and the satisfaction process H_t up to time t, since $(u_t)_{t\geq 0}$ and $(B_t)_{t\geq 0}$ are both adapted to $\mathbb{F}$.

The firm cares about customer satisfaction because the customer is likely to quit the system if

[4] If $\mu_R = \mu_S$, both service modes are classified as superior. Section 4 provides a discussion for this boundary case.

unsatisfied. We model churn via a hazard rate function $Q(\cdot)$ which acts on the current satisfaction:

$$Q(H_t) \triangleq \mathbb{1}\{H_t < q\}, \tag{4}$$

where q is the customer's *satisfaction threshold* such that the customer does not quit the system when his satisfaction is above or equal to it. The customer is called *satisfied* when his satisfaction is above q, *borderline satisfied* at q, and *unsatisfied* below q. The literature on managing customer churn has considered different hazard rate functions, such as logit functions (Rust et. al. [28]) and exponential functions (Berger and Nasr [6]). The step function in Eq. (4) can be thought of as the limit of logit functions and is chosen for the sake of tractability.

We denote the customer lifetime by T. The hazard rate assumption implies that the customer survival probability S_t at time t is equal to

$$S_t \triangleq P\left(T > t \mid \mathcal{F}_t\right) = e^{-\int_0^t Q(H_s)ds}. \tag{5}$$

More formally, we let z be a uniform random variable over $[0,1]$ independent of $\mathcal{F}$ and let the lifetime T be defined as:[5]

$$T \triangleq \inf\left\{t \geq 0 : e^{-\int_0^t Q(H_s)ds} = z\right\}. \tag{6}$$

Here, $\int_0^t Q(H_s)ds$ is a Lebesgue integral that measures the amount of time up to t that the customer satisfaction process has spent in the unsatisfied zone.

The firm's objective is to maximize the customer lifetime value (CLV) it earns from interacting with the customer. For a given starting satisfaction value x and the firm's choice of action $u_t \in \{R, S\}$ at any time t (according to some admissible policy π which we will define next), the CLV is

$$V(x,\pi) = \mathbb{E}\left[Y_T \mid H_0 = x\right] = \mathbb{E}\left[\int_0^\infty \mu_{u_t}\mathbb{1}\{t < T\}dt + \int_0^\infty \sigma_{u_t}\mathbb{1}\{t < T\}dB_t \,\middle|\, H_0 = x\right]. \tag{7}$$

For the CLV and the firm's problem to be properly defined, we restrict ourselves to the set of admissible policies defined below.

Definition 1 (Admissible policy)**.** *A policy π, which defines a mapping from $\mathcal{F}_t$ to the action space $\{R, S\}$ for all $t \leq T$, is admissible if the following conditions are satisfied:*

1. *The firm's action process $(u_t)_{t\geq 0}$ (by following this policy) is adapted to the filtration $\mathbb{F}$ and takes value in $\{R, S\}$.*
2. *The corresponding reward process $(Y_t)_{t\geq 0}$, satisfaction process $(H_t)_{t\geq 0}$ and survival probability process $(S_t)_{t\geq 0}$ defined respectively by the solutions to Eqs. (1) and (3) and by Eq. (5), exist and are $\mathbb{F}$-adapted semimartingales specified uniquely in law.*
3. *The integrals and expectation that define the value function and expectation in Eq. (7) exist.*

We denote the space of admissible policies by Π. The firm wants to find an admissible policy $\pi \in \Pi$ that maximizes the CLV. The optimal CLV given a starting satisfaction value x is given by

$$V^*(x) = \sup_{\pi\in\Pi} V(x,\pi). \tag{8}$$

This completes our definition of the firm's problem. Note that in our definition of admissible policies, the policy does not know whether or not the customer has already left. This clearly has no impact on the CLV since rewards stop accruing after T.

We say a policy is *stationary Markov* if it is a time-invariant mapping from the current satisfaction state to the service mode $\pi : \mathbb{R} \to \{R, S\}$. Given that the platform is facing a stationary Markov control problem, one would expect there to be a stationary Markov optimal control policy, and indeed in Section 3 we will formally establish that this is the case. Note that under a stationary

[5] Even more formally, from here on we consider the product probability space $(\Omega, \mathcal{F}, \mathbb{F}, P) \otimes ([0,1], \mathcal{B}, U)$, where $\mathcal{B}$ is the σ-algebra of Borel sets of $[0,1]$, U denotes the uniform probability measure on $[0,1]$, and $z \in [0,1]$ is the "state of the world" in the second component probability space.

Markov policy π, we can write $u_t = \pi(H_t)$, and the CLV in Eq. (7) can be restated as:[6]

$$V(x,\pi) = \mathbb{E}\left[\int_0^\infty \mu_{\pi(H_t)} e^{-\int_0^t Q(H_s)ds} dt \Big| H_0 = x\right]. \tag{9}$$

Observe from Eq. (9) that the expectation is with respect to the probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ only.

Interval policies: a subset of admissible policies. Our setting presents some atypical technical features as a result of our assumption that the volatility under the Safe mode σ_S is zero. However, we are able to formally specify the stochastic processes for reward and satisfaction resulting from a broad set of control policies by drawing on the work of Salins and Spiliopoulos [29] on Markov processes with spatial delay. In Lemma 1 and Corollary 1 below we establish admissibility for this set of control policies, which we call *interval* policies as defined below.

Definition 2 (Interval policy)**.** *A policy π is an* interval *policy if it is stationary Markov (that is, the corresponding action process is a function of the current satisfaction $u_t = \pi(H_t)$) such that it divides the satisfaction real line into a countable number of alternating Risky and Safe intervals (the policy uses the Risky (Safe) action for satisfaction levels which lie in the Risky (Safe) intervals), and such that the Safe mode is adopted at each boundary point between intervals.*

Recall the definition of admissible policies from Definition 1. The firm's action process under an interval policy as given by Definition 2 is clearly adapted to the filtration $\mathbb{F}$ and takes values in $\{R, S\}$. If we can show that the corresponding reward and satisfaction processes exist and are $\mathbb{F}$-adapted semimartingales specified uniquely in law, and that the integrals and expectation in Eq. (7) that define the value function exist, we can conclude that an interval policy is admissible.

The next lemma formally establishes that the first condition is satisfied by interval policies, and further deduces that we can apply the Itô-Tanaka formula to the satisfaction process $(H_t)_{t\geq 0}$ under any interval policy. The proof can be found in Section 6.1.

Lemma 1. *Let π be an interval policy and $(u_t)_{t\geq 0}$ the corresponding action process. Then under π and for any starting satisfaction level x, the following results hold:*

1. *The processes $(Y_t)_{t\geq 0}$, $(H_t)_{t\geq 0}$ and $(S_t)_{t\geq 0}$ defined by the solutions to Eqs. (1), (3) and (5) exist and are $\mathbb{F}$-adapted semimartingales specified uniquely in law.*
2. *Assume f is a function that is continuously differentiable on $\mathbb{R}$ and twice continuously differentiable on $\mathbb{R} \setminus \mathcal{E}$ for some countable set $\mathcal{E}$. Then $f(H_t)$ is also an $\mathbb{F}$-adapted semimartingale, with:*[7]

$$f(H_t) = f(x) + \int_0^t (\mu_{u_s} - H_s) f'(H_s) ds + \int_0^t \sigma_{u_s} f'(H_s) dB_s + \frac{1}{2}\int_0^t \sigma_{u_s}^2 f''(H_s) \mathbb{1}\{H_s \notin \mathcal{E}\} ds. \tag{10}$$

In fact, the Itô-Tanaka formula in Eq. (10) holds for the satisfaction process under any admissible policy, since by definition of admissibility H_t is a semimartingale. Also note that the only remaining requirement for admissibility — the integrals and expectation in Eq. (7) that define the value function exist — is naturally satisfied by any interval policy. This is because $\mu_S > 0$ and $\mu_R > 0$, so by the Monotone Convergence Theorem, the CLV is equivalent to the summation of μ_S (μ_R) multiplied by the total Lebesgue measure of time before churn that the satisfaction process spends in the Safe (Risky) intervals. We summarize this in the next corollary.

[6] The second integral in Eq. (7) has zero expectation since the integrand is bounded. By the Dominated Convergence Theorem and the tower property, we can rewrite $V(x,\pi) = \lim_{t\to\infty} \mathbb{E}\left[\int_0^t \mu_{\pi(H_s)} \mathbb{1}\{s < T\} ds \Big| H_0 = x\right] = \lim_{t\to\infty} \mathbb{E}\left[\mathbb{E}\left[\int_0^t \mu_{\pi(H_s)} \mathbb{1}\{s<T\} ds \Big| \mathcal{F}_t\right] \Big| H_0 = x\right] = \lim_{t\to\infty} \mathbb{E}\left[\int_0^t \mu_{\pi(H_s)} P(s < T|\mathcal{F}_s) ds \Big| H_0 = x\right] = \lim_{t\to\infty} \mathbb{E}\left[\int_0^t \mu_{\pi(H_s)} e^{-\int_0^s Q(H_v)dv} ds \Big| H_0 = x\right]$. The last step follows from Eq. (5).

[7] We define $f''(H_s)\mathbb{1}\{H_s \notin \mathcal{E}\}$ to be zero if $H_t \in \mathcal{E}$.

Corollary 1. *Any interval policy as defined by Definition 2 is admissible.*

Next we impose an assumption on the model primitives throughout this paper to ensure that the CLV is finite.[8]

Assumption 1. *The satisfaction threshold q is greater than the Safe mode's drift μ_S, i.e., $q > \mu_S$.*

To understand the importance of this assumption, note that if $q \leq \mu_S$, then for starting satisfaction level $x = \mu_S$, a simple policy which always uses the Safe mode would make $H_t \equiv \mu_S$ and $Y_t = \mu_S t$. Therefore, the satisfaction process is always in the satisfied zone $[q, \infty)$, and hence the CLV is infinity. In other words, Assumption 1 ensures that the firm cannot satisfy the customer forever under the Safe service mode.

The next proposition formally establishes that under Assumption 1 and under interval policies, the expected customer lifetime and the expected customer lifetime value (CLV) are both finite.[9]

Proposition 1. *Under Assumption 1, for any $x \in \mathbb{R}$ and any interval policy π with a finite number of intervals, both the customer lifetime and the CLV are finite, i.e., $\mathbb{E}[T^{x,\pi}] < \infty$ and $V(x, \pi) < \infty$.*

To prove Proposition 1, we only need to show that the satisfaction process H_t eventually will spend enough time in the unsatisfied zone $(-\infty, q)$, where the customer churns at hazard rate 1. Intuitively speaking, this is true since the satisfaction process H_t drifts toward $\mu_S < q$ deterministically under the Safe mode, and moves stochastically over the entire real line under the Risky mode. The proof is formalized in Section 6.1.

Discussion of model assumptions. We briefly discuss specific assumptions in our model. Our model assumes that customer satisfaction is an exponentially-weighted moving average of recent rewards with the window size parameter $w = 1$; that is, rewards from Δt time ago are given a weight $(1/w)\exp(-\Delta t/w)$ for $w = 1$ in determining satisfaction. If instead a customer uses an exponentially-weighted moving average with arbitrary window size $w \in (0, \infty)$, then we can rescale time to ensure that Eq. (3) still holds, so our assumption $w = 1$ is without loss of generality. However, having fixed the scaling of time, we point out that Eq. (4) represents a specific assumption that the height of the step in the hazard rate function is 1. This assumption enables us to obtain analytical results for our stochastic control problem, but numerical results demonstrate that our structural insights hold for other step height values as well[10]. Also note that we apply a separate treatment for each customer in this paper. This allows us to personalize service mode control based on the customer's latent characteristics such as his satisfaction threshold, his recency bias window, and his hazard rate of churn as a function of his satisfaction state. This decoupling of the problem implicitly assumes that no capacity constraints are present, and that the firm's overall utility corresponds to the sum of expected rewards it earns from different customers. While we do not explicitly model a firm's preference across aggregate customer outcomes, such as a firm-level risk measure, our policy is personalized to each customer's satisfaction threshold and this organically generates some degree of cross-customer diversification.

4 Analysis of the Model

In this section, we solve the firm's optimization problem posed in Eq. (8) and describe the structure of an optimal policy. We find that there is an optimal policy that achieves the optimal CLV that is

[8] Proposition 1 formalizes that the CLV is finite under a big class of admissible policies. Also, in Theorem 1 (in Section 4) we establish that there exists an optimal policy in that big admissible policy class. This also means that no other admissible policy can produce a larger CLV. Therefore the CLV must be finite for any admissible policy.

[9] In fact, if *any* stationary Markov policy (not necessarily an interval policy) is admissible, the associated CLV is finite. This follows from our main result (Theorem 1 below) which says that an interval policy is optimal, since we do not restrict admissible policies to be interval policies.

[10] See Section 4.3 and Online Appendix E.

a simple interval policy. In particular, the optimal policy will consist of at most three intervals.

In Section 4.1, we present the structure of an optimal policy. In Section 4.2, we state some formal comparative statics results and use numerics to understand both the structure of the optimal policy (including comparative statics) as well as the CLV increase from the optimal policy relative to the myopic policy. We find that the benefit can be very large. In Section 4.3 we present a summary of robustness checks of our model in four variations.

4.1 Structure of an Optimal Policy

In our main theorem of this section, we prove that there exists an optimal policy where the firm should always use the myopically superior service mode except (possibly) for some intermediate customer satisfaction levels. In fact, we show that regardless of which service mode is the superior one, the firm should choose the Safe service mode if the customer satisfaction level is in some region (possibly empty) $[q, \theta_b]$ just above the satisfaction threshold q, and choose the Risky service mode if the customer satisfaction level is in some nonempty region (θ_G, q) just below q (see Figure 1 for a stylized representation), where θ_b and θ_G are functions of the model primitives μ_R, μ_S, σ_R and q. The values θ_b and θ_G are defined in Lemma 2 and Lemma 3 below (proofs in Section 6). We name the satisfaction region $[q, \theta_b]$ where the firm should always use the Safe mode the *risk-averse* region. Likewise, we name (θ_G, q) where the firm should always use the Risky mode the *risk-seeking* region. We follow the conventional definition of the error function $\mathrm{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

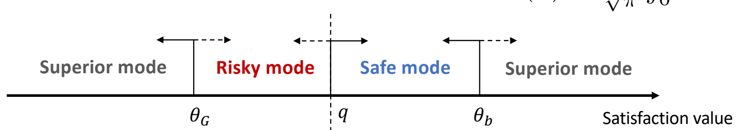


Figure 1: Stylized representation of the firm's optimal policy as per Theorem 1.

Lemma 2. *If $\mu_R \leq \mu_S$, let $\theta_b = \infty$. If $\mu_R > \mu_S$, let Θ be the set of values of θ that satisfy*

$$\frac{\mu_R \sqrt{\pi}(\mu_S - \theta)}{\sigma_R} \exp\left(\frac{(\theta - \mu_R)^2}{\sigma_R^2}\right)\left(1 - \mathrm{erf}\left(\frac{\theta - \mu_R}{\sigma_R}\right)\right) + \mu_S = 0.$$

Then, the set $\Theta \cap (\mu_S, \infty)$ contains a single element, which we label θ_b.

From Lemma 2, note that the risk-averse region $[q, \theta_b]$ is bounded only if the Risky mode is superior $\mu_R > \mu_S$. Moreover, in a subset of cases where it is bounded, the risk-averse region is in fact empty; this occurs if $\theta_b < q$.

Lemma 3. *If $\mu_R \geq \mu_S$, let $\theta_G = -\infty$. If $\mu_R < \mu_S$, let Θ_G^{small} be the set of values of θ that satisfy*

$$\exp\left(\frac{(q - \mu_R)^2}{\sigma_R^2}\right) \frac{\sqrt{\pi}(\theta - \mu_R)(q - \mu_R)}{\sigma_R^2} \left(\mathrm{erf}\left(\frac{q - \mu_R}{\sigma_R}\right) - \mathrm{erf}\left(\frac{\theta - \mu_R}{\sigma_R}\right)\right)$$
$$+ \exp\left(\frac{(q - \mu_R)^2 - (\theta - \mu_R)^2}{\sigma_R^2}\right) \frac{q - \mu_R}{\sigma_R} - \frac{\theta - \mu_R}{\sigma_R} - \frac{\mu_S \sigma_R}{2(\mu_S - \mu_R)(q - \mu_S)} = 0, \tag{11}$$

and let Θ_G^{big} be the set of values of θ that satisfy

$$2 \exp\left(\frac{(q - \mu_R)^2 - (\theta - \mu_R)^2}{\sigma_R^2}\right)\left(1 + \frac{(\theta - \mu_S)(\theta - \mu_R)}{\sigma_R^2}\right)$$
$$+ \exp\left(\frac{(q - \mu_R)^2}{\sigma_R^2}\right)\sqrt{\pi}\left(\frac{\mu_S + 2\mu_R - 3\theta}{\sigma_R} + \frac{2(\theta - \mu_R)^2(\mu_S - \theta)}{\sigma_R^3}\right)\left(\mathrm{erf}\left(\frac{q - \mu_R}{\sigma_R}\right) - \mathrm{erf}\left(\frac{\theta - \mu_R}{\sigma_R}\right)\right)$$
$$+ \left(\frac{\mu_S + 2\mu_R - 3\theta}{\sigma_R} + \frac{2(\theta - \mu_R)^2(\mu_S - \theta)}{\sigma_R^3}\right)\frac{\sigma_R}{q - \mu_R} - \frac{\mu_S\left(\sigma_R^2 + (\mu_S - \mu_R)(\theta - \mu_S)\right)}{(\mu_S - \mu_R)(q - \mu_R)(q - \mu_S)} = 0. \tag{12}$$

Then, the set $\left(\Theta_G^{\text{big}} \cap (\mu_S - \frac{\sigma_R^2}{\mu_S - \mu_R}, \mu_S)\right) \cup \left(\Theta_G^{\text{small}} \cap [\mu_S, q)\right)$ *contains exactly one element, which we label* θ_G.

Observe from Lemma 3 that the risk-seeking region (θ_G, q) is always nonempty. It is bounded only when the Safe mode is the superior mode $\mu_S > \mu_R$. In fact, for any drift values such that $\mu_S < q$ and $\mu_R \neq \mu_S$, exactly one of the two regions $[q, \theta_b]$ and (θ_G, q) is bounded, while the other is unbounded. The bounded region corresponds to the intermediate satisfaction level where the firm should use the inferior mode. Everywhere else (including the other unbounded region) the firm should use the myopically superior service mode. In the special case of $\mu_R = \mu_S$, both the risk-averse region and the risk-seeking region are unbounded, and they partition the entire satisfaction real line. We postpone a description of how we arrive at the above θ_b and θ_G until just before Section 4.2.

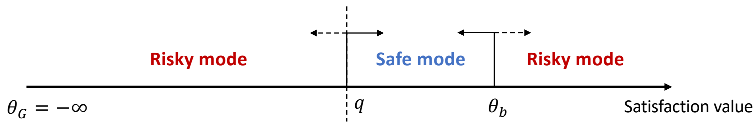


Figure 2: Optimal policy when $\mu_R > \mu_S$.

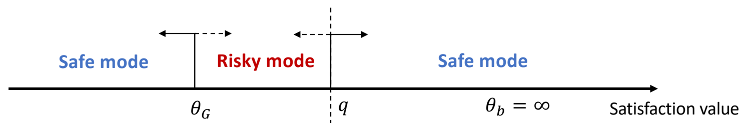


Figure 3: Optimal policy when $\mu_R < \mu_S$.

We are now ready to formally state our first theorem.

Theorem 1. *Fix μ_R, μ_S, σ_R and satisfaction threshold q (recall Eq. (4)) such that $\mu_S < q$. Consider the firm's problem as presented in Eq. (8). Let θ_b be as defined in Lemma 2 and let θ_G be as defined in Lemma 3. Then there is an optimal policy where the firm chooses the Safe mode on $[q, \theta_b]$, the Risky mode on (θ_G, q), and the superior of the two modes elsewhere.*[11] *Also, the customer's expected lifetime and the CLV are finite under the optimal policy (hence the CLV is finite under any admissible policy).*

In the optimal policy described in Theorem 1, the firm chooses the Safe mode on $[q, \theta_b]$, the Risky mode on (θ_G, q), and the superior mode elsewhere (see Figure 1). Now let us take a closer look at this policy under different cases of μ_R and μ_S. When the Risky service mode is the superior one ($\mu_R > \mu_S$), the optimal policy uses the Risky mode for all satisfaction values below q. Moreover, depending on the value of θ_b (see Lemma 2), the risk-averse region $[q, \theta_b]$ might degenerate to an empty set. This happens if $\theta_b < q$, in which case the optimal policy becomes a myopic one where the firm always uses the Risky service mode (see Figure 2). On the other hand when $\theta_b \geq q$, the risk-averse region $[q, \theta_b]$ is nonempty and bounded, and the optimal policy is a *sandwich* policy, i.e., an interval policy with three exact intervals, where the firm uses the Safe mode for satisfaction values in $[q, \theta_b]$ just above the satisfaction threshold (see Figure 2), and the Risky mode elsewhere. In the other case where the Safe mode is the superior one ($\mu_R < \mu_S$), the optimal policy uses the Safe mode for all satisfaction values (weakly) above q. Meanwhile, the risk-seeking region (θ_G, q) is always nonempty and bounded (see Lemma 3). Therefore in this case the firm's optimal policy is always a sandwich policy (see Figure 3), where the Safe mode is used everywhere except for satisfaction levels in (θ_G, q) just below the satisfaction threshold. When customer satisfaction value is in this region, the firm should switch to the Risky service mode.

[11] Note that if $\mu_R = \mu_S$ then $\theta_G = -\infty$ and $\theta_b = +\infty$, so the specified policy uses Risky on $(-\infty, q)$ and Safe on $[q, \infty)$. In $\mu_R \neq \mu_S$ then there is a unique superior mode. Hence the policy is uniquely specified in all cases.

In the special case of $\mu_R = \mu_S$, we have $\theta_b = \infty$ and $\theta_G = -\infty$. In this case both the Risky mode and the Safe mode are superior, and the optimal policy becomes a particular myopic one, where the firm uses the Risky mode for satisfaction values on $(-\infty, q)$ and the Safe mode on $[q, \infty]$.

Establishing the optimal control is essentially equivalent to determining the value function V^* (see Eq. (8)). A classical technique for determining a continuous-time, continuous-space value function V^* such as ours is called the *verification* technique, which involves first obtaining a candidate value function and subsequently proving its optimality. There are many papers in the literature that use the verification technique for solving stochastic control problems, including Borkar (1989), Mirică (1992), Radner and Shepp (1996), and Touzi (2002). We cannot immediately use a method from the literature for two reasons. First, our problem involves stochastic abandonment rather than deterministic discounting. Second, as will become clear, the value function V^* in our problem does not satisfy the standard smoothness condition, which is that V^* should be twice continuously differentiable everywhere. We modify the standard approach from the literature in order to create a methodology that works for our problem. More details are provided in Section 5.

Trade off between immediate payoffs and customer lifetime. Observe that except for the possible optimal myopic policy when $\mu_R > \mu_S$, in all other cases, the firm's optimal policy exhibits a risk-averse region just above the satisfaction threshold. In this region, the firm should use the Safe mode even if it generates lower immediate rewards. Consider the optimal sandwich policy for $\mu_R > \mu_S$. When $x \in (q, \theta_b)$, the firm uses the inferior service mode. If the optimal policy is sacrificing rewards in the short run, it must be that it confers some long-term benefits. The intuition is that when the satisfaction level is close to the unsatisfied zone, the Safe mode prolongs customer lifetime by delaying entry into the unsatisfied zone, compared with the Risky mode. When the volatility σ_R is high, the mean first passage time into the unsatisfied zone from above will be longer under the Safe mode than under the Risky mode. Therefore, using the Safe mode in the risk-averse region just above q serves to delay the inevitable — entry into the unsatisfied zone — when the volatility of the Risky mode is high.[12] A preference for low volatility in relatively low states is familiar from other settings, for example Radner and Shepp [25].

Perhaps more unexpected is the existence of a risk-seeking region. In particular, surprisingly, no matter by how much the Risky mode is inferior to the Safe mode, there is some satisfaction region below the satisfaction threshold q in which the firm should use the Risky mode. This challenges the customary idea that a high volatility and low return combination should be dominated by a low volatility and high return one. In fact, it turns out that the Risky mode outperforms the Safe mode in that region via its ability to push the customer out of the unsatisfied zone (with positive probability) when his satisfaction is below but close to q, thereby extending his lifetime. We highlight that the risk-seeking region is always non-empty, regardless of model primitives.

An immediate consequence of the sandwich policy featuring both a risk-averse region just above the satisfaction threshold and a risk-seeking region just below it is that, surprisingly, once it enters the unsatisfied zone, the customer satisfaction will never again reemerge above the satisfaction threshold q (see Figure 4). Recall that by assumption, $\mu_S < q$, so that switching to the Safe mode when the satisfaction process hits q from below leads to negative drift $dH_t/dt = \mu_S - q$, and so H_t immediately drops back into the unsatisfied zone. This might seem counterintuitive. After all, we want the customer to stay satisfied with the firm to prolong his lifetime. However, the optimal (sandwich) policy seems to advise us to keep his satisfaction low and prevent him from being satisfied. How can such a policy maximize customer lifetime value? Remember that the myopic policy always produces the maximum payoff rate. Therefore we can deduce that the reason for a non-myopic policy to be optimal must be that the customer lives longer under this policy. In other words, the customer spends more time in the satisfied zone under this policy. Counterintuitively, the

[12]In Theorem 2 later, we show that θ_b is increasing in σ_R, holding other primitives fixed.

optimal sandwich policy indeed increases the customer's time spent in the satisfied zone even *after* he enters the unsatisfied zone. This occurs because it causes the customer satisfaction to spend a *positive measure of time exactly at the borderline satisfied level of q*.

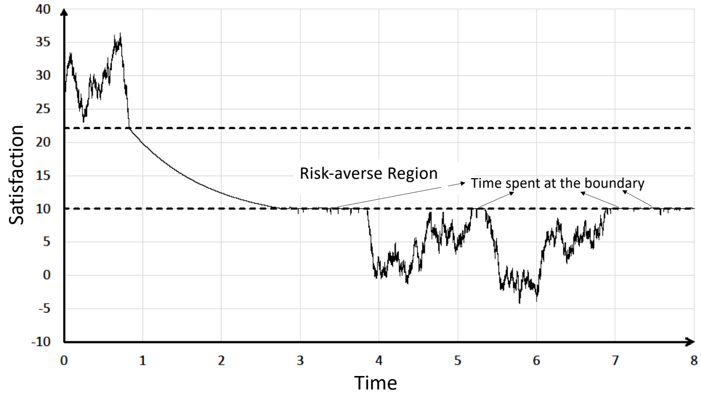


Figure 4: Sample path of satisfaction process under the optimal sandwich policy: $\mu_S = 8$, $\mu_R = 9$, $\sigma_R = 10$, $q = 10$. The risk-averse region where the optimal policy uses the Safe mode is $[10, 22.1]$.

Comparison to a reflected Ornstein-Uhlenbeck process. Again consider the case $\mu_R > \mu_S$. As in Section 3, let us consider the satisfaction process as an infinite horizon stochastic process. That is, let us ignore the customer abandonment at time $T^{x,\pi}$ and continue to track the evolution of $H_t^{x,\pi}$ under policy π. Assume that the optimal policy is a sandwich policy and consider a starting satisfaction level $x \leq q$, so that the satisfaction level never exceeds q for any time t. At first sight, the satisfaction process appears to be the well-understood reflected O-U process reflected from above at q (Reed et al. [26]). However, this is not the case. The reflected O-U process and the reflected Brownian motion both spend a measure zero of time at the reflecting boundary. In our process, the time spent at the reflecting boundary has a positive measure with probability one. While this process is less well known, such a *delayed reflected* process was first introduced by Skorokhod [31] and was recently shown to be a semimartingale by Salins and Spiliopoulos [29]. The measure of time spent by a delayed reflected process at a reflection boundary is proportional to the local time of the process at the boundary, and the constant of proportionality is the inverse of the drift at the boundary. In our setting, the drift is $-(q - \mu_S)$ under the Safe mode at the threshold q. The standard reflected O-U process corresponds to a negative infinity drift at the reflecting boundary.

The switching thresholds of the sandwich policy. We now turn our attention to the values of θ_b and θ_G, as defined in Lemmas 2 and 3. We provide here a very informal argument to explain how these values arise. We present the key ideas used to build a formal proof in Section 5, with the details of the argument deferred to Section 6.

Suppose $\mu_R > \mu_S$. The value of θ_b can be established informally by comparing the value functions of two stationary Markov policies which make different decisions only on a tiny interval of satisfaction values. Intuitively, the optimal policy should choose the Risky mode for a sufficiently high levels of satisfaction, given that the Risky mode is myopically superior. We want to find the satisfaction level $\theta_b > q$ below which the firm should switch to the Safe mode. Let x be some starting satisfaction level and let Δx be some value such that $x - \Delta x \geq q$. Consider two stationary policies, π_S and π_R, that both choose the Risky mode above x and make identical decisions for values of satisfaction below $x - \Delta x$ (for our argument, it does not matter what decisions the policies make below $x - \Delta x$; we are only concerned with the expected reward accumulated in the period it takes for the satisfaction to drop from x to $x - \Delta x$). Within the interval $[x - \Delta x, x]$, the two policies choose different modes, with π_S choosing the Safe mode and π_R choosing the Risky mode.

We now compare across the policies $\pi \in \{\pi_S, \pi_R\}$, the incremental value $V(x, \pi) - V(x - \Delta x, \pi)$

the policy π generates from satisfaction being x rather than $x - \Delta x$, which is equal to the reward (per unit Δx) accumulated by the policy while the satisfaction falls from x to $x - \Delta x$. This is fairly easy to compute: it is simply the product of the expected reward rate and the expected first passage time of an O-U process from x to $x - \Delta x$. Note that the satisfaction reaches $x - \Delta x$ in finite time with probability 1, and the customer does not depart before this occurs. Define the *marginal value of satisfaction* as

$$m(x, \pi) \triangleq \lim_{\Delta x \to 0} \frac{V(x, \pi) - V(x - \Delta x, \pi)}{\Delta x}.$$

If there exists a level of satisfaction $x \geq q$ such that $m(x, \pi_S) \geq m(x, \pi_R)$, then at satisfaction x, the firm weakly prefers the Safe action, assuming it uses the Risky action above it. In fact, θ_b can be determined by finding x such that the marginal benefits of these two policies are equal; that is

$$m(\theta_b, \pi_S) = m(\theta_b, \pi_R).$$

Lemma 2 encodes this condition in terms of model primitives, which uniquely specifies θ_b.

Similarly, to derive the value of θ_G, the conditions in Eqs. (11) and (12) come from guessing a sandwich policy that uses the Risky mode in (θ_G, q) and the Safe mode elsewhere and equating the marginal value of satisfaction (i.e., the slope of the value function) just to the left and right of θ_G. As evidenced by the complexity of the definition of θ_G, matching marginal values in the unsatisfied zone is a harder task here than in the satisfied zone. The extra difficulty arises from having to account for the customer abandonment risk while computing first passage times.

4.2 Comparative Statics and Numerics

In the previous section, we constructed an optimal policy that is either the myopic policy (superior mode everywhere) or a sandwich policy (superior mode everywhere except for some intermediate satisfaction range). In this section, we explore how the optimal policy and the optimal CLV depend on model primitives. There are two findings we highlight in this section: (1) the improvement in CLV under the optimal policy compared to the CLV under the myopic policy can be very large; (2) an increase in volatility of the Risky mode substantially *increases* the CLV in some cases.

We first look at some numerical examples regarding CLVs under the optimal policy versus under the myopic policy. Theorem 1 tells us that the myopic policy is not always optimal. Figure 5 below shows how much improvement in CLV the optimal policy provides over the myopic policy for $\mu_R = 9$, $q = 10$, initial satisfaction q (as an example), and for different μ_S and σ_R. The figure shows that the magnitude of improvement can be very large, especially when the Safe mode is the superior mode. Also, as the Risky mode's volatility increases, the improvement in CLV gets larger. Note that this benifit would reduce if the initial satisfaction level is far away from q.

It is interesting to investigate how the optimal policy varies with problem primitives. Figure 6 below shows that both the risk-seeking region described by value θ_G, and the risk-averse region by value θ_b vary monotonically in σ_R, holding μ_S, μ_R and q fixed. In the plot, the horizontal coordinate is the Safe mode's drift μ_S (a problem primitive), and the vertical coordinate is the customer's current satisfaction value (the "state" of the customer). For each μ_S (a fixed location on the horizontal axis), Theorem 1 gives us an optimal policy. The shading in the figure is used to represent when the firm should use the Risky mode under the optimal policy. Therefore, the boundaries between the shaded and unshaded areas are the switching thresholds between the two service modes. The top right unshaded block with a left-pointing tail corresponds to the risk-averse region in the optimal policy. The bottom left shaded block with a right-pointing tail corresponds to the risk-seeking region in the optimal policy. The plot shows that as σ_R decreases, for $\mu_S < \mu_R$ the risk-averse region is getting smaller, and for $\mu_S > \mu_R$ the risk-seeking region is getting smaller. This is intuitive because the firm wants to avoid volatility in the risk-averse region and seek it in the risk-seeking region. The plot also shows that when μ_S increases, the size of the risk-averse region

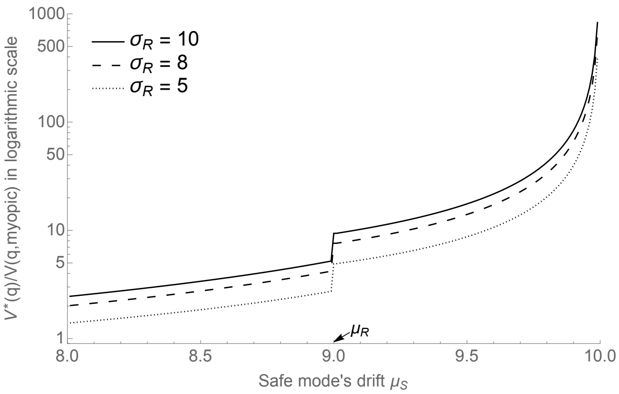


Figure 5: The ratio of CLV under the optimal policy to CLV under the myopic policy (on a logarithmic scale) versus μ_S, for $\mu_R = 9$, $q = 10$, and for initial satisfaction at q.

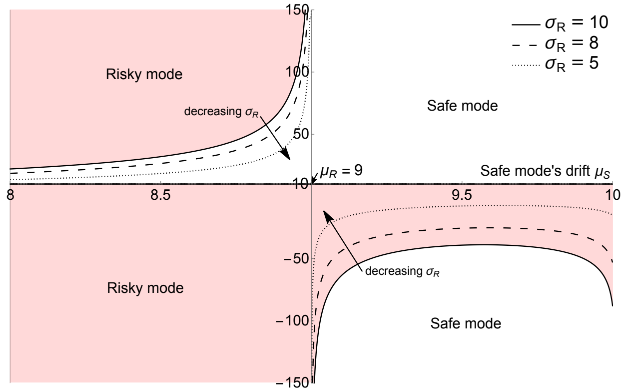


Figure 6: The optimal sandwich policies for different model primitives. Fix $\mu_R = 9$, $q = 10$. The horizontal axis corresponds to the value of μ_S, and the vertical line marks the satisfaction value. The two curves are the switching boundaries between the Risky mode and Safe mode.

increases. In fact, the risk-averse region would disappear if μ_S is sufficiently small, making the myopic policy optimal. However, as illustrated by the right half of Figure 6, the risk-seeking region is nonmonotone in μ_S. This is because as μ_S increases, there are two competing forces in terms of whether or not the firm should use the Risky mode more often. On one hand, as μ_S increases, the Risky mode becomes more inferior, incentivizing the firm to use it less often. On the other hand, a larger μ_S makes the risk-seeking region more effective since the delayed reflected O-U process would spend more time at the satisfaction threshold q, prolonging customer lifetime even more. The nonmonotonicity of the risk-seeking region indicates that the former effect dominates when μ_S is just slightly bigger than μ_R, while the latter effect dominates when μ_S becomes close to q.

We formalize our monotonicity results in the next theorem. To prove these results, we naturally investigate the corresponding derivatives (for example, $\frac{d\theta_b}{d\mu_S}$ can be computed from the definition of θ_b in Lemma 2), and draw upon L'Hospital-type rules for monotonicity [24] (see Lemma 5, which is applied throughout the proof of Theorem 2). The proof is in Section 6.

Theorem 2. *Denote by V^* the optimal value function. Consider θ_b and θ_G defined in Lemmas 2 and 3 respectively. Then if the Risky mode is superior, i.e., $\mu_R > \mu_S$,*

1. *the value θ_b is strictly increasing in μ_S, and for any $x \in \mathbb{R}$, V^* is increasing in μ_S;*
2. *the value θ_b is strictly decreasing in μ_R, and for any $x \in \mathbb{R}$, V^* is increasing in μ_R;*
3. *the value θ_b is strictly increasing in σ_R and, for any $x > q$, we have that $\frac{\partial V^*}{\partial x}$ is weakly decreasing in σ_R.*

On the other hand, when the Safe mode is superior, i.e., $\mu_S > \mu_R$,

4. *the value θ_G is decreasing in σ_R.*

Parts 1 and 2 of Theorem 2 show that a higher mean reward rate makes the corresponding service mode more attractive (so the risk-averse region, where the Safe mode is used, is larger if μ_S is larger, and smaller if μ_R is larger) and improves the overall CLV. Part 3 and 4 show that the greater the volatility of the Risky mode, the larger the size of the risk-averse region in the optimal sandwich policy when $\mu_S < \mu_R$, and also the larger the size of the risk-seeking region in the optimal sandwich policy when $\mu_S > \mu_R$. In part 3 we also show that the marginal value of customer satisfaction in the satisfied zone is lower when the Risky mode is more volatile. This is due to the fact that the mean first passage time to the unsatisfied zone is smaller (or unchanged if $x \in (q, \theta_b)$) when the Risky mode is more volatile. We also numerically observe that θ_G is decreasing in μ_R.

At the end of this section, we present a corollary showing that the benefit of the optimal sandwich policy with respect to the myopic policy is unbounded when the Safe mode is superior. This can be observed from the upper tail of the curve in Figure 5.

Corollary 2. *Fix q, μ_R (which satisfies $\mu_R < q$) and σ_R. Consider the regime where $\mu_S > \mu_R$. Let $V^*(q)$ be the CLV under the optimal policy (see Eq. (8)) when the customer's starting satisfaction level is q. Also, let $V(q, \text{myopic})$ be the CLV under the myopic policy (always using the Safe mode in this case, see Eq. (9)) when the customer's starting satisfaction level is q. Then we have*

$$\lim_{\mu_S \to q} \frac{V^*(q)}{V(q, \text{myopic})} = \infty.$$

Corollary 2 can be proved by utilizing the explicit expressions for $V^*(q)$ and $V(q, \text{myopic})$ in the proof of Theorem 1. The full proof is given in Online Appendix B.

4.3 Summary of Results under Variations of our Model

We tested the robustness of our findings by analyzing four variations of the original model from Section 3. These variations are 1) allowing mixed policies; 2) letting the reward process to be a geometric Brownian motion; 3) incorporating positive switching costs; 4) using alternative churn hazard rate functions. We find that in most cases of our interest, either a sandwich policy or a myopic policy is optimal, and our results under the original model are remarkably robust on all these dimensions. Below we summarize our findings under each of the four variations. The detailed analysis of each variation can be found in Online Appendix C–F.

Mixed policies. An investment manager can mix his investment between the Risky asset and the Safe asset. We consider and solve the firm's problem under $\mu_R > \mu_S$ that allows mixed policies; see Online Appendix C for a complete description and analysis. We find that (see Theorem 3 in Online Appendix C) the essence of the optimal sandwich policy in the original model is preserved under mixed policies: the firm should be myopic except for some intermediate satisfaction region $[q, \theta_I]$ just above the satisfaction threshold q. In this region, the firm should be risk-averse, i.e., mix between the two service modes. Theorem 4 in Online Appendix C shows two additional results: First, as the satisfaction value approaches the unsatisfied zone from above, the firm should be more risk-averse, i.e., give more weight to the Safe asset. Second, a monotonicity result similar to Theorem 2 holds; specifically, the upper threshold of the sandwich θ_I is increasing in μ_S and σ_R.

Geometric Brownian motion rewards. Motivated by an investment problem where a risk-free asset generates returns (interest) percentage drift μ_S deterministically and a risky asset generates returns (capital gains and dividends, with automatic reinvestment of dividends) with percentage drift μ_R and percentage volatility σ_R, we consider a model that uses geometric Brownian motion (GBM) reward process instead of arithmetic Brownian motion reward (see Online Appendix D). We formally develop optimality conditions for finding the optimal value function and policy, and numerically solve for the optimal policy for a large variety of (random) problem instances. The results show that, as in the original model (see Theorem 1 for the $\mu_R > \mu_S$ case), the (numerically solved) optimal policies among all the considered problem instances are either a myopic policy that always uses the Risky mode everywhere, or a sandwich policy that uses the Risky mode for all satisfaction states except for some intermediate satisfaction range $[q, \tilde{\theta}_b]$ (for some numerically specified $\tilde{\theta}_b$). The optimal sandwich policy once again provides a substantial CLV increase over the myopic policy. Also, the optimal switching threshold $\tilde{\theta}_b$ exhibits similar monotonicity in model primitives as in the original model (see Theorem 2): $\tilde{\theta}_b$ decreases in μ_R and increases in σ_R.

Switching costs. One might notice that the optimal sandwich policy in the original model switches infinitely many times (with positive probability and in expectation) between the two service modes. However, in several applications, switching between service modes may be costly. Motivated by this, we consider a model that assumes each transition from one service mode to the other incurs a fixed cost K and numerically solve for the firm's optimal policy under various values of K (see Online Appendix E). Notably, focusing on the case $\mu_R > \mu_S > 0$, we find that when switching cost is small, each optimal switching threshold in the original model (see Theorem 1) is replaced by a *buffer* interval, above and below which the firm should use the service mode as prescribed by the optimal policy in the original model, but within which the firm should *not* switch service modes.

Imperfect measurement of customer satisfaction: Not only is the above mentioned finding interesting in its own right, it has favorable implications for a setting where the firm cannot perfectly measure satisfaction. In particular, our findings suggest that small to medium-sized errors in estimating customer satisfaction would not significantly impair the CLV benefits of the optimal policy (relative to the myopic policy). In each case where the optimal policy for the case of switching costs is one where switching is postponed by a buffer, by definition this policy produces higher CLV than the myopic policy, and performs even better if there were no switching costs. This gives us confidence that our proposed policies still substantially increase the CLV in the face of small to medium-sized errors in estimating customer satisfaction. Along similar lines, interpreting the effect of the buffer intervals as delays in switching, one can argue that (small) delays in estimating customer satisfaction do not erode the benefits of using our proposed policies.

Other hazard rate functions. So far, we have restricted the customer's hazard rate of churn to be a step function of the customer's satisfaction state with step height 1 (recall the original step function in Eq. (4)). What about other specifications of the hazard rate function? We investigate this problem for a variety of non-increasing hazard rate functions including alternate step heights, nth power, exponential and logit functions in the unsatisfied zone (see Online Appendix F). We also consider hazard rate functions that decrease to some positive value instead of zero in the satisfied zone. Extensive numerical evidence under various hazard rate function specifications show that, in the case $\mu_R > \mu_S > 0$, the optimal policy is still either myopic or a sandwich policy, and moreover, the switching thresholds in the optimal sandwich policy are fairly robust to the shape of hazard rate functions. The gap in CLVs between under the optimal policy (which is a sandwich policy) and the myopic policy remains large for different hazard rate functions.

5 Proof of Theorem 1

To start with, the HJB equation (formally derived in the proof of Proposition 2, see Section 6) yields the following ordinary differential equation (ODE):

$$\max_{i=S,R}\left\{-Q(x)V(x)+(\mu_i-x)V'(x)+\frac{1}{2}\sigma_i^2V''(x)+\mu_i\right\}=0 \tag{13}$$

for all $x \in \mathbb{R}$ where V'' exists. We cannot deduce that Eq. (13) has a twice continuously differentiable solution, since $Q(x)$ is discontinuous and the second order term can be zero. In fact, the value function of this problem is never twice continuously differentiable at q. We shall conclude, by the end of this section, that V^* is continuously differentiable everywhere on $\mathbb{R}$, and twice continuously differentiable everywhere on $\mathbb{R} \setminus \mathcal{E}$, with $\mathcal{E}$ containing the switching boundaries and q. In this section, we first state a proposition containing the optimality conditions for verifying the value function (see Proposition 2). In Section 5.2, we build a key lemma (Lemma 5) about properties of the error function $\mathrm{erf}(x)$ (see the beginning of Section 4.1 for its definition) and its complement $\mathrm{erfc}(x) \triangleq 1-\mathrm{erf}(x)$. This key lemma helps with constructing a candidate value function (see Proposition 3) that satisfies the optimality conditions established in Proposition 2.

5.1 Optimality Conditions

The following proposition provides conditions for optimality.

Proposition 2. *Suppose a function $\bar{V}:\mathbb{R}\to\mathbb{R}$ satisfies*

1. *the function $\bar{V}$ is non-negative;*
2. *the function $\bar{V}$ is continuously differentiable on $\mathbb{R}$ and twice continuously differentiable on $\mathbb{R}\setminus\mathcal{E}$ for some countable set $\mathcal{E}$;*
3. *the function $\bar{V}'$ is bounded;*
4. *for any $x\in\mathbb{R}$ for $i=S$ and for any $x\in\mathbb{R}\setminus\mathcal{E}$ for $i=R$, the following inequality holds*[13]
$$-Q(x)\bar{V}(x)+(\mu_i-x)\bar{V}'(x)+\frac{1}{2}\sigma_i^2\bar{V}''(x)+\mu_i\le 0; \tag{14}$$
5. *for some interval policy $\bar{\pi}$ (see Definition 2) such that $\bar{\pi}(y)=S$ for all $y\in\mathcal{E}$, the process $\bar{V}(H_t^{x,\bar{\pi}})$ is an $\mathbb{F}$-adapted semimartingale, and for all $x\in\mathbb{R}$ it holds that*
$$-Q(x)\bar{V}(x)+(\mu_{\bar{\pi}(x)}-x)\bar{V}'(x)+\frac{1}{2}\sigma_{\bar{\pi}(x)}^2\bar{V}''(x)+\mu_{\bar{\pi}(x)}=0. \tag{15}$$
6. *for all $x\in\mathbb{R}$ and the policy $\bar{\pi}$ in Condition 5,*
$$\lim_{t\to\infty}\mathbb{E}\left[\left(1+\left|H_t^{x,\bar{\pi}}\right|\right)e^{-\int_0^t Q(H_s^{x,\bar{\pi}})ds}\right]=0. \tag{16}$$

Then, the function $\bar{V}$ is the value function V^*, and $\bar{\pi}$ is an optimal policy.

The formal proof is in Section 6. We now provide a short summary of the proof of Proposition 2. Conditions 1–4 imply that the function $\bar{V}$ is an upper bound of the optimal value function V^*. This is established by constructing a stochastic process
$$X_t^{x,\pi}=\bar{V}(H_t^{x,\pi})e^{-\int_0^t Q(H_s^{x,\pi})ds}+\int_0^t e^{-\int_0^s Q(H_u^{x,\pi})du}dY_s^{x,\pi}, \tag{17}$$
and showing $\bar{V}(x)\ge\limsup_{t\to\infty}\mathbb{E}X_t^{x,\pi}\ge V(x,\pi)$ for any admissible policy $\pi\in\Pi$. Note that Condition 3 serves to ensure that the local martingale component of $X_t^{x,\pi}$ is a martingale and hence has zero expectation. Conditions 5 and 6 further imply that the bound is tight under the interval policy $\bar{\pi}$, i.e., $\bar{V}$ is the optimal value function and the optimal policy is $\bar{\pi}$. In particular, we use Condition 5 to show that $\bar{V}(x)=\limsup_{t\to\infty}\mathbb{E}X_t^{x,\bar{\pi}}$. We then use Condition 6 together with Condition 3 to deduce $\lim_{t\to\infty}\bar{V}(H_t^{x,\pi})e^{-\int_0^t Q(H_s^{x,\pi})ds}=0$ and hence $\limsup_{t\to\infty}\mathbb{E}X_t^{x,\bar{\pi}}=V(x,\bar{\pi})$.

Proposition 2 tells us that if an interval policy $\bar{\pi}$ together with a function $\bar{V}$ satisfies the stated conditions, then $\bar{\pi}$ is optimal among the class of all admissible policies Π (not just among interval policies), and $\bar{V}=V^*$. With Proposition 2 in place, our task is to find $\bar{V}$ and $\bar{\pi}$ satisfying conditions 1–6 there. The following lemma shows that Condition 6 is satisfied by any interval policy with a finite number of intervals. Note that the sandwich policy is an interval policy with three intervals.

Lemma 4. *For any starting satisfaction level $x\in\mathbb{R}$ and any interval policy with a finite number of intervals, we have*
$$\lim_{t\to\infty}\mathbb{E}\left[(1+|H_t^{x,\pi}|)e^{-\int_0^t Q(H_s^{x,\pi})ds}\right]=0.$$

This lemma implies that the policy from Theorem 1 satisfies Condition 6 in Proposition 2.

[13]For any x including $x\in\mathcal{E}$, we define $\frac{1}{2}\sigma_S^2\bar{V}''(x)\triangleq 0$ consistent with $\sigma_S=0$.

5.2 Properties of the error function

To prove Lemmas 2 – 3 and to further solve for and verify the optimal value function require deep understanding of the error function $\mathrm{erf}(x)$ and its complement $\mathrm{erfc}(x)$. Lemma 5 in this section establish the limit, sign, and monotonicity of complicated functions involving $\mathrm{erf}(\cdot)$ and $\mathrm{erfc}(\cdot)$. To achieve this, we draw upon existing literature on the Chernoff-type bound (see Chang et al. [8]) and asymptotic expansion (see Ren et al. [27]) for the $\mathrm{erfc}(\cdot)$ function, as well as "L'Hospital-type rules for monotonicity" (see Pinelis [24]).

Lemma 5. *Consider the error function* $\mathrm{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ *and its complement* $\mathrm{erfc}(x) \triangleq 1 - \mathrm{erf}(x)$. *They satisfy the following properties:*

1. $e^{x^2}\mathrm{erfc}(x)(1+2x^2) - \frac{2x}{\sqrt{\pi}} \geq 0$ *for any* $x \in \mathbb{R}$*;*
2. $e^{x^2}\mathrm{erfc}(x) \to 0$ *as* $x \to \infty$*;*
3. $xe^{x^2}\mathrm{erfc}(x) \to \frac{1}{\sqrt{\pi}}$ *as* $x \to \infty$*;*
4. $xe^{x^2}\mathrm{erf}(x) > -\frac{1}{\sqrt{\pi}}$ *for all* $x \in \mathbb{R}$.
5. *Fix* $\alpha \in \mathbb{R}$. *Let* $r(x) \triangleq (x+\alpha)e^{x^2}\mathrm{erfc}(x)$. *The following statements are true:*
   - *(i) If* $\alpha > 0$, $r(x)$ *is first increasing then decreasing on* $(-\alpha, \infty)$*; and if* $\alpha \leq 0$, $r(x)$ *is increasing on* $(-\alpha, \infty)$.
   - *(ii) Fix any* $0 < \beta < 1$. *Then* $r(x) = \frac{\beta}{\sqrt{\pi}}$ *has a unique root (denote by* x_0*) on* $(-\alpha, \infty)$. *Moreover,* $r(x) < \frac{\beta}{\sqrt{\pi}}$ *on* $(-\alpha, x_0)$, $r(x) > \frac{\beta}{\sqrt{\pi}}$ *on* (x_0, ∞).
   - *(iii) If* $\alpha = 0$, *then* $r(x) < \frac{1}{\sqrt{\pi}}$ *for any* $x \in \mathbb{R}$.
   - *(iv)* $e^{x^2}\mathrm{erfc}(x)$ *is monotonically decreasing in* x.
6. *Suppose* $\alpha > 0$ *and* $0 < \beta < 1$. *Then* x_0 *defined in property 5(ii) satisfies* $x_0 \leq \frac{\alpha(\beta-2)+\sqrt{\alpha^2\beta^2+2\beta(1-\beta)}}{2(1-\beta)}$.
7. $\frac{C}{\sqrt{\pi}} + xe^{x^2}(1+C\mathrm{erf}(x))$ *is non-decreasing in* x, *where* $C \in [-1, 1]$ *is some constant.*
8. *The derivative of function* $F(x) \triangleq \log\left(xe^{x^2}\mathrm{erfc}(-x) + \frac{1}{\sqrt{\pi}}\right)$ *is increasing in* x.
9. *Consider* $q > a > 0$. *Then* $(a^2-2)\left(1 + e^{q^2}\sqrt{\pi}q\left(\mathrm{erf}(q) - \mathrm{erf}\left(a - \frac{1}{a}\right)\right)\right) - 2ae^{q^2 - \left(a - \frac{1}{a}\right)^2} q < 0$.
10. *Consider* $q > 0$ *and* $\theta < q$. *Then* $e^{q^2}q - e^{\theta^2}\theta - e^{\theta^2+q^2}\theta q\sqrt{\pi}\left(\mathrm{erfc}(\theta) - \mathrm{erfc}(q)\right) \geq 0$.
11. *Consider* $q > 0$ *and* $\theta < q$. *Then* $2e^{q^2-\theta^2}q\theta - \left(1+2\theta^2\right)\left(1 + e^{q^2}q\sqrt{\pi}\left(\mathrm{erfc}(\theta) - \mathrm{erfc}(q)\right)\right) < 0$.
12. *Consider* $q > 0$ *and* $\theta < q$. *Then* $\theta(1+2q^2) - e^{q^2-\theta^2}q(1+2q^2) + 2e^{q^2}q(1+q^2)\theta\sqrt{\pi}\left(\mathrm{erf}(q) - \mathrm{erf}(\theta)\right) \leq 0$.
13. *Consider* $a > 0$ *and* $x \in (a - \frac{1}{a}, a)$. *Then* $2 - 2ax + 2x^2 + \left(3x - a - 2ax^2 + 2x^3\right)e^{x^2}\sqrt{\pi}\mathrm{erf}(x) > 0$.
14. *Consider* $a > 0$ *and* $x \in (a - \frac{1}{a}, a)$. *Then* $2 - 2ax + 2x^2 - \left(3x - a - 2ax^2 + 2x^3\right)e^{x^2}\sqrt{\pi}\mathrm{erfc}(x) > 0$.
15. *Consider* $a > 0$ *and* $x \in (a - \frac{1}{a}, a)$. *Then* $2 - 2ax + 2x^2 + \left(3x - a - 2ax^2 + 2x^3\right)e^{x^2}\sqrt{\pi}(1 + \mathrm{erf}(x)) > 0$.

16. *Consider* $a > 0$, $\theta \in (a - \frac{1}{a}, a)$, *and* $x \in (\theta, a]$. *Then* $2(a-x)\left(a - 3\theta + 2a\theta^2 - 2\theta^3\right) - 2e^{x^2-\theta^2}\left(1 - a\theta + \theta^2\right)(1-2ax+2x^2) + e^{x^2}\sqrt{\pi}\left(a - 3\theta + 2a\theta^2 - 2\theta^3\right)(1-2ax+2x^2)(\text{erf}(\theta) - \text{erf}(x)) \leq -2 + 2a^2 - 2a\theta$.

Lemma 2 follows from property 5(ii) (proofs in Section 6). Properties 9 – 11 are used to prove Lemma 3. The rest are used in the verification step (proof of Proposition 3). We give the proofs of properties 1 – 3, 5 and 8 here. In particular, properties 5 and 8 rely on utilizing the L'Hospital-type rules for monotonicity ([24]). The proof of property 4 is quite standard, as well as the rest of the proofs which utilize the first five properties. They can be found in online appendix A.

Proof of Lemma 5 (Part I). We first prove 1. The Chernoff-type lower bound for the erfc() function by Chang et al. [8] gives: $\text{erfc}(x) \geq \sqrt{\frac{2e}{\pi}\frac{\sqrt{\beta-1}}{\beta}}e^{-\beta x^2}, \forall \beta > 1$. Choose $\beta(x) = 1 + \frac{1}{2x^2}$. Then,

$$e^{x^2}\text{erfc}(x)(1+2x^2) - \frac{2x}{\sqrt{\pi}} \geq \frac{1}{\sqrt{\pi}}\left(\frac{\sqrt{2e(\beta(x)-1)}}{\beta(x)}e^{x^2(1-\beta(x))}(1+2x^2) - 2x\right) = 0.$$

Properties 2 and 3 follow immediately from the asymptotic expansion in Ren & MacKenzie [27]: $\text{erfc}(x) = \frac{e^{-x^2}}{x\sqrt{\pi}}(1 + o(\frac{1}{x^2}))$.

We now prove 5. First consider (i). Clearly $r(-\alpha) = 0$. Also from properties 2 and 3, $r(\infty) = \frac{1}{\sqrt{\pi}}$ follows easily. Let $f(x) \triangleq \text{erfc(x)}$, $g(x) \triangleq \frac{e^{-x^2}}{x+\alpha}$, so that $r(x) = \frac{f(x)}{g(x)}$. Consider the "derivative ratio" $\rho(x) \triangleq \frac{f'(x)}{g'(x)} = \frac{2(x+\alpha)^2}{\sqrt{\pi}(1+2\alpha x+2x^2)}$. Then $\rho'(x) = \frac{4(\alpha+x)(1-\alpha^2-\alpha x)}{\sqrt{\pi}(2x^2+2\alpha x+1)^2}$.

<u>Case 1: $\alpha > 0$.</u> $\rho'(x) > 0$ on $(-\alpha, -\alpha + \frac{1}{\alpha})$, and $\rho'(x) < 0$ on $(-\alpha + \frac{1}{\alpha}, \infty)$. Hence $\rho(x)$ first increases then decreases on $(-\alpha, \infty)$. Also both f and g vanish at ∞. Using the "L'Hospital-type rule for monotonicity" [24, Proposition 4.3], $r(x)$ first increases then decreases on $(-\alpha, \infty)$.

<u>Case 2: $\alpha \leq 0$.</u> $\rho'(x) > 0$ on $(-\alpha, \infty)$. Hence $\rho(x)$ increases on $(-\alpha, \infty)$. Again, both f and g vanish at ∞. Using the "L'Hospital-type rule for monotonicity" [24, Proposition 4.1], $r(x)$ also increases on $(-\alpha, \infty)$.

We have proved (i). Observe that since $r(-\alpha) = 0$ and $r(\infty) = \frac{1}{\sqrt{\pi}}$, (ii) – (iii) simply follow from (i). Also (iv) follows from (iii) since the derivative of $e^{x^2}\text{erfc}(x)$ is $2xe^{x^2}\text{erfc}(x) - \frac{2}{\sqrt{\pi}}$.

Next we prove 8. By property 5(iii), $F(x)$ is well-defined on $\mathbb{R}$. Let $r(x) \triangleq F'(x) = \frac{f(x)}{g(x)}$, where $f(x) \triangleq \sqrt{\pi}\left(2x^2+1\right)\text{erfc}(-x) + 2e^{-x^2}x$ and $g(x) \triangleq \sqrt{\pi}x\text{erfc}(-x) + e^{-x^2}$. Consider the first and second order "derivative ratio":

$$r_1(x) \triangleq \frac{f'(x)}{g'(x)} = \frac{4\left(\sqrt{\pi}y\text{erfc}(-x) + e^{-x^2}\right)}{\sqrt{\pi}\text{erfc}(-x)}, \qquad r_2(x) \triangleq \frac{f''(x)}{g''(x)} = 2\sqrt{\pi}\text{erfc}(-x)e^{x^2}.$$

$r_2(x)$ is increasing in x by property 5(iv). One can also verify, by applying $\text{erfc}(x) = \frac{e^{-x^2}}{x\sqrt{\pi}}(1+o(\frac{1}{x^2}))$ from Ren & MacKenzie [27], that $\lim_{x\to-\infty} f(x) = \lim_{x\to-\infty} f'(x) = \lim_{x\to-\infty} g(x) = \lim_{x\to-\infty} g'(x) = 0$. Therefore we can apply the "L'Hospital-type rules for monotonicity" [24, Proposition 4.1] on $r_1(x)$ and $r(x)$ to get the desired result.

□

5.3 Verification step

Next we find a candidate optimal value function by solving the HJB equation (13). Observe that Conditions 4–5 in Proposition 2 imply $\bar{V}$ and $\bar{\pi}$ solve the HJB equation (13). We obtain a candidate value function by the following informal reasoning. First of all, we conjecture that the superior mode should be used myopically for sufficiently large and sufficiently small satisfaction values. This is a straightforward conjecture since the only possible reason to use the inferior mode is to prolong customer lifetime, but this benefit (if there is such a benefit) is minimal when the customer's satisfaction value is far away from the satisfaction threshold q. In particular, when the satisfaction state is far below, there is almost no chance to escape the unsatisfied zone before the customer departs. On the other hand, when the satisfaction state is far above q, it descends slower under the superior mode. Thus inspired, we start by solving Eq. (15) for a control policy that always uses the superior mode, while asking for continuity of V' at q and boundedness of V' everywhere. Next, we check at which x values the superior mode achieves the maximum in the LHS of the HJB equation (13). We then update our policy by retaining the superior mode at such satisfaction locations and replacing it with the inferior mode at the complementary locations where the inferior mode achieves the maximum. For the updated policy, we compute the updated V by solving Eq. (15) again, while satisfying the boundary conditions — continuity of V' as well as boundedness of V' everywhere. We then verify if this new control policy achieves the maximum in the LHS of Eq.(15). Indeed, we find that this new policy and its associated candidate value function satisify Eq. (15). The candidate value function that we denote by W^* is defined in Proposition 3, where we also verify that it satisfies not only the HJB equation (13), but all conditions in Proposition 2.

Proposition 3. *Let $W(x, C_1, C_2, C_3, C_4, C_5)$ be defined as*

$$W(x, C_1, C_2, C_3, C_4, C_5) = \begin{cases} V_1(x, C_1) & \text{if } x \le \theta_G; \\ V_2(x, C_2, C_3) & \text{if } \theta_G < x < q; \\ V_3(x, C_4) & \text{if } q \le x \le \theta_b; \\ V_4(x, C_5) & \text{if } x > \max\{q, \theta_{\mathrm{b}}\}, \end{cases}$$

where q is the satisfaction threshold (see Eq. (4)), θ_b is as defined in Lemma 2, θ_G is as defined in Lemma 3, and

$$V_1(x, C_1) = \frac{C_1}{x - \mu_S} + \mu_S; \tag{18}$$

$$V_2(x, C_2, C_3) = C_2 e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} + C_3 \mathrm{erf}\Big(\frac{x - \mu_R}{\sigma_R}\Big) e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} + \mu_R; \tag{19}$$

$$V_3(x, C_4) = C_4 + \mu_S \log(x - \mu_S). \tag{20}$$

$$V_4(x, C_5) = C_5 + \int_0^x \frac{\mu_R \sqrt{\pi}}{\sigma_R} e^{\frac{(z-\mu_R)^2}{\sigma_R^2}} \Big(1 - \mathrm{erf}\Big(\frac{z - \mu_R}{\sigma_R}\Big)\Big) dz. \tag{21}$$

Then there exist $C_1^*, C_2^*, C_3^*, C_4^*$, and C_5^* (defined explicitly in the proof of this proposition) such that $W^*(\cdot) \triangleq W(\,\cdot\,, C_1^*, C_2^*, C_3^*, C_4^*, C_5^*)$ together with the policy specified in Theorem 1 satisfies all conditions in Proposition 2.

The proof of this proposition is in Section 6. Note that verifying the conditions in Proposition 2 is a challenging task, the hardest component being to show non-positivity Eq. (14) of parametric functions involving four free parameters (μ_S, μ_R, σ_R, q) plus a variable x representing the satisfaction level. We achieve this by decomposing the function of interest into different pieces and showing non-positivity for each piece, utilizing the properties established in Lemma 5).

The proof of Theorem 1 is now immediate.

Proof of Theorem 1. The main part of the theorem (the structure of the optimal policy) is an immediate consequence of Proposition 3. The upper bound on the value function (see Proposition 2 Conditions 1-4) is a bound on the CLV under any admissible policy. A bound on expected customer lifetime under any admissible policy also follows using Wald's identity, since the expected reward rate is at least $\min(\mu_S, \mu_R)$ at every instant while the customer is alive. $\square$

6 Other Proofs

6.1 Proofs of the Model Section

To prove Lemma 1, we first establish a technical lemma, which will also be used later in the proof of Proposition 2.

Lemma 6. *Fix any admissible policy $\pi \in \Pi$ and starting satisfaction level $x \in \mathbb{R}$. Let H_t be the corresponding satisfaction process as defined in Eq.* (2) *and let u_t be the corresponding action process. For $y \in \mathbb{R}$, denote by $L^H(t,y)$ the symmetric local time of H_t at y. Then for any $t \geq 0$ and any $y \in \mathbb{R}$, it holds that $L^H(t,y) < \infty$ almost surely. Let $\mathcal{E}$ be any countable set in $\mathbb{R}$. Then for any $t \geq 0$ it holds that $\int_0^t \mathbb{1}\{H_s \in \mathcal{E} \ \&\ u_s \neq S\} ds = 0$ almost surely.*

Proof. By the definition of local time (see Definition A.1 in Salins and Spiliopoulos [29]), we have $L^H(t,y) = |H_t - y| - (x-y) - \int_0^t \operatorname{sign}(H_s - y)(\mu_{u_s} - H_s) ds - \int_0^t \operatorname{sign}(H_s - y)\sigma_{u_s} dB_s$, which implies $\mathbb{E}L^H(t,y) \leq \mathbb{E}|H_t| + |y| + |x| + |y| + \int_0^t (\mu_R + \mathbb{E}|H_s|) ds$. If we take expectations on both sides of Eq. (2), one can easily see that $\mathbb{E}H_t \leq x + \mu_R t$. Therefore the above inequality implies that $\mathbb{E}L^H(t,y) < \infty$, and hence $L^H(t,y) < \infty$ almost surely. To complete the proof it remains to show the occupation time $\int_0^t \mathbb{1}\{H_s \in \mathcal{E} \ \&\ u_s \neq S\} ds = 0$, or equivalently, $\int_0^t \mathbb{1}\{H_s \in \mathcal{E}\}(dH_s)^2 = 0$ almost surely. This is indeed true, since if otherwise, there exists $y \in \mathcal{E}$ such that with positive probability we have $\int_0^t \mathbb{1}\{H_s = y\}(dH_s)^2 > 0 \Rightarrow L^H(t,y) = \infty$, where the implication follows from the definition of local time. But this contradicts $L^H(t,y) < \infty$ almost surely. □

Next we prove Lemma 1. Note that we would have been able to use Salins and Spiliopoulos [29] directly to specify the satisfaction process for any interval policy except for one wrinkle — Salins and Spiliopoulos [29] assumes that the volatility of the stochastic process is bounded below by a positive constant everywhere. So the strategy we adopt is to use Salins and Spiliopoulos [29] to specify the satisfaction process during the time intervals when it is in the closure of a Risky piece, and to combine this specification with the obvious, deterministic trajectory converging exponentially at rate 1 to μ_S when the satisfaction is in the interior of a Safe piece.

Proof of Lemma 1. **(Part 1.)** Note that by Eq. (1), under control process $u_t \in \{R, S\}$, Eq. (3) can be equivalent written as

$$dH_t = \mu_{u_t} dt + \sigma_{u_t} dB_t - H_t dt.$$

Next we first establish that the solution H_t to this SDE (along with S_t as defined by Eq. (5)) exists and is an $\mathbb{F}$-adapted semimartingale specified uniquely in law.

We specify the satisfaction process for any interval policy inductively on the number of pieces of intervals. The simplest case is where there is just one piece: If the policy is to use the Risky mode everywhere, the satisfaction process is simply an O-U process with parameters μ_R and σ_R. If the policy is to use the Safe mode everywhere, the satisfaction process is deterministic and converges exponentially at rate 1 to μ_S.

Now let us consider the case where the policy has a Risky piece and a Safe piece. One possible case (among four possibilities, we will consider the other three cases below) is that the policy uses the Risky mode on $(-\infty, \theta)$ and the Safe mode on $[\theta, \infty)$ for some $\theta > \mu_S$. Now if the satisfaction starts at $x > \theta$, there is an initial deterministic transient

$$H_t = \mu_S + (x - \mu_S)e^{-t}, \qquad \text{for all } t \in \left[0, \log\left(\frac{x - \mu_S}{\theta - \mu_S}\right)\right], \tag{22}$$

where the satisfaction decays exponentially at rate 1 from x to θ under the Safe mode, arriving at θ at time $t_0 = \log\left(\frac{x-\mu_S}{\theta-\mu_S}\right)$. After this transient, the satisfaction process resembles a reflected O-U process with parameters μ_R and σ_R that lives in (∞, θ) and is reflected downwards at θ, but with a crucial difference. The reflected O-U process spends a measure zero of time at the reflecting boundary. In our process, the time spent at the reflecting boundary has positive measure with probability one, conditioned on the customer satisfaction being q at some time. While this process is less well-known, such a *delayed reflected* process was introduced by Skorokhod [31], and shown to be a semimartingale by Salins and Spiliopoulos [29]. The measure of time spent by a delayed reflected process at a reflection boundary is proportional to the local time of the process at the boundary, and the constant of proportionality (termed the *delay parameter* in Salins and Spiliopoulos [29]) is the inverse of the drift at the boundary. In our setting the drift is negative with magnitude $\theta - \mu_S$ under the Safe mode at the boundary satisfaction θ, and hence the delay parameter is $1/(\theta - \mu_S)$. Having established this connection, we rely on Theorem 3.4 from Salins and Spiliopoulos [29] to conclude that the distribution of the satisfaction process for $[t_0, \infty)$ can be defined as the law of the unique solution (guaranteed to exist) of the SDE and local time pair

$$dH_t = (\mu_R - H_t)\mathbb{1}\{H_t < \theta\}dt + \sigma_R \mathbb{1}\{H_t < \theta\}dB_t - L^H(dt, \theta), \qquad H_{t_0} = \theta\,,$$

$$\frac{1}{\theta - \mu_S} L^H(t, \theta) = \int_{t_0}^{t} \mathbb{1}\{H_t = \theta\}ds\,,$$

where $L^H(t, \theta)$ is the symmetric local time of the satisfaction process at θ. And being a (weak) solution to an SDE, the satisfaction process is guaranteed to be a semimartingale. Notice that we also obtain a unique specification of (the distribution of) time spent at the boundary: this will be helpful when the boundary is exactly at q, so that time spent at the boundary is identical to time spent in the satisfied zone. For example, integrals such as $\int_0^t Q(H_s)ds$ will be meaningful even when the boundary is at q; this integral will simply be the time spent away from the boundary inside the unsatisfied zone (since the hazard rate there is 1, whereas it is 0 in the satisfied zone). As a result, we can, in fact, conclude that $(H_t)_{t\in[0,\bar{t}]}$ and $(S_t)_{t\in[0,\bar{t}]}$ are semimartingales adapted to $(\mathcal{F}_t)_{t\geq 0}$ for any $\bar{t} < \infty$, since the survival probability process $S_t = e^{-\int_0^t Q(H_s)ds}$ is uniquely specified in terms of the satisfaction process.

Now we discuss the other three cases for one Risky and one Safe piece along similar lines. In the interest of space, we focus on the satisfaction process H_t and skip Y_t and S_t, though they are immediate to specify.

- If the policy uses the Risky mode on $(-\infty, \theta)$ and the Safe mode on $[\theta, \infty)$ for some $\theta \leq \mu_S$: If the satisfaction starts at $x \geq \theta$, then it decays exponentially at rate 1 towards μ_S as per Eq. (22) for all $t \in [0, \infty)$. If the satisfaction starts at $x < \theta$ it follows an O-U process with parameters μ_R and σ_R until the (finite w.p. 1) time at which it hits θ, after which it remains in the Safe piece forever and decays exponentially at rate 1 towards μ_S.
- If the policy uses the Safe mode on $(-\infty, \theta]$ and the Risky mode on (θ, ∞) for some $\theta \leq \mu_S$: this is analogous to the case discussed at length above. The trajectory is deterministic upwards inside the Safe piece, whereas it is a delayed reflected O-U process inside the Risky piece with upward reflection with delay parameter $1/(\mu_S - \theta)$ at the boundary θ.
- If the policy uses the Safe mode on $(-\infty, \theta]$ and the Risky mode on (θ, ∞) for some $\theta > \mu_S$: this is analogous to the first bullet with a deterministic trajectory for all time inside the Safe piece, and an O-U process in the Risky piece such that after the first instant when it hits the boundary θ, the satisfaction thereafter remains within the Safe piece forever.

At this point, it is straightforward to see how the construction extends inductively to any countable number of pieces with the Risky pieces being open. The fact that $(Y_t)_{t\geq 0}$ as the solution to Eq. (1) exists and is uniquely specified in law follows, since from Eq. (3) we have $Y_t = H_t - H_0 + \int_0^t H_s ds$.

(Part 2.) Since H_t is a semimartingale, we can apply the Itô-Tanaka formula (see Theorem A.3 in Salins & Spiliopolous [29]) for any function f that is continuously differentiable on $\mathbb{R}$ and twice continuously differentiable on $\mathbb{R} \setminus \mathcal{E}$ for some countable set $\mathcal{E}$: $f(H_t) = f(H_0) + \int_0^t f'(H_s)dH_s + \frac{1}{2}\int_0^t f''(H_s)\mathbb{1}\{H_s \notin \mathcal{E}\}(dH_s)^2 + \frac{1}{2}\sum_{y\in\mathcal{E}}(f'_r(y) - f'_l(y))L^H(t, y)$, where f'_r and f'_l are the right and left derivatives of f, and $L^H(t, y)$ is the symmetric local time of H_t at value y. Since f is continuously differentiable everywhere on $\mathbb{R}$, and since local time $L^H(t,y) < \infty$ almost surely by Lemma 6, we can conclude that $\sum_{y\in\mathcal{E}}(f_r'(y) - f_l'(y))L^H(t,y) = 0$ almost surely. Therefore, from the above equation and by definition of dH_t in Eq. (3), almost surely, $f(H_t) = f(H_0) + \int_0^t f'(H_s)(\mu_{u_s} - H_s)ds + \int_0^t f'(H_s)\sigma_{u_s}dB_s + \frac{1}{2}\int_0^t f''(H_s)\mathbb{1}\{H_s \notin \mathcal{E}\}\sigma_{u_s}^2 ds$. This completes the proof of Lemma 1. □

Towards the proof of Proposition 1, we start with the definition of a reflected O-U process for reader's reference. This is from Reed et al. [26], Definition 3.1 (we fix $\gamma = 1$ in that definition).

Definition 3 (Reflected O-U Process). *Let $B = (B_t : t \geq 0)$ be a standard Brownian motion, and let $\sigma > 0$, and $\theta \in \mathbb{R}$. We say that the process Z is a (σ, θ) reflected O-U process starting from $x \geq 0$, if the following four conditions are satisfied.*

1. $Z_t = x + \theta t - \int_0^t Z_s ds + \sigma B_t + L_t$ *for* $t \geq 0$,
2. $Z_t \geq 0$ *for* $t \geq 0$,
3. L *is non-decreasing with* $L_{0-} = 0$,
4. $\int_0^\infty \mathbb{1}\{Z_t > 0\} dL_t = 0$.

The following lemma will be used in the proof of Proposition 1.

Lemma 7. *Let X_t be a (σ_0, θ_0) reflected O-U process on $[0,\infty)$ starting from $x > 0$. Also define the unreflected process Y_t starting from $y \leq x$:*

$$Y_t = y + \theta_0 t - \int_0^t Y_s ds + \sigma_0 B_t. \tag{23}$$

Then, $X_t \geq^{\mathrm{st}} Y_t$ and $X_t \geq^{\mathrm{st}} Y_t^+$ for all $t \geq 0$, where $\geq^{\mathrm{st}}$ is the usual stochastic order and $Y_t^+ \triangleq$ $\max(Y_t, 0)$.

Proof of Lemma 7. If $X_t \geq^{\mathrm{st}} Y_t$ then $X_t \geq^{\mathrm{st}} Y_t^+$ follows easily since $X_t \geq 0$. Let $B = (B_t : t \geq 0)$ be a standard Brownian motion and let $X_t = x + \theta_0 t - \int_0^t X_s ds + \sigma_0 B_t + L_t$, $Y_t = y + \theta_0 t - \int_0^t Y_s ds + \sigma_0 B_t$, where L_t is the local time processes by Definition 3. It suffices to show $X_t \geq Y_t$ for all $t \geq 0$.

Suppose there exists $t > 0$ such that $X_t < Y_t$. Then by continuous paths there exists $0 \leq s < t$ such that $X_s = Y_s$, and $0 \leq X_u < Y_u$ for all $s < u < t$. By definition, we have $X_t - Y_t = X_s + \theta_0(t-s) - \int_s^t X_u du + \sigma_0(B_t - B_s) + (L_t - L_s) - Y_s - \theta_0(t-s) + \int_s^t Y_u du - \sigma_0(B_t - B_s) = (X_s - Y_s) + \int_s^t (Y_u - X_u)du + (L_t - L_s) > L_t - L_s$. Observe the last inequality follows from $X_s = Y_s$ and $Y_u > X_u \geq 0$ for all $s < u < t$. Also the last term is nonnegative since L_t is nondecreasing. This is a contradiction with $X_t < Y_t$. □

Now we are ready to prove Proposition 1.

Proof of Proposition 1. The time customer satisfaction spends in the unsatisfied zone prior to the customer's departure is exponentially distributed and has a finite expectation, given that the hazard rate of customer departure is 1 in the unsatisfied zone and zero in the satisfied zone. We want to show the time customer satisfaction spends in the satisfied zone also has finite expectation. Our approach will be to show that the long-run average ratio of time in satisfied zone to time in unsatisfied zone is finite (or zero).

Case (i): Policy π uses the Safe mode somewhere in $[\mu_S, q)$. First, consider the case where there the policy π uses the Safe mode at some satisfaction value in $[\mu_S, q)$, i.e., the set $\{x \in [\mu_S, q) : \pi(x) = S\}$ is nonempty. Define $l \triangleq \max\{x \in [\mu_S, q-\epsilon] : \pi(x) = 0\}$ where $\epsilon \in (0, (q-\mu_S)/2)$ is chosen to be small enough that $\{x \in [\mu_S, q-\epsilon] : \pi(x) = 0\}$ is nonempty. If the starting satisfaction $x \leq l$, then it is easy to see that the satisfaction level never rises above l, hence the customer is always unsatisfied and we are done. So suppose $x > l$. Let τ be the first time at which the satisfaction reaches l. We will show that $\mathbb{E}\tau < \infty$. This will complete our proof for this case, since the satisfaction will never rise above l thereafter.

It is easy to bound the total time spent in the interior of Safe intervals of policy π prior to τ, since the satisfaction process decays exponentially towards μ_S at such satisfaction values, and these satisfaction values exceed $q-\epsilon > \mu_S+\epsilon$ by definition of l. It remains to bound the total time spent traversing Risky intervals of π above l, including the time spent on delayed reflections at the upper boundary of such intervals (at the lower boundary of a Risky interval above l, the process enters a Safe interval and never again returns to the Risky interval). But this bound is also easy, since the satisfaction process inside each Risky interval is merely a (σ_R, μ_R) O-U process, possibly with a delayed reflecting upper boundary (if the interval has an upper boundary) b where the drift is $-(b-\mu_S)$. Any standard (unreflected) O-U process has first passage times with finite expectation (see Thomas [35]). Also the fraction of time a delayed reflected O-U process spends at its reflecting boundary is bounded away from one. Hence using the fact that the unreflected O-U process dominates the reflected O-U process as per Lemma 7, we conclude that the time for our reflected O-U process to reach the lower end of the Risky interval (or l, whichever is larger) has finite expectation. Hence, we have shown $\mathbb{E}\tau < \infty$.

Case (ii): Policy π uses the Risky mode everywhere in $[\mu_S, q)$. In this case, policy π has a Risky interval (a, b) that contains $[\mu_S, q)$. By the argument above, the time taken to enter this Risky interval has finite expectation. Having entered this interval, the satisfaction process remains within the closure of this interval, and is simply a (σ_R, μ_R) delayed reflected O-U process with drift $-(b-\mu_S) < 0$ at the upper boundary and drift $\mu_S - a > 0$ at the lower boundary. This process has a steady state distribution [38] with positive measure everywhere in (a, b) and atoms at a and b. Since $a < \mu_S \Rightarrow a < q$, it follows that the satisfaction process spends a positive fraction of its time in the unsatisfied zone in the long run. This completes our proof. □

6.2 Proofs for the Main Result

Proof of Lemma 2. Define $\alpha \triangleq \frac{\mu_R-\mu_S}{\sigma_R}, \beta \triangleq \frac{\mu_S}{\mu_R}, z \triangleq \frac{\theta-\mu_R}{\sigma_R}$. Then we have $\alpha > 0$ and $0 < \beta < 1$. Also define $f(z) \triangleq (z+\alpha)e^{z^2}\operatorname{erfc}(z) - \frac{\beta}{\sqrt{\pi}}$. Then our goal is to show that $r(z)$ has a unique root on $(-\alpha, \infty)$. This is true by property 5(ii) in Lemma 5. □

Proof of Lemma 3. Let $a \triangleq \frac{\mu_S-\mu_R}{\sigma_R}$, $b = \mu_S - \mu_R$, $\hat{q} \triangleq \frac{q-\mu_R}{\sigma_R}$, and $\hat{\theta} \triangleq \frac{\theta-\mu_R}{\sigma_R}$. Also define

$$F_{\text{small}}(\hat{\theta}) \triangleq 2\left(e^{\hat{q}^2-\hat{\theta}^2}\hat{q} - \hat{\theta} + e^{\hat{q}^2}\sqrt{\pi}\hat{q}\hat{\theta}\left(\operatorname{erf}(\hat{\theta}) - \operatorname{erf}(\hat{q})\right)\right) - \frac{\mu_S}{b(\hat{q}-a)}$$

and

$$F_{\text{big}}(\hat{\theta}) \triangleq \left(a - 3\hat{\theta} + 2a\hat{\theta}^2 - 2\hat{\theta}^3\right) + 2e^{\hat{q}^2-\hat{\theta}^2}\hat{q}\left(1 - a\hat{\theta} + \hat{\theta}^2\right)$$
$$+ e^{\hat{q}^2}\sqrt{\pi}\hat{q}\left(a - 3\hat{\theta} + 2a\hat{\theta}^2 - 2\hat{\theta}^3\right)\left(\operatorname{erf}(\hat{q}) - \operatorname{erf}(\hat{\theta})\right) - \frac{\mu_S\left(1 - a^2 + a\hat{\theta}\right)}{b(\hat{q}-a)}.$$

Observe that $F_{\text{small}}(\hat{\theta})$ corresponds to the LHS of Eq. (11), multiplied by a factor of 2. Also $F_{\text{big}}(\hat{\theta})$ corresponds to the LHS of Eq. (12), multiplied by a factor $\frac{q-\mu_R}{\sigma_R}$ ($q - \mu_R > 0$ since $q > \mu_S > \mu_R$). Then proving this lemma is equivalent to showing that exactly one of the following two cases is true: (1) $F_{\text{small}}(\hat{\theta})$ has a unique root on $[a, \hat{q})$ and $F_{\text{big}}(\hat{\theta})$ has no root on $\left(a - \frac{1}{a}, a\right)$; (2)$F_{\text{small}}(\hat{\theta})$ has no root on $[a, \hat{q})$ and $F_{\text{big}}(\hat{\theta})$ has a unique root on $\left(a - \frac{1}{a}, a\right)$.

First, observe that $F_{\text{small}}(\hat{q}) = -\frac{\mu_S}{b(\hat{q}-a)} < 0$. Also, apply Lemma 5 property 9, we have $F_{\text{big}}(a - \frac{1}{a}) = \frac{(a^2-2)\left(1+e^{\hat{q}^2}\sqrt{\pi}\hat{q}\left(\operatorname{erf}(\hat{q})-\operatorname{erf}\left(a-\frac{1}{a}\right)\right)\right)-2ae^{\hat{q}^2-\left(a-\frac{1}{a}\right)^2}\hat{q}}{-a^3} > 0$. Since $a > 0$, for $\hat{\theta} > a - \frac{1}{a}$, we have $1 - a^2 + a\hat{\theta} > 0$. Also, since $F_{\text{big}}(\cdot)$ is continuous at $a - \frac{1}{a}$, we have $\lim_{x \to (a-1/a)^+} \frac{F_{\text{big}}(\hat{\theta})}{1-a^2+a\hat{\theta}} > 0$ as well. Consider $F_{\text{small}}(a)$ and $F_{\text{big}}(a)$. In fact, one can check that $F_{\text{small}}(a) = F_{\text{big}}(a) = 2\left(e^{\hat{q}^2-a^2}\hat{q} - a + e^{\hat{q}^2}\sqrt{\pi}\hat{q}a\left(\operatorname{erf}(a) - \operatorname{erf}(\hat{q})\right)\right) - \frac{\mu_S}{b(\hat{q}-a)}$. Therefore, if we can show that both $F_{\text{small}}(\hat{\theta})$

and $\hat{F}_{\text{big}}(\hat{\theta}) \triangleq \frac{F_{\text{big}}(\hat{\theta})}{1-a^2+a\hat{\theta}}$ are strictly decreasing functions of $\hat{\theta}$ on $(a-\frac{1}{a},\hat{q})$, then we are done. In fact, one can check that $F''_{\text{small}}(\hat{\theta}) = 2\hat{q}e^{\hat{q}^2-\hat{\theta}^2} > 0$. Thus, $F'_{\text{small}}(\hat{\theta}) < F'_{\text{small}}(\hat{q}) = -1 < 0$ for $\hat{\theta} \in (a-\frac{1}{a},\hat{q})$, and hence $F_{\text{small}}(\hat{\theta})$ is strictly decreasing for $\hat{\theta} \in (a-\frac{1}{a},\hat{q})$.

It remains to show that $\hat{F}_{\text{big}}(\hat{\theta})$ is strictly decreasing in $\hat{\theta}$ for $\hat{\theta} \in (a-\frac{1}{a},\hat{q})$. We have $\hat{F}'_{\text{big}}(\hat{\theta}) = -\frac{e^{-\hat{\theta}^2}\left(3-2a^2+2a\hat{\theta}\right)}{\left(1-a^2+a\hat{\theta}\right)^2} \cdot f(\hat{\theta},a)$, where $f(\hat{\theta},a) = e^{\hat{\theta}^2}\left(1-2a\hat{\theta}+2\hat{\theta}^2\right)\left(1+e^{\hat{q}^2}\sqrt{\pi}\hat{q}\left(\text{erfc}(\hat{\theta})-\text{erfc}(\hat{q})\right)\right) + 2e^{\hat{q}^2}\hat{q}(a-\hat{\theta})$. We want to show that $\hat{F}'_{\text{big}}(\hat{\theta}) \le 0$ for $\hat{\theta} \in (a-\frac{1}{a},\hat{q})$. Observe that $3-2a^2+2a\hat{\theta} > 0$ since $\hat{\theta} > a-\frac{1}{a}$ and $a > 0$. Therefore, we only need to show $f(\hat{\theta},a) > 0$ for $\hat{\theta} \in (a-\frac{1}{a},\hat{q})$. In fact, $\frac{\partial f}{\partial x_2}(\hat{\theta},a) = 2e^{\hat{q}^2}\hat{q} - 2e^{\hat{\theta}^2}\hat{\theta} - 2e^{\hat{q}^2+\hat{\theta}^2}\hat{q}\hat{\theta}\sqrt{\pi}\left(\text{erfc}(\hat{\theta})-\text{erfc}(\hat{q})\right) \ge 0$ for $\hat{\theta} < \hat{q}$ by an application of Lemma 5 property 10. Thus, since $a > 0$, if we can show $f(\hat{\theta},0) > 0$ for $\hat{\theta} \in (a-\frac{1}{a},\hat{q})$, then we are done. Applying Lemma 5 property 11 we get $f(\hat{\theta},0) = -2e^{\hat{q}^2}\hat{q}\hat{\theta} + e^{\hat{\theta}^2}\left(1+2\hat{\theta}^2\right)\left(1+e^{\hat{q}^2}\hat{q}\sqrt{\pi}\left(\text{erfc}(\hat{\theta})-\text{erfc}(\hat{q})\right)\right) > 0$. □

Proof of Proposition 2. To show that a function $\bar{V}$ as described in Proposition 2 is the optimal value function V^*, we will first show that it is an upper bound for V^*, and then show that the bound is tight.

To show that $\bar{V}$ is an upper bound for V^*, it suffices to show $\bar{V}(x) \ge V(x,\hat{\pi})$ for $\forall x \in \mathbb{R}$ and for any admissible policy $\hat{\pi} \in \Pi$. Now fix any $x \in \mathbb{R}$ and any $\hat{\pi} \in \Pi$. Define a process $X_t, t \ge 0$ by

$$X_t = \bar{V}(H_t)e^{-\int_0^t Q(H_s)ds} + \int_0^t e^{-\int_0^s Q(H_z)dz}dY_s, \tag{24}$$

where H_t is the satisfaction process under policy $\hat{\pi}$ with $H_0 = x$, and Y_t the corresponding cumulative reward (conditional on no quitting) up to time t. Next we will rewrite X_t in integral form.

Since $\bar{V}$ is continuously differentiable everywhere and twice continuously differentiable almost everywhere except for a countable set $\mathcal{E}$ (Condition 2 in the proposition) and since H_t is a semimartingale under the admissible policy, we can apply the Itô-Tanaka formula to obtain that $\bar{V}(H_t) = \bar{V}(x) + \int_0^t \bar{V}'(H_s)(\mu_{u_s}-H_s)ds + \int_0^t \bar{V}'(H_s)\sigma_{u_s}dB_s + \frac{1}{2}\int_0^t \mathbb{1}\{H_s \notin \mathcal{E}\}\bar{V}''(H_s)\sigma_{u_s}^2 ds$ is also a semimartingale (see Eq. (10), denote u_t the corresponding action process under policy $\hat{\pi}$).

Since both $\bar{V}(H_t)$ and $e^{-\int_0^t Q(H_s)ds}$ are semimartingales (the latter follows from the fact that H_t is a semimartingale), we can then apply the multidimensional Itô formula on semimartingales to the function $g(\bar{V}(H_t), e^{-\int_0^t Q(H_s)ds}) = \bar{V}(H_t)e^{-\int_0^t Q(H_s)ds}$ and rewrite X_t as

$$\begin{aligned} X_t = \bar{V}(x) &+ \int_0^t e^{-\int_0^s Q(H_v)dv}\bar{V}'(H_s)(\mu_{u_s}-H_s)ds + \int_0^t e^{-\int_0^s Q(H_v)dv}\bar{V}'(H_s)\sigma_{u_s}dB_s \\ &+ \int_0^t \bar{V}(H_s)e^{-\int_0^s Q(H_v)dv}(-Q(H_s))ds + \int_0^t e^{-\int_0^s Q(H_v)dv}\frac{1}{2}\sigma_{u_s}^2\mathbb{1}\{H_s \notin \mathcal{E}\}\bar{V}''(H_s)ds \\ &+ \int_0^t e^{-\int_0^s Q(H_v)dv}\mu_{u_s}ds + \int_0^t e^{-\int_0^s Q(H_v)dv}\sigma_{u_s}dB_s. \end{aligned} \tag{25}$$

Since $\bar{V}'$ is bounded (Condition 3 in the proposition) and $\sigma_{u_s} \in \{0,\sigma_R\}$, the two stochastic integral terms in the above equation have bounded integrands hence have zero expectations. Now take expectations on both sides of Eq. (25) and write 1 as $\mathbb{1}\{H_s \notin \mathcal{E}\} + \mathbb{1}\{H_s \in \mathcal{E}\ \&\ u_s = S\} + \mathbb{1}\{H_s \in \mathcal{E}\ \&\ u_s = R\}$, we get:

$$\begin{aligned} \mathbb{E}X_t = \bar{V}(x) + \mathbb{E}\int_0^t e^{-\int_0^s Q(H_v)dv}\Bigg[&\left(-Q(H_s)\bar{V}(H_s) + \bar{V}'(H_s)(\mu_{u_s}-H_s) + \frac{1}{2}\sigma_{u_s}^2\bar{V}''(H_s) + \mu_{u_s}\right)\mathbb{1}\{H_s \notin \mathcal{E}\} \\ &+ \left(-Q(H_s)\bar{V}(H_s) + \bar{V}'(H_s)(\mu_S-H_s) + \mu_S\right)\mathbb{1}\{H_s \in \mathcal{E}\ \&\ u_s = S\} \\ &+ \left(-Q(H_s)\bar{V}(H_s) + \bar{V}'(H_s)(\mu_R-H_s) + \mu_R\right)\mathbb{1}\{H_s \in \mathcal{E}\ \&\ u_s = R\}\Bigg]ds \end{aligned}$$

$$\leq \bar{V}(x) + \mathbb{E}\int_0^t e^{-\int_0^s Q(H_v)dv}\Big(- Q(H_s)\bar{V}(H_s) + \bar{V}'(H_s)(\mu_R - H_s) + \mu_R\Big)\mathbb{1}\{H_s \in \mathcal{E} \ \& \ u_s = R\}ds$$
$$= \bar{V}(x). \tag{26}$$

The inequality results from a direct application of Condition 4. The last step follows from the fact that $-Q(H_s)\bar{V}(H_s) + \bar{V}'(H_s)(\mu_R - H_s) + \mu_R$ is bounded for $H_s \in \mathcal{E}$, and $\int_0^t \mathbb{1}\{H_s \in \mathcal{E} \ \& \ u_s = R\}ds = 0$ almost surely by Lemma 6. Since Eq. (26) holds for any $t \geq 0$, it also holds in the limit:

$$\limsup_{t\to\infty} \mathbb{E}X_t \leq \bar{V}(x). \tag{27}$$

We will now show $\limsup_{t\to\infty} \mathbb{E}X_t \geq V(x, \hat{\pi})$. Observe that since $\mu_R > 0$ and $\mu_S > 0$, the integral $\int_0^t e^{-\int_0^s Q(H_v)dv}\mu_{u_s}ds$ is pathwise monotone increasing in t and hence converges pathwise to $\int_0^\infty e^{-\int_0^s Q(H_v)dv}\mu_{u_s}ds$ as $t \to \infty$. It follows from Eq. (9) that

$$V(x, \hat{\pi}) = \mathbb{E}\left[\int_0^\infty e^{-\int_0^s Q(H_v)dv}\mu_{u_s}ds\right] = \lim_{t\to\infty}\mathbb{E}\left[\int_0^t e^{-\int_0^s Q(H_v)dv}dY_s\right]. \tag{28}$$

Take expectations on both sides of Eq. (24) and let $t \to \infty$, we get

$$\limsup_{t\to\infty} \mathbb{E}X_t = \limsup_{t\to\infty} \mathbb{E}\left[\bar{V}(H_t)e^{-\int_0^t Q(H_s)ds}\right] + V(x, \hat{\pi}) \geq V(x, \hat{\pi}), \tag{29}$$

where we used Eq. (28) and that $\bar{V}$ in non-negative (Condition 1 in the proposition). Combining Eqs. (27) and (29), we obtain the desired result $\bar{V}(x) \geq V(x, \hat{\pi})$ for $\forall x \in \mathbb{R}$ and any admissible policy $\hat{\pi} \in \Pi$. That is, $\bar{V}$ is an upperbound of the optimal value function V^*.

Now it remains to show this upper bound is tight, i.e., the inequalities in Eqs. (27) and (29) are binding under policy $\bar{\pi}$ ($\bar{\pi}$ defined in Condition 5). By Condition 5 of the proposition, the inequality in Eq. (26) is binding under policy $\bar{\pi}$ (where $u_t = \pi(H_t)$). Hence its limit, Eq. (27) is also binding under policy $\bar{\pi}$. On the other hand, by Condition 3 and 6 of the proposition, (Condition 3 implies that $\bar{V}(x) \leq K(1 + |x|)$ for some $K \leq \infty$) we have $\lim_{t\to\infty}\mathbb{E}\left[\bar{V}(H_t)e^{-\int_0^t Q(H_s)ds}\right] + V(x, \bar{\pi}) \leq \lim_{t\to\infty}\mathbb{E}\left[K(1 + |H_t|)e^{-\int_0^t Q(H_s)ds}\right] + V(x, \bar{\pi}) = V(x, \bar{\pi})$. Combining this inequality with Eq. (29), we have that the inequality in Eq. (29) is binding. Therefore we have proved that $\bar{V}$ is the value function V^* and that $\bar{\pi}$ is an optimal policy. □

Proof of Lemma 4. We want to show $\lim_{t\to\infty}\mathbb{E}\left[(1 + |H_t|)e^{-\int_0^t Q(H_s)ds}\right] = 0$ for the described policies. This is equivalent to showing $\lim_{t\to\infty}\mathbb{E}\left[e^{-\int_0^t Q(H_s)ds}\right] = 0$ and $\lim_{t\to\infty}\mathbb{E}\left[|H_t|e^{-\int_0^t Q(H_s)ds}\right] = 0$. By the Cauchy-Schwarz inequality, we have $\lim_{t\to\infty}\mathbb{E}\left[|H_t|e^{-\int_0^t Q(H_s)ds}\right] \leq \lim_{t\to\infty}\sqrt{\mathbb{E}[H_t^2]\mathbb{E}\left[e^{-2\int_0^t Q(H_s)ds}\right]}$. Hence it suffices to show

$$\lim_{t\to\infty}\mathbb{E}\left[e^{-\int_0^t Q(H_s)ds}\right] = 0, \tag{30}$$

$$\limsup_{t\to\infty}\mathbb{E}[H_t^2] < \infty, \tag{31}$$

and

$$\limsup_{t\to\infty}\mathbb{E}\left[e^{-2\int_0^t Q(H_s)ds}\right] = 0. \tag{32}$$

First consider Eqs. (30) and (32). Since (Eq. (5)) $P(T > t|\mathcal{F}_t) = e^{-\int_0^t Q(H_s)ds}$, we have LHS of Eq. (30) equivalent to $\lim_{t\to\infty} P(T > t) = P(T > \infty)$, and LHS of Eq. (32) is $\limsup_{t\to\infty}\mathbb{E}\left[e^{-2\int_0^t Q(H_s)ds}\right] = \limsup_{t\to\infty}\mathbb{E}\left[P(T > t|\mathcal{F}_t)^2\right] = \limsup_{t\to\infty} P(T > t)^2 = P(T > \infty)^2$, both of which are zero since by Proposition 1, $\mathbb{E}T < \infty$ under interval policies with finite intervals.

It only remains to show Eq. (31). Fix any starting satisfaction value $x \in \mathbb{R}$ and any interval policy with finite intervals. Denote by u_t the corresponding action process. By definition of H_t in

Eq. (2), we have $H_t = xe^{-t} + \int_0^t e^{-(t-s)}\mu_{u_s} ds + \int_0^t e^{-(t-s)}\sigma_{u_s} dB_s$, which gives

$$H_t^2 = x^2 e^{-2t} + \int_0^t e^{-(t-s)}\mu_{u_s}\int_0^t e^{-(t-v)}\mu_{u_v} dv ds + \int_0^t e^{-2(t-s)}\sigma_{u_s}^2 ds + 2xe^{-t}\int_0^t e^{-(t-s)}\mu_{u_s} ds$$
$$+ 2xe^{-t}\int_0^t e^{-(t-s)}\sigma_{u_s} dB_s + 2\int_0^t e^{-(t-s)}\mu_{u_s}\int_0^t e^{-(t-s)}\sigma_{u_s} dB_s ds.$$

Take expectation of H_t^2, we can remove both stochastic integral terms since the their integrands are both bounded. Note that $\mu_{u_s} \le \mu \triangleq \max\{\mu_S, \mu_R\}$ and $\sigma_{u_s} \le \sigma_R$ for any $t \ge 0$. Therefore we have

$$\mathbb{E}\left[H_t^2\right] \le x^2 e^{-2t} + \int_0^t e^{-(t-s)}\mu\int_0^t e^{-(t-v)}\mu dv ds + \int_0^t e^{-2(t-s)}\sigma_R^2 ds + 2xe^{-t}\int_0^t e^{-(t-s)}\mu ds$$
$$= x^2 e^{-2t} + \mu^2(e^{-2t} + 1 - 2e^{-t}) + \frac{\sigma_R^2(1-e^{-2t})}{2} + 2x\mu(e^{-t} - e^{-2t}).$$

Let $t \to \infty$ in this equation, we get $\limsup_{t\to\infty} \mathbb{E}\left[H_t^2\right] \le \mu^2 + \frac{\sigma_R^2}{2} < \infty$, and hence Eq. (31) is true. Therefore we have proved Lemma 4. $\square$

Proof of Proposition 3. The policy associated with the W function belongs to the policy types in Lemma 4. Therefore by Lemma 4, Condition 6 of Proposition 2 is satisfied. Also one can check that for any C_1, C_2, C_3, C_4, C_5 in $\mathbb{R}$, $W(\cdot, C_1, C_2, C_3, C_4, C_5)$ and the corresponding policy satisfies Condition 5. Moreover $W(\cdot, C_1, C_2, C_3, C_4, C_5)$ is twice continuously differentiable everywhere except possibly at θ_G, q and θ_b. Therefore to prove Proposition 3 we only need to show Conditions 1–4 are satisfied, where Condition 2 only concerns continuous differentiability at θ_G, q and θ_b.

First consider the case $\mu_R \ge \mu_S$. By the definition of θ_G in Lemma 3, we know that when $\mu_R \ge \mu_S$, $\theta_G = -\infty$. Therefore the function W as defined in Proposition 3 reduces to:

$$W(x, C_2, C_3, C_4, C_5) = \begin{cases} V_2(x, C_2, C_3) & \text{if } x < q; \\ V_3(x, C_4) & \text{if } q \le x \le \theta_{\rm b}; \\ V_4(x, C_5) & \text{if } x > \max\{q, \theta_{\rm b}\}, \end{cases} \tag{33}$$

where θ_b is as defined in Lemma 2 and $V_2(x, C_2, C_3), V_3(x, C_4), V_4(x, C_5)$ as defined in the statement of Proposition 3. For easier writing, we introduce the following notations:

$$B_i(x, V) \triangleq -Q(x)V(x) + (\mu_i - x)V'(x) + \frac{1}{2}\sigma_i^2 \mathbb{1}\{x \notin \mathcal{E}\}V''(x) + \mu_i, \quad \text{for } i = R, S; \tag{34}$$

$$\delta(x, V) \triangleq B_R(x, V) - B_S(x, V) = (\mu_R - \mu_S)V'(x) + \frac{1}{2}\sigma_R^2 \mathbb{1}\{x \notin \mathcal{E}\}V''(x) + \mu_R - \mu_S, \tag{35}$$

where $\mathcal{E}$ is the set of values at which V'' doesn't exist. We want to find $C_2^*, C_3^*, C_4^*, C_5^*$ such that $W^*(\cdot) \triangleq W(\cdot, C_2^*, C_3^*, C_4^*, C_5^*)$ satisfy the following statements:

Statement 1. $W^*(\cdot)$ *is non-negative;*

Statement 2. $W^*(\cdot)$ *is continuously differentiable at* q *and* θ_b*;*

Statement 3. $W^*(\cdot)$ *has bounded first derivative;*

Statement 4. $\delta(x, V_2(x, C_2^*, C_2^*)) \ge 0$ *for any* $x \in (-\infty, q)$*,*

Statement 5. $\delta(x, V_3(x, C_4^*)) \le 0$ *for any* $x \in (q, \max\{q, \theta_b\})$*,*

Statement 6. $B_S(x, V_4(x, C_5^*)) \le 0$ *for any* $x \in (\max\{q, \theta_b\}, \infty)$*.*

Statements 1–3 correspond to Conditions 1–3, respectively. Also Statements 4–6 correspond to Condition 4. We utilize Statement 2 to find the values of C_2^*, C_3^*, C_4^* and C_5^*. Observe that the first derivatives of $V_3(\cdot, C_4)$ and $V_4(\cdot, C_5)$ does not depend on C_4 and C_5. We therefore denote them $V_3'(\cdot)$ and $V_4'(\cdot)$, respectively. Define

$$C_3^* \triangleq C_2^* \triangleq \begin{cases} \dfrac{V_4'(q)}{\frac{2}{\sigma_R}\left(\frac{1}{\sqrt{\pi}}+e^{\frac{(q-\mu_R)^2}{\sigma_R^2}}\frac{q-\mu_R}{\sigma_R}\left(1+\text{erf}\left(\frac{q-\mu_R}{\sigma_R}\right)\right)\right)} & \text{if } \theta_b < q \\ \dfrac{V_3'(q)}{\frac{2}{\sigma_R}\left(\frac{1}{\sqrt{\pi}}+e^{\frac{(q-\mu_R)^2}{\sigma_R^2}}\frac{q-\mu_R}{\sigma_R}\left(1+\text{erf}\left(\frac{q-\mu_R}{\sigma_R}\right)\right)\right)} & \text{if } \theta_b \geq q \end{cases}, \tag{36}$$

$$C_4^* \triangleq C_2^* e^{\frac{(q-\mu_R)^2}{\sigma_R^2}}\left(1+\text{erf}\left(\frac{q-\mu_R}{\sigma_R}\right)\right)+\mu_R-\mu_S\log(q-\mu_S), \tag{37}$$

and

$$C_5^* \triangleq \begin{cases} V_2(q, C_2^*, C_2^*) - \int_0^q \frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(z-\mu_R)^2}{\sigma_R^2}}\left(1-\text{erf}\left(\frac{z-\mu_R}{\sigma_R}\right)\right)dz & \text{if } \theta_b < q \\ V_3(\theta_b, C_4^*) - \int_0^{\theta_b} \frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(z-\mu_R)^2}{\sigma_R^2}}\left(1-\text{erf}\left(\frac{z-\mu_R}{\sigma_R}\right)\right)dz & \text{if } \theta_b \geq q \end{cases}. \tag{38}$$

We define $V_2^*(\cdot) \triangleq V_2(\cdot, C_2^*, C_2^*)$, $V_3^*(\cdot) \triangleq V_3(\cdot, C_4^*)$, $V_4^*(\cdot) \triangleq V_4(\cdot, C_5^*)$ and $W^*(\cdot) \triangleq W(\cdot, C_2^*, C_2^*, C_4^*, C_5^*)$. It's easy to verify that Statement 2 is satisfied, that is, $V_2^*(q) = V_4^*(q), {V_2^*}'(q) = {V_4^*}'(q)$ if $\theta_b < q$, and $V_2^*(q) = V_3^*(q), {V_2^*}'(q) = {V_3^*}'(q), V_3^*(\theta_b) = V_4^*(\theta_b), {V_3^*}'(\theta_b) = {V_4^*}'(\theta_b)$ if $\theta_b < q$.

Now consider Statement 3. Since we have just proved Statement 2 that ${W^*}'(\cdot)$ is continuous in $\mathbb{R}$, to show that ${W^*}'(\cdot)$ is bounded, it suffices to show $\left|\lim_{x\to-\infty} {W^*}'(x)\right| < \infty$ and $\left|\lim_{x\to\infty} {W^*}'(x)\right| < \infty$. This is equivalent to showing $\left|\lim_{x\to-\infty} {V_2^*}'(x)\right| < \infty$ and $\left|\lim_{x\to\infty} {V_4^*}'(x)\right| < \infty$. By definition of V_2^* and V_4^*, we have $\lim_{x\to-\infty} {V_2^*}'(x) = \lim_{x\to-\infty} \frac{2C_2^*}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} + e^{\frac{(x-\mu_R)^2}{\sigma_R^2}}\frac{x-\mu_R}{\sigma_R}\left(1+\text{erf}\left(\frac{x-\mu_R}{\sigma_R}\right)\right)\right) = \lim_{z\to\infty}\frac{2C_2^*}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} - ze^{z^2}\text{erfc}(z)\right)$ and $\lim_{x\to\infty} {V_4^*}'(x) = \lim_{x\to\infty} \frac{\mu_R\sqrt{\pi}}{\sigma_R}e^{\frac{(x-\mu_R)^2}{\sigma_R^2}}\left(1-\text{erf}\left(\frac{x-\mu_R}{\sigma_R}\right)\right) = \lim_{z\to\infty}\frac{\mu_R\sqrt{\pi}}{\sigma_R}e^{z^2}\text{erfc}(z)$. From Lemma 2, both limits are 0.

It only remains to prove Statements 1, 4–6. We only present the proof of Statement 4 here to demonstrate the use of Lemma 5 and leave the proofs of Statement 1, 5–6, which are similar to the proof of Statement 4, in Online Appendix B.

Proof of Statement 4. For $x \leq q$, substitute $V_2(x, C_2^*, C_2^*)$ (see Proposition 3 for V_2 and Eq. (36) for C_2^*) into $\delta(x, V_2)$ (see Eq. (35) for $\delta(x, V)$), we get $\delta(x, V_2^*) = (\mu_R - \mu_S){V_2^*}'(x) + \frac{1}{2}\sigma_R^2 {V_2^*}''(x) + \mu_R - \mu_S$. Want to show $\delta(x, V_2^*) \geq 0$ for any $x < q$. Compute ${V_2^*}'(x)$ and ${V_2^*}''(x)$:

$${V_2^*}'(x) = \frac{2C_2^*}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} + e^{\frac{(x-\mu_R)^2}{\sigma_R^2}}\frac{x-\mu_R}{\sigma_R}\text{erfc}\left(-\frac{x-\mu_R}{\sigma_R}\right)\right);$$

$${V_2^*}''(x) = \frac{2C_2^*}{\sigma_R^2}\left(\frac{2(x-\mu_R)}{\sigma_R\sqrt{\pi}} + e^{\frac{(x-\mu_R)^2}{\sigma_R^2}}\left(1+\frac{2(x-\mu_R)^2}{\sigma_R^2}\right)\text{erfc}\left(-\frac{x-\mu_R}{\sigma_R}\right)\right).$$

By property 1, and property 5(iii) in Lemma 5, the terms in both parenthesis above are non-negative. Therefore, if we can show $C_2^* > 0$, then both ${V_2^*}'(x)$ and ${V_2^*}''(x)$ are non-negative, which implies (since $\mu_R \geq \mu_S$) $\delta(x, V_2^*) \geq 0$ for $x \leq q$. The next claim establishes this result.

Claim 1. *The value C_2^* as defined in Eq.* (36) *is strictly positive.*

Proof. By Eq. (36) and Eqs. (20)–(21), we have

$$C_2^* = \begin{cases} \dfrac{\frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} \left(1-\mathrm{erf}\left(\frac{q-\mu_R}{\sigma_R}\right)\right)}{\frac{2}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} - e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} \frac{-q+\mu_R}{\sigma_R}\left(1-\mathrm{erf}\left(\frac{-q+\mu_R}{\sigma_R}\right)\right)\right)} & \text{if } \theta_b < q; \\ \dfrac{\frac{\mu_S}{q-\mu_S}}{\frac{2}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} - e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} \frac{-q+\mu_R}{\sigma_R}\left(1-\mathrm{erf}\left(\frac{-q+\mu_R}{\sigma_R}\right)\right)\right)} & \text{if } \theta_b \geq q. \end{cases}$$

Since $q > \mu_S$, the numerators in both cases are positive. In fact the denominator, which is the same for both cases, is also positive by an application of property 5(iii) in Lemma 5. Hence $C_2^* > 0$. □

□

Next we prove the proposition for the other scheme, $\mu_R < \mu_S$. By the definition of θ_b in Lemma 2, we know that when $\mu_R < \mu_S$, $\theta_b = \infty$. Therefore in this case, the function W as defined in Proposition 3 reduces to:

$$W(x, C_1, C_2, C_3, C_4) = \begin{cases} V_1(x, C_1) & \text{if } x \leq \theta_G \\ V_2(x, C_2, C_3) & \text{if } \theta_G < x < q; \\ V_3(x, C_4) & \text{if } x \geq q, \end{cases}$$

where $V_1(x, C_1), V_2(x, C_2, C_3), V_3(x, C_4)$ are as defined in the statement of Proposition 3. We need to show that there exists $C_1^G, C_2^G, C_3^G, C_4^G$ such that $W^G(\cdot) \triangleq W(\cdot, C_1^G, C_2^G, C_3^G, C_4^G)$ satisfy the following statements:

Statement 7. *$W^G(\cdot)$ is non-negative;*

Statement 8. *$W^G(\cdot)$ is continuously differentiable at q and θ_G;*

Statement 9. *$W^G(\cdot)$ has bounded first derivative;*

Statement 10. *$B_R(x, V_1(x, C_1^G)) \leq 0$ for any $x < \theta_G$;*

Statement 11. *$B_S(x, V_2(x, C_2^G, C_3^G)) \leq 0$ for any $x \in (\theta_G, q)$;*

Statement 12. *$B_R(x, V_3(x, C_4)) \leq 0$ for any $x > q$.*

Statements 7–9 correspond to Conditions 1–3, respectively. Also Statements 10–12 completes Condition 4. We utilize Statement 8 to find C_1^G, C_2^G, C_3^G and C_5^G. To shorten the expressions a bit, we introduce the following notations:

$$a \triangleq \frac{\mu_S - \mu_R}{\sigma_R}, \quad b \triangleq \mu_S - \mu_R, \quad \hat{\theta}_G \triangleq \frac{\theta_G - \mu_R}{\sigma_R}.$$

Define

$$C_1^G = \begin{cases} 0 & \text{if } \theta_G \geq \mu_S; \\ \frac{-b^2(a-\hat{\theta}_G)^3}{a(1-a^2+a\hat{\theta}_G)} & \text{if } \theta_G < \mu_S, \end{cases} \tag{39}$$

$$C_2^G = \begin{cases} be^{-\hat{\theta}_G^2} + \sqrt{\pi} b\hat{\theta}_G \mathrm{erf}(\hat{\theta}_G) & \text{if } \theta_G \geq \mu_S; \\ \frac{be^{-\hat{\theta}_G^2}\left(2-2a\hat{\theta}_G+2\hat{\theta}_G^2+e^{\hat{\theta}_G^2}\sqrt{\pi}\left(-a+3\hat{\theta}_G-2a\hat{\theta}_G^2+2\hat{\theta}_G^3\right)\mathrm{erf}(\hat{\theta}_G)\right)}{2(1-a^2+a\hat{\theta}_G)} & \text{if } \theta_G < \mu_S, \end{cases} \tag{40}$$

$$C_3^G = \begin{cases} -b\sqrt{\pi}\hat{\theta}_G & \text{if } \theta_G \geq \mu_S; \\ \frac{b\sqrt{\pi}\left(a-3\hat{\theta}_G+2a\hat{\theta}_G^2-2\theta_G^3\right)}{2(1-a^2+a\hat{\theta}_G)} & \text{if } \theta_G < \mu_S, \end{cases} \tag{41}$$

and

$$C_4^G = V_2(q, C_2^G, C_3^G) - \mu_S \log(q - \mu_S). \tag{42}$$

Also define $V_1^G(\cdot) \triangleq V_1(\cdot, C_1^G)$, $V_2^G(\cdot) \triangleq V_2(\cdot, C_2^G, C_3^G)$ and $V_3^G(\cdot) \triangleq V_3(\cdot, C_4^G)$. It's easy to verify that Statement 8 is satisfied, i.e., $V_2^G(q) = V_3^G(q), V_1^G(\theta_G) = V_2^G(\theta_G), V_1^{G'}(\theta_G) = V_2^{G'}(\theta_G)$, and $V_2^{G'}(q) = V_3^{G'}(q)$.

Next consider Statement 9, i.e., want to show $W^{G'}$ is bounded everywhere. By Statement 8, $W^{G'}$ exists and is continuous everywhere. Therefore to show $W^{G'}$ is bounded everywhere, it suffices to show $\lim_{x\to\infty} W^{G'}(x)$ and $\lim_{x\to-\infty} W^{G'}(x)$ are bounded. In fact, $\lim_{x\to\infty} W^{G'}(x) = \lim_{x\to\infty} \frac{\mu_S}{x-\mu_S} = 0$ and $\lim_{x\to-\infty} W^{G'}(x) = \lim_{x\to-\infty} -\frac{C_1^G}{(x-\mu_S)^2} = 0$, since C_1^G is a constant. Hence, we have proved Statement 9.

It only remains to prove Statements 7, 10–12. The proofs are similar to the proof of Statement 4 and are relegated to Online Appendix B. This completes the proof of Proposition 3. □

Proof of Theorem 2. We first prove the monotonicity results when $\mu_R > \mu_S$. Consider the parts regarding θ_b. Define function

$$F(\mu_S, \mu_R, \sigma_R, x) \triangleq \frac{(x-\mu_S)}{\sigma_R} e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \left(1 - \operatorname{erf}\left(\frac{x-\mu_R}{\sigma_R}\right)\right) - \frac{\mu_S}{\mu_R\sqrt{\pi}}.$$

By Lemma 2, $F(\mu_S, \mu_R, \sigma_R, \theta_b) = 0$. Moreover, by evaluating the function $r(\cdot)$ from property 5(ii) in Lemma 5 at $(x-\mu_R)/\sigma_R$ with parameters $\alpha = (\mu_R - \mu_S)/\sigma_R$ and $\beta = \mu_S/\mu_R$, we obtain that $F(\mu_S, \mu_R, \sigma_R, x) < 0$ for all $x \in (\mu_S, \theta_b)$, and $F(\mu_S, \mu_R, \sigma_R, x) > 0$ for all $x \in (\theta_b, \infty)$. Hence to prove θ_b is strictly decreasing in μ_R, it suffices to show that $\frac{\partial F}{\partial \mu_R}(\mu_S, \mu_R, \sigma_R, x) > 0$ for all $x > \mu_S$. Similarly, to show θ_b is strictly increasing in μ_S and σ_R, it suffices to show $\frac{\partial F}{\partial \mu_S}(\mu_S, \mu_R, \sigma_R, x) < 0$ for all $x > \mu_S$ and $\frac{\partial F}{\partial \sigma_R}(\mu_S, \mu_R, \sigma_R, x) < 0$ for all $x > \mu_S$. We have the partial derivatives:

$$\frac{\partial F}{\partial \mu_R}(\mu_S, \mu_R, \sigma_R, x) = \frac{2(x-\mu_S)}{\sigma_R}\left[-\frac{x-\mu_R}{\sigma_R} e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \operatorname{erfc}\left(\frac{x-\mu_R}{\sigma_R}\right) + \frac{1}{\sqrt{\pi}}\right] + \frac{\mu_S}{\mu_R^2\sqrt{\pi}};$$

$$\frac{\partial F}{\partial \mu_S}(\mu_S, \mu_R, \sigma_R, x) = -\frac{1}{\sigma_R} e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \operatorname{erfc}\left(\frac{x-\mu_R}{\sigma_R}\right) - \frac{1}{\sqrt{\pi}\mu_R};$$

$$\frac{\partial F}{\partial \sigma_R}(\mu_S, \mu_R, \sigma_R, x) = -\frac{x-\mu_S}{\sigma_R}\left[e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \operatorname{erfc}\left(\frac{x-\mu_R}{\sigma_R}\right)\left(1 + 2\frac{(x-\mu_R)^2}{\sigma_R^2}\right) - \frac{2\frac{x-\mu_R}{\sigma_R}}{\sqrt{\pi}}\right].$$

Fix any $x > \mu_S$. Apply change of variable $\hat{x} \triangleq \frac{x-\mu_R}{\sigma_R}$. By property 5(iii) in Lemma 5, we get $\frac{\partial F}{\partial \mu_R}(\mu_S, \mu_R, \sigma_R, x) > 0$. Also $\frac{\partial F}{\partial \mu_S}(\mu_S, \mu_R, \sigma_R, x) < 0$ since $\operatorname{erfc}(\cdot) \geq 0$. Finally, property 1 in Lemma 5 implies $\frac{\partial F}{\partial \sigma_R}(\mu_S, \mu_R, \sigma_R, x) < 0$. This completes the proof of monotonicity regarding θ_b.

Next, we show monotonicity results of the value function and its first order derivative. We already know the functional form of the value function. Hence, we can check the partial derivatives with regard to model parameters directly. Consider the monotonicity of V^* with regard to μ_S. Consider two possible valid values for the Safe mode drift $\mu_S^L < \mu_S^H$. Let θ_b^L and θ_b^H be the cutoffs defined in Lemma 2 for μ_S^L and μ_S^H, respectively. Also, denote the corresponding value functions by V_L and V_H. We have just proved that θ_b is strictly increasing in μ_S, hence $\theta_b^L < \theta_b^H$. There are three cases based on the location of q: (1) $\theta_b^L < \theta_b^H < q$, (2) $\theta_b^L < q \leq \theta_b^H$, (3) $q \leq \theta_b^L < \theta_b^H$. In the first case, the myopic policy is optimal for both μ_S^L and μ_S^H, hence the value functions V_L and V_H are identical. In the second case, V_L is the value function for the myopic policy regardless of the value of the Safe mode drift. Hence, $V_L \leq V_H$ by the optimality of V_H under Safe mode drift μ_S^H. We now consider the last case, $q \leq \theta_b^L < \theta_b^H$. It is easy to verify that for any $x \geq q$, $\frac{\mu_S^L}{x-\mu_S^L} < \frac{\mu_S^H}{x-\mu_S^H}$, i.e., $V_L'(x) < V_H'(x)$ on $[q, \theta_b^L]$. In particular, $V_L'(q) < V_H'(q)$. As a result, $V_L(x) < V_H(x)$ for $x < q$ (see Eq. (19)). By continuity, $V_L(q) < V_H(q)$. Therefore $V_L(x) < V_H(x)$ (see Eq. (20)) for $q \leq x \leq \theta_b^L$ since $V_L'(x) < V_H'(x)$ for $q \leq x \leq \theta_b^L$ and $V_L(q) < V_H(q)$. From Lemma 2 and property 5(ii) in Lemma 5, we know that $\frac{\mu_S^H}{x-\mu_S^H} \geq \frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \operatorname{erfc}\left(\frac{x-\mu_R}{\sigma_R}\right)$ for $q \leq x \leq \theta_b^H$. Hence $V_L'(x) \leq V_H'(x)$

(see Eqs. (20) and (21)) for $\theta_b^L < x \le \theta_b^H$, which implies $V_L(x) < V_H(x)$ for $\theta_b^L < x \le \theta_b^H$. Finally, for $x > \theta_b^H$, we have $V_L'(x) = V_H'(x)$, therefore $V_L(x) < V_H(x)$ since $V_L(\theta_b^H) < V_H(\theta_b^H)$.

We have just proved that the value function is increasing in μ_S. Next, we will similarly prove the monotonicity of the value function in μ_R. Consider two valid values for the Risky mode drift $\mu_R^L < \mu_R^H$. Let θ_b^L and θ_b^H be as defined in Lemma 2 for μ_R^L and μ_R^H, respectively, and denote V_L and V_H as the corresponding value functions. We have proved that θ_b is strictly decreasing in μ_R, hence $\theta_b^H < \theta_b^L$. Again, there are three cases based on the location of q: (1) $\theta_b^H < \theta_b^L < q$, (2) $\theta_b^H < q \le \theta_b^L$, (3) $q \le \theta_b^H < \theta_b^L$. First consider case (1). In both the high Risky reward and low Risky reward scenarios, the myopic policy is optimal. Apply property 5(iv) in Lemma 5 to $V_L'(x)$ (see Eq. (21)) for $x \ge q$, we obtain that $V_L'(x) < V_H'(x)$ for $x \ge q$. Next we will show that $V_L'(x) < V_H'(x)$ for $x < q$, so that since $\lim_{x\to-\infty} V_L(x) = \mu_R^L < \mu_R^H = \lim_{x\to-\infty} V_H(x)$, we can get $V_L(x) < V_H(x)$ on $\mathbb{R}$.

The explicit expression for $V_L'(x)$ and $V_H'(x)$ on $(-\infty, q)$ is given by

$$V_L'(x) = \frac{\sqrt{\pi}\mu_R^L e^{\hat{q}_L^2}\mathrm{erfc}(\hat{q}_L)\left(e^{y_L^2}y_L\mathrm{erfc}(-y_L)+\frac{1}{\sqrt{\pi}}\right)}{\sigma_R\left(e^{\hat{q}_L^2}\hat{q}_L\mathrm{erfc}(-\hat{q}_L)+\frac{1}{\sqrt{\pi}}\right)}$$

and

$$V_H'(x) = \frac{\sqrt{\pi}\mu_R^H e^{\hat{q}_L^2}\mathrm{erfc}(\hat{q}_H)\left(e^{y_H^2}y_H\mathrm{erfc}(-y_H)+\frac{1}{\sqrt{\pi}}\right)}{\sigma_R\left(e^{\hat{q}_H^2}\hat{q}_H\mathrm{erfc}(-\hat{q}_H)+\frac{1}{\sqrt{\pi}}\right)},$$

where $\hat{q}_L \triangleq \frac{q-\mu_R^L}{\sigma_R}$, $y_L \triangleq \frac{x-\mu_R^L}{\sigma_R}$, $\hat{q}_H \triangleq \frac{q-\mu_R^H}{\sigma_R}$ and $y_H \triangleq \frac{x-\mu_R^H}{\sigma_R}$. Observe that $\hat{q}_L > \hat{q}_H$ and $y_L > y_H$. Since $\mu_R^L < \mu_R^H$ and $e^{\hat{q}_L^2}\mathrm{erfc}(\hat{q}^L) < e^{\hat{q}_H^2}\mathrm{erfc}(\hat{q}^H)$ by property (iv) in Lemma 5, to show $V_L'(x) < V_H'(x)$ for $x < q$, it is sufficient if we can show $\frac{e^{y_L^2}y_L\mathrm{erfc}(-y_L)+\frac{1}{\sqrt{\pi}}}{e^{\hat{q}_L^2}\hat{q}_L\mathrm{erfc}(-\hat{q}_L)+\frac{1}{\sqrt{\pi}}} < \frac{e^{y_H^2}y_H\mathrm{erfc}(-y_H)+\frac{1}{\sqrt{\pi}}}{e^{\hat{q}_H^2}\hat{q}_H\mathrm{erfc}(-\hat{q}_H)+\frac{1}{\sqrt{\pi}}}$, or equivalently, $\log\left(e^{y_L^2}y_L\mathrm{erfc}(-y_L)+\frac{1}{\sqrt{\pi}}\right) - \log\left(e^{\hat{q}_L^2}\hat{q}_L\mathrm{erfc}(-\hat{q}_L)+\frac{1}{\sqrt{\pi}}\right)$ is increasing in μ_R^L. Define $G(y_L) \triangleq \log\left(e^{y_L^2}y_L\mathrm{erfc}(-y_L)+\frac{1}{\sqrt{\pi}}\right)$. By the chain rule, the derivative of $\log\left(e^{y_L^2}y_L\mathrm{erfc}(-y_L)+\frac{1}{\sqrt{\pi}}\right)$ with respect to μ_R^L is $-\frac{G'(y_L)}{\sigma_R}$, which is decreasing in y_L by property 8 in Lemma 5. Therefore, since $y_L < \hat{q}_L$, the derivative of $\log\left(e^{y_L^2}y_L\mathrm{erfc}(-y_L)+\frac{1}{\sqrt{\pi}}\right) - \log\left(e^{\hat{q}_L^2}\hat{q}_L\mathrm{erfc}(-\hat{q}_L)+\frac{1}{\sqrt{\pi}}\right)$ with respect to μ_R^L is non-negative.

Now consider case (2). Here, the myopic policy is optimal in the low Risky reward scenario but suboptimal in the high Risky reward scenario. We just showed that the value function corresponding to the myopic policy in the low Risky reward scenario, V_L, is lower than the value function corresponding to the myopic policy in the high Risky reward scenario. By optimality, the latter must be lower than the optimal value function in the high Risky reward scenario, V_H. Hence, $V_L \le V_H$.

Now consider case (3), where for both low Risky reward scenario and high Risky reward scenarios, sandwich policies are optimal. By property 5(iv) in Lemma 5, we have $V_L'(x) < V_H'(x)$ for $x > \theta_b^L$. Also $V_L'(x) = V_H'(x)$ on $[q, \theta_b^H]$. We only need to consider $x < q$ and $\theta_b^H < x \le \theta_b^L$. For the latter, we can refer to the definition of θ_b^L in Lemma 2 and property 5(ii) in Lemma 5 to get $\frac{\mu_S}{x-\mu_S} \le \frac{\mu_R^H\sqrt{\pi}}{\sigma_R}e^{\frac{(x-\mu_R^H)^2}{\sigma_R^2}}\mathrm{erfc}\left(\frac{x-\mu_R^H}{\sigma_R}\right)$ for $x \ge \theta_b^H$, which implies $V_L'(x) \le V_H'(x)$ for $x \in [\theta_b^H, \theta_b^L]$ (see Eqs. (20) and (21)). For the former, we will now show $V_L'(x) < V_H'(x)$ for $x < q$. We have

$$V_L'(x) = \frac{\frac{\mu_S}{(q-\mu_S)}\left(e^{y_L^2}y_L\mathrm{erfc}(-y_L)+\frac{1}{\sqrt{\pi}}\right)}{e^{\hat{q}_L^2}\hat{q}_L\mathrm{erfc}(-\hat{q}_L)+\frac{1}{\sqrt{\pi}}},$$

where $\hat{q}_L \triangleq \frac{q-\mu_R^L}{\sigma_R}$ and $y_L \triangleq \frac{x-\mu_R^L}{\sigma_R}$. Therefore to show $V_L'(x) < V_H'(x)$ for $x < q$, it is equivalent to show $\log\left(e^{y_L^2}y_L\mathrm{erfc}(-y_L)+\frac{1}{\sqrt{\pi}}\right) - \log\left(e^{\hat{q}_L^2}\hat{q}_L\mathrm{erfc}(-\hat{q}_L)+\frac{1}{\sqrt{\pi}}\right)$ is increasing in μ_R^L. By the same argument in case (1), we know that this is true. Since $\lim_{x\to-\infty} V_L(x) = \mu_R^L < \mu_R^H = \lim_{x\to-\infty} V_H(x)$, and $V_L'(x) < V_H'(x)$ on $\mathbb{R}$, we obtain $V_L(x) < V_H(x)$ on $\mathbb{R}$, i.e., the monotonicty with regard to μ_R.

Next we show monotonicity results of the first order derivative of the value function for $x \geq q$. Consider two valid Risky volatility values $\sigma_R^L < \sigma_R^H$. Let θ_b^L and θ_b^H be as defined in Lemma 2 for σ_R^L and σ_R^H, respectively, and denote V_L and V_H as the corresponding value functions. We have proved that θ_b is strictly increasing in σ_R, hence $\theta_b^L < \theta_b^H$. Again, there are three cases: (1) $\theta_b^L < \theta_b^H < q$, (2) $\theta_b^L < q \leq \theta_b^H$, (3) $q \leq \theta_b^L < \theta_b^H$. In case (1), for $x \geq q$, we have

$$f(\sigma_R^L, x) \triangleq V_L'(x) = \frac{\mu_R\sqrt{\pi}}{\sigma_R^L} e^{\frac{(x-\mu_R)^2}{(\sigma_R^L)^2}} \operatorname{erfc}\left(\frac{x-\mu_R}{\sigma_R^L}\right).$$

Take its derivative with respect to σ_R, we get

$$\frac{\partial f(\sigma_R^L, x)}{\partial \sigma_R^L} = -\frac{\mu_R \hat{f}(\sigma_R^L)}{(\sigma_R^L)^2},$$

where

$$\hat{f}(\sigma_R^L) \triangleq e^{\frac{(x-\mu_R)^2}{(\sigma_R^L)^2}} \sqrt{\pi}\left(1 + \frac{2(x-\mu_R)^2}{(\sigma_R^L)^2}\right) \operatorname{erfc}\left(\frac{x-\mu_R}{\sigma_R^L}\right) - \frac{2(x-\mu_R)}{\sigma_R^L}.$$

By property 1 in Lemma 5, $\hat{f}(\sigma_R^L) \geq 0$, which implies that $\frac{\partial f(\sigma_R^L, x)}{\partial \sigma_R^L} \leq 0$, i.e., $V_L'(x) \geq V_H'(x)$ for $x \geq q$. Similarly, in case (2), we have $V_L'(x) \geq V_H'(x)$ for $x \geq \theta_b^H$. Hence, we only need to consider when $q \leq x < \theta_b^H$. In fact, by the definition of θ_b^L, we get $\frac{\mu_S}{x-\mu_S} \leq \frac{\mu_R\sqrt{\pi}}{\sigma_R^L} e^{\frac{(x-\mu_R)^2}{\sigma_R^{L\,2}}} \operatorname{erfc}\left(\frac{x-\mu_R}{\sigma_R^L}\right)$ for $x \geq \theta_b^L$, which implies $V_H'(x) \leq V_L'(x)$ for $x \in [q, \theta_b^H]$ (see Eqs. (20) and (21)). In case (3), by the same analysis in case (1), we have $V_L'(x) \geq V_H'(x)$ for $x \geq \theta_b^H$. Moreover $V_L'(x) = V_H'(x)$ for $q \leq x < \theta_b^L$. Hence, we only need to consider when $\theta_b^L \leq x < \theta_b^H$. By the definition of θ_b^L, we get $\frac{\mu_S}{x-\mu_S} \leq \frac{\mu_R\sqrt{\pi}}{\sigma_R^L} e^{\frac{(x-\mu_R)^2}{\sigma_R^{L\,2}}} \operatorname{erfc}\left(\frac{x-\mu_R}{\sigma_R^L}\right)$ for $x \geq \theta_b^L$, which implies $V_H'(x) \leq V_L'(x)$ for $x \in [\theta_b^L, \theta_b^H]$ (see Eqs. (20) and (21)).

Next we prove the monotonicity result when $\mu_S > \mu_R$. First we want to show that θ_G is monotonically decreasing in σ_R. Recall the definition of θ_G in Lemma 3. We already know that θ_G is the unique root of either of two monotone strictly-decreasing functions (see $F_{\text{small}}(\cdot)$ and $\hat{F}_{\text{big}}(\cdot)$ in the proof of Lemma 3) on $(\mu_S - \frac{\sigma^2}{\mu_S-\mu_R}, q)$. Therefore if we can show the two function values are decreasing in σ_R, then we are done. Consider a fixed value of θ such that $\theta \in (\mu_S - \frac{\sigma^2}{\mu_S-\mu_R}, q)$. We first put down the values of the two functions, both evaluated at θ:

$$\begin{aligned} f_{\text{small}}(\sigma_R) = &-\frac{\mu_S\sigma_R}{(\mu_S-\mu_R)(q-\mu_S)} + 2\Bigg(e^{\frac{(q-\mu_R)^2-(\theta-\mu_R)^2}{\sigma_R^2}} \frac{q-\mu_R}{\sigma_R} - \frac{\theta-\mu_R}{\sigma_R} \\ &+ e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} \frac{(q-\mu_R)(\theta-\mu_R)}{\sigma_R^2}\sqrt{\pi}\left(\operatorname{erf}\left(\frac{\theta-\mu_R}{\sigma_R}\right) - \operatorname{erf}\left(\frac{q-\mu_R}{\sigma_R}\right)\right)\Bigg), \end{aligned}$$

$$\begin{aligned} f_{\text{big}}(\sigma_R) = &-\frac{\mu_S\sigma_R\left(1-\frac{(\mu_S-\mu_R)(\mu_S-\theta)}{\sigma_R^2}\right)}{(\mu_S-\mu_R)(q-\mu_S)} + \Bigg(\frac{\mu_S-\mu_R}{\sigma_R} - 3\cdot\frac{\theta-\mu_R}{\sigma_R} \\ &+ 2\cdot\frac{(\mu_S-\theta)(\theta-\mu_R)^2}{\sigma_R^3}\Bigg) + 2e^{\frac{(q-\mu_R)^2-(\theta-\mu_R)^2}{\sigma_R^2}} \frac{q-\mu_R}{\sigma_R} \\ &\cdot\left(1-\frac{(\theta-\mu_R)(\mu_S-\theta)}{\sigma_R}\right) + e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} \frac{q-\mu_R}{\sigma_R}\sqrt{\pi}\Bigg(\frac{\mu_S-\mu_R}{\sigma_R} \end{aligned}$$

$$- 3 \cdot \frac{\theta - \mu_R}{\sigma_R} + 2 \cdot \frac{(\mu_S - \theta)(\theta - \mu_R)^2}{\sigma_R^3} \Bigg) \left(\operatorname{erf}\left(\frac{q - \mu_R}{\sigma_R}\right) - \operatorname{erf}\left(\frac{\theta - \mu_R}{\sigma_R}\right) \right).$$

Want to show that $f'_{\text{small}}(\sigma_R)$ and $f'_{\text{big}}(\sigma_R)$ are non-positive. Consider first $f'_{\text{small}}(\sigma_R)$:

$$f'_{\text{small}}(\sigma_R) = \frac{1}{\sigma_R} \cdot g_1\Big(\frac{\theta - \mu_R}{\sigma_R}, \frac{q - \mu_R}{\sigma_R}, \frac{\mu_S - \mu_R}{\sigma_R}\Big),$$

where

$$g_1(\hat{\theta}, \hat{q}, a) = \frac{\mu_S}{(\mu_S - \mu_R)(a - \hat{q})} - 2\hat{q}(1 + 2\hat{q}^2)e^{\hat{q}^2 - \hat{\theta}^2} + 2\hat{\theta}(1 + 2\hat{q}^2) + 4\hat{q}\hat{\theta}(1 + \hat{q}^2)\sqrt{\pi}e^{\hat{q}^2} \left(\operatorname{erf}(\hat{q}) - \operatorname{erf}(\hat{\theta} \right).$$

To show $f'_{\text{small}}(\sigma_R) \le 0$, it suffices to show $g_1(\hat{\theta}, \hat{q}, a) \le 0$ for $\hat{q} > a > 0$, $\hat{\theta} \in (a - \frac{1}{a}, \hat{q})$. Since $\mu_S > \mu_R > 0$ and $\hat{q} > a > 0$, we have $\frac{\mu_S}{\mu_S - \mu_R} > 1$ and $\frac{1}{a - \hat{q}} < \frac{1}{-\hat{q}}$, and hence $\frac{\mu_S}{(\mu_S - \mu_R)(a - \hat{q})} < \frac{1}{-\hat{q}}$ and

$$g_1(\hat{\theta}, \hat{q}, a) < -\frac{1}{\hat{q}} - 2\hat{q}(1 + 2\hat{q}^2)e^{\hat{q}^2 - \hat{\theta}^2} + 2\hat{\theta}(1 + 2\hat{q}^2) + 4\hat{q}\hat{\theta}(1 + \hat{q}^2)\sqrt{\pi}e^{\hat{q}^2} \left(\operatorname{erf}(\hat{q}) - \operatorname{erf}(\hat{\theta} \right).$$

Observe that the RHS of the above inequality only depends on $\hat{\theta}$ and $\hat{q}$, hence we denote it by $\bar{g}_1(\hat{\theta}, \hat{q})$. It suffices to show $\bar{g}(\hat{\theta}, \hat{q}) \le 0$. If $\hat{\theta} \le 0$, then this is true, since $\hat{q} > 0 \ge \hat{\theta}$ and $\operatorname{erf}(\cdot)$ is an increasing function. On the other hand, if $\hat{\theta} > 0$, since $\hat{\theta} < \hat{q}$, it must be that $\hat{\theta} \in (0, \hat{q})$. In this case, we have $\frac{\partial^2 \bar{g}_1}{\partial \hat{\theta}^2}(\hat{\theta}, \hat{q}) = -4e^{\hat{q}^2 - \hat{\theta}^2}\hat{q}\left(3 + 2\hat{q}^2 - 2\hat{\theta}^2\right) < 0$, and hence $\frac{\partial \bar{g}_1}{\partial \hat{\theta}}(\hat{\theta}, \hat{q}) = 2 + 4\hat{q}^2 - 4e^{\hat{q}^2 - \hat{\theta}^2}\hat{q}\hat{\theta} + 4e^{\hat{q}^2}\hat{q}(1 + \hat{q}^2)\sqrt{\pi}\left(\operatorname{erf}(\hat{q}) - \operatorname{erf}(\hat{\theta}\right) > \frac{\partial \bar{g}_1}{\partial \hat{\theta}}(\hat{q}, \hat{q}) = 2 > 0$. Therefore $\bar{g}_1(\hat{\theta}, \hat{q}) < \bar{g}_1(\hat{q}, \hat{q}) = -\frac{1}{\hat{q}} < 0$.

Now it only remains to show $f'_{\text{big}}(\sigma_R) \le 0$. Since $f_{\text{big}}(\sigma_R)$ is only relevant to the value of θ_G when $\theta_G < \mu_S$, it suffices to consider $\theta \in \left(\mu_S - \frac{\sigma_R^2}{\mu_S - \mu_R}, \mu_S\right)$. Compute $f'_{\text{big}}(\sigma_R)$, we get

$$f'_{\text{big}}(\sigma_R) = \frac{a}{b} \cdot h_1\left(\frac{\theta - \mu_R}{\sigma_R}, \frac{q - \mu_R}{\sigma_R}, \frac{\mu_S - \mu_R}{\sigma_R}\right),$$

where

$$\begin{aligned} h_1(\hat{\theta}, \hat{q}, a) =& \frac{\mu_S}{\mu_S - \mu_R} \cdot \frac{a^2 - a\hat{\theta} + 1}{a - \hat{q}} - a + 3\hat{\theta} - 6a\hat{\theta}^2 + 6\hat{\theta}^3 - 2\hat{q}^2\left(a - 3\hat{\theta} + 2a\hat{\theta}^2 - 2\hat{\theta}^3\right) \\ &+ e^{\hat{q}^2 - \hat{\theta}^2}\hat{q}\left(6a\hat{\theta} - 6\hat{\theta}^2 - (1 + 2\hat{q}^2)\left(2 - 2a\hat{\theta} + 2\hat{\theta}^2\right)\right) + e^{\hat{q}^2}\hat{q}\sqrt{\pi}\Big(- a + 3\hat{\theta} \\ &- 6a\hat{\theta}^2 + 6\hat{\theta}^3 - (1 + 2\hat{q}^2)\left(a - 3\hat{\theta} + 2a\hat{\theta}^2 - 2\hat{\theta}^3\right)\Big)\left(\operatorname{erf}(\hat{q}) - \operatorname{erf}(\hat{\theta})\right). \end{aligned}$$

To show $f'_{\text{big}}(\sigma_R) \le 0$, it is equivalent to show $h_1(\hat{\theta}, \hat{q}, a) \le 0$ for $\hat{q} > a > 0$, $\hat{\theta} \in (a - \frac{1}{a}, a)$. Observe that $a^2 - a\hat{\theta} + 1 > 0$ since $\hat{\theta} < a$ and $a > 0$. Also $a - \hat{q} < 0$ and $\frac{\mu_S}{\mu_S - \mu_R} > 1$. Hence we have $\frac{\mu_S}{\mu_S - \mu_R} \cdot \frac{a^2 - a\hat{\theta} + 1}{a - \hat{q}} < \frac{a^2 - a\hat{\theta} + 1}{a - \hat{q}}$, and

$$\begin{aligned} h_1(\hat{\theta}, \hat{q}, a) <& \frac{a^2 - a\hat{\theta} + 1}{a - \hat{q}} - a + 3\hat{\theta} - 6a\hat{\theta}^2 + 6\hat{\theta}^3 - 2\hat{q}^2\left(a - 3\hat{\theta} + 2a\hat{\theta}^2 - 2\hat{\theta}^3\right) \\ &+ e^{\hat{q}^2 - \hat{\theta}^2}\hat{q}\left(6a\hat{\theta} - 6\hat{\theta}^2 - (1 + 2\hat{q}^2)\left(2 - 2a\hat{\theta} + 2\hat{\theta}^2\right)\right) + e^{\hat{q}^2}\hat{q}\sqrt{\pi}\Big(- a + 3\hat{\theta} \\ &- 6a\hat{\theta}^2 + 6\hat{\theta}^3 - (1 + 2\hat{q}^2)\left(a - 3\hat{\theta} + 2a\hat{\theta}^2 - 2\hat{\theta}^3\right)\Big)\left(\operatorname{erf}(\hat{q}) - \operatorname{erf}(\hat{\theta})\right). \end{aligned}$$

Denote the RHS of the above by $\bar{h}_1(a)$. If we can show $\bar{h}_1(a) \le 0$ then we are done. Next we will show $\bar{h}_1(\cdot)$ is monotone decreasing on $(0, \hat{q})$. Since $\hat{q} > a > 0$ and $\hat{q} > a > \hat{\theta}$, we have $\bar{h}''_1(a) = -\frac{2(1 + \hat{q}(\hat{q} - \hat{\theta}))}{(\hat{q} - a)^3} < 0$, which implies

$$\begin{aligned} \bar{h}'_1(a) =& -1 - 6\hat{\theta}^2 - 2\hat{q}^2(1 + 2\hat{\theta}^2) + \frac{a^2 - 2a\hat{q} + \hat{\theta}\hat{q} - 1}{(a - \hat{q})^2} + e^{\hat{q}^2 - \hat{\theta}^2}\hat{q}\left(6\hat{\theta} + 2\hat{\theta}(1 + 2\hat{q}^2)\right) \\ &- e^{\hat{q}^2}\hat{q}\left((1 + 2\hat{q}^2)(1 + 2\hat{\theta}^2) + 1 + 6\hat{\theta}^2\right)\sqrt{\pi}\left(\operatorname{erf}(\hat{q}) - \operatorname{erf}(\hat{\theta})\right) \\ <& \bar{h}'_1(0) = -1 - 6\hat{\theta}^2 - 2\hat{q}(1 + 2\hat{\theta}^2) + \frac{\hat{\theta}}{\hat{q}} - \frac{1}{\hat{q}^2} + e^{\hat{q}^2 - \hat{\theta}^2}\hat{q}\left(6\hat{\theta} + 2\hat{\theta}(1 + 2\hat{q}^2)\right) \end{aligned}$$

$$-e^{\hat{q}^2}\hat{q}\left((1+2\hat{q}^2)(1+2\hat{\theta}^2)+1+6\hat{\theta}^2\right)\sqrt{\pi}\left(\operatorname{erf}(\hat{q})-\operatorname{erf}(\hat{\theta})\right):=\hat{h}_1(\hat{\theta}).$$

We will now show $\hat{h}_1(\hat{\theta})<0$ for $\hat{\theta}<a<\hat{q}$. First one can verify that $\hat{h}_1(\hat{q})=-\frac{1}{\hat{q}^2}<0$. Therefore we will only show $\hat{h}_1'(\hat{\theta})\geq 0$ for $\hat{\theta}<\hat{q}$. Since $\hat{\theta}<\hat{q}$ and that $\operatorname{erf}(\cdot)$ is a monotone increasing function, if $\hat{\theta}\leq 0$, then $\hat{h}_1''(\hat{\theta})=-12-8\hat{q}^2+8e^{\hat{q}^2-\hat{\theta}^2}\hat{q}\hat{\theta}-8e^{\hat{q}^2}\hat{q}(2+\hat{q}^2)\sqrt{\pi}\left(\operatorname{erf}(\hat{q})-\operatorname{erf}(\hat{\theta})\right)<0$. On the other hand, if $\hat{\theta}>0$, then for $\hat{\theta}\in(0,\hat{q})$, we have $\hat{h}_1'''(\hat{\theta})=8e^{\hat{q}^2-\hat{\theta}^2}\hat{q}\left(5+2\hat{q}^2-2\hat{\theta}^2\right)>0$, which implies that $\hat{h}_1''(\hat{\theta})<\hat{h}_1''(\hat{q})=-12<0$. Therefore $\hat{h}_1'(\hat{\theta})$ is monotone decreasing for $\hat{\theta}<\hat{q}$, and we have $\hat{h}_1'(\hat{\theta})=\frac{1}{\hat{q}}+e^{\hat{q}^2-\hat{\theta}^2}\left(12\hat{q}+8\hat{q}^3\right)-12\hat{\theta}-8\hat{q}^2\hat{\theta}-8e^{\hat{q}^2}\hat{q}(2+\hat{q}^2)\hat{\theta}\sqrt{\pi}\left(\operatorname{erf}(\hat{q})-\operatorname{erf}(\hat{\theta})\right)>\hat{h}_1'(\hat{q})=\frac{1}{\hat{q}}>0$, which is what we want in order to have $\hat{h}_1(\hat{\theta})<0$ for $\hat{\theta}<a<\hat{q}$, and thus $\bar{h}_1'(a)<0$. Recall that we want to show $\bar{h}_1(a)\leq 0$ so that $f_{\text{big}}'(\sigma_R)\leq 0$. Since we now know that $\bar{h}_1'(a)<0$ for $a\in(\hat{\theta},\hat{q})$, we have

$$\begin{aligned}\bar{h}_1(a)<\bar{h}_1(\hat{\theta})&=2\hat{\theta}(1+2\hat{q}^2)-\frac{1}{\hat{q}-\hat{\theta}}-2e^{\hat{q}^2-\hat{\theta}^2}\hat{q}(1+2\hat{q}^2)+4e^{\hat{q}^2}\hat{q}(1+\hat{q}^2)\hat{\theta}\sqrt{\pi}\left(\operatorname{erf}(\hat{q})-\operatorname{erf}(\hat{\theta})\right)\\&<2\hat{\theta}(1+2\hat{q}^2)-2e^{\hat{q}^2-\hat{\theta}^2}\hat{q}(1+2\hat{q}^2)+4e^{\hat{q}^2}\hat{q}(1+\hat{q}^2)\hat{\theta}\sqrt{\pi}\left(\operatorname{erf}(\hat{q})-\operatorname{erf}(\hat{\theta})\right)\leq 0,\end{aligned}$$

where the last step follows from property 12 in Lemma 5. This completes the proof. □

References

[1] Daniel Adelman and Adam J. Mersereau. Dynamic capacity allocation to customers who remember past service. *Management Science*, 59(3):592–612, 2013.

[2] Sam Aflaki and Ioana Popescu. Managing retention in service relationships. *Management Science*, 60(2):415–433, 2014.

[3] Eva Ascarza, Oded Netzer, and Bruce GS Hardie. Some customers would rather leave without saying goodbye. *Marketing Science*, 37(1):54–77, 2018.

[4] Bariş Ata, JM Harrison, Lawrence A Shepp, et al. Drift rate control of a brownian processing system. *The Annals of Applied Probability*, 15(2):1145–1160, 2005.

[5] Rami Atar and Anat Lev-Ari. Workload-dependent dynamic priority for the multiclass queue with reneging. *Mathematics of Operations Research*, 43(2):494–515, 2018.

[6] Paul D. Berger and Nada I. Nasr. Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*, 12(1):17–30, 1998.

[7] Vivek S. Borkar. Optimal control of diffusion processes. *Pitman Research Notes in Math.*, 203(36):213–244, 1989.

[8] S. H. Chang, P. C. Cosman, and L. B. Milstein. Chernoff-type bounds for the gaussian error function. *IEEE Transactions on Communications*, 59(11):2939–2944, November 2011.

[9] M Aslam Chaudhry and Syed M Zubair. *On a class of incomplete gamma functions with applications.* CRC press, 2001.

[10] Aparupa Das Gupta, Uday S Karmarkar, and Guillaume Roels. The design of experiential services with acclimation and memory decay: Optimal sequence and duration. *Management Science*, 62(5):1278–1296, 2015.

[11] Vishal Gaur and Young-Hoon Park. Asymmetric consumer learning and inventory competition. *Management Science*, 53(2):227–240, 2007.

[12] J Michael Harrison and Nur Sunar. Investment timing with incomplete information and multiple means of learning. *Operations Research*, 63(2):442–457, 2015.

[13] Teck-Hua Ho, Young-Hoon Park, and Yong-Pin Zhou. Incorporating satisfaction into customer value analysis: Optimal investment in lifetime value. *Marketing Science*, 25(3):260–277, 2006.

[14] Ramesh Johari and Sven Schmit. Learning with abandonment. *arXiv preprint arXiv:1802.08718*, 2018.

[15] Daniel Kahneman, Peter P Wakker, and Rakesh Sarin. Back to bentham? explorations of experienced utility. *The Quarterly Journal of Economics*, 112(2):375–406, 1997.

[16] Qian Liu and Garrett van Ryzin. Strategic capacity rationing when customers learn. *Manufacturing & Service Operations Management*, 13(1):89–107, 2011.

[17] Melda Ormeci Matoglu, John Vande Vate, and Huizhu Wang. Solving the drift control problem. *Stochastic Systems*, 5(2):324–371, 2015.

[18] Stefan Mirică. *Verification Theorems of Dynamic Programming Type in Optimal Control*, pages 181–191. Birkhäuser Basel, Basel, 1992.

[19] Ricardo Montoya, Oded Netzer, and Kamel Jedidi. Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Science*, 29(5):909–924, 2010.

[20] Marc Nerlove and Kenneth J Arrow. Optimal advertising policy under dynamic conditions. *Economica*, pages 129–142, 1962.

[21] Oded Netzer, James M Lattin, and Vikram Srinivasan. A hidden markov model of customer relationship dynamics. *Marketing science*, 27(2):185–204, 2008.

[22] Melda Ormeci Matoglu and John Vande Vate. Drift control with changeover costs. *Operations research*, 59(2):427–439, 2011.

[23] Anton Ovchinnikov and Joseph M. Milner. Revenue management with end-of-period discounts in the presence of customer learning. *Production and Operations Management*, 21(1):69–84, 2012.

[24] Iosif Pinelis. On l'hospital-type rules for monotonicity. *JIPAM. Journal of Inequalities in Pure & Applied Mathematics*, 7(2):1–19, 2006.

[25] Roy Radner and Larry Shepp. Risk vs. profit potential: A model for corporate strategy. *Journal of Economic Dynamics and Control*, 20(8):1373 – 1393, 1996.

[26] Josh Reed, Amy Ward, and Dongyuan Zhan. On the generalized drift skorokhod problem in one dimension. *Journal of Applied Probability*, 50(1):16–28, 2013.

[27] C. Ren and A. R. MacKenzie. Closed-form approximations to the error and complementary error functions and their applications in atmospheric science. *Atmospheric Science Letters*, 8(3):70–73, 2007.

[28] Roland T. Rust, Katherine N. Lemon, and Valarie A. Zeithaml. Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing*, 68(1):109–127, 2004.

[29] Michael Salins and Konstantinos Spiliopoulos. Markov processes with spatial delay: Path space characterization, occupation time and properties. *Stochastics and Dynamics*, 17(06):1750042, 2017.

[30] Grant Schneider, Peter F Craigmile, and Radu Herbei. Maximum likelihood estimation for stochastic differential equations using sequential kriging-based optimization. *arXiv preprint arXiv:1408.2441*, 2014.

[31] A. V. Skorokhod. Stochastic equations for diffusion processes in a bounded region. *Theory of Probability & Its Applications*, 6(3):264–274, 1961.

[32] Bruno Strulovici and Martin Szydlowski. On the smoothness of value functions and the existence of optimal strategies in diffusion models. *Journal of Economic Theory*, 159:1016–1055, 2015.

[33] Nur Sunar, John R Birge, and Sinit Vitavasiri. Optimal dynamic product development and launch for a network of customers. *Operations Research*, 67(3):770–790, 2019.

[34] Nur Sunar, Siyun Yu, and Vidyadhar G Kulkarni. Competitive investment with bayesian learning: Choice of business size and timing. *Operations Research*, 69(5):1430–1449, 2021.

[35] Marlin U. Thomas. Some mean first-passage time approximations for the ornstein-uhlenbeck process. *Journal of Applied Probability*, 12(3):600–604, 1975.

[36] Nizar Touzi. *Stochastic Control Problems, Viscosity Solutions and Application to Finance*. Edizioni della Normale, 2002.

[37] Zhengli Wang and Stefanos Zenios. New venture creation: A drift-variance diffusion control model. *Operations Research*, 2021.

[38] Amy R. Ward and Peter W. Glynn. Properties of the reflected ornstein–uhlenbeck process. *Queueing Systems*, 44(2):109–123, Jun 2003.

[39] Valarie A Zeithaml. Service quality, profitability, and the economic worth of customers: what we know and what we need to learn. *Journal of the Academy of Marketing Science*, 28(1):67–85, 2000.

Online Appendix for "Managing Customer Churn via Service Mode Control"

This online appendix is an e-companion for paper "Managing Customer Churn via Service Mode Control" and provides additional proofs and materials in supplement to it. This online appendix is subdivided into six sections. In Online Appendix A, we prove results related to the error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ and its complement $\text{erfc}(x) = 1 - \text{erf}(x)$. These are used to support and complete the proof of Lemma 5 in the paper. Online appendix B provides the remaining proofs of the Base model, including the remaining proof of Proposition 3 and the proof of Corollary 2. Online appendices C–F demonstrate robustness of our main findings by analyzing four variations of the Base model: allowing mixed policies, the reward process being a geometric Brownian motion, the inclusion of positive switching costs, and alternative hazard rate functions, respectively.

A Properties of the error function

This section provides the proofs for properties 4, 6, 7, and 9 – 16 in Lemma 5.

Proof of Lemma 5 (Part II). We provide the proofs for properties 4, 6, 7, 9 – 16.

To prove 4, it is equivalent to show $f(x) \triangleq x\text{erf}(x) + \frac{e^{-x^2}}{\sqrt{\pi}} > 0$ for all $x \in \mathbb{R}$. In fact, $f'(x) = \text{erf}(x)$. Therefore $f(x)$ is decreasing on the interval $(-\infty, 0)$, increasing on the interval $(0, \infty)$, and the minimum value is obtained at $x = 0$, with $f(0) = \frac{1}{\sqrt{\pi}} > 0$.

Now we prove 6. Denote $x_1 \triangleq \frac{\alpha(\beta-2)+\sqrt{\alpha^2\beta^2+2\beta(1-\beta)}}{2(1-\beta)}$. Since $\alpha > 0$ and $0 < \beta < 1$, we have $x_1 > \frac{\alpha(\beta-2)+\alpha\beta}{2(1-\beta)} = -\alpha$. Also by property 5 (ii), $\{x > -\alpha : r(x) \geq \frac{\beta}{\sqrt{\pi}}\} = [x_0, \infty)$. Hence to show $x_1 \geq x_0$, we only need to show $r(x_1) \geq \frac{\beta}{\sqrt{\pi}}$. Since $x_1 > -\alpha$, we can divide both sides by $(x_1 + \alpha)e^{x_1^2}$ and equivalently show $f(\beta) \geq 0$ for any $\beta \in (0, 1)$, where

$$f(\beta) \triangleq -\frac{\left(\alpha\beta + \sqrt{\alpha^2\beta^2 + 2\beta(1-\beta)}\right) e^{-\frac{\left(\alpha(\beta-2)+\sqrt{\alpha^2\beta^2+2\beta(1-\beta)}\right)^2}{4(\beta-1)^2}}}{\sqrt{\pi}} + \text{erfc}\left(\frac{\alpha(\beta-2) + \sqrt{\alpha^2\beta^2 + 2\beta(1-\beta)}}{2(1-\beta)}\right).$$

One can verify that $\lim_{\beta \to 1^-} \frac{\alpha(\beta-2)+\sqrt{\alpha^2\beta^2+2\beta(1-\beta)}}{2(\beta-1)} = \alpha - \frac{1}{\alpha}$. Therefore, $g(\alpha) \triangleq \lim_{\beta \to 1^-} f(\beta) = \text{erfc}(\frac{1}{2\alpha} - \alpha) - \frac{2\alpha e^{-\left(\alpha - \frac{1}{2\alpha}\right)^2}}{\sqrt{\pi}}$. In fact, $g(\alpha) > 0$ for all $\alpha > 0$, since $\lim_{\alpha \to 0^+} g(\alpha) = 0$ and $g'(\alpha) = \frac{4\alpha^2 e^{-\left(\alpha - \frac{1}{2\alpha}\right)^2}}{\sqrt{\pi}} > 0$. Also, for any $\alpha > 0$ and $\beta \in (0, 1)$, obviously

$$f'(\beta) = -\frac{\left(\alpha\beta + \sqrt{\alpha^2\beta^2 + 2\beta(1-\beta)}\right) e^{-\frac{\left(\alpha(\beta-2)+\sqrt{\alpha^2\beta^2+2\beta(1-\beta)}\right)^2}{4(\beta-1)^2}}}{\beta\sqrt{\pi}} < 0.$$

Therefore $f(\beta)$ is decreasing on $(0, 1)$ with $\lim_{\beta \to 1^-} f(\beta) > 0$, and the result follows.

Now we prove 7. Define $F(x) \triangleq \frac{C}{\sqrt{\pi}} + xe^{x^2}(1 + C\text{erf}(x))$, with $F'(x) = e^{x^2}(2x^2 + 1)(1 + C\text{erf}(x)) + \frac{2Cx}{\sqrt{\pi}}$. If $C \in \{0, 1\}$, then the result trivially holds. If $C = -1$, $F'(x) \geq 0$ follows from property 1. It only remains to consider $C \in (-1, 1)\backslash\{0\}$. Let $f(x) \triangleq F'(x)e^{-x^2}$. We have $f''(x) = 4C\text{erf}(x) + 4 > 0$, $\lim_{x \to -\infty} f'(x) = -\infty$, and $\lim_{x \to \infty} f'(x) = \infty$ if $C \in (-1, 1)$. Thus $f(x)$ is first decreasing then increasing on $\mathbb{R}$ if $C \in (-1, 1)$. Consider its global minimizer x_0, with

$$f'(x_0) = 4\left(Cx_0\text{erf}(x_0) + \frac{Ce^{-x_0^2}}{\sqrt{\pi}} + x_0\right) = 0. \tag{43}$$

Also one can check that $f'(0) = \frac{4C}{\sqrt{\pi}}$, so that $f'(0) > 0$ if $C \in (0,1)$, and $f'(0) < 0$ if $C \in (-1,0)$. Recall that $f''(x) > 0$, hence $x_0 < 0$ if $C \in (0,1)$ and $x_0 > 0$ if $C \in (-1,0)$. In both cases $Cx_0 < 0$, and thus we can divide both sides of Eq. (43) by Cx_0 and obtain $\operatorname{erf}(x_0) = \frac{\frac{Ce^{-x_0^2}}{\sqrt{\pi}}+x_0}{-Cx_0}$. Then we have $\min_{x\in\mathbb{R}} f(x) = f(x_0) = \left(2Cx_0^2 + C\right)\operatorname{erf}(x_0) + \frac{2Ce^{-x_0^2}x_0}{\sqrt{\pi}} + 2x_0^2 + 1 = \frac{2Ce^{-x_0^2}(x_0^2-x_0+1)}{-x_0}$. Observe that $x_0^2 - x_0 + 1 > 0$, and that $\frac{C}{-x_0} > 0$ from the above argument. The result follows.

Next we prove 9. We can divide both sides by $e^{q^2}q$ and show the following inequality instead: $F(q) \triangleq (a^2-2)\left(\frac{e^{-q^2}}{q} + \sqrt{\pi}\left(\operatorname{erf}(q) - \operatorname{erf}\left(a-\frac{1}{a}\right)\right)\right) - 2ae^{-\left(a-\frac{1}{a}\right)^2} < 0$. In fact, if $a^2 - 2 \le 0$, then since $\operatorname{erf}(\cdot)$ is an increasing function and that $q > a > a - \frac{1}{a}$, $F(q) < 0$ is obviously true. Now consider when $a^2 - 2 > 0$. Compute $F'(q) = -\frac{(a^2-2)e^{-q^2}}{q^2} < 0$ for any $q \in \mathbb{R}$, which implies that for $q > a$ (note that $a > a - \frac{1}{a}$), $F(q) < F(a - \frac{1}{a}) = -\frac{a^3 e^{-\left(\frac{1}{a}-a\right)^2}}{a^2-1} < 0$.

Next we prove 10. Denote $F(\theta) \triangleq e^{q^2}q - e^{\theta^2}\theta - e^{\theta^2+q^2}\theta q\sqrt{\pi}\left(\operatorname{erfc}(\theta) - \operatorname{erfc}(q)\right)$. If $\theta \le 0$, then the result easily holds since each term is non-negative. Now consider $\theta \in (0,q)$. To show $F(\theta) \ge 0$, it is equivalent to show $\frac{F(\theta)}{e^{q^2+\theta^2}q\theta} \ge 0$. In fact, $\frac{F(\theta)}{e^{q^2+\theta^2}q\theta} = f(\theta) - f(q)$, where $f(\theta) \triangleq \frac{1-\sqrt{\pi}e^{\theta^2}\theta\operatorname{erfc}(\theta)}{e^{\theta^2}\theta}$. Clearly, $e^{\theta^2}\theta$ is positive and increasing in θ for $\theta \in (0,q)$. Also, $1 - \sqrt{\pi}e^{\theta^2}\theta\operatorname{erfc}(\theta)$ is decreasing in θ by applying property 5 (i). Therefore $f(\theta)$ is decreasing in θ for $\theta \in (0,q)$, and the result follows.

Next we prove 11. Denote $F(\theta) \triangleq 2e^{q^2-\theta^2}q\theta - \left(1+2\theta^2\right)\left(1 + e^{q^2}q\sqrt{\pi}\left(\operatorname{erfc}(\theta) - \operatorname{erfc}(q)\right)\right)$. One can verify that $F(q) = F'(q) = 0$. Since $q > \theta$, we have $\operatorname{erf}(q) > \operatorname{erf}(\theta)$ and hence $F''(\theta) = -4 - 4e^{q^2}q\sqrt{\pi}\left(\operatorname{erf}(q) - \operatorname{erf}(\theta)\right) < 0$. This implies $F'(\theta) > F'(q) = 0$ on $(-\infty, q)$. Therefore $F(\theta) < F(q) = 0$ for $\theta < q$.

Next we prove 12. Denote $F(\theta) \triangleq \theta(1+2q^2) - e^{q^2-\theta^2}q(1+2q^2) + 2e^{q^2}q(1+q^2)\theta\sqrt{\pi}\left(\operatorname{erf}(q) - \operatorname{erf}(\theta)\right)$. One can verify that $F(q) = 0$ and $F'(q) = 1$. Since $q > \theta$, we have $\operatorname{erf}(q) > \operatorname{erf}(\theta)$ and hence $F(\theta) < 0$ on $(-\infty, 0]$. Now consider $\theta \in (0,q)$. It's not hard to see that $F''(\theta) = -2e^{q^2-\theta^2}q\left(3 + 2q^2 - 2\theta^2\right) < 0$, which implies $F'(\theta) > F'(q) > 0$ for $\theta < q$, and hence $F(\theta) < F(q) = 0$ for $\theta \in (0,q)$ as well.

Next we prove 13 – 15. Define

$$F_1(x,a) \triangleq 2 - 2ax + 2x^2 + \left(3x - a - 2ax^2 + 2x^3\right)e^{x^2}\sqrt{\pi}\operatorname{erf}(x),$$
$$F_2(x,a) \triangleq 2 - 2ax + 2x^2 - \left(3x - a - 2ax^2 + 2x^3\right)e^{x^2}\sqrt{\pi}\operatorname{erfc}(x),$$
$$F_3(x,a) \triangleq 2 - 2ax + 2x^2 + \left(3x - a - 2ax^2 + 2x^3\right)e^{x^2}\sqrt{\pi}(1 + \operatorname{erf}(x)).$$

If $a > 0$ and $x \in (a - \frac{1}{a}, a)$, then equivalently $a \in \left(\max\{x,0\}, \frac{x}{2} + \sqrt{1 + \frac{x^2}{4}}\right)$. Therefore it is sufficient to prove $F_1(x,a) > 0$, $F_2(x,a) > 0$, and $F_3(x,a) > 0$ for $a \in \left(x, \frac{x}{2} + \sqrt{1 + \frac{x^2}{4}}\right)$. Obvious that all three functions are linear in a. Consider the partial derivatives:

$$\frac{\partial F_1}{\partial a}(x,a) = -\sqrt{\pi}e^{x^2}\left(2x^2+1\right)\operatorname{erf}(x) - 2x, \qquad \frac{\partial F_2}{\partial a}(x,a) = \sqrt{\pi}e^{x^2}\left(2x^2+1\right)\operatorname{erfc}(x) - 2x,$$
$$\frac{\partial F_3}{\partial a}(x,a) = -\sqrt{\pi}e^{x^2}\left(2x^2+1\right)\operatorname{erfc}(-x) - 2x.$$

Property 1 implies that $\frac{\partial F_2}{\partial a}(x,a) \ge 0$ and $\frac{\partial F_3}{\partial a}(x,a) \le 0$. To prove the results, it then suffices to show $F_2(x,x) > 0$, $F_3(x, \frac{x}{2} + \sqrt{1 + \frac{x^2}{4}}) > 0$, $F_1(x,x) > 0$ and $F_1(x, \frac{x}{2} + \sqrt{1 + \frac{x^2}{4}}) > 0$ for all $x \in \mathbb{R}$. Property 4 implies $F_1(x,x) = 2\sqrt{\pi}e^{x^2}x\operatorname{erf}(x) + 2 > 0$. Property 5(iii) implies $F_2(x,x) =$

$2-2\sqrt{\pi}e^{x^2}x\mathrm{erfc}(x)>0$. Define

$$f_1(x)\triangleq F_1(x,\frac{x}{2}+\sqrt{1+\frac{x^2}{4}})=-\frac{1}{2}\sqrt{\pi}e^{x^2}\left(-2x^3+2\sqrt{x^2+4}x^2+\sqrt{x^2+4}-5x\right)\mathrm{erf}(x)+x^2-\sqrt{x^2+4}x+2$$

and

$$f_3(x)\triangleq F_3(x,\frac{x}{2}+\sqrt{1+\frac{x^2}{4}})=-\frac{1}{2}\sqrt{\pi}e^{x^2}\left(-2x^3+2\sqrt{x^2+4}x^2+\sqrt{x^2+4}-5x\right)\mathrm{erfc}(-x)+x^2-\sqrt{x^2+4}x+2.$$

It remains to show $f_1(x)>0$ and $f_3(x)>0$ on $\mathbb{R}$. In fact, one can check that $-2x^3+2\sqrt{x^2+4}x^2+\sqrt{x^2+4}-5x$ has a single root at $\frac{1}{\sqrt{2}}$, and that $-2x^3+2\sqrt{x^2+4}x^2+\sqrt{x^2+4}-5x>0$ on $(-\infty,\frac{1}{\sqrt{2}})$ and $-2x^3+2\sqrt{x^2+4}x^2+\sqrt{x^2+4}-5x<0$ on $(\frac{1}{\sqrt{2}},\infty)$. Now we consider the cases separately.

First check that $f_1(\frac{1}{\sqrt{2}})=1>0$ and $f_3(\frac{1}{\sqrt{2}})\approx 2>0$. Consider $x\neq\frac{1}{\sqrt{2}}$. Define $g_1(x)\triangleq\frac{e^{-x^2}f_1(x)}{-2x^3+2\sqrt{x^2+4}x^2+\sqrt{x^2+4}-5x}$ and $g_3(x)\triangleq\frac{e^{-x^2}f_3(x)}{-2x^3+2\sqrt{x^2+4}x^2+\sqrt{x^2+4}-5x}$. Compute their derivatives:

$$g_1'(x)=g_3'(x)=\frac{2e^{-x^2}\left(\sqrt{x^2+4}-x\right)}{\sqrt{x^2+4}\left(-2x^3+2\sqrt{x^2+4}x^2+\sqrt{x^2+4}-5x\right)^2}>0.$$

Then $g_1(x)>\lim_{x\to-\infty}g_1(x)=\frac{\sqrt{\pi}}{2}$ for $x<\frac{1}{\sqrt{2}}$, $g_1(x)<\lim_{x\to\infty}g_1(x)=-\frac{\sqrt{\pi}}{2}$ for $x>\frac{1}{\sqrt{2}}$, $g_3(x)>\lim_{x\to-\infty}g_3(x)=0$, and $g_3(x)<\lim_{x\to\infty}g_3(x)=-\sqrt{\pi}$. This means $f_1(x)>0$ and $f_3(x)>0$ everywhere.

Finally we prove 16. Denote the LHS of the inequality-to-show by $F(x)$. Observe that $F(x)+2-2a^2+2a\theta$ is quadratic in a. In fact, as we will show next, it is convex in a. Consider the second order partial derivative of $G(a,\theta,x)\triangleq F(x)+2-2a^2+2a\theta$ with respect to a:

$$\frac{\partial^2G(a,\theta,x)}{\partial a^2}=4\sqrt{\pi}\left(2\theta^2+1\right)e^{x^2}x(\mathrm{erf}(x)-\mathrm{erf}(\theta))+8\theta\left(\theta-xe^{x^2-\theta^2}\right).$$

We will show next that $g(\theta)\triangleq\frac{\partial^2G(a,\theta,z)}{\partial a^2}\geq 0$ for $\theta\leq x$. One can easily verify that $g(x)=g'(x)=0$. Compute $g''(\theta)=16\sqrt{\pi}e^{x^2}x(\mathrm{erf}(x)-\mathrm{erf}(\theta))+16$. When $x\geq 0$, it is obvious that $g''(\theta)>0$ for all $\theta\leq x$. When $x<0$, since $-\mathrm{erf}(\cdot)<1$, property 5(iii) implies that $g''(\theta)>16\sqrt{\pi}\left(\frac{1}{\sqrt{\pi}}-xe^{x^2}\mathrm{erfc}(x)\right)\geq 0$, which then implies that $g'(\theta)\leq g'(x)=0$ for $\theta\leq x$. Therefore, $g(\theta)\geq g(x)=0$ for $\theta\leq x$, and G is convex in a. Recall that we want to show $G(a,\theta,x)\leq 0$ for $a\in\left[\max\{0,x\},\frac{\theta}{2}+\sqrt{1+\frac{\theta^2}{4}}\right]$, $x\in\left(\theta,\frac{\theta}{2}+\sqrt{1+\frac{\theta^2}{4}}\right]$, and $\theta\in\mathbb{R}$. Since G is convex in a and $x\leq\max\{0,x\}$, it is sufficient to show $G(x,\theta,x)\leq 0$ and $G(\frac{\theta}{2}+\sqrt{1+\frac{\theta^2}{4}},\theta,x)\leq 0$ for all $\theta<x$.

First consider $f_1(\theta,x)\triangleq G(x,\theta,x)$. One can verify that $f_1(x,x)=0$, $\left.\frac{\partial f_1(\theta,x)}{\partial\theta}\right|_{\theta=x}=0$ and $\left.\frac{\partial^2 f_1(\theta,x)}{\partial\theta^2}\right|_{\theta=x}=-12$. Compute $\frac{\partial^3 f_1(\theta,x)}{\partial\theta^3}=4e^{x^2}\left(3\sqrt{\pi}(\mathrm{erf}(x)-\mathrm{erf}(\theta))+2e^{-\theta^2}x\right)$. If $x\geq 0$, then obviously $\frac{\partial^3 f_1(\theta,x)}{\partial\theta^3}>0$ for all $\theta<x$, and hence $\frac{\partial^2 f_1(\theta,x)}{\partial\theta^2}<\left.\frac{\partial^2 f_1(\theta,x)}{\partial\theta^2}\right|_{\theta=x}<0$. On the other hand, if $x<0$, then for all $\theta<x<0$, $\frac{\partial^4 f_1(\theta,x)}{\partial\theta^4}=-8e^{x^2-\theta^2}(2\theta x+3)<0$. Also $\lim_{\theta\to-\infty}\frac{\partial^3 f_1(\theta,x)}{\partial\theta^3}=12\sqrt{\pi}e^{x^2}\mathrm{erfc}(-x)>0$, $\frac{\partial^3 f_1(x,x)}{\partial\theta^3}=8x<0$, hence there is a unique root $\theta=\theta_0$ of $\frac{\partial^3 f_1(\theta,x)}{\partial\theta^3}$ on $(-\infty,x)$, which is also the maximizer of $\frac{\partial^2 f_1(\theta,x)}{\partial\theta^2}$ on $\theta\in(-\infty,x)$. θ_0 satisfies $\left.\frac{\partial^3 f_1(\theta,x)}{\partial\theta^3}\right|_{\theta=\theta_0}=0$, i.e., $\mathrm{erf}(\theta_0)-\mathrm{erf}(x)=\frac{2e^{-\theta_0^2}x}{3\sqrt{\pi}}$. Then for $\theta<x<0$, $\frac{\partial^2 f_1(\theta,x)}{\partial\theta^2}\leq\left.\frac{\partial^2 f_1(\theta,x)}{\partial\theta^2}\right|_{\theta=\theta_0}=4e^{x^2-\theta_0^2}\left(\frac{2x(x-3\theta_0)}{3}-3\right)<4e^{x^2-\theta_0^2}\left(\frac{-4x\theta_0}{3}-3\right)<0$. We have just showed that $\frac{\partial^2 f_1(\theta,x)}{\partial\theta^2}<0$ for all $\theta<x$ and $x\in\mathbb{R}$. This

implies that for $\theta < x$, $\frac{\partial f_1(\theta,x)}{\partial\theta} > \frac{\partial f_1(\theta,x)}{\partial\theta}\Big|_{\theta=x} = 0$, and hence $f_1(\theta,x) < f_1(x,x) = 0$.

We have shown $f_1(\theta,x) = G(x,\theta,x) < 0$ for $\theta < x$. It only remains to show $G(\frac{\theta}{2}+\sqrt{1+\frac{\theta^2}{4}},\theta,x) \le 0$ for $\theta < x$. This is equivalent to showing $f_2(a,x) \triangleq G(a,a-\frac{1}{a},x)a^3e^{-x^2} \le 0$ for all $x \in (a-\frac{1}{a},a]$ and $a > 0$. One can verify that $f_2(a,a-\frac{1}{a}) = \frac{\partial f_2(a,x)}{\partial x}\Big|_{x=a-\frac{1}{a}} = 0$. Compute $\frac{\partial^3 f_2(a,x)}{\partial x^3} = 8\left(a^2-2\right)e^{-x^2}(ax+1)$. Observe that since $x > a-\frac{1}{a}$ and $a > 0$, we have $ax+1 > a^2 > 0$, hence $\frac{\partial^3 f_2(a,x)}{\partial x^3} \le 0$ if $a \in (0,\sqrt{2}]$, $\frac{\partial^3 f_2(a,x)}{\partial x^3} > 0$ if $a > \sqrt{2}$. Compute $\frac{\partial^2 f_2(a,x)}{\partial x^2} = -4\sqrt{\pi}\left(a^2-2\right)\left(\mathrm{erf}\left(a-\frac{1}{a}\right)-\mathrm{erf}(x)\right) - 4\left(a^2-2\right)ae^{-x^2} - 8e^{2-\frac{a^4+1}{a^2}}a$. If $a \in (0,\sqrt{2}]$, $\frac{\partial^2 f_2(a,x)}{\partial x^2} \le \frac{\partial^2 f_2(a,x)}{\partial x^2}\Big|_{x=a-\frac{1}{a}} = -4a^3e^{-(a-\frac{1}{a})^2} < 0$ for all $x > a-\frac{1}{a}$. If $a > \sqrt{2}$, $\frac{\partial^2 f_2(a,x)}{\partial x^2} < \hat{f}_2(a) \triangleq \lim_{x\to\infty}\frac{\partial^2 f_2(a,x)}{\partial x^2} = 4\sqrt{\pi}(a^2-2)\mathrm{erfc}(a-\frac{1}{a}) - 8ae^{-(a-\frac{1}{a})^2} < 0$, where the last inequality follows from $\hat{f}_2'(a) = 8a\left(e^{2-\frac{a^4+1}{a^2}}a + \sqrt{\pi}\left(\mathrm{erf}\left(\frac{1}{a}-a\right)+1\right)\right) > 0$ on $(\sqrt{2},\infty)$ and $\lim_{a\to\infty}\hat{f}_2(a) = 0$. Since $\frac{\partial^2 f_2(a,x)}{\partial x^2} < 0$ for all $a > 0$ and $x \in (a-\frac{1}{a},a]$, we have $\frac{\partial f_2(a,x)}{\partial x} < \frac{\partial f_2(a,x)}{\partial x}\Big|_{x=a-\frac{1}{a}} = 0$, and thus $f_2(a,x) < f_2(a,a-\frac{1}{a}) = 0$. Proof complete. □

B Remaining proofs for the Base model

This section provides the remaining proofs for Proposition 3 as well as the proof for Corollary 2.

We first complete the proof for Proposition 3, including Statements 1, 5–7, 10–12.

Proof of Statement 6. The case $\mu_R = \mu_S$ is trivial, since $\theta_b = \infty$ by Lemma 2. Suppose $\mu_R > \mu_S$. By the definitions of $B_S(\cdot,\cdot)$, $V_4(\cdot,\cdot)$, and C_5^* (see Eqs. (34), (21) and (38)), and the definition of q in Eq. (4), for any $x > \max\{q,\theta_{\mathrm{b}}\}$, $B_S(x,V_4(x,C_5^*))$ becomes

$$B_S(x,V_4^*) = \frac{\mu_R\sqrt{\pi}(\mu_S-x)}{\sigma_R}e^{\frac{(x-\mu_R)^2}{\sigma_R^2}}\mathrm{erfc}\left(\frac{x-\mu_R}{\sigma_R}\right) + \mu_S.$$

Observe that the RHS of the above equation is the same as the LHS of the equation in Lemma 2. Therefore $B_S(\theta_b,V_4^*) = 0$. By property 5(ii) in Lemma 5, we get $B_S(x,V_4^*) < 0$ on (θ_b,∞). □

Proof of Statement 5. By the definitions of $\delta(\cdot,\cdot)$, $V_3(\cdot,\cdot)$, and C_4^* (see Eqs. (35), (20), and (37)), for any $x \in (q,\max\{q,\theta_b\})$, $\delta(x,V_3(x,C_4^*))$ becomes

$$\delta(x,V_3^*) = \frac{2(\mu_R-\mu_S)x^2 - 2\mu_S(\mu_R-\mu_S)x - \mu_S\sigma_R^2}{2(\mu_S-x)^2}.$$

If $\mu_R = \mu_S$, then $\delta(x,V_3^*) \le 0$ easily follows. Suppose $\mu_R > \mu_S$. Solving $\delta(x,V_3^*) \le 0$, we get

$$\frac{\mu_S}{2} - \sqrt{\frac{\mu_S^2}{4} + \frac{\mu_S\sigma_R^2}{2(\mu_R-\mu_S)}} \le x \le \frac{\mu_S}{2} + \sqrt{\frac{\mu_S^2}{4} + \frac{\mu_S\sigma_R^2}{2(\mu_R-\mu_S)}}.$$

Since $\mu_S \le q$, we have $\frac{\mu_S}{2} - \sqrt{\frac{\mu_S^2}{4} + \frac{\mu_S\sigma_R^2}{2(\mu_R-\mu_S)}} \le \mu_S \le q$. Hence if we can show $\theta_{\mathrm{b}} \le \frac{\mu_S}{2} + \sqrt{\frac{\mu_S^2}{4} + \frac{\mu_S\sigma_R^2}{2(\mu_R-\mu_S)}}$, then this means $\delta(x,V_3^*) \le 0$ for any $x \in (q,\max\{q,\theta_b\})$. Let $\alpha = \frac{\mu_R-\mu_S}{\sigma_R}$, $\beta = \frac{\mu_S}{\mu_R}$, and $\hat{\theta}_b = \frac{\theta_b-\mu_R}{\sigma_R}$. We can equivalently show

$$\hat{\theta}_b \le \frac{\alpha(\beta-2) + \sqrt{\alpha^2\beta^2 + 2\beta(1-\beta)}}{2(1-\beta)}. \tag{44}$$

By Lemma 2, we have $\hat{\theta}_b > -\alpha$ and $\hat{\theta}_b$ satisfies $(\hat{\theta}_b+\alpha)e^{\hat{\theta}_b^2}\mathrm{erfc}(\hat{\theta}_b) - \frac{\beta}{\sqrt{\pi}} = 0$. Using property 6 in Lemma 5 we deduce Eq. (44). □

Proof of Statement 1. By construction (see Eq. (19)), $V_2(x, C_2^*, C_2^*) > 0$ if $C_2^* > 0$. By Claim 1, we have $C_2^* > 0$, which implies that $W^*(x) > 0$ for all $x \le q$. In particular, $W^*(q) > 0$. Also by definition of $V_3(\cdot,\cdot)$ (see Eq. (20)), $V_3(x, C_4^*)$ increases in x on $[q, \infty)$. Therefore by continuity of W^* at q, we have $W^*(x) > 0$ for all $x \le \max\{q, \theta_b\}$. Similarly $V_4(x, C_5^*)$ increases in x as well (see Eq. (21) in Proposition 3), hence by continuity of W^* at θ_b, we have $W^*(x) > 0$ for all $x \in \mathbb{R}$. $\square$

Proof of Statement 10. By the definitions of $B_R(\cdot,\cdot)$, $V_1(\cdot,\cdot)$, and C_1^G (see Eqs. (34), (18) and (39)), for any $x < \theta_G < q$, $B_R(x, V_1^G)$ becomes

$$B_R(x, V_1^G) = \begin{cases} -b & \text{if } \theta_G \ge \mu_S; \\ \frac{\sigma_R^3 b f(\frac{x-\mu_R}{\sigma_R})}{(1-a^2+a\hat{\theta}_G)(\mu_S - x)^3} & \text{if } \theta_G < \mu_S, \end{cases}$$

where $f(z) = (a - \hat{\theta}_G)^3(1 - a^2 + az) - (a - z)^3(1 - a^2 + a\hat{\theta}_G)$. Note that $a > 0$ and $\hat{\theta}_G > a - \frac{1}{a}$ imply $1 - a^2 + a\hat{\theta}_G > 0$. Also $x < \theta_G < \mu_S$, hence to show $B_R(x, V_1^G) \le 0$ we only need to show $f(\frac{x-\mu_R}{\sigma_R}) \le 0$ for $x < \theta_G$, or equivalently, $f(z) \le 0$ for $z < \hat{\theta}_G$, where $\hat{\theta}_G \in (a - \frac{1}{a}, a)$. Observe that $f(\hat{\theta}_G) = 0$ and $f'(z) = a(a - \hat{\theta}_G)^3 + 3(1 - a^2 + a\hat{\theta}_G)(a - z)^2$. In fact, $f'(z) > 0$ since $\hat{\theta}_G < a$ and $1 - a^2 + a\hat{\theta}_G > 0$. Therefore $f(z) \le f(\hat{\theta}_G) = 0$ for $z < \hat{\theta}_G$. $\square$

Proof of Statement 7. If we can show $W^{G'}(x) \ge 0$ for all $x \in \mathbb{R}$, then Statement 7 easily follows since $\lim_{x \to -\infty} W^G(x) = \mu_S > 0$. Now we show $W^{G'}(x) \ge 0$ for all $x \in \mathbb{R}$ is true.

Lemma 3 implies that $\hat{\theta}_G > a - \frac{1}{a}$ (recall that $a = \frac{\mu_S - \mu_R}{\sigma_R}$), i.e., $1 - a^2 + a\hat{\theta}_G > 0$. Then one can check to see that $V_1^{G'}(\cdot) \ge 0$ on $(-\infty, \theta_G]$ (see Proposition 3 for V_1 and Eq. (39) for C_1^G). Also $V_4^{G'}(\cdot) \ge 0$ on $[q, \infty)$. Therefore it only remains to show $V_2^{G'}(\cdot) \ge 0$ on (θ_G, q), i.e., $\frac{C_3^G}{\sqrt{\pi}} + ze^{z^2}(C_2^G + C_3^G \text{erf}(z)) \ge 0$ for $z \in (\hat{\theta}_G, \hat{q})$. Recall the definition of C_2^G in Eq. (40) and apply properties 4, 13 – 15 in Lemma 5, we can get $C_2^G > 0$. Therefore we can equivalently show $\frac{C}{\sqrt{\pi}} + ze^{z^2}(1 + C\text{erf}(z)) \ge 0$ for $z \in (\hat{\theta}_G, \hat{q})$, where $C \triangleq \frac{C_3^G}{C_2^G}$. Compute

$$C_3^G + C_2^G = \begin{cases} b\left(e^{-\hat{\theta}_G^2} - \sqrt{\pi}\hat{\theta}_G \text{erfc}(\hat{\theta}_G)\right) & \text{if } \theta_G \ge \mu_S; \\ \frac{be^{-\hat{\theta}_G^2}\left(2 - 2a\hat{\theta}_G + 2\hat{\theta}_G^2 + e^{\hat{\theta}_G^2}\sqrt{\pi}\left(a - 3\hat{\theta}_G + 2a\hat{\theta}_G^2 - 2\hat{\theta}_G^3\right)\text{erfc}(\hat{\theta}_G)\right)}{2(1-a^2+a\hat{\theta}_G)} & \text{if } \theta_G < \mu_S, \end{cases}$$

and

$$C_3^G - C_2^G = \begin{cases} b\left(-e^{-\hat{\theta}_G^2} - \sqrt{\pi}\hat{\theta}_G \text{erfc}(-\hat{\theta}_G)\right) & \text{if } \theta_G \ge \mu_S; \\ \frac{be^{-\hat{\theta}_G^2}\left(-2 + 2a\hat{\theta}_G - 2\hat{\theta}_G^2 + e^{\hat{\theta}_G^2}\sqrt{\pi}\left(a - 3\hat{\theta}_G + 2a\hat{\theta}_G^2 - 2\hat{\theta}_G^3\right)(1 + \text{erf}(\hat{\theta}_G))\right)}{2(1-a^2+a\hat{\theta}_G)} & \text{if } \theta_G < \mu_S. \end{cases}$$

By property 5(iii) in Lemma 5, if $\theta_G \ge \mu_S$, then $C_3^G + C_2^G > 0$ and $C_3^G - C_2^G < 0$, hence $C = \frac{C_3^G}{C_2^G} \in (-1, 1)$. This is also true if $\theta_G < \mu_S$, since $1 - a^2 + a\hat{\theta}_G > 0$ and properties 13 – 15 in Lemma 5 together imply $C_3^G + C_2^G > 0$ and $C_3^G - C_2^G < 0$. Since $C \in (-1, 1)$, we can apply property 7 in Lemma 5 to get $\frac{C}{\sqrt{\pi}} + ze^{z^2}(1 + C\text{erf}(z))$ is increasing in z. Therefore to show $\frac{C}{\sqrt{\pi}} + ze^{z^2}(1 + C\text{erf}(z)) \ge 0$ for $z \in (\hat{\theta}_G, \hat{q})$, we only need to show $\frac{C}{\sqrt{\pi}} + \hat{\theta}_G e^{\hat{\theta}_G^2}(1 + C\text{erf}(\hat{\theta}_G)) \ge 0$. Since $C = \frac{C_3^G}{C_2^G}$ is explicitly characterized by a, b and $\hat{\theta}_G$, we denote

$$\begin{aligned} F(\hat{\theta}_G) &\triangleq \frac{C}{\sqrt{\pi}} + \hat{\theta}_G e^{\hat{\theta}_G^2}(1 + C\text{erf}(\hat{\theta}_G)) \\ &= \begin{cases} 0 & \text{if } \hat{\theta}_G \ge a; \\ \frac{e^{\hat{\theta}_G^2}(a - \hat{\theta}_G)}{2 - 2a\hat{\theta}_G + 2\hat{\theta}_G^2 + e^{\hat{\theta}_G^2}\sqrt{\pi}(-a + 3\hat{\theta}_G - 2a\hat{\theta}_G^2 + 2\hat{\theta}_G^3)\text{erf}(\hat{\theta}_G)} & \text{if } \hat{\theta}_G \in (a - \frac{1}{a}, a). \end{cases} \end{aligned}$$

It remains to show $F(\hat{\theta}_G) \geq 0$. If $\hat{\theta}_G \geq a$ this is trivial. Suppose $\hat{\theta}_G \in (a - \frac{1}{a}, a)$. By properties 13 – 15 in Lemma 5, the above denominator is nonnegative. The statement follows. $\square$

Proof of Statement 11. By definitions of $B_S(\cdot,\cdot)$, $V_2(\cdot,\cdot,\cdot)$, C_2^G and C_3^G (see Eqs. (34), (19), and (40) – (41)), for any $x \in (\theta_G, q)$, $B_S(x, V_2(x, C_2^G, C_3^G))$ becomes

$$B_S(x, V_2^G) = -V_2^G(x) + (\mu_S - x)V_2^{G'}(x) + \mu_S.$$

Since $\lim_{x \to -\infty} W^G(x) = \mu_S$ and $W^{G'}(\cdot) \geq 0$ on $\mathbb{R}$, it follows that if $\theta_G \geq \mu_S$, then $B_S(x, V_2^G) \leq 0$ for all $x \in (\theta_G, q)$. It only remains to consider the case where $\theta_G < \mu_S$. Similarly, for all $x \in (\mu_S, q)$, the desired result $B_S(x, V_2^G) \leq 0$ easily holds. Consider the remaining region $x \in (\theta_G, \mu_S]$. In this case ($\theta_G < \mu_S$), we have

$$B_S(x, V_2^G) = \frac{b}{2\left(1 - a^2 + a\hat{\theta}_G\right)} \cdot \left(2 - 2a^2 + 2a\hat{\theta}_G + F\left(\frac{x - \mu_R}{\sigma_R}\right)\right),$$

where

$$\begin{aligned} F(z) = {} & 2(a - z)\left(a - 3\hat{\theta}_G + 2a\hat{\theta}_G^2 - 2\hat{\theta}_G^3\right) - 2e^{z^2 - \hat{\theta}_G^2}\left(1 - a\hat{\theta}_G + \hat{\theta}_G^2\right)(1 - 2az + 2z^2) \\ & + e^{z^2}\sqrt{\pi}\left(a - 3\hat{\theta}_G + 2a\hat{\theta}_G^2 - 2\hat{\theta}_G^3\right)(1 - 2az + 2z^2)\left(\mathrm{erf}(\hat{\theta}_G) - \mathrm{erf}(z)\right). \end{aligned}$$

Since $\theta_G < \mu_S$, then $\hat{\theta}_G \in (a - \frac{1}{a}, a)$ where $a > 0$, and hence $1 - a^2 + a\hat{\theta}_G > 0$. Also we know $b = \mu_S - \mu_R > 0$. Therefore to complete the proof, we only need to show $F(z) \leq -2 + 2a^2 - 2a\hat{\theta}_G$ for all $z \in (\hat{\theta}_G, a]$. Applying property 16 in Lemma 5, we obtain the desired result. $\square$

Proof of Statement 12. For any $x > q$, we have $Q(x) = 0$ and hence (see Eq. (34) for definition of B_R, Proposition 3 for V_3 and Eq. (42) for C_4^G)

$$\begin{aligned} B_R(x, V_3^G) &= (\mu_R - x)V_3^{G'}(x) + \frac{1}{2}\sigma_R^2 V_3^{G''}(x) + \mu_R \\ &= -\frac{b}{(\mu_S - x)^2}\left(x(x - \mu_S) + \frac{\mu_S \sigma_R^2}{2b}\right). \end{aligned}$$

Since $x > q > \mu_S > 0$ and $b > 0$, it is clear that $B_R(x, V_3^G) \leq 0$. The result hence follows. $\square$

We now prove Corollary 2.

Proof of Corollary 2. First consider the CLV under the myopic policy. It is intuitive to see that $V(q, \text{myopic}) = \mu_S < q$, since under the Safe mode, the customer's satisfaction state will stay in the unsatisfied zone until he churns (which happens according to a hazard rate of one). Now consider the optimal policy. From the proof of Proposition 3, we know that when $\mu_S > \mu_R$,

$$V^*(q) = V_3(x, C_4^G)$$

where V_3 is defined in Eq. (20) and C_4^* defined in Eq. (42). One can therefore verify that

$$V^*(q) = C_2^G e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} + C_3^G \mathrm{erf}\left(\frac{q - \mu_R}{\sigma_R}\right) e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} + \mu_R \tag{45}$$

where C_2^G and C_3^G are defined in Eqs. (40) – (41), respectively. We now show that as $\mu_S \to q$, it must be the case $\theta_G < \mu_S$, where θ_G is defined in Lemma 3. Suppose this is not true, then by Lemma 3, θ_G must be the root of Eq. (11) and satisfies $\theta_G \in [\mu_S, q)$. This indicates that the LHS of Eq. (11) must be zero as $\theta \to q$. Compute the LHS, we have

$$\begin{aligned} \lim_{\theta \to q} \Bigg(& \exp\left(\frac{(q - \mu_R)^2}{\sigma_R^2}\right) \frac{\sqrt{\pi}(\theta - \mu_R)(q - \mu_R)}{\sigma_R^2} \left(\mathrm{erf}\left(\frac{q - \mu_R}{\sigma_R}\right) - \mathrm{erf}\left(\frac{\theta - \mu_R}{\sigma_R}\right)\right) \\ & + \exp\left(\frac{(q - \mu_R)^2 - (\theta - \mu_R)^2}{\sigma_R^2}\right) \frac{q - \mu_R}{\sigma_R} - \frac{\theta - \mu_R}{\sigma_R} - \frac{\mu_S \sigma_R}{2(\mu_S - \mu_R)(q - \mu_S)} \Bigg) \end{aligned}$$

$$= -\frac{\mu_S\sigma_R}{2(\mu_S-\mu_R)(q-\mu_S)} < 0,$$

hence a contradiction. Therefore, as $\mu_S \to q$, it must be the case $\theta_G < \mu_S$, and by Eqs. (40) – (41),

$$C_2^G = \frac{be^{-\hat{\theta}_G^2}\left(2-2a\hat{\theta}_G+2\hat{\theta}_G^2+e^{\hat{\theta}_G^2}\sqrt{\pi}\left(-a+3\hat{\theta}_G-2a\hat{\theta}_G^2+2\hat{\theta}_G^3\right)\operatorname{erf}(\hat{\theta}_G)\right)}{2(1-a^2+a\hat{\theta}_G)}$$

and

$$C_3^G = \frac{b\sqrt{\pi}\left(a-3\hat{\theta}_G+2a\hat{\theta}_G^2-2\theta_G^3\right)}{2(1-a^2+a\hat{\theta}_G)},$$

where

$$a \triangleq \frac{\mu_S-\mu_R}{\sigma_R}, \quad b \triangleq \mu_S-\mu_R, \quad \hat{\theta}_G \triangleq \frac{\theta_G-\mu_R}{\sigma_R}.$$

From Eq. (45), to show $\lim\limits_{\mu_S\to q} V^*(q) \to \infty$, it suffices to show $\lim\limits_{\mu_S\to q} C_3^G \to \infty$, i.e., $\lim\limits_{\mu_S\to q} \hat{\theta}_G \to a-\frac{1}{a}$. Recall in the proof of Lemma 3, $\hat{\theta}_G$ must be the root of the following equation:

$$\begin{aligned} F_{\text{big}}(\hat{\theta}) \triangleq &\left(a-3\hat{\theta}+2a\hat{\theta}^2-2\hat{\theta}^3\right)+2e^{\hat{q}^2-\hat{\theta}^2}\hat{q}\left(1-a\hat{\theta}+\hat{\theta}^2\right) \\ &+e^{\hat{q}^2}\sqrt{\pi}\hat{q}\left(a-3\hat{\theta}+2a\hat{\theta}^2-2\hat{\theta}^3\right)\left(\operatorname{erf}(\hat{q})-\operatorname{erf}(\hat{\theta})\right)-\frac{\mu_S\left(1-a^2+a\hat{\theta}\right)}{b(\hat{q}-a)}=0 \end{aligned}$$

on interval $\in (a-1/a, a)$, where $\hat{q} = \frac{q-\mu_R}{\sigma_R}$. Note that as $\mu_S \to q$, or equivalently, as $a \to \hat{q}$, $\hat{\theta}_G \in (a-1/a, a)$ is bounded. Thus the first three terms in the above equation have a finite limit. At the same time, the denominator of the last term goes to zero. Therefore, to satify $F_{\text{big}}(\hat{\theta}_G) = 0$, we must have $\lim\limits_{a\to\hat{q}} \hat{\theta}_G \to a-\frac{1}{a}$, or equivalently, $\lim\limits_{\mu_S\to q} \hat{\theta}_G \to a-\frac{1}{a}$. □

C Robustness check: mixed policies

In this online appendix, we consider a variant of the original model where the firm is allowed to mix between the two service modes. We first setup the model. Then we give comparative statics regarding the optimal sandwich policy (see Theorem 4) and numerical evidence about the optimal sandwich policy outperforming the myopic one. In the last part of this online appendix, we provide proofs of the optimal policy structure in Theorem 3 and the monotonicity results in Theorem 4.

Model setup. At each point in time $t \geq 0$, the firm chooses the proportion of the Risky mode $p_t \in [0,1]$ in its service, so that the reward is generated according to

$$dY_t = ((1-p_t)\mu_S + p_t\mu_R)dt + p_t\sigma_R dB_t. \tag{46}$$

We restrict attention to $\mu_R > \mu_S > 0$. This is motivated by the real world application of an investment manager, where he can mix between assets and where the riskier asset usually provides better rewards. We also require $\mu_S < q$ as in the original model to ensure finite lifetime.

Note that over time, the investor may dynamically adjust the fraction of assets invested in the stock market. The investment manager, who is paid proportionally to returns, would prefer the higher-return option — to be fully invested in the stock market at all times — but is aware that a period of poor returns could cause the customer to leave.

Analogous to the original model, a policy λ is admissible if the firm's action process $(p_t)_{t\geq 0}$ (by following this policy) is adapted to the filtration $\mathbb{F}$, takes value in $[0,1]$, and is such that the corresponding satisfaction processes is an $\mathbb{F}$-adapted semimartingale specified uniquely in law. We denote the set of admissible policies by Λ.

The optimal value function under the new policy space Λ is given by

$$V^I(x) = \sup_{\lambda \in \Lambda} \mathbb{E}\left[\int_0^\infty (\mu_S + p_t(\mu_R - \mu_S))\mathbb{1}\{t < T\}dt + \int_0^\infty p_t \sigma_R \mathbb{1}\{t < T\}dB_t \;\middle|\; H_0 = x \right]. \tag{47}$$

We call this model under policy space Λ the *Investor model*.

As in the original model, we expect that there is a stationary Markov optimal policy $\lambda : \mathbb{R} \to [0, 1]$. Next we show that interval policies, suitably generalized, are a subclass of stationary Markov policies that are admissible. First we extend the definition of interval policies to allow mixed policies.

Definition 4 (Interval policy in Investor model)**.** *In the Investor model, a policy λ is an interval policy if:*

- *it is stationary Markov, that is, the corresponding action process is given by a mapping from current satisfaction to the proportion of the Risky mode, which we denote by $p_t = \lambda(H_t)$.*
- *there is a partition of the satisfaction real line into a countable number of intervals, such that $\lambda(\cdot)$ is Lipschitz continuous within each interval, and that there exists some $c > 0$ such that $\lambda(x) \in \{0\} \cup [c, 1]$ for all $x \in \mathbb{R}$.*

Interval policies for the Investor model are admissible. The argument is similar to the proof of Lemma 1. Here, the policy may choose an arbitrary blend of the service modes at different points in the "Risky" pieces as long as the fraction of the Risky mode $\lambda(x)$ is Lipschitz continuous in the satisfaction level x within each piece, and $\lambda(x)$ is uniformly bounded below everywhere on the union of all "Risky" pieces (so that Salins and Spiliopoulos [29] still applies on the closure of each "Risky" piece). It turns out (see Theorem 3 below) that the optimal policy for the Investor model belongs to the class of interval policies.

Theorem 3. *Suppose $\mu_S < \mu_R$ and $\mu_S < q$. Consider the firm's problem as presented in Eq. (47). Let θ_I be as defined in Lemma C.1. If $\theta_I \le q$, then the myopic (pure Risky-everywhere) policy is optimal. If $\theta_I > q$, then a sandwich policy is optimal, where the proportion of the Risky mode is $\lambda^*(x)$ as defined in Lemma C.2 for satisfaction levels $x \in [q, \theta_I]$, and $\lambda^*(x) = 1$ for $x \notin [q, \theta_I]$.*

Comparative statics and numerics. Recall in Theorem 3 the structure of the optimal sandwich policy in the Investor model: the firm still uses the Risky mode for low and high levels of satisfaction, but instead of using purely the Safe mode for an intermediate satisfaction interval, the firm mixes the Risky mode with the Safe mode in $[q, \theta_I]$, where θ_I is defined in Lemma C.1 later in this online appendix. We call the interval $[q, \theta_I]$ the risk-averse region. Moreover, inside the risk-averse region, the proportion of the Risky mode which the firm employs at satisfaction level x is specified by $\lambda^*(x)$ in Lemma C.2. The next theorem provides monotonicity results regarding θ_I and $\lambda^*(x)$.

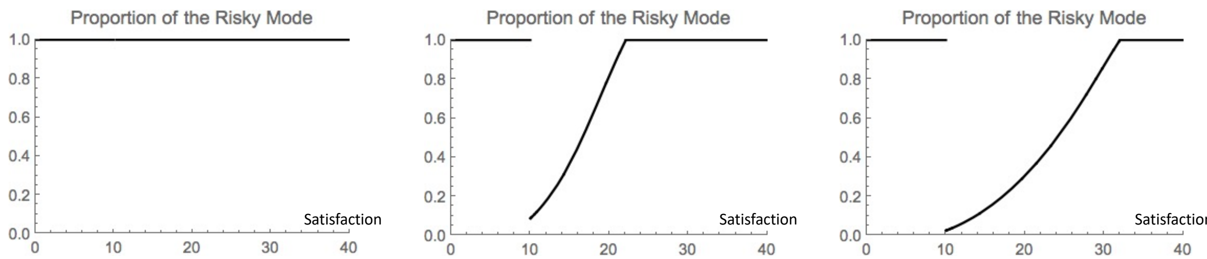
Theorem 4. *Let θ_I be as defined in Lemma C.1 and $\lambda^*(\cdot)$ be as defined in Lemma C.2. Then, the following properties hold:*

1. *The threshold θ_I increases in μ_S and σ_R.*
2. *Assume the optimal policy for a given set of parameters (see Theorem 3) is a sandwich policy. Then, the proportion of the Risky mode $\lambda^*(\cdot)$ in the risk-averse region $[q, \theta_I]$ is strictly increasing with satisfaction, and the firm strictly mixes at the satisfaction threshold q but not at θ_I; that is, we have $\lambda^*(q) > 0$ and $\lambda^*(\theta_I) = 1$.*

The first part of this theorem simply states that the size of the risk-averse region in the Investor model possesses the same monotonicities with regard to μ_S and σ_R as in the original model. The second part of Theorem 4 states that when the customer satisfaction level is in the unsatisfied zone, the firm is more risk-averse closer to the satisfaction threshold q. That is, the firm prefers a lower risk profile closer to q even though this generates a lower current reward rate. Note that $\lambda^*(q-) = 1$, and $\lambda^*(q+)$ has a value such that $(V^I)'$ is continuous at q despite the step in $Q(\cdot)$.

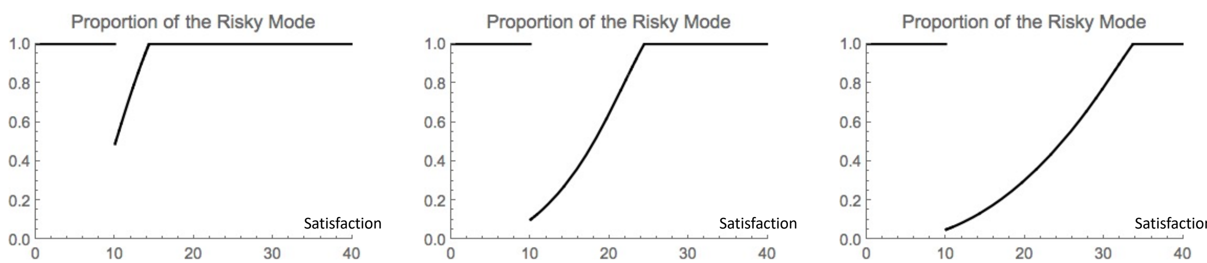
An interesting implication of part 2 of Theorem 4 is that the optimal policy never uses the Safe service mode alone. It always mixes the Risky mode with the Safe mode in the risk-averse region. The intuition is that the Risky mode has a higher drift ($\mu_R > \mu_S$), and as specified in Eq. (46), the variance is only quadratic in $\lambda(\cdot)$ while the drift is the $\lambda(\cdot)$-weighted convex combination of μ_R and μ_S, so it is always beneficial to include at least a small proportion of Risky.

We give a few numerical examples of the optimal policy in the following graphs (Figures 7–9). Notice that the model parameters considered in Figure 7 (c) here are the same as those in the original model in Figure 4. In comparison, the size of the risk-averse region here are larger than in the original model (which is $[10, 22.1]$).



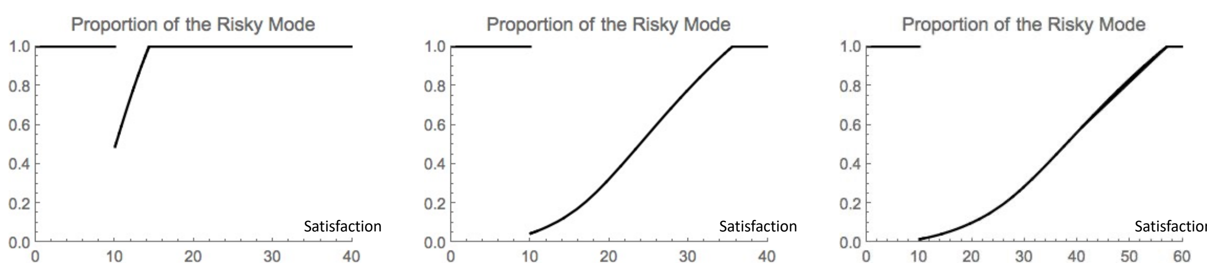
(a) $\mu_S = 3$ (b) $\mu_S = 7$ (c) $\mu_S = 8$

Figure 7: Optimal policies for the Investor model: $\mu_R = 9$, $\sigma_R = 10$, $q = 10$.



(a) $\mu_R = 9$ (b) $\mu_R = 6$ (c) $\mu_R = 5.5$

Figure 8: Optimal policies for the Investor model: $\mu_S = 5$, $\sigma_R = 10$, $q = 10$.



(a) $\sigma_R = 10$ (b) $\sigma_R = 30$ (c) $\sigma_R = 50$

Figure 9: Optimal policies for the Investor model: $\mu_S = 5$, $\mu_R = 9$, $q = 10$.

We omit to provide the improvement in CLV under the optimal policy compared with the myopic policy (pure Risky everywhere). Observe that the Investor model expands the policy space, hence increasing optimal CLV. That is, the gap for the Investor model exceeds that under the original model (Figure 5).

Proof of Theorems 3 and 4. To prove Theorem 3, we first establish two lemmas that characterize the optimal sandwich threshold θ_I and the optimal Risky proportion $\lambda^*(x)$ as a function of the satisfaction value x inside the risk-averse region $[q, \theta_I]$. First we present and prove the lemma on θ_I.

Lemma C.1. *Let Θ_I be the set of values of θ that satisfy*

$$\frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(\theta-\mu_R)^2}{\sigma_R^2}} \left(1 - \operatorname{erf}\left(\frac{\theta-\mu_R}{\sigma_R}\right)\right)(2\theta - \mu_R - \mu_S) - (\mu_R + \mu_S) = 0.$$

Then, the set $\Theta_I \cap (\frac{\mu_S+\mu_R}{2}, \infty)$ contains a single element, which we label θ_I.

Proof of Lemma C.1. Define

$$\alpha \triangleq \frac{\mu_R - \mu_S}{\sigma_R}, \quad \beta \triangleq \frac{\mu_S}{\mu_R} \quad \text{and} \quad z \triangleq \frac{\theta - \mu_R}{\sigma_R}.$$

We have $\alpha > 0$ and $0 < \beta < 1$. Define also

$$r(z) \triangleq (z + \frac{\alpha}{2}) e^{z^2} \operatorname{erfc}(z) - \frac{1+\beta}{2\sqrt{\pi}}.$$

The lemma is equivalent to showing that $r(z)$ has a unique root on $(-\frac{\alpha}{2}, \infty)$. Applying property 5(ii) in Lemma 5, we get that $r(z)$ has a unique root on $(-\frac{\alpha}{2}, \infty)$. □

Next we define the proportion $\lambda^*(x)$ of the Risky mode which the firm employs at satisfaction level x inside the risk-averse region $[q, \theta_I]$.

Lemma C.2. *Let $X(\cdot)$ be the following function:*

$$X(y) = e^{-2a\left(\log(y+1)+\frac{1}{y+1}\right)} K_0 - (-2a)^{1+2a} b e^{-2a\left(\log(y+1)+\frac{1}{y+1}\right)} \Gamma\left(-2a, -\frac{2a}{y+1}\right),$$

where

$$K_0 = \theta_I e^{2a\left(\log(g)+\frac{1}{g}\right)} + b(-2a)^{1+2a}\Gamma\left(-2a, -\frac{2a}{g}\right),$$

$$a = \frac{\sigma_R^2}{(\mu_R-\mu_S)^2}, \quad b = \mu_S, \quad g = \frac{2\theta_I}{2\theta_I - \mu_S - \mu_R},$$

and $\Gamma(s,z)$ is the upper incomplete gamma function (see Chaudry and Zubair [9]). Then, the inverse function of function $X(\cdot)$ is properly defined on $(\mu_S, \theta_I]$ and is represented by $G(\cdot) \triangleq X^{-1}(\cdot)$. Then, on $(\mu_S, \theta_I]$, the function $G(\cdot)$ is positive, strictly decreasing, differentiable, and satisfies the following ODE:

$$(G(x)+1)^2 + 2a(x-b)G(x)G'(x) - 2abG'(x) = 0.$$

Finally, define the function

$$\lambda^*(x) \triangleq \frac{(\mu_S - \mu_R)(G(x)+1)}{\sigma_R^2 G'(x)}. \tag{48}$$

Then, λ^* is strictly increasing on $[q, \theta_I]$ with $\lambda^*(q) > 0$ and $\lambda^*(\theta_I) = 1$.

The function $G(x)$ in Lemma C.2 captures the marginal benefit of satisfaction at x, when the firm uses the conjectured optimal proportion $\lambda^*(\cdot)$ of the Risky mode. The function $X(\cdot)$ is the inverse function of $G(\cdot)$.

Lemmas C.3 and C.4 are key to proving Lemma C.2.

Lemma C.3. *Let $X(y)$ be as defined in Lemma C.2. Then $X(y)$ is strictly decreasing on $\left[\frac{\mu_S+\mu_R}{2\theta_I-\mu_S-\mu_R}, \infty\right)$. Moreover, $X(\frac{\mu_S+\mu_R}{2\theta_I-\mu_S-\mu_R}) = \theta_I$ and $\lim_{y\to\infty} X(y) = \mu_S$.*

Proof of Lemma C.3. First we restate $X(y)$ here:

$$X(y) = e^{-2a\left(\log(y+1)+\frac{1}{y+1}\right)}K_0 - (-2a)^{1+2a}be^{-2a\left(\log(y+1)+\frac{1}{y+1}\right)}\Gamma\left(-2a, -\frac{2a}{y+1}\right). \tag{49}$$

Then one can verify that $X\left(\frac{\mu_S+\mu_R}{2\theta_I-\mu_S-\mu_R}\right) = \theta_I$ and $\lim_{y\to\infty} X(y) = b$. Recall by definition of θ_I in Lemma C.1 that $\theta_I > \frac{\mu_S+\mu_R}{2} > \mu_S$. Next we want to show that $X'(y) < 0$ on $\left[\frac{\mu_S+\mu_R}{2\theta_I-\mu_S-\mu_R}, \infty\right)$. Compute this derivative:

$$X'(y) = 2aye^{-\frac{2a}{y+1}}(y+1)^{-2(a+1)}F(y)$$

where

$$F(y) = \frac{be^{\frac{2a}{y+1}}(y+1)^{2a+1}}{y} + 2^{2a+1}(-a)^{2a+1}b\Gamma\left(-2a, -\frac{2a}{y+1}\right) - K_0.$$

Observe that for $y \in \left[\frac{\mu_S+\mu_R}{2\theta_I-\mu_S-\mu_R}, \infty\right)$, we have $2aye^{-\frac{2a}{y+1}}(y+1)^{-2(a+1)} > 0$. Therefore we want to show $F(y) < 0$ on $\left[\frac{\mu_S+\mu_R}{2\theta_I-\mu_S-\mu_R}, \infty\right)$. In fact, this is true since

$$F'(y) = -be^{\frac{2a}{y+1}}\frac{(1+y)^{2a}}{y^2} < 0$$

for $y \geq \frac{\mu_S+\mu_R}{2\theta_I-\mu_S-\mu_R} > 0$ and

$$F\left(\frac{\mu_S+\mu_R}{2\theta_I-\mu_S-\mu_R}\right) = \frac{e^{\frac{2a}{g}}g^{2a}\left(bg - g\theta_I + \theta_I\right)}{g-1} < 0,$$

where the last step follows from $g = \frac{2\theta_{\mathrm{I}}}{2\theta_{\mathrm{I}}-\mu_S-\mu_R} > 1$ and $b = \mu_S < \frac{\mu_S+\mu_R}{2}$. We have thus completed the proof. $\square$

Lemma C.4. *The function $\lambda^*(x)$ as defined in Lemma C.2 is strictly increasing on $[q, \theta_I]$, and $\lambda^*(q) > 0$, $\lambda^*(\theta_I) = 1$.*

Proof of Lemma C.4. Let $p^*(\cdot)$ be defined as Eq. (48), but on the domain $(\mu_S, \theta_I]$. Then on $(\mu_S, \theta_I]$, we have

$$p^*(x) = \frac{(\mu_S - \mu_R)(G(x)+1)}{\sigma_R^2 G'(x)}.$$

Also from Lemma C.2, $G(x)$ satisfies

$$(G(x)+1)^2 + 2a(x-b)G(x)G'(x) - 2abG'(x) = 0.$$

on this interval. Therefore we have

$$G'(x) = \frac{(G(x)+1)^2}{2a(b-(x-b)G(x))}$$

on $(\mu_S, \theta_I]$, and thus we can rewrite $p^*(x)$ as

$$p^*(x) = \frac{2a(\mu_S-\mu_R)(b-(x-b)G(x))}{\sigma_R^2(G(x)+1)}. \tag{50}$$

Since $\mu_S < \mu_R$, to show $p^*(x)$ is strictly increasing on $(\mu_S, \theta_I]$ is equivalent to showing $\frac{b-(x-b)G(x)}{G(x)+1}$ is strictly decreasing on $(\mu_S, \theta_I]$. Also $G(\cdot)$ is the inverse function of $X(\cdot)$, and by Lemma C.3, X is strictly decreasing on $[g-1, \infty)$, therefore it is equivalent if we can show $\frac{b-y(X(y)-b)}{y+1}$ is strictly increasing on $[g-1, \infty)$.

Denote $L(y) \triangleq \frac{b-y(X(y)-b)}{y+1}$. Compute its derivative:

$$L'(y) = 2ae^{-\frac{2a}{y+1}}(y+1)^{-2a-3}\Bigg(2^{2a+1}a(-a)^{2a}b\left(2ay^2-y-1\right)\Gamma\left(-2a, -\frac{2a}{y+1}\right) \\ - 2abye^{\frac{2a}{y+1}}(y+1)^{2a+1} + K_0\left(2ay^2-y-1\right)\Bigg).$$

We want to show that $L'(y) > 0$ on $[g-1, \infty)$. Consider y sufficiently large such that $2ay^2 - y - 1 > 0$. We want to show

$$f(y) = 2^{2a+1}a(-a)^{2a}b\Gamma\left(-2a, -\frac{2a}{y+1}\right) - \frac{2abye^{\frac{2a}{y+1}}(y+1)^{2a+1}}{2ay^2 - y - 1} + K_0 > 0.$$

One can check that $f'(y) = \frac{2ab(y+2)e^{\frac{2a}{y+1}}(y+1)^{2a+1}}{(-2ay^2+y+1)^2} > 0$. Also

$$\begin{aligned} f(g-1) &= 2^{2a+1}a(-a)^{2a}b\Gamma\left(-2a, -\frac{2a}{g}\right) - \frac{2ab(g-1)e^{\frac{2a}{g}}g^{2a+1}}{2a(g-1)^2 - g} + K_0 \\ &= -\frac{e^{\frac{2a}{g}}g^{2a+1}\left(2a(g-1)^2(\mu_S - \mu_R) + g(\mu_R + \mu_S)\right)}{2(g-1)\left(2a(g-1)^2 - g\right)}. \end{aligned}$$

If we can show g is such that $2a(g-1)^2 - g > 0$ (so that $-2ay^2 + y + 1 > 0$ for all $y \geq g - 1$) and $2a(g-1)^2(\mu_S - \mu_R) + g(\mu_R + \mu_S) \leq 0$, then we are done. In fact, we can check that the second inequality implies the first, therefore we only need to show the second one is true. Now let $\alpha \triangleq \frac{1}{2}\sqrt{\frac{1}{a}}$, $\beta \triangleq \frac{\mu_S + \mu_R}{2\mu_R}$, $\hat{\theta}_I \triangleq \frac{\theta_I - \mu_R}{\sigma_R}$, and recall the definition of $g \triangleq \frac{2\theta_I}{2\theta_I - \mu_S - \mu_R}$. After solving the desired inequality on g, we can get an equivalent desired inequality on $\hat{\theta}_I$:

$$\hat{\theta}_I \leq \frac{\alpha(\beta - 2) + \sqrt{\alpha^2\beta^2 + 2\beta(1-\beta)}}{2(1-\beta)}.$$

By Lemma C.1, $\hat{\theta}_I$ is the unique root of $F(\theta) = (x+\alpha)e^{\theta^2}\mathrm{erfc}(\theta) - \frac{\beta}{\sqrt{\pi}}$on $(-\alpha, \infty)$. Then apply property 6 in Lemma 5, we get that the desired inequality of $\hat{\theta}_I$ is true. Therefore we have proved that $p^*(x)$ is strictly increasing on $(\mu_S, \theta_I]$.

Finally, consider $\lim_{x \to \mu_S^+} p^*(x)$ and $p^*(\theta_I)$. Since $G(\cdot)$ is the inverse function of $X(\cdot)$, then by Lemma C.3 we know that $\lim_{x \to \mu_S^+} G(x) = \infty$, and $G(\theta_I) = \frac{\mu_S + \mu_R}{2\theta_I - \mu_S - \mu_R}$. Combine with Eq. (50), we get $\lim_{x \to \mu_S^+} p^*(x) = 0$, and $p^*(\theta_I) = 1$. Therefore since $p^*(\cdot) = \lambda^*(\cdot)$ on $[q, \theta_I]$ and $q > \mu_S$, we have $\lambda^*(q) = p^*(q) > 0$ and $\lambda^*(\theta_I) = p^*(\theta_I) = 1$. □

Now we are ready to prove Lemma C.2.

Proof of Lemma C.2. By Lemma C.3, the function $X(\cdot) : [g-1, \infty) \to (\mu_S, \theta_I]$ as specified in Lemma C.2 is strictly decreasing and differentiable, and $X(g-1) = \theta_I$, and $\lim_{y \to \infty} X(y) = \mu_S$. Therefore, its inverse function $G(\cdot) : (\mu_S, \theta_I] \to [g-1, \infty)$ is well-defined, strictly decreasing and differentiable, and $G'(x) = \frac{1}{X'(G(x))}$ by the inverse function theorem. Let $G(x) = y$, $x = X(y)$, and $G'(x) = \frac{1}{X'(y)}$, one can easily check that $(y+1)^2 + 2a(X(y) - b)\frac{y}{X'(y)} - \frac{2ab}{X'(y)} = 0$. Therefore $G(\cdot)$ satiesfies the ODE $(G(x)+1)^2 + 2a(x-b)G(x)G'(x) - 2abG'(x) = 0$ on $[\mu_S, \infty)$. Since $g - 1 > 0$, we have that $G(\cdot)$ is strictly positive on $(\mu_S, \theta_I]$. Finally by Lemma C.4, the function $\lambda^*(\cdot)$ is strictly increasing on $[q, \theta_I]$, with $\lambda^*(q) > 0$ and $\lambda^*(\theta_I) = 1$. □

Next we establish the full proof of Theorem 3.

Proof of Theorem 3. The proof technique is similar to one we used prove Theorem 1 (see Section 5). Recall that in Section 5, we first obtain a candidate value function $W(\cdot)$ by solving the HJB equation (13), then we prove its optimality by showing that it satisfies a set of optimality conditions (Conditions 1-6 in Proposition 2). Similarly, for the Investor model, we will first provide a candidate value function $W^I(\cdot)$ by solving the HJB equation

$$\max_{p \in [0,1]} \left\{ -Q(x)V(x) + (\mu_S + p(\mu_R - \mu_S) - x)V'(x) + \frac{1}{2}p^2\sigma_R^2 V''(x) + \mu_S + p(\mu_R - \mu_S) \right\} = 0, \quad (51)$$

(where p denotes the proportion of the Risky mode to invest in) and then prove its optimality by showing that $W^I(\cdot)$ together with the policy stated in Theorem 3 satisfies Conditions 1-6, with Condition 4^I-5^I (see below) replacing Condition 4-5.

4^I. the following inequality[14] is true for any $x \in \mathbb{R}$ when $p = 0$ and for any $x \in \mathbb{R} \setminus \mathcal{E}$ when $p \in (0, 1]$:

$$-Q(x)\bar{V}(x) + (\mu_S + p(\mu_R - \mu_S) - x)\bar{V}'(x) + \frac{1}{2}p^2\sigma_R^2\bar{V}''(x) + \mu_S + p(\mu_R - \mu_S) \leq 0; \tag{52}$$

and

5^I. for all $x \in \mathbb{R}$ and some interval policy (see Definition 4 in Online Appendix C) $\bar{\lambda} \in \Lambda$ such that $\bar{\lambda}(y) = 0$ for all $y \in \mathcal{E}$, the process $(\bar{V}(H_t^{x,\bar{\lambda}}))_{t\geq 0}$ is an $\mathbb{F}$-adapted semimartingale, and

$$-Q(x)\bar{V}(x) + (\mu_S + \bar{\lambda}(x)(\mu_R - \mu_S) - x)\bar{V}'(x) + \frac{1}{2}\bar{\lambda}(x)^2\sigma_R^2\bar{V}''(x) + \mu_S + \bar{\lambda}(x)(\mu_R - \mu_S) = 0 \tag{53}$$

Definition of $W^I(\cdot)$. Define

$$W^I(x) = \begin{cases} V_1(x, C_1^I) \text{ if } x < q; \\ V_2^I(x, C_2^I) \text{ if } q \leq x \leq \theta_I; \\ V_3(x, C_3^I) \text{ if } x > \max\{q, \theta_I\}; \end{cases} \tag{54}$$

for some uniquely specified C_1^I, C_2^I and C_3^I, where V_1 and V_3 are as defined in Eqs. (19) and (21), restated here:

$$V_1(x, C_1) = C_1 e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \left(1 + \text{erf}\left(\frac{x - \mu_R}{\sigma_R}\right)\right) + \mu_R;$$

$$V_3(x, C_3) = C_3 + \int_0^x \frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(z-\mu_R)^2}{\sigma_R^2}} (1 - \text{erf}(\frac{z - \mu_R}{\sigma_R}))dz.$$

V_2^I is defined as follows:

$$V_2^I(x, C_2) = C_2 + \int_q^x G(z)dz, \tag{55}$$

where G is as defined in Lemma C.2. Observe that $V_2^{I'}(\cdot, C_2)$ and $V_3'(\cdot, C_3)$ are independent of C_2 and C_3, hence we will use $V_2^{I'}(\cdot)$ and $V_3'(\cdot)$ to denote the two derivatives, respectively.

We now provide the explicit expressions of C_1^I, C_2^I and C_3^I:

$$C_1^I = \begin{cases} \dfrac{\frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} (1 - \text{erf}(\frac{q-\mu_R}{\sigma_R}))}{\frac{2}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} + e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} \frac{q-\mu_R}{\sigma_R}\left(1 + \text{erf}\left(\frac{q-\mu_R}{\sigma_R}\right)\right)\right)} & \text{if } \theta_I < q; \\[2ex] \dfrac{G(q)}{\frac{2}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} + e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} \frac{q-\mu_R}{\sigma_R}\left(1 + \text{erf}\left(\frac{q-\mu_R}{\sigma_R}\right)\right)\right)} & \text{if } \theta_I \geq q, \end{cases} \tag{56}$$

$$C_2^I = C_1^I e^{\frac{(q-\mu_R)^2}{\sigma_R^2}} \left(1 + \text{erf}\left(\frac{q - \mu_R}{\sigma_R}\right)\right) + \mu_R, \tag{57}$$

and

$$C_3^I = \begin{cases} V_1(q, C_1^I) - \int_0^q \frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(z-\mu_R)^2}{\sigma_R^2}} (1 - \text{erf}(\frac{z-\mu_R}{\sigma_R}))dz & \text{if } \theta_I < q; \\ V_2^I(\theta_I, C_2^I) - \int_0^{\theta_I} \frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(z-\mu_R)^2}{\sigma_R^2}} (1 - \text{erf}(\frac{z-\mu_R}{\sigma_R}))dz & \text{if } \theta_I \geq q. \end{cases} \tag{58}$$

[14] When $p = 0$, we define $\frac{1}{2}p^2\sigma_R^2\bar{V}''(x)$ to be zero for any x including $x \in \mathcal{E}$.

To reduce the burden of notation, we define $V_1^I(\cdot) \triangleq V_1(\cdot, C_1^I)$, $V_2^I(\cdot) \triangleq V_2^I(\cdot, C_2^I)$ and $V_3^I(\cdot) \triangleq V_3(\cdot, C_3^I)$.

Conditions 1**,** 2**,** 3**,** 4^I**,** 5^I **and** 6**.** Now we have an explicitly defined candidate value function $W^I(\cdot)$ and an explicit stationary Markov policy $\lambda^*(\cdot)$ (see Lemma C.2). Next we want to show that $W^I(\cdot)$ and $\lambda^*(\cdot)$ together satisfy Conditions 1-3 (see Proposition 2), 4^I and 5^I (see (52) and (53)) and 6 (see Proposition 2).

We start with Condition 1. We want to show that $W^I(\cdot) \geq 0$. By construction of V_1 (see Eq. (19)), if $C_1^I > 0$ then $V_1^I(\cdot) > 0$. Indeed this is true since $G(q) > 0$ (see Lemma C.2). Then, in particular $V_1^I(q) > 0$ and $W^I(\cdot) > 0$, since $W^I(\cdot)$ is continuous everywhere (see below) including at q and increases on $[q, \infty)$, see Eq. (54).

Condition 2 requires $W^I(\cdot)$ to be continuously differentiable everywhere and twice continuously differentiable almost everywhere. By construction, $W^I(\cdot)$ is continuously differentiable and twice continuously differentiable everywhere except possibly at q and θ_I (note that by Lemma C.2, G is differentiable on $[q, \theta_I]$). Hence, we only need to show that $W^I(\cdot)$ is differentiable at q and θ_I. Equivalently, we want to show

$$V_1^I(q) = \begin{cases} V_3^I(q) & \text{if } \theta_I < q; \\ V_2^I(q) & \text{if } \theta_I \geq q, \end{cases} \tag{59}$$

$$V_1^{I'}(q) = \begin{cases} V_3^{I'}(q) & \text{if } \theta_I < q; \\ V_2^{I'}(q) & \text{if } \theta_I \geq q, \end{cases} \tag{60}$$

and if $\theta_I \geq q$,

$$V_2^I(\theta_I) = V_3^I(\theta_I), \tag{61}$$

$$V_2^{I'}(\theta_I) = V_3^{I'}(\theta_I). \tag{62}$$

Eq. (59) is implied by the definitions of C_2^I and C_3^I (see Eqs. (57) and (58)). Eq. (60) is implied by the definition of C_1^I (see Eq. (56)). Eq. (61) is implied by the definition of C_3^I (see Eq. (58)). Eq. (62) is implied by the fact that $G^{-1}(V_3^{I'}(\theta_I)) = \theta_I$ (see Lemma C.2).

Condition 3 requires that $W^{I'}$ be bounded. Since we have just proved that $W^{I'}(\cdot)$ is continuous in $\mathbb{R}$, to show that $W^{I'}(\cdot)$ is bounded, it suffices to show $\left|\lim_{x\to-\infty} W^{I'}(x)\right| < \infty$ and $\left|\lim_{x\to\infty} W^{I'}(x)\right| < \infty$. This is equivalent to showing $\left|\lim_{x\to-\infty} V_1^{I'}(x)\right| < \infty$ and $\left|\lim_{x\to\infty} V_3^{I'}(x)\right| < \infty$. By the definitions of V_1^I (see Eqs. (19) and (56)) and V_3^I (see Eqs. (21) and (58)), we have

$$\begin{aligned}\lim_{x\to-\infty} V_1^{I'}(x) &= \lim_{x\to-\infty} \frac{2C_1^I}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} + e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \frac{x-\mu_R}{\sigma_R}\left(1 + \operatorname{erf}\left(\frac{x-\mu_R}{\sigma_R}\right)\right)\right) \\ &= \lim_{z\to\infty} \frac{2C_1^I}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} - ze^{z^2}\operatorname{erfc}(z)\right)\end{aligned}$$

and

$$\begin{aligned}\lim_{x\to\infty} V_3^{I'}(x) &= \lim_{x\to\infty} \frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \left(1 - \operatorname{erf}\left(\frac{x-\mu_R}{\sigma_R}\right)\right) \\ &= \lim_{z\to\infty} \frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{z^2}\operatorname{erfc}(z)\,,\end{aligned}$$

for $z = (x-\mu_R)/\sigma_R$. By property 2 in Lemma 5, the limits above are zero.

Conditions 5^I and 4^I respectively require $W^I(\cdot)$ and $\lambda^*(\cdot)$ to satisfy

$$\begin{aligned}-Q(x)W^I(x) + (\mu_S(1-\lambda^*(x)) + \mu_R\lambda^*(x) - x)W^{I'}(x) + \frac{1}{2}(\sigma_R\lambda^*(x))^2 W^{I''}(x) \\ + \mu_S(1-\lambda^*(x)) + \mu_R\lambda^*(x) = 0\end{aligned} \tag{63}$$

and

$$-Q(x)W^I(x) + (\mu_S(1-p) + \mu_R p - x)W^{I\prime}(x) + \frac{1}{2}(\sigma_R p)^2 W^{I\prime\prime}(x)$$
$$+ \mu_S(1-p) + \mu_R p \leq 0 \quad \text{for all } p \in [0,1] \tag{64}$$

for all $x \in \mathbb{R}$, except possibly at q and θ_I. Eq. (63) is true by the construction of $W^I(\cdot)$ and $\lambda^*(\cdot)$. Specifically, $V_1^I(x)$ satisfies

$$-V_1^I(x) + (\mu_R - x)V_1^{I\prime}(x) + \frac{1}{2}\sigma_R^2 V_1^{I\prime\prime}(x) + \mu_R = 0 \text{ for all } x < q;$$

$V_3^I(x)$ satisfies

$$(\mu_R - x)V_3^{I\prime}(x) + \frac{1}{2}\sigma_R^2 V_3^{I\prime\prime}(x) + \mu_R = 0 \text{ for all } x \geq \theta_I;$$

and $G(x)$ satisfies (see Lemma C.2)

$$(G(x)+1)^2 + 2a(x-b)G(x)G'(x) - 2abG'(x) = 0,$$

which is equivalent to (since $G'(\cdot)$ is nonzero)

$$(\mu_S(1-\lambda^*(x)) + \mu_R\lambda^*(x) - x)G(x) + \frac{1}{2}(\sigma_R\lambda^*(x))^2 G'(x)$$
$$+ \mu_S(1-\lambda^*(x)) + \mu_R\lambda^*(x) = 0 \text{ for all } x \in [q, \theta_I].$$

Next we will show Eq. (64) to complete Condition 4^I. Denote the LHS of Eq. (64) by $B(p,x,W^I)$. Then, we need to show $B(p,x,V_1^I) \leq 0$ for $x < q$, $B(p,x,V_2^I) \leq 0$ for $x \in [q,\theta_I]$, and $B(p,x,V_3^I) \leq 0$ for $x > \max\{q,\theta_I\}$; each for all $p \in [0,1]$.

Let us first start with $x < q$. We already know that $B(1,x,V_1^I) = 0$. Therefore, to show $B(p,x,V_1^I) \leq 0$ is equivalent to showing $B(1,x,V_1^I) - B(p,x,V_1^I) \geq 0$. Rearranging the terms, we can get

$$B(1,x,V_1^I) - B(p,x,V_1^I)$$
$$= (1-p)\left((\mu_R - \mu_S)V_1^{I\prime}(x) + \frac{(1+p)\sigma_R^2}{2}V_1^{I\prime\prime}(x) + \mu_R - \mu_S\right), \tag{65}$$

In fact, by properties 1, 5(iii) in Lemma 5, and the fact that $C_1^I > 0$, we have

$$V_1^{I\prime}(x) = \frac{2C_1^I}{\sigma_R}\left(\frac{1}{\sqrt{\pi}} + e^{\frac{(x-\mu_R)^2}{\sigma_R^2}}\frac{x-\mu_R}{\sigma_R}\text{erfc}\left(-\frac{x-\mu_R}{\sigma_R}\right)\right) \geq 0$$

and

$$V_1^{I\prime\prime}(x) = \frac{2C_1^I}{\sigma_R^2}\left(\frac{2(x-\mu_R)}{\sigma_R\sqrt{\pi}} + e^{\frac{(x-\mu_R)^2}{\sigma_R^2}}\left(1 + \frac{2(x-\mu_R)^2}{\sigma_R^2}\right)\text{erfc}\left(-\frac{x-\mu_R}{\sigma_R}\right)\right) \geq 0.$$

Therefore since $p \in [0,1]$ and $\mu_R > \mu_S$, following Eq. (65), we obtain the desired inequality $B(1,x,V_1^I) - B(p,x,V_1^I) \geq 0$, and we have proved Eq. (64) for $x < q$.

Now consider the case $x > \max\{q,\theta_I\}$. We know that $B(1,x,V_3^I) = 0$. Therefore, to show $B(p,x,V_3^I) \leq 0$ is equivalent to proving $B(1,x,V_3^I) - B(p,x,V_3^I) \geq 0$. Rearranging the terms, we get

$$B(1,x,V_3^I) - B(p,x,V_3^I)$$
$$= (1-p)\left((\mu_R - \mu_S)V_3^{I\prime}(x) + \frac{(1+p)\sigma_R^2}{2}V_3^{I\prime\prime}(x) + \mu_R - \mu_S\right) = (1-p)\Lambda(p), \tag{66}$$

where

$$\Lambda(p) \triangleq \frac{\mu_R\sqrt{\pi}}{\sigma_R}e^{\frac{(x-\mu_R)^2}{\sigma_R^2}}\text{erfc}\left(\frac{x-\mu_R}{\sigma_R}\right)(-p\mu_R - \mu_S + (p+1)x) - p\mu_R - \mu_S.$$

By property 5(iii) in Lemma 5, we obtain

$$\Lambda'(p) = \frac{\mu_R\sqrt{\pi}}{\sigma_R}e^{\frac{(x-\mu_R)^2}{\sigma_R^2}}\text{erfc}\left(\frac{x-\mu_R}{\sigma_R}\right)(x-\mu_R) - \mu_R \leq 0.$$

Therefore, since $p \in [0,1]$, following Eq. (66) we have

$$\begin{aligned}
B(1,x,V_3^I) - B(p,x,V_3^I) &= (1-p)\Lambda(p) \\
&\geq (1-p)\Lambda(1) \\
&= (1-p)\left(\frac{\mu_R\sqrt{\pi}}{\sigma_R} e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \operatorname{erfc}\left(\frac{x-\mu_R}{\sigma_R}\right)(2x-\mu_R-\mu_S) - \mu_R - \mu_S\right) \\
&\geq 0\ ,
\end{aligned}$$

where the last step follows from the definition of θ_I (see Lemma C.1) and property 5(ii) in Lemma 5.

Now we are only left with $x \in [q, \theta_I]$. Similar to the argument above, we want to show $B(\lambda^*(x), x, V_2^I) - B(p, x, V_2^I) \geq 0$. Rearranging the terms, we get

$$\begin{aligned}
&B(\lambda^*(x), x, V_2^I) - B(p, x, V_2^I) \\
&\quad = (\lambda^*(x) - p)\left((\mu_R - \mu_S)G(x) + \frac{\lambda^*(x)+p}{2}\sigma_R^2 G'(x) + \mu_R - \mu_S\right) \\
&\quad = -\frac{(\lambda^*(x)-p)^2\sigma_R^2}{2} G'(x) \qquad \geq 0.
\end{aligned}$$

The second step follows from the definition of $\lambda^*(x)$ (see Lemma C.2). The last step follows from the fact that $G(\cdot)$ is nonincreasing (see Lemma C.2). Finally, it remains to be shown that Condition 6 holds. Condition 6 holds straightforwardly by an application[15] of Lemma 4.

□

Finally we prove Theorem 4.

Proof of Theorem 4. The monotonicity of $\lambda^*(x)$ on $[q, \theta_I]$ and the boundary values follow from Lemma C.2. It remains to show the monotonicity of θ_I. Let

$$\begin{aligned}
F(\mu_S, \mu_R, \sigma_R, x) &\triangleq \frac{2x - \mu_S - \mu_R}{2\sigma_R} e^{\frac{(x-\mu_R)^2}{\sigma_R^2}} \left(1 - \operatorname{erf}\left(\frac{x-\mu_R}{\sigma_R}\right)\right) - \frac{\mu_S + \mu_R}{2\mu_R\sqrt{\pi}} \\
&= (y + \frac{a}{2}) e^{y^2}(1 - \operatorname{erf}(y)) - \frac{b+1}{2\sqrt{\pi}},
\end{aligned}$$

where

$$y \triangleq \frac{x-\mu_R}{\sigma_R}, \quad a \triangleq \frac{\mu_R - \mu_S}{\sigma_R}, \quad b \triangleq \frac{\mu_S}{\mu_R}.$$

By Lemma C.1, θ_I is the only root of $F(\mu_S, \mu_R, \sigma_R, \cdot)$ on $(\frac{\mu_S+\mu_R}{2}, \infty)$. From property 5(i) in Lemma 5, we obtain that $F(\mu_S, \mu_R, \sigma_R, \cdot) < 0$ for all $x \in (\frac{\mu_S+\mu_R}{2}, \theta_I)$, and $F(\mu_S, \mu_R, \sigma_R, \cdot) > 0$ for all $x \in (\theta_I, \infty)$. Hence to prove θ_I is increasing in μ_S and σ_R, it suffices to show $F(\cdot, \mu_R, \sigma_R, x)$ is decreasing and $F(\mu_S, \mu_R, \cdot, x)$ is decreasing. Equivalently, we want to show $\frac{\partial F}{\partial \mu_S}(\mu_S, \mu_R, \sigma_R, x) < 0$, and $\frac{\partial F}{\partial \sigma_R}(\mu_S, \mu_R, \sigma_R, x) < 0$ for any $\mu_R > \mu_S > 0$ and $x > \frac{\mu_S+\mu_R}{2}$. Suppose now we fix such μ_R, μ_S, σ_R and x. Then we have $y > \frac{a}{2}$. Compute the partial derivatives:

$$\begin{aligned}
\frac{\partial F}{\partial \mu_S}(\mu_S, \mu_R, \sigma_R, x) &= -\frac{1}{2\sigma_R} e^{y^2} \operatorname{erfc}(y) - \frac{1}{2\mu_R\sqrt{\pi}}; \\
\frac{\partial F}{\partial \sigma_R}(\mu_S, \mu_R, \sigma_R, x) &= -\frac{1}{2\sigma_R}\left[e^{y^2}\operatorname{erfc}(y)(2y^2+1) - \frac{2y}{\sqrt{\pi}}\right](a + 2y).
\end{aligned}$$

[15] Lemma 4 applies here as long as the customer lifetime is finite in expectation. Recall the proof of Proposition 1 for finite lifetime. Consider the satisfaction process under the policy stated in Theorem 3. If the initial satisfaction level is below q, then we are done since satisfaction process spends positive measure of time below q. On the other hand, if the initial satisfaction level is above q, then the first passage time to q must have finite expectation since both μ and σ terms are Lipschitz and bounded below, and hence we are done.

The first result needed, $\frac{\partial F}{\partial \mu_S}(\mu_S, \mu_R, \sigma_R, x) < 0$, follows from the fact that $\mathrm{erfc}(\cdot) > 0$ everywhere. The second, $\frac{\partial F}{\partial \sigma_R}(\mu_S, \mu_R, \sigma_R, x) \leq 0$, follows from the fact that $x > \frac{\mu_S+\mu_R}{2}$ (so that $a + 2y > 0$) and using a Chernoff-type bound of the error function (Chang et al. [8]), $\mathrm{erfc}(y) \geq \frac{2y}{\sqrt{\pi}(2y^2+1)} e^{-y^2}$. This completes the proof. □

D Robustness check: Geometric Brownian Motion reward process

Motivated by an investment problem where a risk-free asset generates returns (interest) with percentage drift μ_S deterministically and a risky asset generates returns (capital gains and dividends, with automatic reinvestment of dividends) with percentage drift μ_R and percentage volatility σ_R, in this online appendix, we consider a model that uses geometric Brownian motion (GBM) reward process instead of arithmetic Brownian motion reward. We will first define the model. Then we will establish optimality conditions (Proposition 4), explain how we numerically solve for the optimal value function, and finish by presenting our numerical findings.

Model setup. In this model, under the firm's choice of service mode $u_t \in \{R, S\}$, the total reward $\tilde{Y}_t$ evolves according to a Geometric Brownian Motion (GBM)

$$d\tilde{Y}_t = \mu_{u_t}\tilde{Y}_t dt + \sigma_{u_t}\tilde{Y}_t dB_t \tag{67}$$

with[16] $\tilde{Y}_0 = 1$. We are interested in cases where $\mu_R > \mu_S > 0$. (However, we do not impose the restriction $\mu_R > \mu_S$ for our analytical development.) Customer satisfaction $\tilde{H}_t$ follows a stochastic differential equation:

$$d\tilde{H}_t = d\tilde{Y}_t/\tilde{Y}_t - \tilde{H}_t dt, \tag{68}$$

where $\tilde{H}_0 = x$ is the initial customer satisfaction. Comparing Eqs. (67) and (68) with Eqs. (1) and (3), we see that under the same action process u_t, customer satisfaction $\tilde{H}_t$ follows the same dynamics as H_t does in the original model. We assume that the hazard rate of customer churn is still a step function but is positive even for $\tilde{H}_t \geq q$. Specifically,

$$\tilde{Q}(\tilde{H}_t) \triangleq Q_1 \mathbb{1}\{\tilde{H}_t < q\} + Q_2 \mathbb{1}\{\tilde{H}_t \geq q\} \tag{69}$$

with $Q_1 > Q_2 > 0$. Then the customer's survival probability $\tilde{S}_t$ at time t is given by

$$\tilde{S}_t \triangleq P\left(\tilde{T} > t \mid \tilde{\mathcal{F}}_t\right) = e^{-\int_0^t \tilde{Q}(\tilde{H}_s)ds}. \tag{70}$$

Denote by $\tilde{T}$ the customer lifetime:

$$\tilde{T} \triangleq \inf\left\{t \geq 0 : e^{-\int_0^t \tilde{Q}(\tilde{H}_s)ds} = w\right\}, \tag{71}$$

where w is a uniform random variable over $[0, 1]$ independent of filtration $\mathbb{F}$.

We require the next condition on Q_1 and Q_2 to ensure that the expected CLV (which will be defined next) is finite[17].

Condition 1. $Q_1 > Q_2 > \max\{\mu_S, \mu_R\}$.

Denote by $\tilde{\Pi}$ the space of admissible policies that satisfy the usual conditions as in the original model, which is that under the policy, the corresponding action process u_t should be adapted to filtration $\mathbb{F}$, takes value in $\{S, R\}$, and the corresponding $\tilde{H}_t$ is an $\mathbb{F}$-adapted semimartingale uniquely specified in law. For a given starting satisfaction x and admissible policy π, let $\tilde{Y}_t^{x,\pi}$ denote the reward gained up to time t and $\tilde{T}^{x,\pi}$ be the customer lifetime. Then the CLV is equal to

$$\tilde{V}(x, \pi) = \mathbb{E}\left[1 + \int_0^\infty \mathbb{1}\{t < T^{x,\pi}\} d\tilde{Y}_t^{x,\pi} \;\middle|\; \tilde{H}_0 = x\right]. \tag{72}$$

[16]The solution to Eq. (67) is $\tilde{Y}_t = \exp\left(\int_0^t \left(\mu_{u_s} - \sigma_{u_s}^2/2\right) ds + \int_0^t \sigma_{u_s} dB_s\right)$.

[17]Suppose the hazard rate is Q for all satisfaction states and the firm always uses the Risky service mode, then it is not hard to show that the expected CLV is $\frac{Q}{Q-\mu_R}$ if $Q > \mu_R$, and ∞ if $Q \leq \mu_R$.

The firm's objective is to maximize the CLV it earns from interacting with the customer. The optimal CLV given a starting satisfaction x is given by

$$\tilde{V}^*(x) = \sup_{\pi \in \tilde{\Pi}} \tilde{V}(x, \pi). \tag{73}$$

Next we establish optimality conditions in order to find the optimal value function.

Optimality conditions. As in the original model, we present the optimality conditions for a function $\tilde{W} : \mathbb{R} \to \mathbb{R}$ to be the optimal value function.

Proposition 4. *Suppose a function $\tilde{W} : \mathbb{R} \to \mathbb{R}$ satisfies*

1. *the function value $\tilde{W}(x) > 1$ for any $x \in \mathbb{R}$;*
2. *the function $\tilde{W}$ is continuously differentiable everywhere on $\mathbb{R}$ and twice continuously differentiable everywhere on $\mathbb{R} \setminus \mathcal{E}$ for some countable set $\mathcal{E}$;*
3. *the function $\tilde{W}$ is bounded;*
4. *the function $\tilde{W}'$ is bounded;*
5. *for any $x \in \mathbb{R}$ for $i = S$ and for any $x \in \mathbb{R} \setminus \mathcal{E}$ for $i = R$ the following holds[18]:*
$$\tilde{Q}(x) + (\mu_i - \tilde{Q}(x))\tilde{W}(x) + \left(\mu_i + \sigma_i^2 - x\right)\tilde{W}'(x) + \frac{1}{2}\sigma_i^2 \tilde{W}''(x) \leq 0;$$
6. *for some interval policy $\tilde{\pi}$ (see Definition 2) such that $\tilde{\pi}(y) = S$ for all $y \in \mathcal{E}$, the process $\tilde{W}(\tilde{H}_t)$ is an $\mathbb{F}$-adapted semimartingale, and for all $x \in \mathbb{R}$ it holds that*
$$\tilde{Q}(x) + (\mu_{\tilde{\pi}(x)} - \tilde{Q}(x))\tilde{W}(x) + \left(\mu_{\tilde{\pi}(x)} + \sigma_{\tilde{\pi}(x)}^2 - x\right)\tilde{W}'(x) + \frac{1}{2}\sigma_{\tilde{\pi}(x)}^2 \tilde{W}''(x) = 0. \tag{74}$$

Then, the function $\tilde{W}$ is the optimal value function $\tilde{V}^*$, and $\tilde{\pi}$ is an optimal policy.

Note that Conditions 5 and 6 together establish the HJB equation for this new setting:

$$\max_{i \in \{S,R\}} \left\{ \tilde{Q}(x) + (\mu_i - \tilde{Q}(x))\tilde{V}(x) + \left(\mu_i + \sigma_i^2 - x\right)\tilde{V}'(x) + \frac{1}{2}\sigma_i^2 \tilde{V}''(x) \right\} = 0 \tag{75}$$

for all $x \in \mathbb{R}$ where $\tilde{V}''$ exists.

Before we present the proof of Proposition 4, we want to give an overview of the remainder of the GBM online appendix. Recall from the analysis of the original model (Section 5), we would like to solve the HJB equation to get a candidate value function such that all optimality conditions are satisfied. A general solution to the HJB equation (75) (for a single service mode) is of the following form:

$$\begin{aligned}\tilde{W}(x, C_1, C_2) = \frac{\tilde{Q}(x)}{\tilde{Q}(x) - \mu_R} &+ C_1 H\left(\mu_R - \tilde{Q}(x), \frac{x - \mu_R - \sigma_R^2}{\sigma_R}\right) \\ &+ C_2 M\left(\frac{\tilde{Q}(x) - \mu_R}{2}, \frac{1}{2}, \frac{\left(x - \mu_R - \sigma_R^2\right)^2}{\sigma_R^2}\right)\end{aligned} \tag{76}$$

under the Risky service mode, and

$$\tilde{W}(x, C_3, S) = \frac{\tilde{Q}(x)}{\tilde{Q}(x) - \mu_S} + C_3(\mu_S - x)^{\mu_S - \tilde{Q}(x)}$$

under the Safe service mode, where C_1, C_2, C_3 are free parameters, and in Eq. (76) $H(\cdot, \cdot)$ is a Hermite polynomial function, and $M(\cdot, \cdot, \cdot)$ is the Kummer confluent hypergeometric function.[19] We

[18] For any $x \in \mathcal{E}$, we define the term $\frac{1}{2}\sigma_S^2 \tilde{W}''(x)$ to be zero consistent with $\sigma_S = 0$.

[19] The two functions $H(\lambda, x)$ and $M(-\frac{\lambda}{2}, \frac{1}{2}, x^2)$ are the two linearly independent solutions to the Hermite Differential Equation $y''(x) - 2xy'(x) + 2\lambda y(x) = 0$.

then numerically find the values of C_1, C_2 and C_3 such that the optimality conditions in Proposition 4 are satisfied. Later in this online appendix, we describe how we numerically solve for the optimal policy and value function, and provide our numerical findings, which shows that our main structural results for the original model also hold in the GBM setting. We defer the proof of Proposition 4 to the end of the GBM online appendix.

How we numerically solve for the optimal policy and value function. We need to numerically find the values of C_1, C_2 and C_3 of the following functions (that solves Eq. (74) for the Risky service mode and for the Safe mode, respectively,)

$$\tilde{W}(x, C_1, C_2, R) = \frac{\tilde{Q}(x)}{\tilde{Q}(x) - \mu_R} + C_1 H\left(\mu_R - \tilde{Q}(x), \frac{x - \mu_R - \sigma_R^2}{\sigma_R}\right) + C_2 M\left(\frac{\tilde{Q}(x) - \mu_R}{2}, \frac{1}{2}, \frac{\left(x - \mu_R - \sigma_R^2\right)^2}{\sigma_R^2}\right)$$

and

$$\tilde{W}(x, C_3, S) = \frac{\tilde{Q}(x)}{\tilde{Q}(x) - \mu_S} + C_3(\mu_S - x)^{\mu_S - \tilde{Q}(x)}$$

in different satisfaction regions where the firm chooses different service modes, such that the optimality conditions in Proposition 4 are satisfied.

In the above expressions, $H(\cdot, \cdot)$ is a Hermite Polynomial, and $M(\cdot, \cdot, \cdot)$ is the Kummer confluent hypergeometric function: the two functions $H(\lambda, x)$ and $M(-\frac{\lambda}{2}, \frac{1}{2}, x^2)$ are the two linearly independent solutions to the Hermite Differential Equation $y''(x) - 2xy'(x) + 2\lambda y(x) = 0$. One challenge of calculating this $\tilde{W}$ is that, as the satisfaction value x decreases, both $H()$ and $M()$ grow exponentially. Since both functions cannot be evaluated to their exact values, the error in calculation $\tilde{W}$ (which is the difference of two very large numbers) can get large for negative x with large magnitude. To control for this error, we place a reflecting boundary at $q - B$ in the unsatisfied zone for some large $B > 0$ and let the satisfaction process only evolve on $[q - B, \infty)$. Recall that the satisfaction process is an Ornstein-Uhlenbeck (O-U) process in the unsatisfied zone, if the firm always utilizes the Risky service mode there. With a reflecting boundary $q - B < q$, it becomes a reflected O-U process. To preserve our insights, we want to choose B large enough that the customer churns before hitting the reflecting boundary $q - B$ with probability close to 1.

We use the following method to choose the reflecting boundary $q - B$. Consider a reflected O-U process $\tilde{X}_t$ on $(-\infty, q]$ with infinitesimal drift $\mu_R - \tilde{X}_t$, infinitesimal volatility σ_R, initial value $\tilde{X}_0 = q$, and reflecting boundary at the satisfaction threshold q. Note that this is an approximation of the satisfaction process $\tilde{H}_t$ under the sandwich policy with the risk-averse region right above the unsatisfied zone, which we conjecture to be optimal, if not the myopic policy. Under this policy, the satisfaction process becomes a delayed reflected O-U process on $(-\infty, q]$ once it hits the unsatisfied zone. Notice the difference between process $\tilde{X}_t$ and the satisfaction process just stated — $\tilde{X}_t$ has instantaneous reflection at q while $\tilde{H}_t$ has delayed reflection at q. Nevertheless, this means that the stationary probability of $\tilde{X}_\infty < q - B$ is an overestimation of the stationary probability of $\tilde{H}_\infty < q - B$, and we can bound the latter by bounding $\Pr\{\tilde{X}_\infty < q - B\}$. From Ward and Glynn's paper [38], we know the stationary distribution of $\tilde{X}_t < q - B$ is $\Pr\left[N\left(\mu_R, \frac{\sigma_R^2}{2}\right) < q - B \middle| N\left(\mu_R, \frac{\sigma_R^2}{2}\right) \leq q\right]$. Hence we would like to choose B such that this probability is small.

Numerical findings. Next we give details of how random instances are generated to compute the optimal policies. We are interested in the regime where $q > \mu_S$, which implies that the customer is eventually not satisfied if the firm uses the Safe mode all the time. We also let $\mu_R > \mu_S$, so that the Risky asset accumulates higher rewards on average. We solve for the optimal policy for random instances and make careful use of a reflecting boundary $q - B$ for some large positive B

to ensure numerical stability while ensuring that the effect on CLV is very small. In particular, in each iteration, the parameters are randomly generated in the following sequence:

1. randomly generate $\sigma_R \sim \text{Uniform}[0.2, 1]$;
2. randomly generate $\mu_R \sim \text{Uniform}\,[0, 0.8]$;
3. randomly generate $\mu_S \sim \text{Uniform}[0, \mu_R]$;
4. randomly generate $q = \mu_S + \text{Uniform}[0, 0.8]$;
5. assign $B = q - \mu_R + 5\sigma_R$;
6. randomly generate $Q_1 \sim \text{Uniform}[\mu_R, \mu_R + 5]$;
7. randomly generate $Q_2 \sim \text{Uniform}[\mu_R, Q_1]$.

Note that the range of the specifications of μ_S, μ_R and σ_R are chosen to have a similar magnitude with the GBM drift and volatility estimations from the financial market (for example, see Schneider et al. [30]). Also note that the choice of B is made to ensure that the second argument $\frac{x-\mu_R-\sigma_R^2}{\sigma_R}$ inside the Hermite Polynomials is always bounded below on $[q-B, \infty)$, in order to ensure numerical stability. We numerically solve 1000 random instances generated by the above procedure, by solving for the free parameters C_1 and C_2 in Eq. (76) and verify that all the conditions in Proposition 4 are satisfied. In each instance, the stationary probability of a reflected O-U process $\tilde{H}_t < q - B$ is calculated. In fact, they are all less than 10^{-8}. Therefore we can say that with very high probability, the customer churns before hitting the boundary $q - B$, and that placing a boundary at $q - B$ will very likely not affect the firm's optimal policy. In fact, we also check this by perturbing the choice of B within $[q-\mu_R+4\sigma_R, q-\mu_R+5\sigma_R]$ and showing that both the value function and the sandwich structure are extremely insensitive to the choice of B. Remarkably, in all the randomly generated instances, the optimal policy is either a myopic policy (Risky mode everywhere) or a sandwich policy (Safe mode only in an interval just above q). More details are presented next.

Optimal policy is either myopic or sandwich. As in the original model (see Theorem 1 for the $\mu_R > \mu_S$ case), the optimal policy is either a myopic one or a sandwich policy. In particular, among all the randomly generated instances, the (numerically solved) optimal solutions are either a myopic policy that always uses the Risky mode everywhere, or a sandwich policy that uses the Risky mode for all satisfaction states except for some intermediate satisfaction range $[q, \tilde{\theta}_b]$ (for some numerically specified $\tilde{\theta}_b$). It is worth noting that in the GBM setting, the optimal sandwich policy once again provides substantial CLV increase over the myopic policy. For example, consider the model primitives $\mu_S = 0.12$, $\mu_R = 0.14$, $\sigma_R = 0.3$, $q = 0.13$, $Q_1 = 1.5$ and $Q_2 = 0.5$. (Here one may think of 1 time unit in the model as being a period of about 2 years.) This means the Safe asset's rate of return is 12% (with continuous compounding), and the Risky asset's expected rate of return is 14%, with volatility 30%. Also the customer is not satisfied with a rate of return below 13%, and his hazard rate of churn increases from 0.5 to 1.5 if he estimates (as quantified by his satisfaction) that the rate of return is below 13%. Under this set of model primitives, the optimal sandwich policy (see the dotted vertical line in Figure 10) is to use the Safe mode for satisfaction values on $[0.13, 0.147]$, and use the Risky mode elsewhere. The CLV increase from using the optimal sandwich policy relative to the myopic policy is 7.0%. Though we do not permit mixed strategies in this section, we briefly observe that the CLV increase relative to the myopic policy will be even larger if mixed strategies are permitted, since the myopic strategy remains unaffected.

Optimal switching threshold exhibits similar monotonicity as in the original model. The optimal switching threshold $\tilde{\theta}_b$ shows similar monotonicity in model primitives as in the original model (see Theorem 2). That is, $\tilde{\theta}_b$ decreases as we increase μ_R, and increases as we increase σ_R.

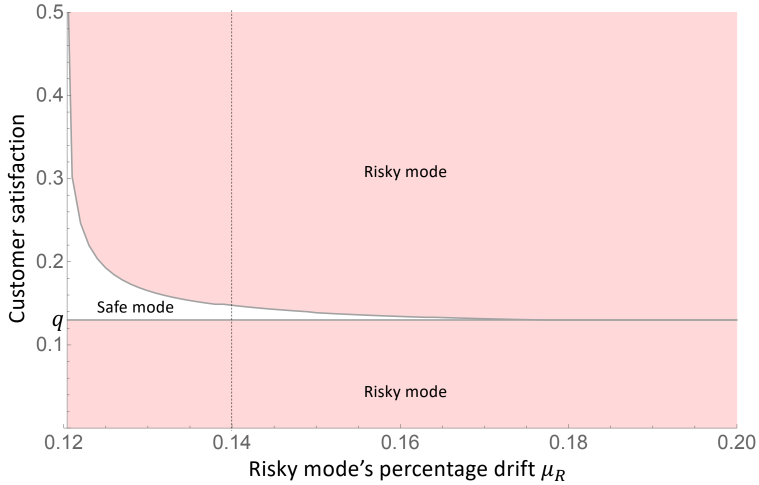


Figure 10: The optimal sandwich policies for different values of μ_R, under $\mu_S = 0.12$, $\sigma_R = 0.3$, $q = 0.13$, $Q_1 = 1.5$ and $Q_2 = 0.5$. The horizontal axis corresponds to the value of μ_R, and the vertical line marks the satisfaction value. The two curves are the switching boundaries between the Risky mode and the Safe mode.

Figure 10 plots the prescribed optimal sandwich policy for various model primitives. In particular, we fix $\mu_S = 0.12$, $\sigma_R = 0.3$, $q = 0.13$, $Q_1 = 1.5$, $Q_2 = 0.5$, and vary μ_R in $[0.12, 0.20]$. The horizontal axis is μ_R, and the vertical axis is the satisfaction value. The two curves on the plot are the two switching thresholds, separating the satisfaction regions where the firm should use different service modes. The white region represents the risk-averse region where the firm should choose the Safe service mode. Notice that $\tilde{\theta}_b$ here has the same monotonicity (i.e., decreasing in μ_R) as θ_b in the original model (see Theorem 2).

We end this online appendix with the proof of Proposition 4.

Proof of Proposition 4. In order to prove Proposition 4, we first establish the following lemma:

Lemma D.1. *For any admissible policy $\pi \in \tilde{\Pi}$, any starting satisfaction $x \in \mathbb{R}$ and any $t > 0$, the following holds:*

$$\mathbb{E}\left[\int_0^t \left\{\tilde{Y}_s^{x,\pi}\right\}^2 ds\right] < \infty,$$

where $\tilde{Y}_s$ is as defined in Eq. (67). In other words, the process $\tilde{Y}_s^{x,\pi} \in L^2[0,t]$, and the stochastic integral $\int_0^t \tilde{Y}_s^{x,\pi} dB_s$ is a martingale for any $t > 0$, which then implies that

$$\mathbb{E}\left[\int_0^t \tilde{Y}_s^{x,\pi} dB_s\right] = 0.$$

Proof. Fix an admissible policy $\pi \in \tilde{\Pi}$, a starting satisfaction $x \in \mathbb{R}$, and a time $t > 0$. Denote u_t the corresponding action process. Since the solution to Eq. (67) is

$$\tilde{Y}_t = \exp\left(\int_0^t \left(\mu_{u_s} - \sigma_{u_s}^2/2\right) ds + \int_0^t \sigma_{u_s} dB_s\right),$$

we have

$$\begin{aligned}\mathbb{E}\left[\int_0^t \left\{\tilde{Y}_s^{x,\pi}\right\}^2 ds\right] &= \int_0^t \mathbb{E}\left[\left\{\tilde{Y}_s^{x,\pi}\right\}^2\right] ds \\ &= \int_0^t \mathbb{E}\left[e^{\int_0^s \left(2\mu_{u_z} - \sigma_{u_z}^2\right) dz + \int_0^s 2\sigma_{u_z} dB_z}\right] ds\end{aligned}$$

$$
\begin{aligned}
&= \int_0^t \mathbb{E}\left[e^{\int_0^s \left(2\mu_{u_z}+\sigma_{u_z}^2\right)dz}\right] ds \\
&\le \int_0^t e^{\left(2\mu_R+\sigma_R^2\right)s} ds \\
&= \frac{e^{\left(2\mu_R+\sigma_R^2\right)t}-1}{2\mu_R+\sigma_R^2} < \infty,
\end{aligned}
$$

where the inequality results from the fact that $\mu_S < \mu_R$ and $\sigma_S < \sigma_R$. Hence we have showed that $\tilde{Y}_s^{x,\pi} \in L^2[0,t]$, and that by the Martingale property of stochastic integrals, the process $\int_0^t \tilde{Y}_s^{x,\pi} dB_s$ is a martingale for any $t > 0$. □

Now we are ready to prove Proposition 4.

Proof of Proposition 4. As in the proof of Proposition 2, we will show that a function $\tilde{W}$ as described in Proposition 4 is an upper bound of the optimal CLV $\tilde{V}^*$, and that the gap $\tilde{W} - \tilde{V}^*$ is zero.

To show that a function $\tilde{W}$ as described in Proposition 4 is an upper bound for $\tilde{V}^*$, it suffices to show $\tilde{W}(x) \ge \tilde{V}(x,\pi)$ for $\forall x \in \mathbb{R}$ and for any admissible policy $\pi \in \tilde{\Pi}$. Now fix any $x \in \mathbb{R}$ and any $\pi \in \tilde{\Pi}$. Let u_t denote the action process under policy π. Define a process $X_t, t \ge 0$ by

$$
X_t = \tilde{W}(\tilde{H}_t)\tilde{Y}_t\tilde{S}_t + \int_0^t \tilde{Y}_s\tilde{Q}(\tilde{H}_s)\tilde{S}_s ds, \tag{77}
$$

where $\tilde{H}_t$ is the satisfaction process under policy π and initial satisfaction x (see Eq. (68)), $\tilde{Y}_t$ the corresponding cumulative reward (conditional on no quitting) up to time t (see Eq. (67)), and $\tilde{S}_t$ is the corresponding customer survival probability at time t (see Eq. (70)). Since π is admissible, the corresponding process $\tilde{H}_t$ is a semimartingale (and hence $\tilde{Y}_t$ and $\tilde{S}_t$ are also semimartingales). Next we expand X_t in integral form.

Since $\tilde{W}$ is continuously differentiable everywhere on $\mathbb{R}$ and twice continuously differentiable everywhere on $\mathbb{R} \setminus \mathcal{E}$ for some countable set $\mathcal{E}$ (Condition 2 in Proposition 4), we can apply the Itô-Tanaka formula to conclude that $\tilde{W}(\tilde{H}_t)$ is also a semimartingale:

$$
\begin{aligned}
\tilde{W}(\tilde{H}_t) = \tilde{W}(x) &+ \int_0^t \tilde{W}'(\tilde{H}_s)(\mu_{u_s} - \tilde{H}_s)ds + \int_0^t \tilde{W}'(\tilde{H}_s)\sigma_{u_s} dB_s \\
&+ \frac{1}{2}\int_0^t \mathbb{1}\{\tilde{H}_s \notin \mathcal{E}\}\tilde{W}''(\tilde{H}_s)\sigma_{u_s}^2 ds + \frac{1}{2}\sum_{y\in\mathcal{E}}(\tilde{W}_r'(y) - \tilde{W}_l'(y))L^{\tilde{H}}(t,y),
\end{aligned} \tag{78}
$$

where $\tilde{W}_r'$ and $\tilde{W}_l'$ are the right and left derivatives of $\tilde{W}$, and $L^{\tilde{H}}(t,y)$ is the symmetric local time of $\tilde{H}_t$ at y. In fact, we can still apply the results of Lemma 6 to the GBM setting, since the dynamics of the satisfaction process in GMB setting is the same with the satisfaction process in the original model (if under the same action process). From Lemma 6, we know that $L^{\tilde{H}}(t,y) < \infty$ almost surely, and since $\tilde{W}$ is continuously differentiable everywhere, we have $\tilde{W}_r'(y) = \tilde{W}_l'(y)$ and hence the last term involving the local time in Eq. 78 is zero almost surely.

Since $\tilde{W}(\tilde{H}_t), \tilde{Y}_t$ and $\tilde{S}_t$ are all semimartingales, we can then apply the multi-dimensional Itô's formula on semimartingales to $g(\tilde{W}(\tilde{H}_t), \tilde{Y}_t, \tilde{S}_t) = \tilde{W}(\tilde{H}_t)\tilde{Y}_t\tilde{S}_t$ and rewrite X_t as:

$$
\begin{aligned}
X_t =& \tilde{W}(x) \\
&+ \int_0^t \tilde{Y}_s\tilde{S}_s\tilde{W}'(\tilde{H}_s)\left(\mu_{u_s} - \tilde{H}_s\right)ds + \int_0^t \tilde{Y}_s\tilde{S}_s\tilde{W}'(\tilde{H}_s)\sigma_{u_s} dB_s \\
&- \int_0^t \tilde{Y}_s\tilde{S}_s\tilde{W}(\tilde{H}_s)\tilde{Q}(\tilde{H}_s)ds + \frac{1}{2}\int_0^t \tilde{Y}_s\tilde{S}_s\tilde{W}''(\tilde{H}_s)\mathbb{1}\{\tilde{H}_s \notin \mathcal{E}\}\sigma_{u_s}^2 ds
\end{aligned}
$$

$$
\begin{aligned}
&+ \int_0^t \tilde{Y}_s \tilde{S}_s \tilde{W}(\tilde{H}_s)\mu_{u_s} ds + \int_0^t \tilde{Y}_s \tilde{S}_s \tilde{W}(\tilde{H}_s)\sigma_{u_s} dB_s \\
&+ \int_0^t \tilde{Y}_s \tilde{S}_s \tilde{W}'(\tilde{H}_s)\sigma_{u_s}^2 ds + \int_0^t \tilde{Y}_s \tilde{S}_s \tilde{Q}(\tilde{H}_s) ds. \qquad (79)
\end{aligned}
$$

Note that by the Martingale property of stochastic integrals, the two stochastic integrals above have zero expectations if $\tilde{Y}_s \tilde{S}_s \tilde{W}'(\tilde{H}_s)\sigma_{u_s}$ and $\tilde{Y}_s \tilde{S}_s \tilde{W}(\tilde{H}_s)\sigma_{u_s}$ are in the $L^2[0,t]$ space. This is indeed true since $\tilde{Y}_s^{x,\pi} \in L^2[0,t]$ (see Lemma D.1), $\tilde{S}_s \in [0,1]$, $\tilde{W}$ is bounded (see Condition 3 in Proposition 4), $\tilde{W}'$ is bounded (see Condition 4 in the proposition), and $\sigma_{u_s} \in \{0, \sigma_R\}$. Hence we can take expectation on both sides of Eq. (79) and remove the two stochastic integrals, while replacing 1 with $\mathbb{1}\{\tilde{H}_s \notin \mathcal{E}\} + \mathbb{1}\{\tilde{H}_s \in \mathcal{E}\ \&\ u_s = S\} + \mathbb{1}\{\tilde{H}_s \in \mathcal{E}\ \&\ u_s = R\}$, and get

$$
\begin{aligned}
\mathbb{E}X_t = \tilde{W}(x) + \mathbb{E}\int_0^t \tilde{Y}_s \tilde{S}_s \Bigg[&\Bigg(\tilde{Q}(\tilde{H}_s) + \Big(\mu_{u_s} - \tilde{Q}(\tilde{H}_s)\Big) \tilde{W}(\tilde{H}_s) + (\mu_{u_s} + \sigma_{u_s}^2 - \tilde{H}_s)\tilde{W}'(\tilde{H}_s) \\
&\quad + \frac{1}{2}\sigma_{u_s}^2 \tilde{W}''(\tilde{H}_s) \Bigg) \mathbb{1}\{\tilde{H}_s \notin \mathcal{E}\} \\
&+ \Bigg(\tilde{Q}(\tilde{H}_s) + \Big(\mu_{u_s} - \tilde{Q}(\tilde{H}_s)\Big) \tilde{W}(\tilde{H}_s) + (\mu_{u_s} + \sigma_{u_s}^2 - \tilde{H}_s)\tilde{W}'(\tilde{H}_s) \Bigg) \mathbb{1}\{\tilde{H}_s \in \mathcal{E}\ \&\ u_s = S\} \\
&+ \Bigg(\tilde{Q}(\tilde{H}_s) + \Big(\mu_{u_s} - \tilde{Q}(\tilde{H}_s)\Big) \tilde{W}(\tilde{H}_s) + (\mu_{u_s} + \sigma_{u_s}^2 - \tilde{H}_s)\tilde{W}'(\tilde{H}_s) \Bigg) \mathbb{1}\{\tilde{H}_s \in \mathcal{E}\ \&\ u_s = R\} \Bigg] ds \\
\le \tilde{W}(x) + \mathbb{E}\int_0^t \tilde{Y}_s \tilde{S}_s \Bigg(&\tilde{Q}(\tilde{H}_s) + \Big(\mu_{u_s} - \tilde{Q}(\tilde{H}_s)\Big) \tilde{W}(\tilde{H}_s) \\
&+ (\mu_{u_s} + \sigma_{u_s}^2 - \tilde{H}_s)\tilde{W}'(\tilde{H}_s) \Bigg) \mathbb{1}\{\tilde{H}_s \in \mathcal{E}\ \&\ u_s = R\} ds \\
= \tilde{W}(x). & \qquad (80)
\end{aligned}
$$

The inequality in Eq.(80) results from applying Condition 5 of Proposition 4 and the fact that $\tilde{Y}_s \tilde{S}_s > 0$. The last step in Eq. (80) follows from applying the Cauchy-Schwarz inequality:

$$
\begin{aligned}
&\left| \mathbb{E}\int_0^t \tilde{Y}_s \tilde{S}_s \Bigg(\tilde{Q}(\tilde{H}_s) + \Big(\mu_{u_s} - \tilde{Q}(\tilde{H}_s)\Big) \tilde{W}(\tilde{H}_s) + (\mu_{u_s} + \sigma_{u_s}^2 - \tilde{H}_s)\tilde{W}'(\tilde{H}_s) \Bigg) \mathbb{1}\{\tilde{H}_s \in \mathcal{E}\ \&\ u_s = R\} ds \right| \\
&\le \sqrt{\left(\mathbb{E}\int_0^t \tilde{Y}_s^2 \tilde{S}_s^2 ds \right)} \\
&\cdot \sqrt{\mathbb{E}\int_0^t \Bigg(\tilde{Q}(\tilde{H}_s) + \Big(\mu_{u_s} - \tilde{Q}(\tilde{H}_s)\Big) \tilde{W}(\tilde{H}_s) + (\mu_{u_s} + \sigma_{u_s}^2 - \tilde{H}_s)\tilde{W}'(\tilde{H}_s) \Bigg)^2 \mathbb{1}\{\tilde{H}_s \in \mathcal{E}\ \&\ u_s = R\} ds} \\
&= 0.
\end{aligned}
$$

In the last step above, the first squared root term is bounded since $\tilde{Y}_s \in L^2[0,t]$ by Lemma D.1 and $\tilde{S}_s \in [0,1]$. The second squared root term above is zero, since $\int_0^t \mathbb{1}\{\tilde{H}_s \in \mathcal{E}\ \&\ u_s = R\} ds = 0$ almost surely by Lemma 6, and $\tilde{Q}(\tilde{H}_s) + \Big(\mu_{u_s} - \tilde{Q}(\tilde{H}_s)\Big) \tilde{W}(\tilde{H}_s) + (\mu_{u_s} + \sigma_{u_s}^2 - \tilde{H}_s)\tilde{W}'(\tilde{H}_s)$ is bounded for $\tilde{H}_s$ in a countable set $\mathcal{E}$. Since Inequality (80) holds for any $t \ge 0$, it also holds in the limit:

$$
\limsup_{t\to\infty} \mathbb{E}X_t \le \tilde{W}(x). \qquad (81)
$$

We will now show $\limsup_{t\to\infty} \mathbb{E}X_t \ge \tilde{V}(x,\pi)$ to complete the proof of $\tilde{W}(x) \ge \tilde{V}(x,\pi)$. Observe that since $\tilde{Y}_t > 0$, $\tilde{S}_t > 0$ for any $t > 0$ and $\mu_R > 0$, $\mu_S > 0$, the integral $\int_0^t \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds$ is pathwise monotone increasing in t and hence converges pathwise to $\int_0^\infty \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds$ as $t \to \infty$. Therefore

$\lim_{t\to\infty} \mathbb{E}\left[\int_0^t \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds\right] = \mathbb{E}\left[\int_0^\infty \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds\right]$. Define $\tilde{V}_t(x,\pi) = \tilde{Y}_0 + \mathbb{E}\left[\int_0^t \tilde{S}_s d\tilde{Y}_s\right]$. It follows that

$$\begin{aligned}
\tilde{V}_t(x,\pi) &= \tilde{Y}_0 + \mathbb{E}\left[\int_0^t \tilde{S}_s d\tilde{Y}_s\right] \\
&= 1 + \mathbb{E}\left[\int_0^t \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds + \int_0^t \tilde{Y}_s \tilde{S}_s \sigma_{u_s} dB_s\right] \\
&= 1 + \mathbb{E}\left[\int_0^t \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds\right]
\end{aligned}$$

is monotone increasing in t. Note that the stochastic integral has zero expectation since $\tilde{Y}_s \tilde{S}_s \sigma_{u_s} \in L^2[0,t]$, the same reasoning as before. Then by the Monotone Convergence Theorem, we have

$$\begin{aligned}
\tilde{V}(x,\pi) &= \tilde{V}_\infty(x,\pi) \\
&= \lim_{t\to\infty} \tilde{V}_t(x,\pi) \\
&= 1 + \lim_{t\to\infty} \mathbb{E}\left[\int_0^t \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds\right] \\
&= 1 + \mathbb{E}\left[\int_0^\infty \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds\right], \qquad (82)
\end{aligned}$$

where the first equation follows from Eq. (72) and replacing $\mathbb{1}\{t < T\}$ with $\tilde{S}_t$ by the tower property of conditional expectation and the applying the definition of $\tilde{S}_t$ in Eq. (70). With Eq. 82, we are ready to take expectations on both sides of Eq. (77) and let $t \to \infty$. Apply Itô's formula on $\tilde{Y}_t \tilde{S}_t$, utilize Monotone Convergence Theorem to exchange limits with expectations and Lemma D.1 to get rid of the stochastic integral, we have

$$\begin{aligned}
\limsup_{t\to\infty} \mathbb{E}X_t &= \limsup_{t\to\infty} \mathbb{E}\left[\tilde{W}(\tilde{H}_t)\tilde{Y}_t \tilde{S}_t\right] + \limsup_{t\to\infty} \mathbb{E}\left[\int_0^t \tilde{Y}_s \tilde{Q}(\tilde{H}_s) \tilde{S}_s ds\right] \\
&= \limsup_{t\to\infty} \mathbb{E}\left[\tilde{W}(\tilde{H}_t)\tilde{Y}_t \tilde{S}_t\right] + \mathbb{E}\left[\int_0^\infty \tilde{Y}_s \tilde{Q}(\tilde{H}_s) \tilde{S}_s ds\right] \\
&\geq \limsup_{t\to\infty} \mathbb{E}\left[\tilde{Y}_t \tilde{S}_t\right] + \mathbb{E}\left[\int_0^\infty \tilde{Y}_s \tilde{Q}(\tilde{H}_s) \tilde{S}_s ds\right] \\
&= 1 + \limsup_{t\to\infty} \mathbb{E}\left[\int_0^t \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds + \int_0^t \tilde{Y}_s \tilde{S}_s \sigma_{u_s} dB_s - \int_0^t \tilde{Y}_s \tilde{S}_s \tilde{Q}(\tilde{H}_s) ds\right] \\
&\quad + \mathbb{E}\left[\int_0^\infty \tilde{Y}_s \tilde{Q}(\tilde{H}_s) \tilde{S}_s ds\right] \\
&= 1 + \mathbb{E}\left[\int_0^\infty \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds\right] - \mathbb{E}\left[\int_0^\infty \tilde{Y}_s \tilde{S}_s \tilde{Q}(\tilde{H}_s) ds\right] + \mathbb{E}\left[\int_0^\infty \tilde{Y}_s \tilde{Q}(\tilde{H}_s) \tilde{S}_s ds\right] \\
&= 1 + \mathbb{E}\left[\int_0^\infty \tilde{Y}_s \tilde{S}_s \mu_{u_s} ds\right] \\
&= \tilde{V}(x,\pi). \qquad (83)
\end{aligned}$$

The inequality results from the fact that $\tilde{W}(\cdot) > 1$ on $\mathbb{R}$ (see Condition 1 of Proposition 4) and that $\tilde{Y}_t \tilde{S}_t > 0$. Combining Eqs. (81) and (83), we obtain the desired result $\tilde{W}(x) \geq \tilde{V}(x,\pi)$ for $\forall x \in \mathbb{R}$ and any admissible policy π. Hence $\tilde{W}$ as described in Proposition 4 is an upper bound for the optimal CLV $\tilde{V}^*$.

Now it remains to show that the gap $\tilde{W} - \tilde{V}^*$ is zero when the policy is chosen to be $\tilde{\pi}$, as described in Proposition 4. Observe in the above analysis that it suffices to show inequalities (80) and (83) are tight. Inequality (80) is tight follows from Condition 74 in Proposition 4. Hence it only remains to show Inequality (83) is tight. In fact, since $\tilde{W}$ is bounded (see Condition 3), this is true if we can show $\limsup_{t\to\infty} \mathbb{E}\left[\tilde{Y}_t \tilde{S}_t\right] = 0$. From definitions of $\tilde{Y}_t$ and $\tilde{S}_t$ (see Eqs. (67) and (70)), we know that (since $\tilde{\pi}$ is an interval policy by Condition 6 of Proposition 4, we use $\tilde{\pi}(\tilde{H}_t)$ to denote the policy's choice of service mode at time t)

$$\mathbb{E}\left[\tilde{Y}_t\tilde{S}_t\right] = \mathbb{E}\left[e^{\int_0^t\left(\mu_{\tilde{\pi}(\tilde{H}_z)}-\frac{\sigma^2_{\tilde{\pi}(\tilde{H}_z)}}{2}-\tilde{Q}(\tilde{H}_z)\right)dz+\int_0^t\sigma_{\tilde{\pi}(\tilde{H}_z)}dB_z}\right]$$
$$= \mathbb{E}\left[e^{\int_0^t\left(\mu_{\tilde{\pi}(\tilde{H}_z)}-\tilde{Q}(\tilde{H}_z)\right)dz}\right].$$

Since $Q_1 > Q_2 > \max\{\mu_S, \mu_R\}$ (see Condition 1), the exponent $\int_0^t\left(\mu_{\tilde{\pi}(\tilde{H}_z)} - \tilde{Q}(\tilde{H}_z)\right)dz$ is bounded above by $(\max\{\mu_S, \mu_R\} - Q_2)t$. Hence

$$0 \le \limsup_{t\to\infty}\mathbb{E}\left[e^{\int_0^t\left(\mu_{\tilde{\pi}(\tilde{H}_z)}-\tilde{Q}(\tilde{H}_z)\right)dz}\right] \le \limsup_{t\to\infty}\mathbb{E}\left[e^{\int_0^t(\max\{\mu_S,\mu_R\}-Q_2)dz}\right] = 0\,.$$

Thus we have proved the desired result.

□

E Robustness check: switching costs

In this online appendix we first define the model with switching costs, and then (informally) derive the HJB equations required to numerically find the optimal policy and value function. Finally we present our numerical findings.

Model with switching costs. Assume that each transition from one mode to the other incurs a fixed cost, denoted by K, and keep all other assumptions in the original model unchanged. In this setting, the decision of which service mode to adopt should not only depend on the customer's current satisfaction level, but should also depend on the firm's current mode of service. In other words, the setting is stationary Markov with respect to a two-variable state which includes both the satisfaction and the current mode of service, and so, without loss of optimality, we can restrict attention to stationary Markov policies with respect to this state. Let π be such a stationary Markov policy, which maps from $\mathbb{R} \times \{S, R\}$ to $\{S, R\}$, i.e., $\pi(x, i)$ prescribes which service mode to use when the customer's satisfaction level is x and the firm's current service mode is $i \in \{S, R\}$. Denote by u_t the firm's service mode prescription at time t. Given an initial satisfaction $H_0 = x$ and initial service mode $u_{0-} = i$, under policy π, the firm's service mode prescription at each time $t \ge 0$ is $u_t = \pi(H_t, u_{t-})$. (As before, H_t and u_t are defined for all $t \ge 0$ independent of the customer lifetime T.) We restrict attention to policies such that, with probability 1, the resulting u_t process is right-continuous with left limits (cadlag).

The firm's objective is to find the policy that maximizes the difference between the expected reward earned during the customer's lifetime and the total switching cost incurred. Accordingly, the optimal expected CLV for a starting satisfaction state x and starting service mode $i \in \{S, R\}$ is

$$V_i^*(x) \triangleq \sup_{\pi\in\Pi}\mathbb{E}\left[\int_0^\infty \mathbb{1}\{t < T\}dY_t - K\sum_{k=1}^{\infty}\mathbb{1}\{\tau_k < T\}\ \middle|\ H_0 = x, u_{0-} = i\right],$$

where τ_k denotes the time point of the kth switch in service mode in the u_t sample path, and the customer lifetime T is defined as per Eq. (6) as before. (Since u_t is cadlag under admissible policy π, the set of time points where u_t switches is countable.)

HJB equations. The optimal CLV under switching cost should satisfy the following HJB equations (derived informally below):

$$0 = \max\left\{-Q(x)V_R^*(x) + (\mu_R - x)V_R^{*\prime}(x) + \frac{\sigma_R^2}{2}V_R^{*\prime\prime}(x) + \mu_R, V_S^*(x) - V_R^*(x) - K\right\}$$

$$\forall x \in \mathbb{R} \text{ where } V_R^{*\prime\prime}(x) \text{ exists;} \quad (84)$$

$$0 = \max\left\{-Q(x)V_S^*(x) + (\mu_S - x)V_S^{*\prime}(x) + \mu_S, V_R^*(x) - V_S^*(x) - K\right\} \quad \forall x \in \mathbb{R}. \quad (85)$$

We now summarize how we obtain these equations: We have two equations in this setting, each corresponding to one of the two possible service modes (the Risky mode or the Safe mode) being used currently. In particular, Eq. (84) comes from comparing the continuation value of staying with the Risky mode with the value of switching to the Safe mode while incurring a switching cost K. Similarly, Eq. (85) comes from comparing the continuation value of staying with the Safe mode with the value of switching to the Risky mode while incurring a switching cost K.

Next, we go over the heuristic steps to obtain the HJB equation.

Consider a starting satisfaction value at $H_0 = x$ and the firm's current service mode being the Risky mode. We answer the following question: should the firm stick with the Risky mode for a very short time t and then continue optimally, or should the firm immediately switch to the Safe mode but incur a switching cost of $K > 0$? In the first option, the total reward collected is

$$\int_0^t \mathbb{1}\{T > s\}(\mu_R ds + \sigma_R dB_s) + \mathbb{1}\{T > t\}V_R^*(H_t),$$

where $V_i^*(x)$ is the continuation value when satisfaction level starts at x and the firm's starting service mode is $i \in \{S, R\}$. Take expectation of the above expression and apply Itô's formula[20] on $e^{-\int_0^t Q(H_s)ds}V_R^*(H_t)$, we get

$$\begin{aligned}
&\mathbb{E}\int_0^t \mathbb{1}\{T > s\}(\mu_R ds + \sigma_R dB_s) + \mathbb{E}\left[\mathbb{1}\{T > t\}V_R^*(H_t)\right] \\
&= \mathbb{E}\int_0^t e^{-\int_0^s Q(H_v)dv}\mu_R ds + \mathbb{E}e^{-\int_0^t Q(H_s)ds}V_R^*(H_t) \\
&= \mathbb{E}\int_0^t e^{-\int_0^s Q(H_v)dv}\mu_R ds + \mathbb{E}\left[V_R^*(H_0) + \int_0^t e^{-\int_0^s Q(H_v)dv}(\mu_R - H_s)V_R^{*\prime}(H_s)ds\right. \\
&\quad \left. + \int_0^t e^{-\int_0^s Q(H_v)dv}\frac{\sigma_R^2}{2}V_R^{*\prime\prime}(H_s)ds - \int_0^t Q(H_s)e^{-\int_0^s Q(H_v)dv}V_R^*(H_s)ds\right].
\end{aligned}$$

On the other hand, if the firm immediately switches to the Safe mode by incurring a cost of K, the total reward (minus cost) collected is $V_S^*(H_0) - K$. Since the firm wants to maximize total reward, we must have

$$\begin{aligned}
V_R^*(H_0) = \max\Bigg\{&\mathbb{E}\int_0^t e^{-\int_0^s Q(H_v)dv}\mu_R ds + \mathbb{E}\left[V_R^*(H_0) + \int_0^t e^{-\int_0^s Q(H_v)dv}(\mu_R - H_s)V_R^{*\prime}(H_s)ds\right. \\
&\left. + \int_0^t e^{-\int_0^s Q(H_v)dv}\frac{\sigma_R^2}{2}V_R^{*\prime\prime}(H_s)ds - \int_0^t Q(H_s)e^{-\int_0^s Q(H_v)dv}V_R^*(H_s)ds\right], V_S^*(H_0) - K\Bigg\}.
\end{aligned}$$

Consider the limit as $t \to 0$, the above equation reduces to

$$0 = \max\left\{-Q(x)V_R^*(x) + (\mu_R - x)V_R^{*\prime}(x) + \frac{\sigma_R^2}{2}V_R^{*\prime\prime}(x) + \mu_R, V_S^*(x) - V_R^*(x) - K\right\}.$$

Similarly, if the firm's current service mode is the Safe mode and the customer's satisfaction value is $H_0 = x$, we have

$$0 = \max\left\{-Q(x)V_S^*(x) + (\mu_S - x)V_S^{*\prime}(x) + \mu_S, V_R^*(x) - V_S^*(x) - K\right\}.$$

Findings. We focus on the case $\mu_R > \mu_S$, and numerically solve the HJB equations (84) and (85) to find the firm's optimal policy under switching cost K. In short, we find that adding a small switching cost results in an optimal policy which is very similar to the sandwich policy we find to be optimal in the original model.

Recall our result for the original model for $\mu_R > \mu_S$, namely, that in cases where a sandwich

[20] Fix $i \in \{S, R\}$. Since H_t is a semimartingale, by the Itô-Tanaka formula, if $V_i^*(\cdot)$ is sufficiently smooth, then $V_i^*(H_t)$ is also a semimartingale. Also $e^{-\int_0^t Q(H_s)ds}$ is a semimartingale. Therefore by the multidimensional Itô formula, $e^{-\int_0^t Q(H_s)ds}V_i^*(H_t)$ is also a semimartingle.

policy is optimal, θ_b and q are the thresholds separating the satisfaction values where the firm should use the Safe service mode from those where the firm should use the Risky service mode (Figure 2). With a small positive switching cost K, our numerics reveal that each switching threshold is replaced by a *buffer* interval. Specifically, above and below a buffer the firm should prefer opposite service modes (as in the original model, regardless of which service mode is currently in use), whereas inside a buffer interval the firm should not switch service modes. (Intuitively, the CLV benefit of having buffers in place of sharp switching thresholds for $K > 0$ is to reduce the number of switches between service modes.)

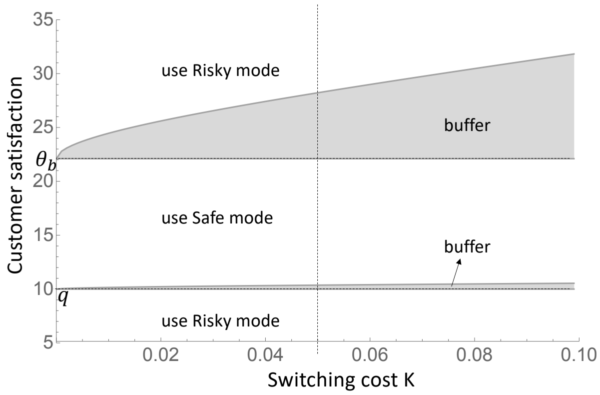
Figure 11 illustrates the optimal policy as a function of switching cost K. In the plot, the horizontal coordinate is the switching cost, and the vertical coordinate is the customer's current satisfaction value. All model primitives besides K are held fixed as per $\mu_S = 8$, $\mu_R = 9$, $\sigma_R = 10$, and $q = 10$. The shaded areas in Figure 11 represents the buffers of the optimal policy. Observe that the buffers grow when we increase the switching costs.[21] The special case $K = 0$ corresponds to the original model, and leads to an interval optimal policy as before with sharp thresholds q and $\theta_b = 22.10$. Consider $K = 0.05$. The optimal policy has buffer zones $(q, 10.37) = (10, 10.37)$ and $(\theta_b, 28.22) = (22.10, 28.22)$, corresponding to the intersection between the $K = 0.05$ line and the shaded areas in Figure 11. Outside the buffer zones, the policy is identical to that in the original model, i.e., it uses the Risky mode for satisfaction values above 28.22 and below q, and it uses the Safe mode for satisfaction values in $(10.37, 22.10)$. To illustrate the role of the buffer zones: if the current service mode is Risky, then the policy prescribes to stay with Risky as long as the satisfaction level is on $(0, 10.37)$ or $(22.10, \infty)$, and switches to Safe once the satisfaction level leaves these regions. If the current service mode is Safe, then stick to Safe while the satisfaction level is on $(10, 28.22)$. The buffer zones $(10, 10.37)$ and $(22.10, 28.22)$ are hence overlapping bands for the two service modes, where the firm may use either the Risky or the Safe service mode depending on the current service mode in use. Despite the positive switching cost, the optimal policy produces a substantially larger expected CLV than the myopic policy. Specifically, the payoff is nearly 79% larger under the optimal policy $\frac{V_R^*(q)}{V(q,\text{Myopic})} = 1.79$, where $V(q, \text{Myopic})$ is the CLV of using the Risky mode always.

In the cases where the myopic (Risky always) policy is optimal under the original model (with $K = 0$), it is clear that the Risky always policy remains optimal even for $K > 0$, since the policy does not incur any switching cost.[22]

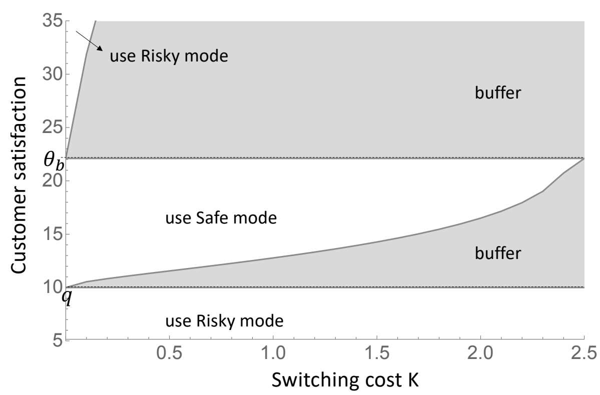
Implications for a setting where the firm cannot perfectly measure satisfaction. So far we have assumed that the firm is able to perfectly estimate the customer's current satisfaction state. A reader might be concerned about the robustness of our findings to estimation errors (or delays). We now argue that the results we have obtained under switching costs suggest that small to medium-sized errors in estimating customer satisfaction would not significantly impair the CLV benefits of the optimal policy (relative to the myopic policy). In each case where the optimal policy for the case of switching costs is one where switching is postponed by a buffer (see Figure 11), by definition this policy produces higher CLV than the myopic policy. We can hence conclude this policy also produces higher CLV than the myopic policy if there were no switching costs (for $K = 0$, the CLV under the former policy is even larger whereas under the CLV under the latter policy stays the same). To give a quantitative example, for $K = 0.70$ the lower buffer interval of the optimal policy is $(10, 12.03)$ as per Figure 11. Thus the policy decisions are loosely similar to those of the

[21] In fact, we also find in numerical solutions that as the switching cost K increases to above $\mu_R - \mu_S$, there emerges a threshold $\theta_0 < q$, increasing with K, such that for satisfaction states under this threshold, the firm should not switch to the Risky mode, if currently using the Safe mode. Here $\mu_R - \mu_S$ is the CLV difference between the Risky-always policy and Safe-always policy in the limit of initial satisfaction $x \to -\infty$, since $Q(x') = 1 \; \forall x' < q$.

[22] There is technical caveat here: if the starting service mode is Safe, then if $K > 0$ there is a nontrivial decision regarding whether to switch to Risky and under what conditions. This case has limited practical relevance, so we avoid discussing it in the interest of space.



(a) Small switching costs



(b) Larger switching costs

Figure 11: Firm's optimal policy as a function of switching cost K. We fix $\mu_S = 8$, $\mu_R = 9$, $\sigma_R = 10$, and $q = 10$. The shaded areas represent the buffers where the optimal policy retains the current service mode. The white areas represents the regions where the policy employs a specific service mode, even if this would incur a switching cost. For each K, the buffers of the optimal policy are given by the intervals where the (horizontal coordinate $= K$) vertical line intersects the shaded areas, e.g, the dotted line in the left subfigure gives the optimal policy for $K = 0.05$.

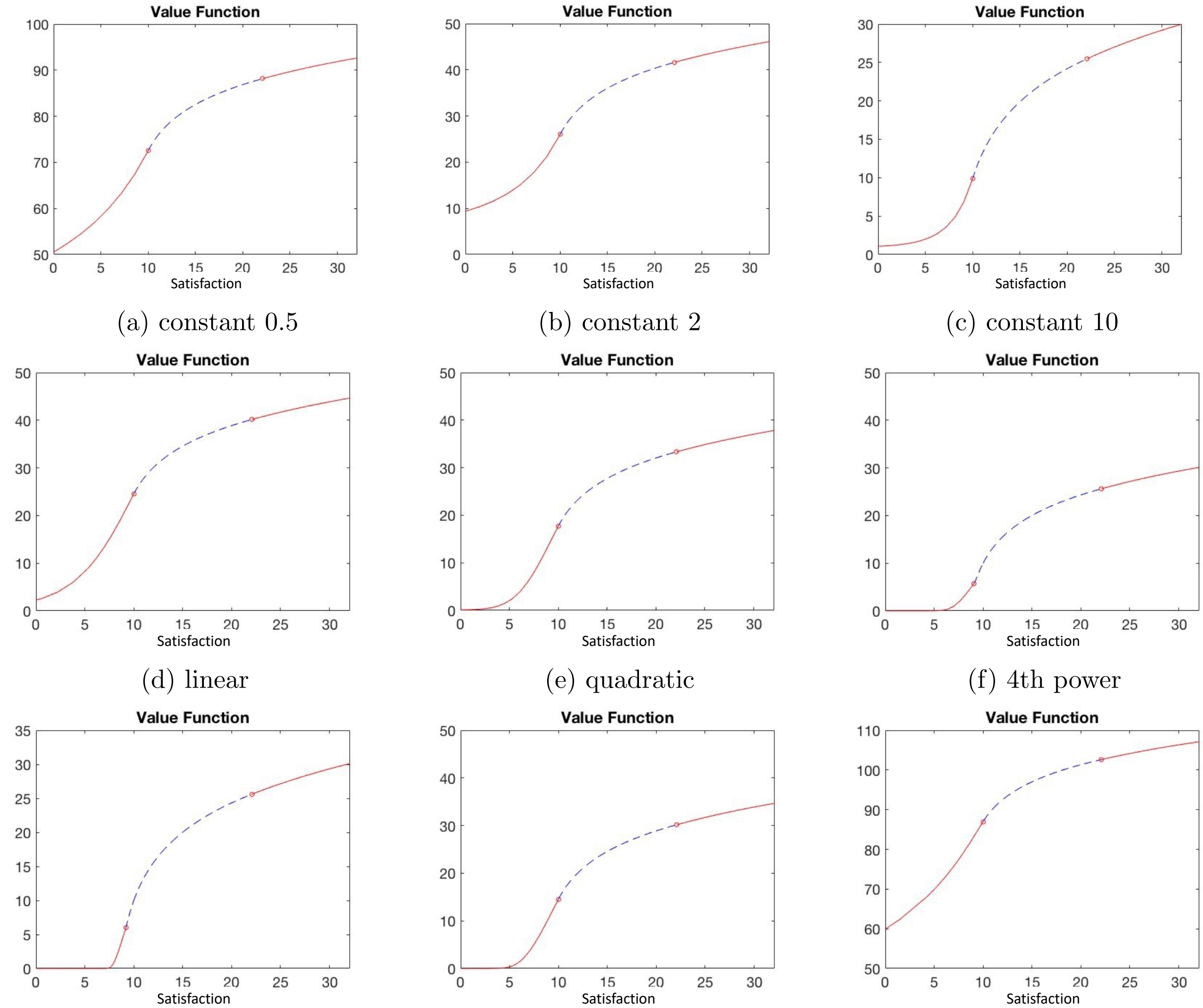
optimal (sandwich) policy under the original model acting on *estimated* customer satisfaction when there are estimation errors of size similar to the size of the lower buffer interval ~ 2 (the upper buffer interval plays only a small role since the customer satisfaction only rarely rises to that level). The CLV increase from using the optimal (buffer) policy relative to the myopic policy is 30.0% for $K = 0.70$, and so the CLV increase from using the same policy in the absence of switching costs is even larger. This gives us confidence that our proposed policies still substantially increase the CLV in the face of small to medium-sized errors in estimating customer satisfaction. Along similar lines, interpreting the effect of the buffer intervals as delays in switching, one can argue that (small) delays in estimating customer satisfaction are unlikely to erode the benefits of using our proposed policies.

F Robustness check: Alternative hazard rate functions

We now provide numerical evidence for our results' robustness to alternative specifications of the hazard rate function (recall the original step function in Eq. (4)) in the case $\mu_R > \mu_S > 0$. The most important takeaway is that, for a variety of hazard rate functions and different model primitives, the optimal policy is still either myopic or a sandwich policy, and moreover, the switching thresholds in the optimal sandwich policy are fairly robust to the shape of hazard rate functions.

First, we consider a variety of different hazard rate specifications in the unsatisfied zone, while keeping the original assumption of zero hazard rate in the satisfied zone. Let q be the satisfaction threshold separating the unsatisfied zone and the satisfied zone. Consider the following four types of hazard rate functions:

1. constant k: $Q(x) = k\mathbb{1}\{x < q\}$;
2. nth power: $Q(x) = (q - x)^n \mathbb{1}\{x < q\}$;
3. exponential: $Q(x) = (e^{q-x} - 1)\mathbb{1}\{x < q\}$;
4. logit: $Q(x) = \left(\frac{e^{q-x}}{1+e^{q-x}} - \frac{1}{2}\right) \mathbb{1}\{x < q\}$.



(a) constant 0.5 (b) constant 2 (c) constant 10

(d) linear (e) quadratic (f) 4th power

(g) 8th power (h) exponential (i) logit

Figure 12: Value function and optimal policies for various forms of $Q(\cdot)$; $\mu_R = 9$, $\mu_S = 8$, $\sigma_R = 10$, $q = 10$. Solid red lines correspond to the Risky service mode, and dashed blue lines correspond to the Safe service mode.

The value function and optimal policy associated with each hazard rate functions can be established by solving the HJB equation (13) and checking that the optimality conditions in Proposition 2 are still satisfied. Under all these different choices of the hazard rate function in the unsatisfied zone (including several different constants, and powers) and different model primitives with $\mu_R > \mu_S$, we find that (similar to Theorem 1) the optimal policy is either myopic or a sandwich policy. Figure 12 presents the firm's optimal policy and the associated CLV for some of the numerical instances.

One interesting observation from Figure 12 is that the size of the risk-averse region depends on how fast the hazard rate of leaving increases as the customer satisfaction level descends into the unsatisfied zone. In the original model, the firm switches from Safe to Risky as soon the customer satisfaction crosses from the satisfied zone into the unsatisfied zone. This is not always the case with arbitrary hazard rate functions. When the hazard rate grows relatively fast as customer satisfaction goes down, the lower switching threshold remains at q. However, in the plotted cases where the hazard rate does not grow swiftly when customer satisfaction crosses into the unsatisfied zone, the switching point to Risky is strictly below q. In our numerics, this occurs for the cases of hazard rate functions growing as the fourth and eighth powers (see Figures 12(f) and 12(g)), where the lower boundary of the risk-averse region is strictly below q.

In Figure 13, we also show that the gap in CLVs between under the optimal policy and the myopic policy remains large for different hazard rate functions. In particular, the second curve from the bottom corresponds to the original step function hazard rate in Eq. (4). The other three curves correspond to the hazard rate function being linear, exponential, and logit in the unsatisfied zone (and zero in the satisfied zone) as introduced earlier in this section.

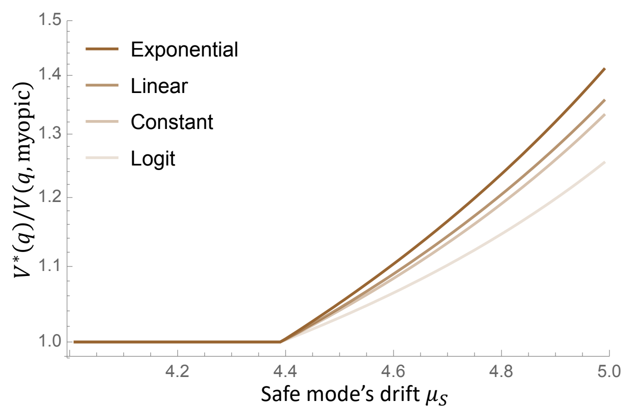


Figure 13: The ratio of CLV under the optimal policy to CLV under the myopic policy versus μ_S under different hazard rate functions), for $\mu_R = 5$, $\sigma_R = 2$, $q = 6$, and initial satisfaction q.

Other than the choice of hazard rate functions listed above, we also numerically examined cases where the hazard rate is strictly positive everywhere (including in the satisfied zone), again in the original model and restricting $\mu_R > \mu_S$, and find our main results remain intact. In particular, we considered hazard rate functions of the form $Q(x) = \mathbb{1}\{x < q\} + \epsilon\mathbb{1}\{x \geq q\}$ for $\epsilon \in (0, 1)$, and various model primitives such that $\mu_R > \mu_S$, and found that the optimal policy is still either myopic or a sandwich policy, where (in the optimal sandwich policy) the upper switching threshold decreases smoothly as we increase the value of ϵ. The CLV increase from using the optimal sandwich policy relative to the myopic policy is still large for small values of ϵ. We omit more details of this robustness check for brevity.