Data-Driven Monitoring and Deterrence in a Changing Environment*

Yeon-Koo Che, Jinwoo Kim, Konrad Mierendorff†

August 13, 2026

Abstract

We study a dynamic model in which a principal monitors agents based on historical data of infractions. This data informs when and at what intensity to monitor; the monitoring decision, in turn, selects the collected data, shaping the principal's future learning. We analyze this feedback loop using a bandit model in which the underlying monitoring environment evolves according to a hidden Markov process. Because data collection is endogenous, how the principal uses this information is critical: surprisingly, a myopic approach renders historical data completely valueless. By endogenizing the agent's incentives, we demonstrate that the principal's purely informational motive to explore serves as an endogenous commitment device. This inherent drive to gather data compels persistent vigilance, strictly lowering the equilibrium infraction rate and restoring the power of deterrence.

Keywords: Dynamic monitoring, endogenous datafication, optimal experimentation, restless bandit, dynamic inspection game, deterrence.

JEL Codes: C73, D82, D83, K42

*We are grateful to the participants of numerous seminars and conferences for valuable comments and suggestions. The authors gratefully acknowledge Mingeon Kim and Tialing Luo for their excellent research assistance. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2024S1A5A2A03038509). Our dear friend and coauthor, Konrad Mierendorff, passed away in August of 2021. All the ideas and results in this paper are the collaborative work of all three authors, but Che and Kim are responsible for any remaining errors.

†Che: Department of Economics, Columbia University (email: yeonkooche@gmail.com); Kim: Department of Economics, Hong Kong University of Science and Technology (email: jikim72@gmail.com); Mierendorff: Department of Economics, University College London.

1 Introduction

The proliferation of big data and algorithmic prediction has fundamentally transformed how organizations monitor compliance, allocate inspection resources, and deter infractions. A prominent example is New York City's inspection of "illegal conversions"—the illicit division of dwellings into smaller units, which poses severe fire hazards. By linking various municipal data sources to predict likely violations, the City's analytics team achieved a five-fold improvement in inspection effectiveness.[1] Similar data-driven monitoring is now ubiquitous across diverse domains: digital platforms utilize machine learning to moderate content and flag compromised accounts distributing malware, environmental protection agencies use data-driven models to inspect facilities likely to bypass emissions controls, and financial regulators deploy anomaly detection algorithms to identify divisions at risk of accounting fraud. In all these settings, the principal's objective is twofold: to mitigate the immediate harm of infractions through active intervention, and to deter strategic agents from committing infractions in the first place.

While data-driven monitoring holds the promise of unprecedented efficiency, it introduces a profound dynamic challenge: the inherent endogeneity of the data generation process. Data is rarely collected randomly; it is selectively generated only where the principal actively looks. Consequently, monitoring serves a dual purpose. It not only halts current infractions but also generates the very information required to predict future ones. This creates a complex feedback loop. A principal who myopically trusts a predictive algorithm may cease monitoring an entity that currently appears safe. However, without ongoing monitoring, the principal creates an inferential blind spot, failing to recognize when the underlying conditions inevitably deteriorate. The fundamental challenge is understanding how to optimally trade off the short-term exploitation of current data against the long-term informational value of continued exploration.

The purpose of this paper is to formally investigate this tension. First, we develop a tractable, continuous-time framework for dynamic monitoring and deterrence that explicitly incorporates the feedback loop between monitoring intensity and endogenous datafication. Second, we extend the literature on optimal experimentation—specifically, the seminal bandit framework of Keller, Rady, and Cripps (2005)—in two critical dimensions to capture the

[1] "Before the big-data analysis, inspectors followed up the complaints they deemed most dire, but only in 13 percent of cases did they find conditions severe enough to warrant a vacate order. Now they were issuing vacate orders on more than 70 percent of the buildings they inspected. By indicating which buildings most needed their attention, big data improved their efficiency five-fold. ... Flowers and his kids looked like wizards with a crystal ball that let them see into the future and predict the most risky places. They took massive quantities of data lying around for years, largely unused after it was collected, and harnessed it in a novel way to extract real value." See Mayer-Schönberger and Cukier (2014).

realities of dynamic enforcement. First, we embed the learning problem in a changing world. In standard bandit models, the unobservable state of the arm is fixed; once the principal learns the true state, the need for experimentation permanently resolves. By modeling the environment as a hidden Markov state, we ensure that the need for learning and vigilance persists indefinitely. Second, we endogenize the value of the arm. While the standard bandit literature treats the payoff of the arm as exogenous, in regulatory and monitoring contexts, the arm represents a strategic agent. We therefore model the agent's infraction rate as a dynamic best response to the principal's anticipated monitoring policy.

We formulate this problem as a continuous-time dynamic inspection game. At each time $t \geq 0$, the environment is in an unobservable state that is either "high risk" (H) or "low risk" (L), transitioning according to a continuous-time Markov chain with stationary belief π_0. Accordingly, in the absence of monitoring, the principal's belief converges to this natural benchmark. Opportunities for infractions arise at a Poisson rate only in the high-risk state. The principal chooses a monitoring intensity $y_t \in [0, 1]$ at a flow cost c, while agents, upon receiving an opportunity, choose to commit an infraction with probability $x_t \in [0, 1]$. Crucially, a detection perfectly reveals that the state is H. However, if no infraction is detected, the principal updates her posterior belief p_t of state H downward. This learning is strictly endogenous: the speed of the downward drift depends entirely on the principal's chosen monitoring intensity y_t and the agent's infraction rate x_t.

To isolate the value of data and the necessity of exploration, we evaluate three distinct monitoring policy regimes: a static policy (SP) that ignores data and relies solely on the stationary prior; a myopic policy (MP) that updates beliefs using data but acts myopically, ignoring the informational option value of monitoring; and the optimal policy (OP), which balances exploitation with the need for persistent learning.

Our analysis yields four principal insights that challenge conventional intuition regarding data-driven monitoring.

First, a striking payoff equivalence emerges: the predictive value of data is entirely neutralized when used myopically. While it is natural to expect that a standard, data-driven algorithm should intuitively outperform a policy that ignores historical data entirely, we show that this is not the case: the myopic use of data yields no long-run performance gain over a static baseline. This equivalence is driven fundamentally by the endogeneity of data—the reality that new data is generated only through active monitoring. Because of this endogeneity, a myopic principal ceases enforcement the moment an entity currently appears safe. This creates a critical inferential blind spot, preventing her from recognizing when the underlying conditions inevitably deteriorate. Ultimately, this equivalence provides a crucial cautionary

lesson: the value of predictive data is not automatic, and simply adopting a data-driven approach does not guarantee an enforcement gain.

Second, unlocking the true value of data requires a dynamically optimal policy that explicitly values exploration. In a changing-world setup, data-driven policies can feature belief freezing at an interior cutoff, where partial monitoring exactly offsets the natural drift of the environment. The optimal policy uses this feature as an optimal "hedging" device. When data are valuable, the principal cannot optimally abandon monitoring, since the environment can revert to a high-risk state at any time. Thus, when the belief reaches the optimal monitoring threshold, the principal does not stop entirely; instead, she chooses an interior monitoring intensity that freezes the belief at the cutoff, trapping the system in a partially absorbing state of persistent vigilance.

Third, we demonstrate that this exploratory motive acts as an endogenous commitment device for deterrence. When we endogenize the agents' behavior, the problem transforms into a dynamic inspection game where the principal lacks commitment power. Under myopic regimes, the principal cannot credibly commit to high monitoring rates, and agents simply adjust their infraction rate upward to keep the principal exactly indifferent. However, under the optimal policy, the principal's inherent informational desire to explore organically pushes her monitoring intensity above what is myopically justified. Knowing the principal is eager to gather data, agents are forced to strictly lower their equilibrium infraction rate to maintain the deterrence condition.

Finally, we establish that the deterrence benefits of data-driven monitoring rely entirely on the principal maintaining an informational advantage. We analyze a scenario in which agents are "data-aware," having access to the same historical detection data, and perfectly tracking the principal's posterior belief. Under symmetric information, agents perfectly best-respond at every instant, which entirely neutralizes the principal's exploratory benefit. Without the advantage of private learning, the endogenous commitment mechanism unravels, and the expected losses across all three policy regimes collapse to the exact same level.

Related Literature

This paper contributes primarily to two strands of economic theory: optimal experimentation in non-stationary environments and dynamic monitoring or inspection. It also relates to the broader literature on dynamic agency with endogenous learning.

Our model builds on the continuous-time experimentation framework pioneered by Keller, Rady, and Cripps (2005). In their canonical setting, the unobservable state of the arm is fixed; the principal learns through Poisson arrivals, and experimentation eventually ceases

once uncertainty is sufficiently resolved. We extend this framework by embedding the learning problem in a changing world governed by a hidden Markov chain. Because the underlying state can continually transition, the need for learning never permanently disappears, fundamentally altering the optimal policy.

This changing-world feature naturally connects our work to the restless bandit literature (Whittle, 1988; Nino-Mora, 2001), recently applied in economics by Fryer and Harms (2018) and Urgun (2021). Whereas Fryer and Harms feature a fixed hidden type and Urgun focuses on indexability in contracting, our emphasis lies on the endogenous generation of data. We highlight a distinctive form of belief "freezing"—an interior monitoring intensity that exactly offsets the natural drift of beliefs. Under the optimal policy, this belief freezing has the nature of hedging, as the principal must continually balance exploitation against the option value of vigilance in a changing environment.

We also depart from the standard bandit paradigm by treating the arm not as a passive source of rewards, but as an endogenous object whose future evolution depends on current actions. From this perspective, our paper is related to Marinovic and Szydlowski (2022), which also studies a dynamic monitoring problem with an endogenous arm. However, their monitoring quality depends on the regulator's effort incentives, whereas our monitoring effectiveness is driven by the agents' infraction incentives. More broadly, our model connects to a recent literature exploring the interplay between endogenous data collection and agent incentives (e.g., Che and Hörner (2018), Che, Kim, and Zhong (2020), and Madsen and Shmaya (2024)). While this literature often focuses on designing mechanisms to incentivize agents to actively provide or reveal data, data in our dynamic inspection game is generated purely as a byproduct of the principal's enforcement and the agents' non-compliance. In our setting, the principal's informational motive to explore serves as an endogenous commitment device for deterrence.

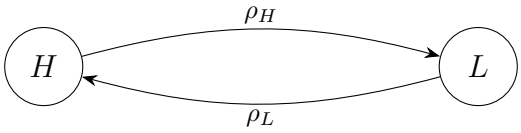
Finally, we contribute to the literature on dynamic inspection games and the economics of deterrence. Traditional enforcement models (e.g., Becker (1968)) typically assume perfect commitment. In settings without commitment, principals are often assumed to optimize myopically, yielding mixed-strategy equilibria where agents merely keep the principal indifferent (Khalil, 1997). While several recent papers explore the optimal timing or frequency of inspections in dynamic settings (Dilmé and Garrett, 2019; Varas, Marinovic, and Skrzypacz, 2020; Wagner and Knoepfle, 2021; Ball and Knoepfle, 2023), we depart from this literature by emphasizing how monitoring endogenously generates data. Incorporating this informational option value into a dynamic inspection game provides a rational foundation for why a principal lacking formal commitment power may nonetheless credibly sustain high levels of

monitoring.

2 The Model

2.1 The Environment and the Hidden State

We consider a continuous-time dynamic inspection game between a principal and a continuum of agents. Time is continuous and indexed by $t \geq 0$. The environment is characterized by an unobservable, time-varying state $\omega_t \in \{H, L\}$. The state evolves according to a continuous-time Markov chain, transitioning from L to H at a Poisson rate $\rho_L > 0$, and from H to L at a Poisson rate $\rho_H > 0$.



The unobservable state ω_t represents a shifting, systemic vulnerability that periodically generates opportunities for infractions. In *financial auditing*, this could be a newly invented tax-evasion scheme or accounting loophole: the high-risk state H indicates the scheme is prevalent and available to agents, while the low-risk state L implies it has been regulated away. Alternatively, in *cybersecurity*, ω_t might represent a zero-day exploit or ransomware strain, where H implies the exploit is actively circulating and creating opportunities for network compromise, while L indicates the threat has been patched or has naturally died out. In both contexts, the hidden state captures an invisible, fluctuating environment that dictates when strategic agents possess the opportunity to offend.

2.2 Infraction Opportunities and Agent Behavior

Opportunities to exploit this vulnerability arise according to a Poisson process with rate $\lambda > 0$ only when the state is H; in state L, the arrival rate is zero. Crucially, we assume these opportunities are drawn from a large population of agents, such that each opportunity is received by a different, randomly selected agent.

When an opportunity arises at time t, the selected agent must decide whether to commit an infraction. We denote by $x_t \in [0, 1]$ the probability that the agent commits the infraction. Because the population is large and any given agent receives an opportunity only rarely, the acting agent does not internalize how their individual choice affects the principal's future beliefs or long-run monitoring policy. Consequently, the agent acts myopically, optimizing

their decision strictly against the principal's instantaneous monitoring intensity at time t. For the initial analysis in Section 3, we will treat the infraction rate x as an exogenous parameter to isolate the principal's learning problem. In Section 4, we will fully endogenize x_t as a strategic best response.

2.3 The Principal's Monitoring Policy and Payoffs

At each instant t, the principal chooses a monitoring intensity $y_t \in [0, 1]$. Maintaining this monitoring intensity imposes a flow cost of $c \cdot y_t$ on the principal, where $c > 0$ represents the marginal cost of deploying audit resources. If an opportunity arises, the agent commits an infraction, and the principal is actively monitoring, the infraction is detected and halted with probability y_t. The principal incurs a fixed loss $\ell > 0$ for every infraction that goes undetected, which we normalize to 1 without loss of generality.

Because the true state ω_t is hidden, the principal's policy must be measurable with respect to her information. Let $\mathcal{H}_t$ denote the history of the game up to time t, which consists of the realized arrival times of all detected infractions prior to t. A monitoring policy is a process $y = \{y_t\}_{t\geq 0}$ that is adapted to the filtration generated by $\mathcal{H}_t$ and admits well-defined belief dynamics.[2] The principal's objective, under a prior belief $p_0 = p$ that $\omega_0 = H$, is to choose an admissible monitoring policy to minimize the expected discounted loss:

$$L(p) := \min_{\{y_t\}_{t\geq 0}} \mathbb{E}\left[\int_0^\infty e^{-rt}\left(cy_t + \lambda x_t(1-y_t)\mathbf{1}_{\{\omega_t=H\}}\right) dt \middle| p_0 = p\right], \tag{1}$$

where $r > 0$ is the discount rate, and the expectation is taken over the stochastic evolution of the hidden state ω_t and the history of detected infractions that determines $\{y_t\}_{t\geq 0}$.

2.4 Endogenous Datafication and the Fixed-State Benchmark

The central friction in this environment is that learning is strictly endogenous: the principal only generates data when she actively monitors and an infraction actually occurs. If the principal exerts monitoring intensity y_t and the agent offends with probability x_t, a detection occurs at a Poisson arrival rate of $\lambda x_t y_t$ conditional on the state being H. Because opportunities only arise in state H, a detection is perfectly revealing. If an infraction is detected at

[2] As will be seen, the belief dynamics are characterized by an ordinary differential equation. The admissibility requirement simply ensures that the solution to this equation is well-defined. In our context, this requirement has an effect in only one circumstance: if a belief p attracts drift from both above and below, $y(p)$ must be chosen at an interior level to ensure that the drift at p vanishes. This prevents the belief process from oscillating infinitely often around p. See further discussion in the remark at the end of this subsection.

time t, the principal's posterior belief that the state is H, denoted by p_t, immediately jumps to 1.

In the absence of a detection, the principal continues to update her belief based on no news. Although no news pushes the belief downward, the overall drift may still be upward when the current belief is sufficiently low, due to the underlying Markov state transitions. By standard Bayesian updating, the evolution of the belief p_t between detections is governed by the deterministic differential equation $\dot{p}_t = f(p_t, y_t)$, where the drift function is defined as:[3]

$$f(p, y) = \rho_L(1 - p) - \rho_H p - p(1 - p)\lambda xy. \tag{2}$$

To appreciate the role of the changing environment, it is instructive to define two natural bounds on the belief. Let $\pi_0 = \rho_L/(\rho_L + \rho_H)$ denote the stationary probability of the high-risk state. Equivalently, π_0 is the unique belief satisfying $f(\pi_0, 0) = 0$. Thus, if the principal stops monitoring entirely ($y_t = 0$), the learning term vanishes, and the belief simply follows the natural rate of the Markov chain, moving upward when $p_t < \pi_0$ and downward when $p_t > \pi_0$. Conversely, let π_1 be the belief where $f(\pi_1, 1) = 0$. If the principal always chooses full enforcement ($y_t = 1$), her belief drifts down if it is above π_1 and drifts up if it is below π_1, as long as no infraction is detected. Naturally, π_1 is strictly lower than the stationary probability π_0.

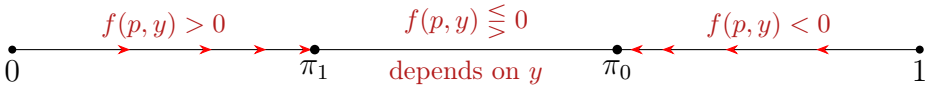


Figure 1: Principal's belief updating given no detection.

Figure 1 illustrates these belief dynamics. Without detection, the belief drifts up if $p_t < \pi_1$ and down if $p_t > \pi_0$. The belief's evolution within the interval (π_1, π_0) depends heavily on the principal's active monitoring choice y_t.

This changing-state framework represents a sharp departure from the classic fixed-state bandit model of Keller, Rady, and Cripps (2005). If the underlying state were permanently fixed ($\rho_L = \rho_H = 0$), the drift equation would reduce to $-p(1 - p)\lambda xy$. In such a world, a sufficiently long period without detections causes the belief p_t to decline monotonically toward zero. Eventually, as long as the monitoring continues, the principal becomes confident the

[3]To derive this drift, consider a small time interval dt. Suppose no detection occurs in $[t, t + dt)$. The probability of this event is $1 - \lambda x_t y_t dt$ if the state is H, and 1 if the state is L. By Bayes' rule, the posterior belief at $t + dt$ is $p_{t+dt} = \frac{p_t(1-\rho_H dt)(1-\lambda x_t y_t dt)+(1-p_t)\rho_L dt}{p_t(1-\lambda x_t y_t dt)+1-p_t} + o(dt)$. Rearranging terms, dividing by dt, and taking the limit as $dt \to 0$ yields the differential equation $\dot{p}_t = \rho_L(1 - p_t) - \rho_H p_t - p_t(1 - p_t)\lambda x_t y_t$.

state is safe, ceases monitoring entirely, and uncertainty is permanently resolved. In our model, however, the state transitions autonomously. The principal can never permanently rule out the high-risk state because the belief is bounded below by π_1. This interplay between a constantly shifting underlying reality and the endogenous generation of data forms the core of the dynamic monitoring challenge.

Admissibility. We focus on Markovian strategies where the principal's action depends solely on her current belief p. This is without loss of generality for exogenous infractions and natural for endogenous infractions, allowing us to suppress time dependence. However, the belief dynamics in equation (2) must be well-defined, imposing an admissibility restriction on the strategy. Essentially, if a belief p attracts drifts from both above and below, it must be a stationary point, implying $y(p) = z(p)$, where $z(p)$ satisfies $f(p, z(p)) = 0$. This ensures the belief process doesn't oscillate arbitrarily frequently around p. While admissibility may seem technical, it captures a substantive property: in a discrete-time approximation, $z(p)$ represents the fraction of time the principal chooses full monitoring (as opposed to no monitoring) in the "neighborhood" of p. The interior solution $z(\hat{p})$ in the continuous-time model corresponds to this relative fraction of time in the limit as the time length approaches zero.

2.5 Alternative Monitoring Regimes

To evaluate the value of endogenous datafication, we define three alternative regimes regarding the principal's use of data for monitoring decisions. These regimes serve as benchmarks to isolate the role of dynamically optimal learning.

Static Policy (SP): The principal does not update her belief based on historical detection data, resulting in a constant monitoring policy dependent solely on the stationary prior, π_0. We assume this prior matches the ground truth. The optimal static policy solves $\min_{y\in[0,1]} \pi_0\lambda x(1-y) + cy$, which yields:

$$y_{\text{SP}}(p) = \begin{cases} 1 & \text{for } \pi_0 > \hat{p}_{\text{M}} \\ 0 & \text{for } \pi_0 < \hat{p}_{\text{M}}, \end{cases} \tag{3}$$

where $\hat{p}_M := c/(\lambda x)$ is the myopic cutoff. That is, the principal enforces fully if and only if the expected long-run harm from an infraction exceeds the monitoring cost.

Myopic Policy (MP): In this benchmark, the principal updates her posterior belief p_t accurately based on the history of detections according to (2), but acts myopically, disregarding

the informational option value of monitoring. The myopic policy solves $\min_{y\in[0,1]} p\lambda x(1-y)+cy$ for the updated posterior p, yielding:

$$y_{\text{MP}}(p) = \begin{cases} 1 & \text{for } p > \hat{p}_{\text{M}} \\ 0 & \text{for } p < \hat{p}_{\text{M}}. \end{cases} \tag{4}$$

While MP utilizes the predictive power of data, it fundamentally neglects the impact of current monitoring decisions on future information generation.

Optimal Policy (OP): Our central focus is the dynamically optimal policy where the principal not only updates her belief p_t accurately but also explicitly internalizes the informational value of monitoring into her decision-making to achieve long-term optimality. The formal derivation of this policy is presented in Section 3.

3 Analysis of Exogenous Infractions

This section examines the scenario of exogenous infractions, where the infraction rate x is fixed within the interval $(0, 1]$. We begin by characterizing the optimal policy (OP) and then compare its outcomes with those of the static policy (SP) and the myopic policy (MP) benchmarks introduced in Section 2.

3.1 The Optimal Policy

This section focuses on the optimal policy (OP) scenario, where the principal dynamically optimizes monitoring decisions, considering both current predictions of the state and the long-term informational option value of monitoring. The problem is formulated as a Markov Decision Problem (MDP) with the belief p as the state variable. The optimal policy, denoted by $p \to y(p)$, is time-independent due to the Markovian nature of the problem.

The analysis employs dynamic programming, focusing on the value function that maximizes the net present value of mitigated infractions minus monitoring costs.[4] The value

[4] This value function, which represents the maximized net present value of mitigated infractions minus monitoring costs, is defined as:

$$V(p) := \max_{\{y_t\}_{t\geq 0}} \mathbb{E}\left[\int_0^\infty (p_t\lambda x - c)y_t e^{-rt}dt \middle| p_0 = p\right].$$

Lemma 8 in Appendix D of the Online Appendix shows that any policy y_t solves this problem if and only if it solves the minimization problem in (1).

function satisfies the Hamilton-Jacobi-Bellman (HJB) equation:

$$rV(p) = \max_{y} (p\lambda x - c)\, y + p\lambda x y[V(1) - V(p)] + f(p, y)V'(p). \tag{5}$$

The HJB equation balances the cost of discounting future value (the LHS) with the immediate benefits of monitoring $(p\lambda x - c)y$ and the value appreciation from Bayesian belief updates. The condition resembles standard bandit frameworks, but the changing state implies that learning remains essential even in the long run. In particular, the value function is not pinned down even for $p = 0$ or $p = 1$, since reaching these extreme states does not mean a permanent resolution of uncertainty. Instead, the value function at any belief depends on the principal's actions at other beliefs, leading to a fixed-point characterization. This presents a new analytical challenge absent in bandit models with a non-changing state.

The optimal policy, derived from the HJB equation, takes a cutoff form: for some $\hat{p} \in [0, 1]$,

$$y(p) = \begin{cases} 1, & \text{for } p > \hat{p} \\ 0, & \text{for } p < \hat{p}. \end{cases} \tag{6}$$

For this structure to hold, the principal must strictly prefer to raise enforcement y as the belief p rises—a standard single-crossing property. However, our model's changing state introduces a novel methodological twist. In standard models (e.g., KRC), this single-crossing condition is slack at the cutoff because abandonment is permanent. Here, because the optimal policy must continually hedge against the risk of the state transitioning back to high risk, the single-crossing condition binds strictly at $\hat{p}$. We utilize this uniquely binding condition as the key analytical tool to pin down the exact value of the cutoff, which plays a central role in characterizing the partial monitoring—i.e., $y(\hat{p}) \in (0, 1)$—seen in Case 3 of Theorem 1.

The following theorem characterizes OP.

Theorem 1. *For each $c > 0$, there exist $\overline{\lambda} > \frac{c}{\pi_0} > \underline{\lambda} > c$ such that the optimal monitoring policy denoted $y^*(p)$ is of the form in (6) where the cutoff is*

Case 1: $\pi_0 \leq \hat{p} < \hat{p}_M$ if $\lambda x \leq \underline{\lambda}$ ("low" infraction rate);

Case 2: $\hat{p} = \hat{p}_M \leq \pi_1$ if $\lambda x \geq \overline{\lambda}$ ("high" infraction rate);

Case 3: $\hat{p} \in (\pi_1, \pi_0)$ and $\hat{p} < \hat{p}_M$ if $\lambda x \in (\underline{\lambda}, \overline{\lambda})$ ("intermediate" infraction rate).

The policy at the cutoff belief, $y^*(\hat{p})$, is arbitrary in Case 1 and Case 2, but in Case 3, $y^*(\hat{p}) = z(\hat{p}) \in (0, 1)$, where $z(p)$ satisfies $f(p, z(p)) = 0$.[5] *Moreover, $\underline{\lambda}$ and $\overline{\lambda}$ are increasing in c.*

[5] In the knife-edge cases where $\hat{p} = \pi_0$ or $\hat{p} = \pi_1$, admissibility pins down the cutoff action: $y^*(\hat{p}) = 0$ when $\hat{p} = \pi_0$, and $y^*(\hat{p}) = 1$ when $\hat{p} = \pi_1$.

Finally, $\hat{p}$ is continuously increasing in c and decreasing in λx, with $\hat{p} = \pi_0$ if $\lambda x = \underline{\lambda}$ and $\hat{p} = \pi_1$ if $\lambda x = \overline{\lambda}$.

Proof. See Appendix A. □

As stated in Theorem 1, there are three cases, depending on where the optimal cutoff $\hat{p}$ lies relative to π_1 and π_0, as depicted previously in Figure 1.

Case 1: Low infraction rates. In this scenario, the infraction rate λx is so low that the optimal cutoff $\hat{p}$ exceeds π_0. The cutoff is set lower than the myopic level $\hat{p}_{\text{M}} := c/(x\lambda)$ to maintain an optimal tradeoff between exploitation and exploration. See Figure 2. While it would be optimal to monitor fully if and only if $p > \hat{p}_{\text{M}}$ from a pure exploitation perspective, there is a long-term benefit in exploring even when p is slightly below the myopic cutoff. Hence, if the initial prior is above $\hat{p}$, the principal starts by monitoring fully. If an infraction is detected, the belief jumps to 1; otherwise, it drifts down. The prospect of irreversible abandonment causes the principal to explore below the myopic cutoff, but the changing world dictates that the principal will eventually cease monitoring as the belief stabilizes at π_0.

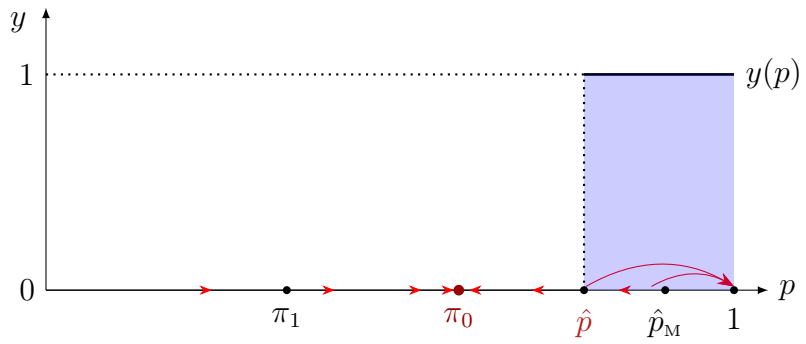


Figure 2: Low Infraction Rates Case

Case 2: High infraction rates. Here, the infraction rate λx is high enough that the optimal cutoff $\hat{p}$ falls below π_1, the lowest possible stationary belief. Hence, the principal monitors fully for any long-run belief above π_1. The belief eventually cycles within the region $[\pi_1, 1]$, jumping to 1 upon detection and drifting down to π_1 otherwise (see Figure 3). Monitoring is always full and never lets up, eliminating the need to trade off exploitation against exploration, resulting in $\hat{p} = \hat{p}_M$. This differs from the fixed-state model, where irreversible abandonment of exploration might occur. The changing world motivates persistent vigilance, ensuring monitoring even in the prolonged absence of infractions.

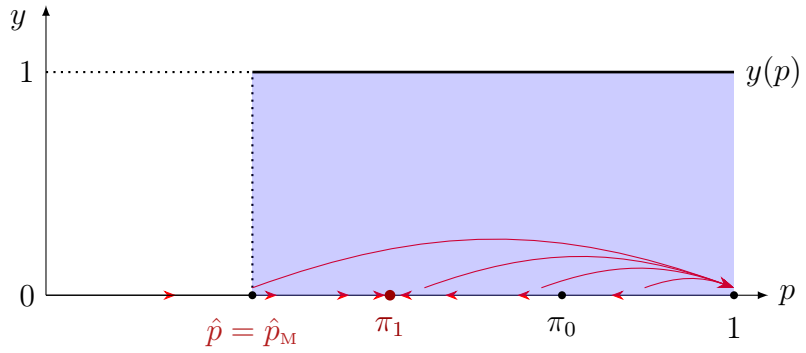


Figure 3: High Infraction Rates Case

Case 3: Intermediate infraction rates. In this case, the monitoring cutoff lies between π_1 and π_0. For beliefs above $\hat{p}$, the principal monitors fully. If no infraction is detected, the belief drifts down, potentially reaching $\hat{p}$. At $\hat{p}$, partial monitoring at an interior level $y = z(\hat{p}) \in (0, 1)$ is employed, freezing the belief at $\hat{p}$ until a detection triggers a jump to 1. See Figure 4.

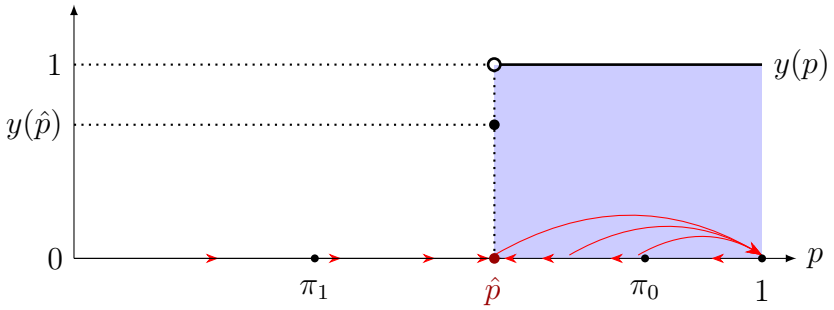


Figure 4: Intermediate Infraction Rates Case

The system spends a significant amount of time at $\hat{p}$, and monitoring alternates between full and partial levels. This pattern deviates from the standard bandit model, where action eventually becomes either full or nonexistent. The changing world leads to this belief-freezing behavior. The optimal policy calibrates this feature to optimally "hedge"—to balance the increasing likelihood of the safe state against the possibility that the environment can become high-risk again. Because partial monitoring suppresses exploration, the standard exploitation-exploration logic dictates pushing the cutoff below the myopic level ($\hat{p} < \hat{p}_M$).

Remark 1 (**Long-run distribution**). One can characterize the stationary distribution of the principal's belief and the monitoring level under OP and MP. If the cutoff lies within (π_1, π_0), monitoring has a two-point distribution with support $\{z(\hat{p}), 1\}$ under either policy.

The long-run proportion of time the monitoring takes each level can be computed precisely; we do this in Section 4.2. Comparing the long-run monitoring levels in states H and L reveals the potential benefit of data-driven monitoring. Under SP, the principal chooses the same monitoring level in both states, but under either MP or OP, the principal chooses, on average, a strictly higher monitoring level in $\omega = H$ than in $\omega = L$.

3.2 Comparison of Regimes

This section compares the optimal policy (OP) with the static (SP) and myopic (MP) benchmarks, focusing on scenarios where OP strictly outperforms them. We are specifically interested in their long-run performance—evaluating the principal's expected discounted loss once the system reaches a steady state and the posterior belief p enters its long-run support.

Because the policies induce varying degrees of exploration and learning, their long-run supports differ. However, they are structurally nested: the support under OP contains the support under MP, which in turn contains the degenerate support under SP, $\{\pi_0\}$.[6] Because expected discounted payoffs depend heavily on the starting belief, meaningful welfare comparisons require evaluating losses from a common baseline: $p_0 = \pi_0$. This represents the unique intersection of the long-run supports across all three regimes. Anchoring the evaluation at π_0 does not render other belief states irrelevant; rather, the divergence in long-run performance is driven precisely by the possibility of the long-run belief visiting these other states across the different regimes.

We initially examine cases where data-driven monitoring, even when optimized, offers no long-run advantage.

Proposition 1. *In Case 1 and Case 2, the monitoring policy is identical in the long run across all three regimes.*

In scenarios with low or high infraction rates, the OP, SP, and MP converge to the same actions in the long run because the prediction problem is relatively simple, and past data does not significantly improve decision-making. The more interesting scenario is Case 3, where prediction is valuable. There are two possibilities:

In Case 3(a), $\hat{p}_M \geq \pi_0$. Under MP, the belief eventually converges to π_0, resulting in the cessation of monitoring and exploration, mirroring the outcome under SP. Consequently, both SP and MP under-monitor relative to OP, which prescribes positive monitoring even in the long run.

[6] In Case 2 and Case 3, the support under OP is $[\hat{p}, 1]$ and the support under MP is $[\hat{p}_M, 1]$, where $\hat{p} \leq \hat{p}_M$. In Case 1, all three regimes degenerate to the exact same singleton support, $\{\pi_0\}$.

In Case 3(b), $\hat{p}_M < \pi_0$. In this case, SP prescribes full monitoring, whereas MP alternates between full monitoring (when $p > \hat{p}_M$) and partial monitoring (when $p = \hat{p}_M$). The comparison reveals that MP under-monitors relative to OP, as the optimal cutoff $\hat{p}$ is set below $\hat{p}_M$ to encourage more exploration (see Figure 5). In contrast, SP results in over-monitoring compared to OP.

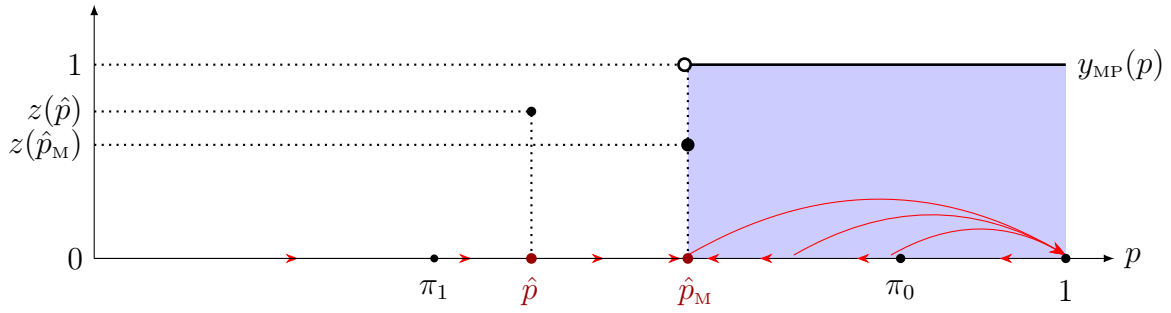


Figure 5: Monitoring Policy under MP: Case 3(b)

Proposition 2. *In Case 3(a), both SP and MP lead to no monitoring, which is too little compared with OP. In Case 3(b), SP leads to full, and thus excessive, monitoring, whereas MP leads to insufficient monitoring, compared with OP.*

Payoff comparisons. How does MP compare with SP in total expected loss? Surprisingly, while MP prescribes distinct monitoring levels based on predictions, the two regimes entail the exact same losses in the long run:

Proposition 3 (Payoff comparison)**.** *The expected loss from MP is always identical to that from SP in the long run. The expected loss is strictly lower under OP in Case 3.*

Proof. It suffices to show the equivalence for Case 3(b), in which $\pi_1 < \hat{p} < \hat{p}_{\text{M}} < \pi_0$. Recall from Proposition 2 that SP prescribes full monitoring in this case, so the intertemporal loss from SP is c/r. Meanwhile, MP implements monitoring that varies with p_t. To prove the equivalence, it suffices to show that the intertemporal loss from MP is c/r when evaluated at our common long-run baseline $p_0 = \pi_0$, which lies within $[\hat{p}_M, 1]$, the long-run support of p under MP. The equivalence follows since the total loss under MP is: for any $p_0 \in [\hat{p}_{\text{M}}, 1]$,

$$\begin{aligned}
&\mathbb{E}\left[\int_0^\infty (p_t \lambda x(1-y_t) + c y_t) e^{-rt} dt \,\middle|\, p_0 = \pi_0\right] \\
&= \int_0^\infty \mathbb{E}\Big[(p_t \lambda x(1-y_t) + c y_t) \Big| p_0 = \pi_0\Big] e^{-rt} dt
\end{aligned}$$

$$= \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\{p_t > \hat{p}_{\mathrm{M}}\}} \cdot c + \mathbf{1}_{\{p_t = \hat{p}_{\mathrm{M}}\}} \cdot (\hat{p}_{\mathrm{M}} \lambda x (1 - z(\hat{p}_{\mathrm{M}})) + c z(\hat{p}_{\mathrm{M}})) \,\Big|\, p_0 = \pi_0\Big] e^{-rt} dt$$
$$= \int_0^\infty c e^{-rt} dt = c/r,$$

where we used the fact that $y_{\mathrm{MP}}(p_t) = \mathbf{1}_{\{p_t > \hat{p}_{\mathrm{M}}\}} + \mathbf{1}_{\{p_t = \hat{p}_{\mathrm{M}}\}} \cdot z(\hat{p}_{\mathrm{M}})$ and that $\hat{p}_{\mathrm{M}} = \frac{c}{\lambda x}$. The last statement follows from the fact that OP prescribes a different action than MP in Case 3, making its expected loss strictly lower. □

The implication is striking: data-driven prediction has no value if it is used in a myopically optimal manner. MP prescribes no monitoring when $p < \hat{p}_{\mathrm{M}}$, whereas SP always prescribes full monitoring; and in this case, no monitoring is strictly better than full monitoring. Why doesn't MP then strictly outperform SP? Because no such $p < \hat{p}_{\mathrm{M}}$ is part of the support of the principal's beliefs. Prediction arises *only* when there is monitoring, but monitoring never occurs when it is undesirable under MP. This makes the prediction obtained under MP redundant.

Passive Learning. The equivalence between MP and SP rests crucially on the assumption that infractions are detected only through active monitoring. Although this assumption is valid for many unobservable regulatory violations, detection may also arise from passive channels, such as whistleblowing by witnesses, victim reporting, or private parties bringing suits to recover damages. Appendix E extends our model to allow for such a possibility; we assume that the principal detects an infraction with probability $w \in (0, 1)$ *passively* even when active monitoring fails to detect one. Let $\pi_w \in (\pi_1, \pi_0)$ denote the unique belief satisfying $f(\pi_w, w) = 0$, i.e. the long-run belief under no active monitoring when passive learning arrives at rate w.

Proposition 4. *If $\hat{p}_M > \pi_w$, MP achieves a strictly lower long-run loss for the principal than SP. Otherwise, MP and SP achieve the same long-run loss.*

Proposition 4 demonstrates that MP benefits from passive learning, while SP does not. Passive learning enables the principal's belief to drop below the myopic cutoff $\hat{p}_M$, where MP recommends no monitoring—a strategy that is strictly optimal. Put simply, prediction occurs even when not monitoring is the best choice from a myopic perspective. However, as $w \to 0$, MP's superiority over SP vanishes. This equivalence highlights the extent to which the endogeneity of learning limits the value of data-driven policies.

4 Endogenous Infractions and Deterrence

By assuming exogenous infraction behavior, the previous section focused on the principal's learning and remedial roles. However, an equally important goal of dynamic monitoring is deterring infractions. To study the deterrence implications of alternative policies, we must endogenize the agent's behavior.

To this end, we adopt a simple model in which an agent commits an infraction only if they anticipate the monitoring rate to be less than some deterrence threshold $\bar{y} \in (0, 1)$. More precisely, an agent offends with probability $x = 0$ if $y > \bar{y}$, with probability $x = 1$ if $y < \bar{y}$, and with any probability $x \in [0, 1]$ if $y = \bar{y}$. This behavior can be micro-founded: suppose an agent gains a private benefit $b > 0$ from an infraction but pays a penalty $\gamma \geq b$ when detected; the assumed behavior emerges with a threshold $\bar{y} := b/\gamma$.[7]

With the endogenous infraction model, the principal's ability to commit to her monitoring policy becomes relevant. If the principal could commit to her policy, she could completely deter infractions by monitoring slightly above $\bar{y}$. More precisely, a simple full-commitment policy is either to deter infractions fully by monitoring arbitrarily above $\bar{y}$ or never to monitor, whichever is cheaper. In practice, however, such formal commitment power is often lacking in regulatory contexts. Accordingly, we henceforth assume the principal cannot commit to a future monitoring level y, meaning she can only choose her policy in response to the likely infraction rate. Effectively, this transforms the bandit problem into a dynamic game in which the infraction rate and the monitoring policy are chosen as mutual best responses to each other.

As before, we consider three distinct regimes: SP, MP, and OP. Unlike the principal, agents do not access past detection data, meaning the infraction level x is a constant (rather than a function of p), except now it is determined endogenously by the agents' equilibrium incentives. (In Section 4.5, we will discuss the implications of agents gaining access to the same data as the principal).

Consistent with the exogenous case, we focus on a Markov perfect equilibrium in which the principal's policy depends only on her posterior, the payoff-relevant state. Since agents choose a constant probability $x \in [0, 1]$, the principal's best response follows the characterization from the previous section. The only difference is that the agents' equilibrium choice of x will depend on the principal's monitoring policy, and thereby on the monitoring regime. Given the endogeneity of x and the lack of commitment power by the principal, the OP is not guaranteed

[7] Note that agents are myopic here. This is because there is a continuum of individuals who may receive an opportunity; any effect of an infraction decision by an individual on future monitoring will be irrelevant since the probability of receiving an opportunity in the future is zero.

to be optimal or even to outperform the other two regimes. We begin our analysis with the SP regime.

4.1 Static Policy

In this regime, neither the agents nor the principal update their beliefs based on past history. Hence, the infraction probability x and the monitoring level y are constants, determined solely based on the long-run prior π_0. The outcome then depends on whether λ is large relative to c for monitoring to be worthwhile for the principal.

Suppose first $\lambda\pi_0 \leq c$. Then, the opportunity arises so rarely that even if the agents choose $x = 1$, it is never optimal for the principal to choose any positive y. So $(x, y) = (1, 0)$ is the only equilibrium. In that equilibrium, the principal's long-run expected loss, evaluated from the baseline belief $p_0 = \pi_0$, equals $\lambda\pi_0/r$.

Suppose next $\lambda\pi_0 > c$. In this case, a classic inspection game ensues. It is well known that there is no pure strategy equilibrium. If agents choose $x = 1$, then the principal will choose $y = 1$. But then agents would deviate to $x = 0$. The principal will, in turn, deviate to $y = 0$. Hence, the only equilibrium is in mixed strategies. In equilibrium, an agent chooses x so that

$$\pi_0\lambda x = c, \tag{7}$$

which keeps the principal indifferent in her monitoring. In turn, the principal chooses $y = \bar{y}$, incentivizing agents to randomize.

The structure of this equilibrium is explained by the principal's lack of commitment power. Unable to commit to a deterring monitoring level, the principal requires infractions to occur on the path to exert monitoring effort. Indeed, the principal's expected loss equals c/r regardless of $\bar{y}$ since the principal is indifferent to full monitoring. We summarize the results.

Proposition 5. *SP admits a unique equilibrium in which:*

- *if* $\lambda \leq \frac{c}{\pi_0}$, *then* $(x_{SP}, y_{SP}) = (1, 0)$;
- *if* $\lambda > \frac{c}{\pi_0}$, *then* $(x_{SP}, y_{SP}) = (\frac{c}{\pi_0\lambda}, \bar{y})$.

Under SP, the expected loss for the principal, evaluated from the baseline belief $p_0 = \pi_0$, *is* $\min\{\pi_0\lambda, c\}/r$.

4.2 Agents' Belief Formation under Data-driven Policies

We next turn to MP and OP. Given an equilibrium infraction level x, our analysis from Section 3 shows that the principal's best response is a cutoff policy in these two regimes. Fix such a policy with any cutoff $\hat{p} \in (\pi_1, \pi_0)$:[8] $y_{\hat{p}}(p) := \mathbf{1}_{\{p>\hat{p}\}} + z(\hat{p}) \cdot \mathbf{1}_{\{p=\hat{p}\}}$.

To understand agents' incentives under such a policy, we must first study their beliefs about the monitoring level $\mathbb{E}_p[y_{\hat{p}}(p)|\omega = H, x]$. The expectation is conditional on state $\omega = H$ (which the agents know upon receiving the opportunity) and the equilibrium infraction level x (which influences the belief about the principal's posterior and, henceforth, the monitoring level). Since the monitoring level depends on the principal's belief p, the agent must form a belief about the relative likelihoods of alternative p's the principal may hold. In the long run, this belief is given by the stationary distribution of the principal's belief.[9] We thus digress briefly on how the distribution is characterized.

4.2.1 Stationary distribution of the principal's belief.

To begin, fix a cutoff policy with an arbitrary cutoff $\hat{p} \in (\pi_1, \pi_0)$ and an arbitrary infraction level $x \in (0, 1)$. Such a policy induces a Markov process on the belief $p \in [\hat{p}, 1]$, which admits a unique stationary distribution $\Phi : [\hat{p}, 1] \to [0, 1]$.[10] Appendix F characterizes the stationary distribution:

Lemma 1. *The stationary distribution Φ for given x and policy $y_{\hat{p}}$, with $\hat{p} \in (\pi_1, \pi_0)$, admits density $\phi(p)$ satisfying:*

$$\phi(p) = \phi(\hat{p}) \exp\left(\int_{\hat{p}}^{p} r(s)ds\right) \tag{8}$$

for each $p \in (\hat{p}, 1)$, where $r(s) := -\frac{s\lambda x + f'(s,1)}{f(s,1)}$, and atom $\Phi(\hat{p})$ at belief $\hat{p}$ satisfying:

$$\Phi(\hat{p}) z(\hat{p}) \hat{p} \lambda x = -\phi(\hat{p}) f(\hat{p}, 1). \tag{9}$$

Proof. See Appendix F. □

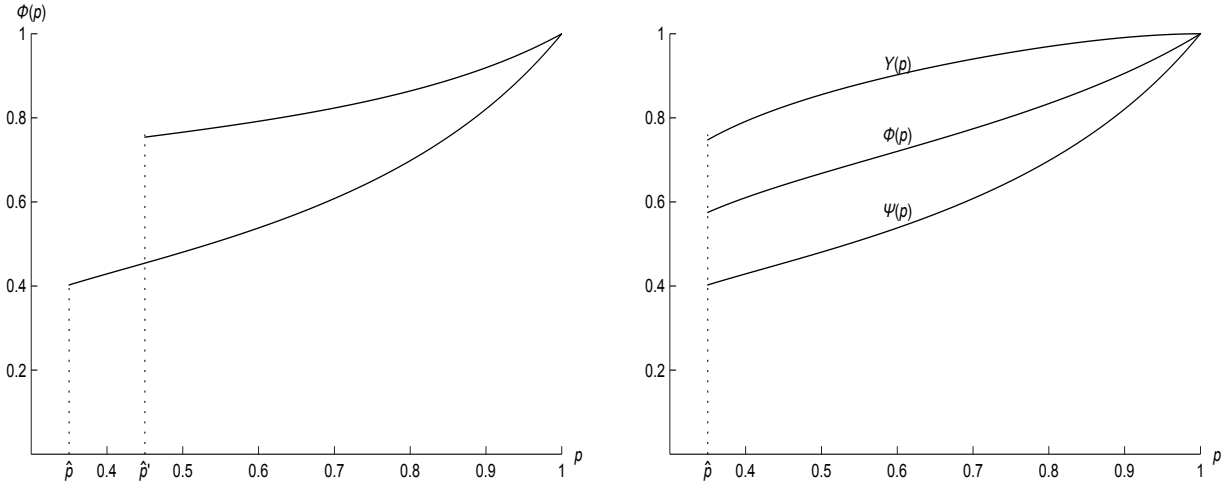
[8] The long-run monitoring will be degenerate otherwise: It is zero if $\hat{p} \geq \pi_0$, and one if $\hat{p} \leq \pi_1$.

[9] The formal justification can be given by the logic of "Poisson arrivals see time averages" (PASTA); see Wolff (1982). Intuitively, for a Poisson-arriving agent, his belief that p falls into some interval $(p', p'') \subset [\hat{p}, 1]$ equals the fraction of time p lies in that interval, once proper conditioning is made on the fact that state $\omega = H$.

[10] The process is Markov since once any $p \in [\hat{p}, 1]$ is reached, the future process is identical and depends only on p, regardless of the prior history. It is also positive-recurrent, with the process returning to the same belief in finite expected time. The regenerative nature of the process means that it admits a unique limit distribution forming a unique stationary distribution (see Asmussen (2003), p. 170, Theorem 1.2).

Intuitively, the posterior belief follows a Poisson process at each $p \in (\hat{p}, 1]$, leading to a well-defined density in the stationary distribution for that region. Meanwhile, the partially absorbing nature of belief $\hat{p}$ means that the system spends real time at that belief, explaining the atom in the stationary distribution.

Figure 6(a) depicts Φ under two cutoffs $\hat{p}$ and $\hat{p}' > \hat{p}$. The policy with a lower cutoff involves more exploration. Hence, the stationary distribution corresponding to $\hat{p}$ is a mean-preserving spread of the one corresponding to $\hat{p}'$.[11]



(a) CDF Φ's under two cutoffs $\hat{p}$ and $\hat{p}' > \hat{p}$

(b) CDF's Φ, Υ, Ψ

Figure 6: Invariant distributions of posterior beliefs

Using the unconditional stationary distribution Φ, we can derive the long-run distribution of the beliefs in each state. The distributions of p conditional on states H and L are respectively:

$$\Psi(p) := \frac{\hat{p}\Phi(\hat{p}) + \int_{\hat{p}}^{p} \tilde{p}\phi(\tilde{p})d\tilde{p}}{\pi_0} \text{ and } \Upsilon(p) := \frac{(1-\hat{p})\Phi(\hat{p}) + \int_{\hat{p}}^{p}(1-\tilde{p})\phi(\tilde{p})d\tilde{p}}{1-\pi_0}$$

if $p \geq \hat{p}$, and zero otherwise. Naturally, Ψ first-order stochastically dominates Φ, which in turn first-order stochastically dominates Υ. Figure 6(b) depicts Φ, Ψ, and Υ, for a given cutoff policy. Observe that Ψ puts a smaller mass at $\hat{p}$ than does Υ. To the extent that a cutoff policy enforces strictly less at $\hat{p}$ than at $p > \hat{p}$, the policy monitors more in state H than in L, revealing how prediction benefits monitoring (both in MP and OP). The relevant belief the agents use to evaluate the monitoring level is $\Psi(p)$, as they know the state is H when making their infraction decisions.

[11] This observation is proved in Lemma 12 in Appendix F.

4.2.2 Agents' belief formation.

We now characterize agents' beliefs about the monitoring level under a data-driven policy with cutoff $\hat{p}$. The answer is trivial if $\hat{p}$ lies outside (π_1, π_0): it is zero if $\hat{p} \geq \pi_0$, and one if $\hat{p} \leq \pi_1$. We therefore assume $\hat{p} \in (\pi_1, \pi_0)$. In this case, monitoring is binary: $y(p) = 1$ for $p > \hat{p}$ and $y(p) = z(p) \in (0, 1)$ for $p = \hat{p}$.

The preceding analysis suggests that agents form their beliefs about monitoring using Ψ as the distribution. Let $\mu(\hat{p}, x) := \Psi(\hat{p})$ denote the mass assigned to $\hat{p}$—the fraction of time the belief is at $\hat{p}$ given state H. Then, the agents face a conditional monitoring level:

$$\mathbb{E}[y_{\hat{p}}(p)|\omega = H, x] = \mathbb{E}_{p\sim\Psi}[y_{\hat{p}}(p)] = \mu(\hat{p}, x)z(\hat{p}, x) + (1 - \mu(\hat{p}, x)) \cdot 1, \tag{10}$$

where we now make explicit the dependence of $z(\hat{p})$ on x. While $z(\hat{p}, x)$ is decreasing in $\hat{p}$, we establish in Appendix F that $\mu(\hat{p}, x)$ is increasing in $\hat{p}$. It then follows from (10) that the anticipated monitoring level $\mathbb{E}_{p\sim\Psi}[y_{\hat{p}}(p)]$ is decreasing in $\hat{p}$. This is intuitive since a lower cutoff policy involves more monitoring. We now show that for any x, there exists a unique cutoff $\hat{p} = \overline{p}(\lambda x)$ that gives rise to the deterring level of monitoring $\overline{y}$.

Lemma 2. *Given any $x \in (0, 1]$, there exists a cutoff $\overline{p}(\lambda x) \in (\pi_1, \pi_0)$ such that the associated cutoff policy $y_{\hat{p}}$ with $\hat{p} = \overline{p}(\lambda x)$ gives rise to a conditional monitoring level of $\overline{y}$; i.e.,*

$$\mathbb{E}_{p\sim\Psi}[y_{\hat{p}}(p)] = \overline{y}. \tag{11}$$

Also, the deterring cutoff $\overline{p}(\lambda x)$ is continuous in x and decreases in $\overline{y}$, with $\lim_{\overline{y}\to 0} \overline{p}(\lambda x) = \pi_0$ and $\lim_{\overline{y}\to 1} \overline{p}(\lambda x) = \pi_1$ for all x.

Proof. See Appendix F. □

This characterization is crucial for our equilibrium analysis of MP and OP. Suppose the equilibrium has an interior infraction level $x \in (0, 1)$ in either the MP or OP regime. Then, agents must be indifferent, meaning they anticipate the monitoring level of $\overline{y}$ upon receiving an opportunity. Lemma 2 means that, for such an equilibrium to obtain, the principal must find it optimal to choose a cutoff $\overline{p}(\lambda x)$. This, in turn, pins down the equilibrium infraction level x.

4.3 Myopic Policy

We are now ready to study the principal's behavior under MP. It is convenient to define $\lambda^M \in (\frac{c}{\pi_0}, \frac{c}{\pi_1})$ to be the smallest λ that satisfies:

$$\hat{p}_{\text{M}}(\lambda) = \frac{c}{\lambda} = \overline{p}(\lambda). \tag{12}$$

If $\lambda = \lambda^M$, then given $x = 1$, the MP policy (with its myopic cutoff $\hat{p}_M(\lambda)$) satisfies (11) and implements the deterring monitoring level $\overline{y}$. Such a λ^M is well-defined.

To characterize the MP equilibrium, consider first $\lambda \leq \lambda^M$. In this case, $\hat{p}_{\text{M}}(\lambda x) = \frac{c}{\lambda x} > \overline{p}(\lambda x)$ for all $x < 1$. In other words, the MP policy does not generate sufficient monitoring to deter infractions. Hence, in equilibrium, agents choose $x_{\text{MP}} = 1$, and the principal responds with the corresponding myopic cutoff $\hat{p}_{\text{M}}(\lambda) = \frac{c}{\lambda}$.

Consider next $\lambda > \lambda^M$. In this case, we have a mixed strategy equilibrium in which agents commit infractions with probability $x \in (0, 1)$. For them to randomize, the anticipated monitoring level must be $\overline{y}$. This requires the principal to find it optimal to choose $\overline{p}(\lambda x)$. Under MP, this requires that:

$$\hat{p}_{\text{M}}(\lambda x) = \frac{c}{\lambda x} = \overline{p}(\lambda x), \tag{13}$$

which determines the equilibrium infraction level x.

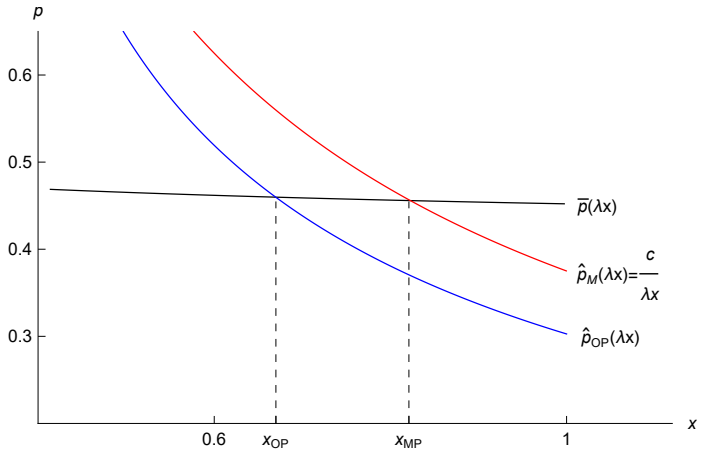


Figure 7: Equilibria under OP and MP with $\lambda = 4$, $c = \frac{3}{2}$, $r = 2$, $\rho_L = \rho_H = 1$, and $\overline{y} = 0.3$.

Figure 7 illustrates how the equilibrium infraction level is determined. The black curve depicts the deterring cutoff $\overline{p}(\lambda x)$ as a function of x.[12] Meanwhile, the red curve depicts the

[12]The curve $\overline{p}(\cdot)$ is negatively sloped in all numerical analyses we conducted, which is intuitive. Suppose x increases. First, $z(\hat{p})$ must decline so that $xz(\hat{p})$ remains constant, implying the principal monitors less at $\hat{p}$.

myopic cutoff $\hat{p}_{\text{M}}(\lambda x) = \frac{c}{\lambda x}$ as a function of x; it is downward sloping since a higher infraction rate calls for increasing monitoring, or a lower cutoff. The equilibrium condition (13) holds at the intersection of the two curves.

Note from (12) that $x = \lambda^M/\lambda$ solves (13). We thus have the following result:

Proposition 6. *MP admits an equilibrium in which:*

- *if* $\lambda \leq \lambda^M$, *then* $x_{MP} = 1 \geq \min\{1, \frac{c}{\pi_0 \lambda}\} = x_{SP}$, *and* $y_{MP}(\cdot)$ *is a cutoff policy with* $\hat{p} = \frac{c}{\lambda}$;
- *if* $\lambda > \lambda^M$, *then* $x_{MP} = \frac{\lambda^M}{\lambda} \in (0,1)$ *and* $y_{MP}(\cdot)$ *is a cutoff policy with* $\hat{p} = \frac{c}{\lambda^M}$.

Under MP, the expected loss for the principal, evaluated from the baseline belief $p_0 = \pi_0$, *is* $\min\{\pi_0 \lambda, c\}/r$.

The payoff characterization follows from a simple observation. When $\pi_0 \lambda \leq c$, the principal chooses zero monitoring in the long run, so the total loss is simply the loss from infractions, $\pi_0 \lambda / r$. If $\pi_0 \lambda > c$, the principal adopts a nontrivial cutoff policy against the infraction level $x_{\text{MP}} = \min\{\frac{\lambda^M}{\lambda}, 1\}$. Since the principal is indifferent at the myopic cutoff, her payoff is the same as if she always monitors fully, making her total loss c/r. Note that the payoff does not depend on the deterring monitoring level $\overline{y}$. This irrelevance follows from the principal's lack of commitment power.

A comparison with Proposition 5 establishes payoff equivalence:

Corollary 1 (Payoff equivalence between SP and MP)**.** *The expected loss for the principal is the same between SP and MP.*

The intuition for this equivalence is the same as before: the information is obtained *only when monitoring is myopically optimal* and hence is redundant.[13]

4.4 Optimal Policy

We now study the Markov perfect equilibrium under OP. From our analysis in Section 3, for any infraction level x, the principal's optimal response is a cutoff policy, and the resulting outcome depends on the optimal cutoff $\hat{p}_{\text{OP}}(\lambda x)$, as characterized in Theorem 1.

The expected monitoring would fall—hence $\bar{p}$ should decrease—if the belief stays longer at $\hat{p}$, that is, if $\Phi(\hat{p})$ increases. This tends to be the case because a higher x causes the belief to drift downward at a faster rate, although a formal proof is unavailable due to the ambiguity in determining the change of $\phi(\hat{p})$.

[13] If $\pi_0 \lambda > c$, the principal chooses full monitoring under SP, but posterior-dependent monitoring under MP. Yet, whenever the monitoring is partial under MP, the principal is indifferent, so she experiences the exact same expected loss as full monitoring.

To characterize the equilibrium, define $\lambda^* \in (\underline{\lambda}, \lambda^M)$ to be the smallest λ that satisfies:

$$\hat{p}_{\text{OP}}(\lambda) = \overline{p}(\lambda). \tag{14}$$

That is, if $\lambda = \lambda^*$, then, given $x = 1$, an optimal cutoff $\hat{p}$ yields a deterring monitoring level $\overline{y}$. Such a λ^* is well defined.

Consider first $\lambda \leq \lambda^*$. In this case, we have $\hat{p}_{\text{OP}}(\lambda x) > \overline{p}(\lambda x)$ for all $x < 1$, meaning monitoring will be insufficient to deter infractions. Hence, in equilibrium, agents choose $x = 1$, and the principal optimally responds with a cutoff $\hat{p}_{\text{OP}}(\lambda)$.

Consider next $\lambda > \lambda^*$. In this case, agents must randomize in equilibrium. This requires that the anticipated monitoring level equal $\overline{y}$, or:

$$\hat{p}_{\text{OP}}(\lambda x) = \overline{p}(\lambda x), \tag{15}$$

which pins down the equilibrium infraction level x_{OP}. Figure 7 depicts x_{OP} as an intersection of the optimal cutoff facing x (the blue curve) and the deterring-cutoff level $\overline{p}(x)$ (the black curve). It follows from (14) that $x = \lambda^*/\lambda$ solves this equation. Hence, in equilibrium, agents choose $x_{\text{OP}} = \lambda^*/\lambda$ and the principal chooses a cutoff $\hat{p}_{\text{OP}}(\lambda^*) = \overline{p}(\lambda^*)$, which yields $\overline{y}$ in expectation, justifying the agents' randomization.

Proposition 7. *The OP regime admits an equilibrium in which:*

- *if $\lambda \leq \lambda^*$, then $x_{OP} = x_{MP} = 1$, and $y_{OP}(\cdot)$ is a cutoff policy with $\hat{p} = \hat{p}(\lambda)$ which is less than the MP cutoff $\hat{p}_M(\lambda) = \frac{c}{\lambda}$;*
- *if $\lambda > \lambda^*$, then $x_{OP} = \frac{\lambda^*}{\lambda} < \min\{\frac{\lambda^M}{\lambda}, 1\} = x_{MP}$, and $y_{OP}(p)$ is a cutoff policy with $\hat{p} = \overline{p}(\lambda^*)$.*

The principal incurs a weakly (resp. strictly) lower expected loss under OP than under MP or SP (if $\lambda > \underline{\lambda}$).

It is instructive to compare the equilibrium outcomes of OP and MP, as illustrated in Figure 7. The crucial difference between the regimes is that the optimal policy is always more aggressive than the myopic one. For any given infraction level x, the OP policy's informational motive leads it to choose a strictly lower monitoring cutoff, i.e., $\hat{p}_{\text{OP}}(\lambda x) < \hat{p}_{\text{M}}(\lambda x)$. This difference in aggressiveness is what drives the infraction rate down under OP.

To see the intuition, imagine the infraction rate is at the myopic equilibrium, x_{MP}. By definition, the myopic cutoff $\hat{p}_{\text{M}}(\lambda x_{\text{MP}})$ generates just enough monitoring to make agents indifferent. At this same infraction rate, however, the OP principal chooses the strictly lower

cutoff $\hat{p}_{\text{OP}}(\lambda x_{\text{MP}})$, leading to a higher expected monitoring level. This over-monitoring incentivizes agents to commit fewer infractions. Consequently, the equilibrium infraction rate must be lower under OP ($x_{\text{OP}} < x_{\text{MP}}$), settling at the point where the OP principal's greater aggressiveness is perfectly balanced by the lower infraction rate to maintain the deterrence threshold.

In short, the informational motive of the principal acts as an endogenous commitment to deter infractions. This explains why the expected loss is smaller under OP than the other two regimes. Specifically, if $\lambda \leq \underline{\lambda}$, all three regimes lead to no monitoring resulting in the identical expected loss of $\pi_0\lambda/r$. If $\lambda > \underline{\lambda}$, the OP entails an expected loss strictly smaller than c/r, the loss under SP and MP.

4.5 Data-aware Agents

So far, we have assumed that agents cannot observe the history of detections in the way the principal can. This assumption is realistic in many regulatory contexts where agencies collect data privately and do not disclose their internal audit histories. In this section, we assume that agents are "data-aware": they have access to the same past detection data as the principal, using it to infer the principal's posterior belief p and respond accordingly at each p. Considering this scenario highlights the role of the informational advantage the principal has over the agents.

In the case of SP, since the principal has no information other than π_0, agents access no additional information, so the analysis remains unchanged. However, unlike the case of uninformed agents, MP and OP become strictly equivalent irrespective of λ.

When agents are data-aware and condition their infraction decisions on the same posterior belief as the principal, the MP and OP regimes admit the same equilibrium monitoring policy. In particular, there is a cutoff $\hat{p} = \frac{c}{\lambda}$ such that $y(p) = 0$ for $p < \hat{p}$ and $y(p) = \bar{y}$ for $p > \hat{p}$.[14] In response, agents choose $x(p) = 1$ for $p < \hat{p}$, and

$$x(p) = \frac{c}{p\lambda} \quad \text{for } p \geq \hat{p},$$

which keeps the principal exactly indifferent at every $p \geq \hat{p}$. Under MP, this strategy profile is the unique equilibrium. Moreover, because the principal is indifferent at every posterior above the cutoff, there is no residual informational gain from experimentation. Hence, the same strategy profile also constitutes an equilibrium under OP. Proposition 8 further shows that this equilibrium outcome is unique under OP within the class of **regular equilibria**,

[14] Any $y(\hat{p}) \in [0, \bar{y}]$ is consistent with equilibrium at the cutoff $\hat{p}$.

where the principal's monitoring policy is discontinuous at finitely many points. Thus, once agents observe the same posterior as the principal, they adjust their infraction decisions so as to eliminate the principal's informational motive for exploration.

Proposition 8. *When data-aware agents condition their infraction decisions on the same posterior belief as the principal, all three regimes yield the same total expected loss in any regular equilibrium under OP.*

Proof. See Appendix B. □

This result, together with Corollary 1 and Proposition 7, establishes that the principal's informational advantage is strictly necessary for her to realize the deterrence value of dynamically optimal monitoring.

5 Discussion

In this section, we discuss two broader implications of our dynamic monitoring framework: the restoration of classic deterrence theory under dynamically optimal policies, and the extension of our model to environments with multiple independent monitoring targets.

5.1 Monitoring Commitment and the Beckerian Prescription

A foundational premise of the economic approach to enforcement, originating with Becker (1968), is that increasing the penalty for an infraction should reduce its equilibrium occurrence. In our model, the principal's ability to manipulate the penalty corresponds to shifting the deterrence threshold $\overline{y}$. Specifically, if the principal increases the penalty γ assessed upon detection, the probability of monitoring required to deter an agent, $\overline{y} = b/\gamma$, strictly decreases. A robust deterrence regime should translate this lower threshold into a reduced equilibrium infraction rate x.

However, our analysis reveals that when monitoring is dynamic and the principal lacks formal commitment power, the effectiveness of this Beckerian prescription depends entirely on how the principal utilizes data. Under the SP and MP regimes, raising the penalty completely fails to reduce infractions. Recall from Proposition 5 and Proposition 6 that the equilibrium infraction rates under these myopic regimes—$x_{\text{SP}} = \min\{1, \frac{c}{\pi_0 \lambda}\}$ and $x_{\text{MP}} = \min\{1, \frac{\lambda^M}{\lambda}\}$—are determined exclusively by the principal's indifference conditions, which are independent of $\overline{y}$. If the principal increases the penalty (lowering $\overline{y}$), agents intuitively require less monitoring to be deterred. Because the myopic principal cannot commit to maintaining a high monitoring

intensity, she simply takes advantage of the higher penalty by monitoring less frequently, allowing the expected monitoring level to drop to the new, lower $\overline{y}$. The infraction rate x remains exactly unchanged, keeping the principal indifferent.

In sharp contrast, the optimal policy (OP) restores the effectiveness of the Beckerian prescription. Because the OP principal explicitly values the informational option value of monitoring, her desire to explore acts as an endogenous commitment to monitor more aggressively than is myopically justified. From Lemma 2, a decrease in $\overline{y}$ implies that the required deterring cutoff $\overline{p}(\lambda x)$ must increase (the principal can afford to wait until the belief is higher before monitoring). In the equilibrium condition for OP, $\hat{p}_{\text{OP}}(\lambda x_{\text{OP}}) = \overline{p}(\lambda x_{\text{OP}})$, an upward shift in the $\overline{p}$ curve intersects the optimally aggressive cutoff $\hat{p}_{OP}$ at a strictly lower infraction rate x_{OP}. Consequently, under the dynamically optimal regime, an increase in the penalty successfully translates into a lower equilibrium infraction rate. This highlights a profound complementarity between the severity of punishments and the dynamic sophistication of the monitoring regime: severe penalties only deter agents if the principal is committed to gathering information.

5.2 Multiple Arms and Whittle Dynamics

Throughout our analysis, we have focused on a single monitoring domain facing a continuum of agents, capturing the scarcity of enforcement resources through a constant shadow cost $c > 0$. In practice, the principal often allocates monitoring resources across many covariate profiles, transaction classes, agents, or network segments. For example, a cybersecurity monitor may distinguish domestic and foreign IP addresses, while a financial auditor may distinguish asset classes, business units, or counterparties. Each profile has its own evolving hidden risk state and its own posterior belief. The principal's prediction problem is then to continuously assess the unknown, changing effect of these profiles on the likelihood of an infraction.

To make the extension concrete, suppose there are several independent risky arms. Arm i corresponds to one covariate profile and has posterior belief p_i, transition rates $\rho_{L,i}$ and $\rho_{H,i}$, opportunity rate λ_i, and infraction rate x_i. Write $\beta_i = \lambda_i x_i$. When monitoring intensity y_i is assigned to arm i, the no-detection drift of its posterior is

$$f_i(p_i, y_i) = \rho_{L,i}(1 - p_i) - \rho_{H,i} p_i - \beta_i p_i (1 - p_i) y_i.$$

The principal has a limited monitoring budget, for instance $\sum_i y_i \leq 1$. Because the structural parameters may differ across covariate profiles, raw posterior beliefs are not directly comparable. A posterior of 0.6 for a persistent arm may represent a very different monitoring priority

from a posterior of 0.6 for a transitory arm.

A standard policy for restless multi-armed bandit problems is the Whittle-index policy.[15] Like the Gittins index policy for classical rested bandits, it assigns to each arm a scalar priority index and then allocates capacity to the arms with the highest indices. The difference is that, in our environment, inactive arms are not frozen: their hidden states continue to change and their posteriors continue to drift. The Whittle policy is therefore best understood as a tractable Lagrangian way to leverage the one-arm solution rather than as a direct solution of the full multidimensional dynamic program.

The Lagrangian interpretation is simple. In the one-arm problem, the monitoring cost c is the opportunity cost of using enforcement capacity on the risky arm. In a multi-arm problem, that opportunity cost is endogenous: using capacity on arm i means not using it on other arms. The Whittle construction replaces this hard, state-dependent opportunity cost by a constant shadow price c, solves the resulting one-arm dynamic problem for each arm, and then asks which arms would still be worth monitoring at the highest shadow price. Thus the policy is dynamic within each arm but price-based across arms: each arm's index incorporates the learning-aware continuation value of monitoring that arm, while the comparison across arms is made through a scalar shadow price rather than through the full multidimensional value function.

Formally, fix an arm i and interpret c as the shadow price of one unit of monitoring capacity. The one-arm problem analyzed in Theorem 1 yields an optimal cutoff $\widehat{p}_i(c)$. The monotonicity of this cutoff is precisely the relevant indexability property: as the shadow price c rises, the set of beliefs at which arm i is worth monitoring shrinks monotonically. Since $\widehat{p}_i(c)$ is continuously increasing in c, we can invert the cutoff rule and define the Whittle index of arm i at posterior p_i by

$$\mathcal{I}_i(p_i) := \widehat{p}_i^{-1}(p_i).$$

Equivalently, $\mathcal{I}_i(p_i)$ is the hypothetical monitoring cost that would make the principal exactly indifferent between monitoring and not monitoring arm i at belief p_i. A high index means that the arm remains worth monitoring even when monitoring capacity is costly; it therefore

[15] Whittle-index policies are the canonical Lagrangian index policies for restless multi-armed bandits (Whittle, 1988) They generalize the logic of the Gittins index for classical rested bandits, but unlike the Gittins index, a Whittle-index policy need not be exactly optimal in a finite restless-bandit problem. The reason is that a passive arm continues to evolve, so the true opportunity cost of monitoring one arm rather than another is itself state-dependent. The Whittle construction replaces this multidimensional opportunity cost by a scalar Lagrange multiplier. This scalar-pricing approximation is what makes the policy tractable and implementable, but it is also why global finite-arm optimality requires additional conditions. Existing results nevertheless provide an important justification: Whittle argued that the policy should be nearly optimal in large systems, and subsequent work proves asymptotic optimality under global-stability or global-attractor conditions; see Weber and Weiss (1990) and Verloop (2016).

has high priority relative to other arms. This is why the index allows the principal to compare arms whose raw posterior beliefs are not directly comparable.

This interpretation also clarifies why the Whittle policy is not myopic. The index of an arm is computed from the dynamic one-arm value problem, so it incorporates the informational option value of monitoring: current mitigation, the possibility that a detection resets the posterior to one, and the way that no detections change future beliefs. The simplification is instead in the comparison across arms. The policy does not compute the full multidimensional HJB value of leaving every other arm passive; it prices that opportunity cost through the scalar index.

The resulting Whittle dynamics are simple. Let

$$\mathcal{T}(p) := \{i : \mathcal{I}_i(p_i) = \max_j \mathcal{I}_j(p_j)\}$$

be the set of arms with the highest current index. If a single arm has the highest index, the principal monitors that arm fully. If several arms share the highest index, the continuous-time implementation must satisfy the same admissibility logic as in the single-arm cutoff problem: it must avoid instantaneous switching among tied arms. Thus the principal splits capacity across tied arms so that, absent a detection, their indices evolve together and remain tied. Formally, arms outside $\mathcal{T}(p)$ receive zero monitoring, while tied arms receive intensities satisfying

$$\sum_{i \in \mathcal{T}(p)} y_i = 1$$

and

$$\mathcal{I}_i'(p_i) f_i(p_i, y_i) = \mathcal{I}_j'(p_j) f_j(p_j, y_j) \qquad \text{for all } i, j \in \mathcal{T}(p).$$

This equal-index-drift rule is the multi-arm analogue of the interior monitoring intensity $z(\widehat{p})$ in the one-arm problem. In the one-arm case, $z(\hat{p})$ keeps the posterior at the cutoff, and equivalently keeps the arm's index fixed at the indifference level because the outside option is constant. In the multi-arm case, there is no single fixed outside index; instead, the fractional allocation keeps the tied top indices moving together until a detection breaks the tie.

The dynamics are intuitive. As an active arm is monitored and no infraction is detected, its posterior and index tend to fall. An inactive arm's posterior continues to evolve toward its stationary benchmark, so its index may rise while the arm is not monitored. When several indices meet at the top, monitoring is shared to maintain the tie. If a detection occurs on one arm, that arm's posterior jumps to one, and its index rises discretely; that arm then receives priority until its index falls back to the common priority level.

This dynamic index policy preserves the central mechanism of the paper. It avoids the purely greedy logic that would permanently neglect profiles that currently look safe, while focusing monitoring on the covariate profiles whose beliefs, transition rates, and detection opportunities make them most valuable to inspect. In this sense, the Whittle construction provides a scalable way to extend the informational commitment logic of the single-arm optimal policy to environments with many evolving sources of risk.

6 Conclusion

The increasing reliance on algorithmic prediction and historical data to allocate monitoring resources presents a profound dynamic mechanism design challenge. While data-driven monitoring promises enhanced efficiency, it is fundamentally constrained by the endogeneity of data generation: the principal only observes the state of the environment when actively monitoring.

This paper provides a rigorous theoretical framework to evaluate the long-run consequences of this endogeneity. We demonstrate that the value of predictive data is not automatic. If a principal utilizes data myopically—monitoring only when the current belief exceeds a static threshold—the data-driven algorithm is rendered functionally redundant. Because the principal ceases monitoring when an entity appears safe, she creates an inferential blind spot, failing to learn when the underlying state inevitably deteriorates. Consequently, myopic monitoring performs no better in the long run than a naive policy that ignores data entirely.

To unlock the true value of data, the principal must behave as a dynamically optimal experimenter in a changing world. Because the underlying state evolves according to a hidden Markov process, the need for learning never permanently resolves. The optimal policy explicitly internalizes the informational option value of monitoring, leading to a novel hedging behavior where the principal maintains persistent, partial vigilance even when the immediate risk appears low.

Crucially, when we embed this learning problem within a dynamic inspection game, we find that the principal's informational motive transcends mere prediction: it acts as an endogenous commitment device. The optimal principal's inherent desire to gather data organically pushes her monitoring intensity above the myopically optimal level. Knowing that the principal is eager to explore, strategic agents are forced to lower their equilibrium infraction rate to maintain the deterrence condition. Thus, dynamically optimal monitoring not only minimizes the principal's expected loss but also restores the effectiveness of classic Beckerian penalties, proving that the deterrence value of predictive data relies entirely on the principal's commitment

to persistent exploration.

References

Asmussen, S. (2003): *Applied Probability and Queues*. Springer-Verlag. 19

Ball, I., and J. Knoepfle (2023): "Should the Timing of Inspections be Predictable?," *arXiv preprint arXiv:2304.01385*. 5

Becker, G. S. (1968): "Crime and punishment: An economic approach," *Journal of political economy*, 76(2), 169–217. 5, 26

Che, Y., K. Kim, and W. Zhong (2020): "Statistical Discrimination in Ratings-Guided Markets," *arXiv preprint arXiv:2004.11531*. 5

Che, Y.-K., and J. Hörner (2018): "Recommender systems as mechanisms for social learning," *The Quarterly Journal of Economics*, 133(2), 871–925. 5

Dilmé, F., and D. F. Garrett (2019): "Residual deterrence," *Journal of the European Economic Association*, 17(5), 1654–1686. 5

Fryer, R., and P. Harms (2018): "Two-armed restless bandits with imperfect information: Stochastic control and indexability," *Mathematics of Operations Research*, 43(2), 399–427. 5

Keller, G., S. Rady, and M. Cripps (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, 73(1), 39–68. 2, 4, 8

Khalil, F. (1997): "Auditing without commitment," *The RAND Journal of Economics*, pp. 629–640. 5

Madsen, E., and E. Shmaya (2024): "Collective Upkeep," *arXiv preprint arXiv:2407.05196*. 5

Marinovic, I., and M. Szydlowski (2022): "Monitoring with career concerns," *The RAND Journal of Economics*, 53(2), 404–428. 5

Mayer-Schönberger, V., and K. Cukier (2014): *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Harper Business. 2

Nino-Mora, J. (2001): "Restless bandits, partial conservation laws and indexability," *Advances in Applied Probability*, 33(1), 76–98. 5

Urgun, C. (2021): "Restless Contracting," Working paper. 5

Varas, F., I. Marinovic, and A. Skrzypacz (2020): "Random inspections and periodic

reviews: Optimal dynamic monitoring," *The Review of Economic Studies*, 87(6), 2893–2937. 5

Verloop, I. M. (2016): "Asymptotically Optimal Priority Policies for Indexable and Non-indexable Restless Bandits," *The Annals of Applied Probability*, 26(4), 1947–1995. 28

Wagner, P., and J. Knoepfle (2021): "Relational enforcement," *Available at SSRN 3832863*. 5

Weber, R. R., and G. Weiss (1990): "On an Index Policy for Restless Bandits," *Journal of Applied Probability*, 27(3), 637–648. 28

Whittle, P. (1988): "Restless bandits: Activity allocation in a changing world," *Journal of applied probability*, 25(A), 287–298. 5, 28

Wolff, R. W. (1982): "Poisson arrivals see time averages," *Operations research*, 30(2), 223–231. 19

A Proof of Theorem 1

For the optimal policy, we consider a cutoff policy of the form: $y(p) = 1$ if $p > \hat{p}$ and $y(p) = 0$ if $p < \hat{p}$, for some threshold $\hat{p}$, and an associated value function that satisfies

$$V(\hat{p}_+) = V(\hat{p}_-) \quad \text{and} \quad V'(\hat{p}_+) = V'(\hat{p}_-), \tag{16}$$

known respectively as the *value-matching* and *smooth-pasting* conditions. We refer to such a policy as a *regular cutoff policy*. Note that we do not claim the necessity of (16). Instead, we will show that the candidate value function and cutoff policy constructed using these conditions satisfy the HJB equation, which, by the verification theorem, implies optimality.

The proof of Theorem 1 then proceeds in three main steps. Our objective is to show that a regular cutoff policy—defined by a single cutoff belief $\hat{p}$—is optimal, and to characterize this cutoff.

1. **Characterize Necessary Conditions**: We first establish a set of necessary conditions that any optimal regular cutoff policy and its associated value function $V(p)$ must satisfy. We consolidate these properties in Proposition 9 in Appendix A.1.

2. **Construct a Candidate Solution**: In Appendix A.2, we construct a candidate solution—a *unique* pair of a cutoff $\hat{p}$ and a value function $V(p)$—that satisfies all the necessary conditions from Step 1. The construction proceeds by analyzing the three

distinct cases for the infraction rate outlined in Theorem 1. This step establishes the existence and uniqueness of a candidate that fits our characterization.

3. **Verify Optimality**: Finally, in Appendix A.3, we verify that the candidate solution constructed in Step 2 is indeed the optimal policy. We do this by showing that our constructed value function and policy satisfy the HJB equation for all beliefs $p \in [0,1]$. By the standard verification theorem, this confirms the policy's optimality.

All omitted proofs in this section are provided in Appendix D of the Online Appendix.

A.1 Necessary Conditions for Optimality

For a cutoff policy, the HJB equation (5) reduces to a two-part ordinary differential equation (ODE):

$$rV(p) = \begin{cases} h(p)V'(p) & \text{for } p < \hat{p}, \quad (17) \\ p\lambda x - c + p\lambda x\big[V(1) - V(p)\big] + g(p)V'(p) & \text{for } p > \hat{p}, \quad (18) \end{cases}$$

which follows from setting $y = 0$ for $p < \hat{p}$ and $y = 1$ for $p > \hat{p}$ in (5).

A solution to this ODE depends on two endogenous variables: the cutoff $\hat{p}$ and the boundary value $V(1)$. To characterize the optimal policy, define the marginal value of enforcement as

$$\Delta(p) := \frac{1}{x}\frac{\partial W(p,y)}{\partial y} = p\lambda - \frac{c}{x} + p\lambda\big[V(1) - V(p) - (1-p)V'(p)\big]. \quad (19)$$

The next proposition provides conditions that determine $\hat{p}$ and $V(1)$ in the subsequent analysis:

Proposition 9. *Any regular cutoff policy that satisfies (5) must satisfy:*[16]

$$\Delta(\hat{p}) = 0 \quad (20)$$

and

$$V'(\hat{p}) = \begin{cases} 0 & \text{if } \hat{p} \geq \pi_0 \text{ (Case 1)} \quad (21) \\ \kappa & \text{if } \hat{p} \leq \pi_1 \text{ (Case 2)} \quad (22) \\ \sigma(\hat{p}) & \text{if } \hat{p} \in (\pi_1, \pi_0) \text{ (Case 3),} \quad (23) \end{cases}$$

[16] In fact, the proof of this proposition establishes stronger properties than (21) and (22): in Case 1, $V(p) = V'(p) = 0, \forall p \leq \hat{p}$; in Case 2, $V'(p) = \kappa, \forall p \geq \hat{p}$.

where

$$\kappa := \frac{\lambda x}{r+\rho_L+\rho_H} \quad \textit{and} \quad \sigma(p) := \frac{h(p)c}{p^2(1-p)\lambda x(r+\rho_L+\rho_H)}.$$

Condition (20) is the first-order condition ensuring that the principal is indifferent at the cutoff $\hat{p}$. The remaining conditions follow from the single-crossing property of $\Delta(\cdot)$ at $\hat{p}$: namely, $\Delta(p) \geq 0$ for $p \geq \hat{p}$ and $\Delta(p) \leq 0$ for $p \leq \hat{p}$. Given (20), this requires $\Delta'(\hat{p}_-) \geq 0$ and $\Delta'(\hat{p}_+) \geq 0$. This condition binds in Case 3, implying $\Delta'(\hat{p}) = 0$, which yields (23). In contrast, it is slack in Case 1 and Case 2, where (21) and (22) follow directly from the structures of the ODEs (17) and (18), respectively.

A.2 Construction of Value Function

We now begin the second step of our roadmap. Our goal is to construct a unique pair $(\hat{p}, V(1))$ that fulfills all the necessary conditions specified in Proposition 9.

To proceed, it is convenient to first focus on the ODE for $p > \hat{p}$ and parametrize it by the unknown boundary value $K := V(1)$. This defines a family of potential value functions, $V(p;K)$, each solving the system we label $D[K]$:

$$D[K] \qquad rV(p) = p\lambda x - c + p\lambda x\,[K - V(p)] + g(p)V'(p) \text{ with } V(1) = K. \tag{24}$$

We will search for a unique value of K and a corresponding belief $\hat{p}$ that jointly satisfy the conditions in Proposition 9. We let $\Delta(p;K)$ denote the function $\Delta(p)$ corresponding to the solution $V(p;K)$.

First, we establish some essential properties of the solutions to $D[K]$ that will be used throughout the construction.

Lemma 3. *For any given $\epsilon > 0$, the ODE system $D[K]$ has a unique solution over $[\pi_1 + \epsilon, 1]$, denoted $V(p;K)$, which satisfies the following properties:*

(i) $V(p;K)$ and $V'(p;K)$ are continuous in (λ, x, c, p, K);

(ii) With $K = \underline{K} := \frac{(\lambda x - c)(r+\rho_L) - c\rho_H}{r(r+\rho_L+\rho_H)}$, $V'(p;K) = \kappa$ for all $p \in [0,1]$;[17]

(iii) $V'(p;K)$ is strictly decreasing in K while $V(p;K)$ is strictly increasing in K;

(iv) $V''(p;K)$ is strictly increasing in K, and $V''(p;K) \gtreqless 0$ whenever $K \gtreqless \underline{K}$.

[17] Note that this solution extends to the entire unit interval, not just $[\pi_1 + \varepsilon, 1]$

Our construction of the pair $(\hat{p}, \hat{K})$ is divided into the three cases. Once we find the value function $V(p)$ on $[\hat{p}, 1]$, it can be extended to $[0, \hat{p}]$ by solving the ODE in (17) with the boundary condition from value-matching (i.e., $V(\hat{p}_-) = V(\hat{p}_+)$), thus completing the construction.

We begin our analysis with Case 2, as it is the most straightforward.

A.2.1 Case 2: $\hat{p} \leq \pi_1$

This case corresponds to scenarios where the infraction rate λx is sufficiently high. The following lemmas formally characterize the parameter region for this case:

Lemma 4. *Defining $\overline{\lambda}$ and $\underline{p}$ as*

$$\overline{\lambda} := \begin{cases} \frac{c(\rho_L+\rho_H-c)}{(\rho_L-c)} & \text{if } \rho_L - c > 0 \\ \infty & \text{otherwise,} \end{cases} \quad \text{and} \quad \sigma(\underline{p}) = \kappa, \tag{25}$$

the following conditions are equivalent: (a) $\lambda x \geq \overline{\lambda}$; (b) $\pi_1 \geq \hat{p}_M$; (c) $\pi_1 \geq \underline{p}$; (d) $\underline{p} \geq \hat{p}_M$. If any one of these inequalities holds with equality, then all of them do.

Lemma 5. *$\hat{p} = \hat{p}_M \leq \pi_1$ if and only if $\lambda x \geq \overline{\lambda}$.*

Given this parameterization, the next proposition establishes the existence and uniqueness of a candidate solution that satisfies the necessary conditions from Proposition 9 for this case:

Proposition 10. *For $\lambda x \geq \overline{\lambda}$, there exists a unique pair $(\hat{p}, \hat{K})$ satisfying the conditions of Proposition 9 for Case 2. This pair is given by:*

$$\hat{p} = \hat{p}_M \quad \text{and} \quad \hat{K} = \underline{K}$$

Furthermore, $\hat{p} = \hat{p}_M = \pi_1$ if $\lambda x = \overline{\lambda}$.

Proof. Let $K = \underline{K}$ and $\hat{p} = \hat{p}_M$. Then, by Lemma 3(ii), we have $V'(\hat{p}) = \kappa$. This linearity of V implies that $V(1) - V(\hat{p}) - (1 - \hat{p})V'(\hat{p}) = 0$. Plugging this into the definition of $\Delta(\hat{p})$ from (19), $\Delta(\hat{p}) = \hat{p}\lambda - \frac{c}{x} = 0$ since $\hat{p} = \hat{p}_M$, which proves the first statement. The second statement follows from Lemma 4. □

A.2.2 Case 3: $\hat{p} \in (\pi_1, \pi_0)$

This is the central and most involved case, where the infraction rate is neither too high nor too low. In this regime, the optimal policy features belief-freezing at the cutoff $\hat{p}$. The following proposition establishes the existence and uniqueness of a candidate solution for this case:

Proposition 11. *There is $\underline{\lambda} \in (c, \frac{c}{\pi_0})$ such that for $\lambda x \in (\underline{\lambda}, \overline{\lambda})$, there is a unique pair $(\hat{p}, \hat{K})$ with $\hat{K} > \underline{K}$ and $\hat{p} \in (\pi_1, \pi_0)$ that satisfies the necessary conditions of Proposition 9 for Case 3. Specifically, it solves:*

$$\Delta(\hat{p}) = 0 \text{ and } V'(\hat{p}) = \sigma(\hat{p}). \tag{26}$$

Furthermore, the cutoff $\hat{p}$ is continuously increasing in c and decreasing in λx, moving from π_0 to π_1 as λx increases from $\underline{\lambda}$ to $\overline{\lambda}$.

Proof. We first establish a couple of lemmas:

Lemma 6. *With $K = \underline{K}$, $V'(p; K) = \kappa = \sigma(\underline{p}), \forall p \in [\underline{p}, 1]$. Moreover, for $\lambda x < \overline{\lambda}$, we have $\underline{p} \in (\pi_1, \min\{\hat{p}_M, \pi_0\})$ that is decreasing in λx and increasing in c.*

Lemma 7. *Let (λ_i, x_i, c_i) for $i = 1, 2$ satisfy $\lambda_1 x_1 \leq \lambda_2 x_2 < \overline{\lambda}$ and $c_1 \geq c_2$, with at least one strict. Let $(\underline{p}_i, \underline{K}_i, V_i(\cdot; K))$ denote $\underline{p}$, $\underline{K}$, and the solution to $D[K]$ associated with (λ_i, x_i, c_i). Then for all $K \geq \underline{K}_1$ and $p \geq \underline{p}_1$, $V_1'(p; K) < V_2'(p; K)$ and $V_1(p; K) > V_2(p; K)$.*

The rest of the proof proceeds in several steps. To begin, let us plug $p = 1$ into (24) to obtain

$$rK = \lambda x - c - \rho_H V'(1; K) \text{ or } V'(1; K) = \frac{\lambda x - c - rK}{\rho_H}. \tag{27}$$

Define $\overline{K}$ to be such that $V'(1; \overline{K}) = 0$ in (27), that is,

$$\overline{K} := \frac{\lambda x - c}{r} > \frac{(\lambda x - c)(r + \rho_L) - c\rho_H}{r(r + \rho_L + \rho_H)} = \underline{K}. \tag{28}$$

Step 1: Define $p(K)$. *Fix (λ, x, c) with $\lambda x < \overline{\lambda}$. Then, there exists $K_0 \in (\underline{K}, \overline{K})$ and a continuous, strictly increasing function $p : [\underline{K}, K_0] \to [\underline{p}, \pi_0]$ such that for each $K \in [\underline{K}, K_0]$, $p(K)$ is the unique solution to $V'(p; K) = \sigma(p)$. Moreover, $p(\underline{K}) = \underline{p}$ and $p(K_0) = \pi_0$, and hence $V'(\pi_0; K_0) = \sigma(\pi_0) = 0$ and $V'(\underline{p}; \underline{K}) = \sigma(\underline{p})$.*

First, $V'(1; \overline{K}) = 0$ implies $V'(p; \overline{K}) < 0$ for all $p < 1$ since $V'(\cdot; \overline{K})$ is increasing by Lemma 3(iv). Thus, $V'(\pi_0; \overline{K}) < \sigma(\pi_0) = 0$. Observe next that we have

$$V'(\pi_0; \underline{K}) = V'(\underline{p}; \underline{K}) = \sigma(\underline{p}) > \sigma(\pi_0),$$

where the equalities hold due to Lemma 6 and the inequality due to the fact that $\pi_0 > \underline{p}$ and σ is decreasing. Since $V'(\pi_0; \underline{K}) > \sigma(\pi_0) > V'(\pi_0; \overline{K})$ and since V' is continuously and strictly decreasing in K, there exists a unique $K_0 \in (\underline{K}, \overline{K})$ such that $V'(\pi_0; K_0) = \sigma(\pi_0) = 0$, meaning $p(K_0) = \pi_0$.

For any $K \in (\underline{K}, K_0)$, we have

$$\sigma(\underline{p}) = V'(\underline{p}; \underline{K}) > V'(\underline{p}; K) \text{ and } \sigma(\pi_0) = V'(\pi_0; K_0) < V'(\pi_0; K), \tag{29}$$

where the first equality follows from Lemma 6 while the inequalities hold due to Lemma 3(iii). These inequalities, together with the monotonicity of σ and V' with respect to p, imply that for each $K \in (\underline{K}, K_0)$, there exists a unique $p(K) \in (\underline{p}, \pi_0)$ such that $V'(p(K); K) = \sigma(p(K))$. As established above, $p(\underline{K}) = \underline{p}$ and $p(K_0) = \pi_0$.

That $p(K)$ is continuously increasing in K is straightforward from the fact that V' continuously increases in p and decreases in K while σ continuously decreases in p.

Step 2: **No solution with $K \leq \underline{K}$.** $\Delta(p(K); K) < 0$ *for any* $K \leq \underline{K}$.

By Lemma 3(ii), Lemma 3(iii), and Step 1, we have for any $K \leq \underline{K}$ and $p > p(\underline{K}) = \underline{p}$,

$$V'(p; K) \geq V'(p; \underline{K}) = V'(\underline{p}; \underline{K}) = \sigma(\underline{p}) > \sigma(p),$$

which implies that $p(K) \leq \underline{p}$ since $V'(p(K); K) = \sigma(p(K))$. Thus, $p(K) \leq \underline{p} < \frac{c}{\lambda x}$. Then,

$$\begin{aligned}\Delta(p(K); K) &= p(K)\lambda - \tfrac{c}{x} + p(K)\lambda \left[V(1; K) - V(p(K); K) - (1 - p(K))V'(p(K); K)\right] \\ &< \lambda p(K) \left[V(1; K) - V(p(K); K) - (1 - p(K))V'(p(K); K)\right] \leq 0,\end{aligned}$$

where the second inequality follows from Lemma 3(iv).

Step 3: **Monotonicity of $\Delta(p(K); K)$ in K.** $\Delta(p(K); K)$ *is strictly increasing in* $K \geq \underline{K}$.

Note first that using $g(p) = h(p) - \lambda x p(1 - p)$, we can rewrite (18) as

$$rV(p) = h(p)V'(p) + x\Delta(p) \text{ for } p \geq \hat{p}. \tag{30}$$

By (30) and the definition of $p(K)$ from Step 1, we have

$$\begin{aligned}x\Delta(p(K); K) &= rV(p(K); K) - h(p(K))V'(p(K); K) \\ &= rV(p(K); K) - h(p(K))\sigma(p(K)).\end{aligned} \tag{31}$$

To prove the desired result, consider any $K_2 > K_1 \geq \underline{K}$ so that $p(K_2) > p(K_1)$. Observe that $h(p(K_1))\sigma(p(K_1)) > h(p(K_2))\sigma(p(K_2))$ since both h and σ are decreasing. Observe also that, by Lemma 3(iv), $V'(p; K_1) \geq V'(p(K_1); K_1) = \sigma(p(K_1)) \geq 0, \forall p \in [p(K_1), p(K_2)]$, which

implies that $V(p(K_2);K_1) \geq V(p(K_1);K_1)$. By these observations and (31),

$$\begin{aligned} x\Delta(p(K_1);K_1) &= rV(p(K_1);K_1) - h(p(K_1))\sigma(p(K_1)) \\ &< rV(p(K_2);K_1) - h(p(K_2))\sigma(p(K_2)) \\ &< rV(p(K_2);K_2) - h(p(K_2))\sigma(p(K_2)) = x\Delta(p(K_2);K_2), \end{aligned}$$

where the second inequality follows from Lemma 3(iii).

Step 4: **Existence of unique solution.** *For $\lambda x \in [\frac{c}{\pi_0}, \overline{\lambda})$, there exists a unique pair $(\hat{p}, \hat{K})$ with $K > \underline{K}$ and $\hat{p} \in (\pi_1, \pi_0)$ satisfying (26).*

Observe first that $p(K_0) = \pi_0 \geq \frac{c}{\lambda x}$ or $\lambda p(K_0) \geq \frac{c}{x}$, so

$$\Delta(p(K_0);K_0) \geq p(K_0)\lambda\left[V(1;K_0) - V(\pi_0;K_0) - (1-\pi_0)V'(\pi_0;K_0)\right] > 0,$$

where the inequality is due to the strict convexity of V as shown in Lemma 3(iv) with $K_0 > \underline{K}$. By this inequality and Step 2, we have $\Delta(p(\underline{K});\underline{K}) < 0 < \Delta(p(K_0);K_0)$. Given this, the continuity and strict monotonicity of $\Delta(p(K),K)$ with respect to K imply that there exists a unique $\hat{K} \in (\underline{K}, K_0)$ such that $\Delta(p(\hat{K}),\hat{K}) = 0$ and $p(\hat{K}) \in (\underline{p}, \pi_0)$. Note also that $V'(p(\hat{K})) = \sigma(p(\hat{K}))$ by definition of $p(K)$. Given Step 2, the strict monotonicity of $\Delta(p(K);K)$ for $K \geq \underline{K}$ implies that a pair $(\hat{p}, \hat{K})$ satisfying (26) is unique. That $\hat{p} > \pi_1$ follows from the fact that $\hat{p} = p(\hat{K}) > \underline{p}$ and $\underline{p} > \pi_1$ if $\lambda x < \overline{\lambda}$ by Lemma 4.

Step 5: **Comparative statics of $\hat{p}$.** *Let (λ_i, x_i, c_i) for $i = 1, 2$ satisfy $\lambda_i x_i < \overline{\lambda}_i$, where $\overline{\lambda}_i$ is $\overline{\lambda}$ with $c = c_i$. Suppose $\lambda_1 x_1 \leq \lambda_2 x_2$ and $c_1 \geq c_2$, with at least one strict inequality. If (λ_1, x_1, c_1) admits a pair $(\hat{p}_1, K_1)$ with $\hat{p}_1 \in (\pi_1, \pi_0)$ satisfying (26), then so does (λ_2, x_2, c_2), with $\hat{p}_2 < \hat{p}_1$.*[18]

Let us use the subscript i to denote functions or variables associated with (λ_i, x_i, c_i).

Consider a pair $(\hat{p}_1, \hat{K}_1)$ solving (26) under (λ_1, x_1, c_1). Observe first that by Lemma 7, $V_2'(\hat{p}_1;\hat{K}_1) > V_1'(\hat{p}_1;\hat{K}_1) = \sigma_1(\hat{p}_1) > \sigma_2(\hat{p}_1)$ (where the second inequality holds since σ is decreasing in λx and increasing in c). Also, $V_2'(\hat{p}_1;\overline{K}_2) < 0$ (recall the argument from Step 1). Thus, one can find $K \in (\hat{K}_1, \overline{K}_2)$ such that $V_2'(\hat{p}_1;K) = \sigma_2(\hat{p}_1)$, which means $\hat{p}_1 = p_2(K)$. Note that $\hat{p}_1 > \underline{p}_1 > \underline{p}_2$, which implies $K > \underline{K}_2$ since $p_2(\underline{K}_2) = \underline{p}_2$ and $p_2(\cdot)$ is increasing. Next, we show that $\Delta_2(p_2(K);K) = \Delta_2(\hat{p}_1;K) > 0$. Consider first the case in which $V_2(\hat{p}_1;K) >$

[18] Note that π_1 depends on (λ_i, x_i, c_i).

$V_1(\hat{p}_1; \hat{K}_1)$. Using (31) and the fact that $\sigma_1(\hat{p}_1) > \sigma_2(\hat{p}_1)$, we obtain

$$0 = x_1\Delta_1(\hat{p}_1; \hat{K}_1) = rV_1(\hat{p}_1; \hat{K}_1) - h(\hat{p}_1)\sigma_1(\hat{p}_1) < rV_2(\hat{p}_1; K) - h(\hat{p}_1)\sigma_2(\hat{p}_1) = x_2\Delta_2(\hat{p}_1; K).$$

Consider next the case in which $V_2(\hat{p}_1; K) \leq V_1(\hat{p}_1; \hat{K}_1)$. Using (19), we again obtain

$$\begin{aligned} 0 = x_1\Delta_1(\hat{p}_1; \hat{K}_1) &= \hat{p}_1\lambda_1 x_1 - c_1 + \hat{p}_1\lambda_1 x_1 \left[V_1(1; \hat{K}_1) - V_1(\hat{p}_1; \hat{K}_1) - (1-\hat{p}_1)V_1'(\hat{p}_1; \hat{K}_1)\right] \\ &< \hat{p}_1\lambda_2 x_2 - c_2 + \hat{p}_1\lambda_1 x_1 \left[V_2(1; K) - V_2(\hat{p}_1; K) - (1-\hat{p}_1)V_2'(\hat{p}_1; K)\right] \\ &< \hat{p}_1\lambda_2 x_2 - c_2 + \hat{p}_1\lambda_2 x_2 \left[V_2(1; K) - V_2(\hat{p}_1; K) - (1-\hat{p}_1)V_2'(\hat{p}_1; K)\right] = x_2\Delta_2(\hat{p}_1; K), \end{aligned}$$

where the first inequality holds due to the fact that $V_1(1; K_1) = K_1 < K = V_2(1; K)$ and $V_1'(\hat{p}_1; \hat{K}_1) = \sigma_1(\hat{p}_1) > \sigma_2(\hat{p}_1) = V_2'(\hat{p}_1; K)$. The second inequality holds since the expression in the square brackets is positive due to the convexity of $V_2(\cdot; K)$. Recalling $\hat{p}_1 = p_2(K)$, we have thus proven $\Delta_2(\hat{p}_1; K) = \Delta_2(p_2(K); K) > 0 > \Delta_2(p_2(\underline{K}_2); \underline{K}_2)$, which implies by STEP 3 that there exists a unique $\hat{K}_2 \in (\underline{K}_2, K)$ such that $\Delta(p_2(\hat{K}_2); \hat{K}_2) = 0$. Moreover, $\hat{p}_2 = p_2(\hat{K}_2) < p_2(K) = \hat{p}_1$ by the monotonicity of $p_2(\cdot)$.

STEP 6: No solution at $\lambda x = c$. *With $\lambda x = c$, there exists no pair $(\hat{p}, \hat{K})$ satisfying (26).*

Given STEP 2 and STEP 3, it suffices to show that with $\lambda x = c$, $\Delta(p(K_0); K_0) < 0$, since it will imply $\Delta(p(K); K) < 0, \forall K \in [\underline{K}, K_0]$. Recall that $p(K_0) = \pi_0$. Then, we have $\sigma(p(K_0)) = 0 = V'(p(K_0); K_0)$. Using this and (24), we also have $V(p(K_0); K_0) = \frac{\pi_0\lambda x(1+K_0)-c}{r+p(K_0)\lambda x}$. Plugging these into the definition of Δ and rearranging, we obtain

$$\Delta(p(K_0); K_0) = \frac{r\left(\pi_0\lambda x(1+K_0) - c\right)}{x\left(r + \pi_0\lambda x\right)} < \frac{r\left(\lambda x\left(1 + \frac{\lambda x - c}{r}\right) - c\right)}{x\left(r + \pi_0\lambda x\right)} = \frac{(\lambda x - c)(r + \lambda x)}{x\left(r + \pi_0\lambda x\right)} = 0,$$

where the first inequality holds since $\pi_0 < 1$ and $K_0 < \overline{K} = \frac{\lambda x - c}{r}$.

STEP 7: Identify $\underline{\lambda}$. *There is $\underline{\lambda} \in (c, \frac{c}{\pi_0})$ such that any $\lambda x \in (\underline{\lambda}, \overline{\lambda})$ admits a unique pair $(\hat{p}, \hat{K})$ with $\hat{p} \in (\pi_1, \pi_0)$ satisfying (26).*

The preceding steps establish that a unique solution for Case 3 exists when λx is in the range $[c/\pi_0, \overline{\lambda})$ but not when $\lambda x = c$. By continuity, there must exist a unique threshold $\underline{\lambda} \in (c, c/\pi_0)$ such that a solution only exists for $\lambda x > \underline{\lambda}$.

STEP 8: Limit behavior of $\hat{p}$ at $\lambda x = \underline{\lambda}$ or $\overline{\lambda}$. *As $\lambda x \searrow \underline{\lambda}$, $\hat{p} \nearrow \pi_0$; as $\lambda x \nearrow \overline{\lambda}$, $\hat{p} \searrow \pi_1$.*

The proof for this step is in Appendix D.5.3 of the Online Appendix. □

A.2.3 Case 1: $\hat{p} \geq \pi_0$

This final case corresponds to scenarios where the infraction rate λx is low, specifically when $\lambda x \leq \underline{\lambda}$, where $\underline{\lambda}$ was defined above.

Proposition 12. *For $\lambda x \in (c, \underline{\lambda}]$, there exists a unique pair $(\hat{p}, \hat{K})$ with $\hat{K} > \underline{K}$ and $\hat{p} \in [\pi_0, 1)$ that satisfies the necessary conditions of Proposition 9 for Case 1. Specifically, it solves*

$$V'(\hat{p}; K) = 0 \quad \textit{and} \quad \Delta(\hat{p}; K) = 0. \tag{32}$$

Furthermore, the cutoff $\hat{p}$ is continuously decreasing in λx and increasing in c.

Proof Sketch. The proof follows a logic analogous to that of Proposition 11. It involves defining a function $p(K)$ for $K \in [K_0, \overline{K})$ where $V'(p(K); K) = 0$, and then showing that the function $\Delta(p(K); K)$ is monotonic in K and crosses zero at a unique value $\hat{K}$, which defines the solution $(\hat{p}, \hat{K})$. Refer to the Online Appendix for details. □

A.3 Comparative Statics and Verification

Using the value functions we have constructed so far, we prove the comparative statics and verification results to complete the proof of Theorem 1.

Proposition 13 (Comparative Statics)**.** *$\hat{p} \leq \hat{p}_M = \frac{c}{\lambda x}$, with the strict inequality if and only if $\hat{p} > \pi_1$. Also, $\hat{p}$ decreases continuously in λx and increases continuously in c.*

Proof. Refer to the Online Appendix. □

Proposition 14 (Verification)**.** *The cutoff policy identified in Appendix A.2 satisfies (5).*

Proof. It suffices to prove $\Delta(p) \geq (\leq) 0$ for $p > (<)\hat{p}$. The proof of this result proceeds in a few steps. To begin, one can derive:[19]

$$\Delta'(p) = \begin{cases} \dfrac{\Delta(p)}{p} + \dfrac{p(1-p)\lambda(r+\rho_L+\rho_H)}{h(p)}\left[\sigma(p) - V'(p)\right] & \text{for } p < \hat{p} \qquad (33) \\ \dfrac{h(p)}{g(p)}\left(\dfrac{\Delta(p)}{p}\right) + \dfrac{p(1-p)\lambda(r+\rho_L+\rho_H)}{g(p)}\left[\sigma(p) - V'(p)\right] & \text{for } p > \hat{p}. \qquad (34) \end{cases}$$

We first prove the following claim:

Claim 1. *If $\Delta(p) \geq 0$ at $p = \max\{\hat{p}, \pi_0\}$, then $\Delta(p) \geq 0, \forall p > \max\{\hat{p}, \pi_0\}$.*

[19] For the derivation, see (SA.13) and (SA.16) in the Online Appendix.

Proof. Note that $g(p) < 0$ and $h(p) \leq 0$ for $p \geq \max\{\hat{p}, \pi_0\}$. Hence, since $V'(p) \geq 0$, (34) implies that $\Delta'(p) > 0$ whenever $\Delta(p) \geq 0$. In particular, if $\Delta(p) \geq 0$ at $p = \max\{\pi_0, \hat{p}\}$, it follows that $\Delta(p) \geq 0$ for all $p > \max\{\pi_0, \hat{p}\}$. □

Verification for Case 1 ($\hat{p} \geq \pi_0$)

- **For $p > \hat{p}$**: In this case, we have $\max\{\hat{p}, \pi_0\} = \hat{p}$, so $\Delta(p) = 0$ at $p = \max\{\hat{p}, \pi_0\}$, which implies by Claim 1 that $\Delta(p) \geq 0, \forall p > \hat{p}$.

- **For $p < \hat{p}$**: To show that $\Delta(p) \leq 0$, note that $V'(p) = 0$ for all $p \leq \hat{p}$, and that $h(\hat{p}) < 0$ and $g(\hat{p}) < 0$. Consequently, by (33), (34), and $\Delta(\hat{p}) = 0$,

$$\Delta'(\hat{p}_-) = \frac{\hat{p}(1-\hat{p})\lambda(r+\rho_L+\rho_H)}{h(\hat{p})}\sigma(\hat{p}) > 0,$$

and

$$\Delta'(\hat{p}_+) = \frac{\hat{p}(1-\hat{p})\lambda(r+\rho_L+\rho_H)}{g(\hat{p})}\sigma(\hat{p}) > 0.$$

This means that there exists $\varepsilon > 0$ such that for $p \in (\hat{p} - \varepsilon, \hat{p})$, $\Delta(p) < 0$, and for $p \in (\hat{p}, \hat{p} + \varepsilon)$, $\Delta(p) > 0$. If $\Delta(p) > 0$ for some $p < \hat{p}$, we must have $\check{p} \in [p, \hat{p}]$ such that $\Delta'(\check{p}) < 0$ and $\Delta(\check{p}) = 0$. But we have

$$\Delta'(\check{p}) = \frac{\check{p}(1-\check{p})\lambda(r+\rho_L+\rho_H)}{h(\check{p})}\sigma(\check{p}) > 0,$$

a contradiction.

Verification for Case 3 ($\hat{p} \in (\pi_1, \pi_0)$)

- **For $p > \hat{p}$**: The result will follow from Claim 1 once we prove $\Delta(p) \geq 0$ for $p \in [\hat{p}, \pi_0]$. Let us thus fix any $p \in [\hat{p}, \pi_0]$. Rewrite (34) for the part of $p \geq \hat{p}$ as

$$\Delta'(p) = \frac{h(p)}{g(p)}\left(\frac{\Delta(p)}{p}\right) - \frac{\delta(p)}{g(p)p}, \tag{35}$$

where

$$\delta(p) := p^2(1-p)\lambda(r+\rho_L+\rho_H)V'(p) - \frac{c}{x}h(p) = p^2(1-p)\lambda(r+\rho_L+\rho_H)\left[V'(p) - \sigma(p)\right].$$

Claim 2. *If* $\Delta(p) \geq 0$ *and* $\delta(p) = 0$ *for* $p \in [\hat{p}, \pi_0)$*, then* $\delta'(p) > 0$.[20]

By Proposition 11, we have $\Delta(\hat{p}) = 0$ and $V'(\hat{p}) = \sigma(\hat{p}) \geq 0$—and hence $\delta(\hat{p}) = 0$. Then, by Claim 2, $\delta'(\hat{p}_+) > 0$. Thus, using (35) and $\Delta(\hat{p}_+) = \Delta'(\hat{p}_+) = \delta(\hat{p}_+) = 0$, we obtain

$$\Delta''(\hat{p}_+) = -\frac{\delta'(\hat{p}_+)}{g(\hat{p})\hat{p}} > 0,$$

from which it follows that $\Delta(p') > 0$ for all $p' \in (\hat{p}, \hat{p} + \varepsilon]$, for some $\varepsilon > 0$.[21] We now prove the following claim:

Claim 3. *If* $\Delta(p'') < 0$ *for some* $p'' \in [\hat{p}, \pi_0]$*, then there exists* $\check{p} \in (\hat{p}, p'')$ *such that* $\Delta(\check{p}) \geq 0$*,* $\delta(\check{p}) = 0$ *and* $\delta'(\check{p}) \leq 0$.

Proof. Let $p_0 = \sup\{p' \in (\hat{p}+\varepsilon, p'') : \Delta(p) > 0, \forall p < p'\}$. Then, we must have $\Delta(p_0) \leq 0$. Given this, it cannot be the case that $\delta(p) > 0, \forall p \in [\hat{p}, p_0]$, since it would imply that whenever $\Delta(p) = 0$, we have $\Delta'(p) > 0$ by (35) and $g(p) < 0$, which in turn implies $\Delta(p_0) > 0$. Thus, we must have some $p_1 \in (\hat{p}, p_0]$ with $\delta(p_1) \leq 0$. Since $\delta(\hat{p}_+) > 0$, this implies that there must be some $\check{p} \in (\hat{p}, p_1]$ such that $\delta(\check{p}) = 0$ and $\delta'(\check{p}) \leq 0$. □

Suppose for contradiction that $\Delta(p'') < 0$ for some $p'' \in [\hat{p}, \pi_0]$. By Claim 3, one can find $\check{p} \in (\hat{p}, p'')$ that $\Delta(\check{p}) \geq 0$, $\delta(\check{p}) = 0$ and $\delta'(\check{p}) \leq 0$. However, this is a contradiction since, by Claim 2, $\Delta(\check{p}) \geq 0$ and $\delta(\check{p}) = 0$ imply $\delta'(\check{p}) > 0$.

- **For** $p < \hat{p}$: Rewriting (33) as

$$\Delta'(p) = \frac{\Delta(p)}{p} - \frac{\delta(p)}{ph(p)},$$

we prove:

Claim 4. *If* $\delta(p) = 0$ *for* $p \in (\pi_1, \hat{p}]$*, then* $\delta'(p) > 0$.[22]

Then, since $\delta(\hat{p}) = 0$, we obtain $\delta'(\hat{p}_-) > 0$. Using this, one can follow an argument analogous to that in STEP 3 to show that $\Delta(p) \leq 0, \forall p < \hat{p}$.

[20] With $p = \hat{p}$, the conclusion should be taken as $\delta'(\hat{p}_+) > 0$. The proof of this result is provided in the Online Appendix.

[21] Note that in Case 3, $V'(\hat{p}) = \sigma(\hat{p})$ is equivalent to $\Delta'(\hat{p}) = 0$, as shown in Lemma 11 in the Online Appendix.

[22] With $p = \hat{p}$, the conclusion should be taken as $\delta'(\hat{p}_-) > 0$.

Verification for Case 2 ($\hat{p} \leq \pi_1$)

- **For $p > \hat{p}$**: In this case, we have $V'(p) = \kappa$ for all $p \in [\hat{p}, 1]$, as seen from the proof of Proposition 11. Thus, $\Delta(p) = p\lambda - \frac{c}{x} > 0$ for $p > \hat{p} = \hat{p}_M = \frac{c}{\lambda x}$.
- **For $p < \hat{p}$**: In this case, $\hat{p} = \hat{p}_{\mathrm{M}} \leq \underline{p}$ by Proposition 10 and Lemma 4, so we have for any $p < \hat{p}$,

$$\sigma(p) - V'(p) > \sigma(\underline{p}) - V'(\hat{p}) = \sigma(\underline{p}) - \kappa = 0,$$

where the inequality follows from the fact that V is strictly convex while σ is decreasing.[23] Thus, by (33), we conclude that for any $p < \hat{p}$, $\Delta'(p) > 0$ whenever $\Delta(p) = 0$. Given that $\Delta(\hat{p}) = 0$, this observation implies that $\Delta(p) \leq 0$ for all $p < \hat{p}$.

This completes the verification. □

B Proof of Proposition 8

Under SP, the principal's equilibrium loss is

$$\frac{\min\{c, \pi_0\lambda\}}{r}.$$

We show that any regular equilibrium under OP yields the same loss. Given an agent response $x(\cdot)$, the principal's value satisfies

$$rV(p) = h(p)V'(p) + \max_{y \in [0,1]} y\,\Gamma(p), \tag{36}$$

where

$$\Gamma(p) := (p\lambda x(p) - c) + p\lambda x(p)\big[V(1) - V(p) - (1-p)V'(p)\big],$$

and $h(p) = \rho_L(1-p) - \rho_H p$.

First, $y(p) > \bar{y}$ cannot occur in equilibrium. If $y(p) > \bar{y}$, then the agents' best response gives $x(p) = 0$, in which case the coefficient of y in the HJB is $-c < 0$; hence $y = 0$ is strictly optimal, a contradiction.

[23] For the strict convexity of V on $[0, \hat{p}]$, we observe that

$$V(p) = \left(\frac{h(p)}{h(\hat{p})}\right)^{-\frac{r}{\rho_L + \rho_H}} V(\hat{p})$$

solves the ODE (17). It is straightforward to check that $V''(p) > 0$, using $h(\hat{p}) > 0$.

We next rule out interior monitoring on a nondegenerate interval. Suppose that $y(p) \in (0, \bar{y})$ for all p in some interval I. Then $x(p) = 1$ on I, and the principal's interior optimality condition implies

$$(p\lambda - c) + p\lambda\big[V(1) - V(p) - (1-p)V'(p)\big] = 0. \tag{37}$$

The HJB then reduces to

$$rV(p) = h(p)V'(p). \tag{38}$$

Differentiating (37) and using (37) again gives

$$V''(p) = \frac{c}{\lambda p^2(1-p)}.$$

Differentiating (38) gives

$$(r + \rho_L + \rho_H)V'(p) = h(p)V''(p),$$

and hence

$$V'(p) = \frac{c\,h(p)}{\lambda p^2(1-p)(r + \rho_L + \rho_H)}.$$

Differentiating this expression and comparing with the previous expression for $V''(p)$ yields

$$(r - \rho_L - \rho_H)p^2 + (3\rho_L - r)p - 2\rho_L = 0 \quad \text{for all } p \in I,$$

which is impossible on a nondegenerate interval. Thus, on every continuity interval, y is either 0 or $\bar{y}$.

On any such interval, V satisfies

$$rV(p) = h(p)V'(p). \tag{39}$$

Indeed, if $y = 0$, this follows directly from the HJB. If $y = \bar{y}$, then $\bar{y}$ is an interior action for the principal, so $\Gamma(p) = 0$, and the same equation follows. Therefore

$$V(p) = K|h(p)|^{-r/(\rho_L+\rho_H)}$$

on each continuity interval. Since V is continuous, the constant K must be the same across adjacent intervals on the same side of π_0. Hence there are constants K_- and K_+ such that this expression holds on $(0, \pi_0)$ and $(\pi_0, 1)$, respectively. Since $|h(p)|^{-r/(\rho_L+\rho_H)} \to \infty$ as $p \to \pi_0$ whereas V is bounded, we must have $K_- = K_+ = 0$. Thus $V \equiv 0$.

With $V \equiv 0$, the HJB coefficient reduces to

$$\Gamma(p) = p\lambda x(p) - c.$$

If $y(p) = 0$, then the agents' best response gives $x(p) = 1$, while optimality of $y = 0$ requires $p\lambda x(p) - c \leq 0$. Hence such points must satisfy $p \leq c/\lambda$. If $y(p) = \bar{y}$, then interior optimality requires

$$p\lambda x(p) - c = 0,$$

so $x(p) = c/(p\lambda)$, which is feasible only if $p \geq c/\lambda$. Therefore the only regular equilibrium policy has cutoff $\hat{p} = c/\lambda$:

$$y(p) = \begin{cases} 0 & \text{for } p < \hat{p}, \\ a & \text{for } p = \hat{p}, \\ \bar{y} & \text{for } p > \hat{p}, \end{cases} \qquad x(p) = \begin{cases} 1 & \text{for } p \leq \hat{p}, \\ \dfrac{c}{p\lambda} & \text{for } p \geq \hat{p}. \end{cases}$$

It remains to compute the loss. If $\hat{p} > \pi_0$, the belief converges to π_0, where the principal does not monitor, so the loss is $\pi_0\lambda/r$. If $\hat{p} \leq \pi_0$, the long-run belief lies in $[\max\{\hat{p}, \pi_1\}, 1]$, where the principal is indifferent and incurs flow loss c. Hence the total loss is

$$\frac{\min\{c, \pi_0\lambda\}}{r},$$

as under SP.

Finally, the same cutoff profile is the unique equilibrium under MP. If $p < \hat{p}$, then $p\lambda x(p) \leq p\lambda < c$, so the myopic principal chooses $y(p) = 0$, and agents choose $x(p) = 1$. If $p > \hat{p}$, then $y(p) > \bar{y}$ is impossible because it induces $x(p) = 0$ and hence $y = 0$ would be optimal; while $y(p) < \bar{y}$ would induce $x(p) = 1$, making $p\lambda > c$ and contradicting the optimality of $y(p) < \bar{y}$. Thus $y(p) = \bar{y}$ and $x(p) = c/(p\lambda)$. At $p = \hat{p}$, any $x(\hat{p}) < 1$ would imply $\hat{p}\lambda x(\hat{p}) < c$, so the principal would choose $y(\hat{p}) = 0$, inducing $x(\hat{p}) = 1$, a contradiction. Hence $x(\hat{p}) = 1$.

Online Appendix for: Predictive Enforcement

C Equivalence between V Formulation and L Formulation

Lemma 8. *Any policy $y(\cdot)$ solves (5) if and only if it minimizes the loss function in (1).*

Proof. Fix any admissible Markov policy $y : [0,1] \to [0,1]$. Let $L_y(p)$ denote the principal's total discounted loss under policy y when the current belief is p. By considering a short interval dt, we can write

$$L_y(p) = \big[p\lambda x(1-y(p)) + cy(p)\big]dt + p\lambda xy(p)dt(1-rdt)L_y(1) \\ + (1-rdt)(1-p\lambda xy(p)dt)L_y(p_{t+dt}) + o(dt).$$

Taking the limit as $dt \to 0$ and using the belief drift in (2), we obtain

$$rL_y(p) = p\lambda x(1-y(p)) + cy(p) + p\lambda xy(p)\big(L_y(1) - L_y(p)\big) + f(p, y(p))L_y'(p). \tag{SA.1}$$

Thus, any optimal policy that minimizes (1) must minimize the RHS of (SA.1).

Next, let $C(p)$ denote the total discounted expected amount of infractions, whether detected or undetected, when the current belief is p. By the same argument,

$$C(p) = p\lambda x\, dt + p\lambda xy(p)\, dt\,(1-rdt)C(1) + (1-rdt)\big(1 - p\lambda xy(p)\, dt\big)C(p_{t+dt}) + o(dt),$$

which yields

$$rC(p) = p\lambda x + p\lambda xy(p)\big(C(1) - C(p)\big) + f(p, y(p))C'(p). \tag{SA.2}$$

It follows immediately from the definition that

$$C(p) = L_y(p) + V_y(p). \tag{SA.3}$$

Subtracting (SA.1) from (SA.2), we obtain

$$rV_y(p) = (p\lambda x - c)y(p) + p\lambda xy(p)\big(V_y(1) - V_y(p)\big) + f(p, y(p))V_y'(p). \tag{SA.4}$$

Since $C(p)$ is independent of the policy, the relationship in (SA.3) implies that any policy y which minimizes the RHS of (SA.1) maximizes the RHS of (SA.4), meaning exactly (5).[24] □

D Omitted Proofs for Theorem 1

D.1 Proof of Lemma 3

The differential equation $D[K]$ can be rewritten as

$$V'(p) = \frac{r + p\lambda x}{g(p)} V(p) + \frac{c - p\lambda x(1+K)}{g(p)},$$

which clearly has a unique solution for $p \in [\pi_1 + \varepsilon, 1]$ that satisfies the desired continuity, proving Part (i).

Part (ii) is established by finding a linear solution to $D[K]$ constructively. To do so, let $V(p) = K + a(p-1)$ for $p \in [\hat{p}, 1]$. Then, the RHS of (24) becomes

$$\begin{aligned} & p\lambda x - c + p\lambda x a(1-p) + [\rho_L(1-p) - \rho_H p - p(1-p)\lambda x]a \\ =& - c + \rho_L a + [\lambda x - (\rho_L + \rho_H)a]p, \end{aligned}$$

which must equal the LHS of (24), $rV(p) = r(K + a(p-1)) = r(K-a) + rap$. We must thus have $\rho_L a = r(K-a)$ and $\lambda x - (\rho_L + \rho_H)a = ra$, which can be solved to yield $a = \frac{\lambda x}{r + \rho_L + \rho_H}$ and $K = \frac{(\lambda x - c)(r + \rho_L) - c\rho_H}{r(r + \rho_L + \rho_H)}$, as desired.

To prove Part (iii), let us consider any $K_2 > K_1$. For simplicity, let $V_i(\cdot)$ denote $V(\cdot; K_i)$ for $i = 1, 2$. To begin, observe that we have $V_1'(1) > V_2'(1)$ by (27). Suppose now for contradiction that there is some $p < 1$ such that $V_1'(p) \leq V_2'(p)$. One can then find some $p' < 1$ such that $V_1'(p') = V_2'(p')$ and $V_1'(p) > V_2'(p), \forall p > p'$. Note first that

$$V_1(p') = V_1(1) - \int_{p'}^{1} V_1'(p)dp < V_2(1) - \int_{p'}^{1} V_2'(p)dp = V_2(p') \tag{SA.5}$$

since $V_1(1) = K_1 < K_2 = V_2(1)$ and $V_1'(p) < V_2'(p), \forall p \in [p', 1]$. We then rewrite (24) with

[24] The function $C(p)$ does not depend on the monitoring policy. The reason is that, when choosing a monitoring policy, the principal treats the infraction rate x as fixed, while infraction opportunities arrive at rate λ only in state H. Hence, the total flow rate of infractions is $\lambda x \mathbf{1}\{\omega_t = H\}$, which does not depend on the monitoring policy. The policy affects only whether an infraction is detected and prevented from causing harm. Hence the total flow rate of infractions is simply $\lambda x\, \mathbf{1}\{\omega_t = H\}$, which is independent of y.

$p = p'$, $V = V_i$, and $K = V_i(1)$ as

$$g(p')V_i'(p') + p'\lambda x - c = rV_i(p') - p'\lambda x[V_i(1) - V_i(p')].$$

With $V_1'(p') = V_2'(p')$, we have the LHS, and thus RHS, of this equation unchanged across $i = 1, 2$. However,

$$\begin{aligned} rV_1(p') - p'\lambda x[V_1(1) - V_1(p')] &= rV_1(p') - p'\lambda x \int_{p'}^{1} V_1'(p)dp \\ &< rV_2(p') - p'\lambda x \int_{p'}^{1} V_2'(p)dp \\ &= rV_2(p') - p'\lambda x[V_2(1) - V_2(p')], \end{aligned}$$

where the inequality holds since $V_2(p') > V_1(p')$ from (SA.5) and $V_2'(p) < V_1'(p), \forall p > p'$. We have thus obtained a contradiction. That $V(p; K)$ is strictly increasing in K can be established by an analogous argument, so its proof is omitted.

For Part (iv), let us differentiate both sides of (24) to obtain

$$rV'(p; K) = \lambda x(1 + K) - \lambda x V(p; K) + (g'(p) - p\lambda x)\, V'(p; K) + g(p)V''(p; K),$$

which can be combined with (24) to remove $V(p)$ and obtain

$$V''(p; K) = \frac{[r^2 + r(\lambda x + \rho_L + \rho_H) + \lambda x \rho_L]\, V'(p; K) - \lambda x(c + r + rK)}{(r + \lambda p x)g(p)}. \tag{SA.6}$$

Since the numerator is strictly decreasing in K and the denominator is negative (recall $g(p) < 0$ for $p > \pi_1$), $V''(p; K)$ is strictly increasing in K. Combining this with Part (ii) proves the rest of Part (iv).

D.2 Proof of Lemma 4

We first prove a couple of claims:

Claim 5. *$\sigma(p)$ strictly decreases from ∞ to 0 as p increases from 0 to π_0.*

Proof. Note first that $\frac{h(p)}{p^2(1-p)}$ is strictly decreasing on $[0, \pi_0]$:

$$\frac{d}{dp}\left(\frac{h(p)}{p^2(1-p)}\right) = \frac{-2p^2(\rho_H + \rho_L) + p(\rho_H + 4\rho_L) - 2\rho_L}{(1-p)^2 p^3}$$

$$= \frac{-2(\rho_H+\rho_L)(p-\pi_0)^2 + \frac{2(\rho_L)^2}{\rho_H+\rho_L} + p\rho_H - 2\rho_L}{(1-p)^2p^3} < 0,$$

since the numerator is maximized at $p = \pi_0$ within the range $[0, \pi_0]$ and the maximized value is equal to $\frac{-\rho_L\rho_H}{\rho_H+\rho_L} < 0$. Thus, $\sigma(p)$ is strictly decreasing. Then, the result follows from observing that $\sigma(0) = \infty$ and $\sigma(\pi_0) = 0$ since $h(0) > 0 = h(\pi_0)$. □

Claim 6. *$\lambda x \geq \overline{\lambda}$ is equivalent to*

$$\lambda x(\rho_L - c) + c(c - \rho_L - \rho_H) \geq 0. \tag{SA.7}$$

Proof. Notice that if $\rho_L - c > 0$, then (SA.7) is just a rearrangement of $\lambda x \geq \overline{\lambda}$. It thus suffices to show that (SA.7) implies $\rho_L > c$. Otherwise, we would have

$$\text{LHS of (SA.7)} \leq c(\rho_L - c) + c(c - \rho_L - \rho_H) = -c\rho_H < 0,$$

where the first inequality holds since $\lambda x > c$. □

To prove the equivalence between (a) and (b), observe first that solving $g(p) = 0$ yields $\pi_1 = \frac{\lambda x+\rho_L+\rho_H-\sqrt{(\lambda x+\rho_L+\rho_H)^2-4\lambda x\rho_L}}{2\lambda x}$. Thus,

$$\begin{aligned}
\frac{c}{\lambda x} \leq \pi_1 &\Leftrightarrow \lambda x + \rho_L + \rho_H - 2c \geq \sqrt{(\lambda x + \rho_L + \rho_H)^2 - 4\lambda x\rho_L}. && \text{(SA.8)}\\
&\Leftrightarrow (\lambda x + \rho_L + \rho_H - 2c)^2 \geq (\lambda x + \rho_L + \rho_H)^2 - 4\lambda x\rho_L.\\
&\Leftrightarrow \lambda x(\rho_L - c) + c(c - \rho_L - \rho_H) \geq 0, && \text{(SA.9)}
\end{aligned}$$

if the LHS of (SA.8) is nonnegative. Given this and Claim 6, the equivalence between $\pi_1 \geq \frac{c}{\lambda x}$ and $\lambda x \geq \overline{\lambda}$ will follow if (SA.7) implies that the LHS of (SA.8) is nonnegative. To see this, recall that (SA.7) implies $\rho_L > c$, so $\lambda + \rho_L + \rho_H - 2c > \lambda + \rho_H - c > 0$ since $\lambda \geq \lambda x > c$.

To prove the equivalence between (b) and (c), observe that since $0 = g(\pi_1) = h(\pi_1) - \pi_1(1-\pi_1)\lambda x$, we have

$$\begin{aligned}
\sigma(\underline{p}) - \sigma(\pi_1) = \frac{\lambda x}{r+\rho_L+\rho_H} - \sigma(\pi_1) &= \frac{\lambda x}{r+\rho_L+\rho_H} - \frac{h(\pi_1)c}{\pi_1^2(1-\pi_1)\lambda x(r+\rho_L+\rho_H)}\\
&= \frac{c}{r+\rho_L+\rho_H}\left(\frac{\lambda x}{c} - \frac{1}{\pi_1}\right) \geq 0. && \text{(SA.10)}
\end{aligned}$$

Since σ is decreasing, we have $\pi_1 \geq \underline{p}$ if and only if $\pi_1 \geq \hat{p}_{\text{M}}$.

For the equivalence between (d) and (a), observe

$$\sigma(\hat{p}_{\text{M}}) - \sigma(\underline{p}) = \frac{\lambda x(\rho_L - c) + c(c - \rho_L - \rho_H)}{c(\lambda x - c)\,(r + \rho_L + \rho_H)}.$$

Thus, given that σ is decreasing, $\underline{p} \geq \hat{p}_{\text{M}}$ is equivalent to $\lambda x(\rho_L - c) + c(c - \rho_L - \rho_H) \geq 0$, which is equivalent to (a) due to Claim 6.

The second statement follows since all inequalities in the above derivations bind simultaneously whenever one of them does.

D.3 Proof of Proposition 9

To first derive the condition (20), let us use (17), (18), (19), and the value matching to write

$$h(\hat{p})V'(\hat{p}_+) + x\Delta(\hat{p}_+) = rV(\hat{p}_+) = rV(\hat{p}_-) = h(\hat{p})V'(\hat{p}_-),$$

which implies $\Delta(\hat{p}_+) = 0$ by the smooth pasting. Also, by value matching and smooth pasting, $\Delta(\hat{p}_+) = \Delta(\hat{p}_-)$, which yields (20).

The conditions (21) to (23) are derived by proving the next three lemmas.

Lemma 9. *If $\hat{p} \geq \pi_0$ (Case 1), then $V(p) = V'(p) = 0$ for all $p \in [0, \hat{p}]$.*

Proof. Observe first that with $\hat{p} \geq \pi_0$, (17) and $h(\pi_0) = 0$ together imply $V(\pi_0) = 0$. Suppose for contradiction that there is some $p < \pi_0$ such that $V(p) < 0$. Then, we must have some $\check{p} \in (p, \pi_0)$ such that $V(\check{p}) < 0$ and $V'(\check{p}) > 0$. Since $h(\check{p}) > 0$ and thus $h(\check{p})V'(\check{p}) > 0$, (17) implies $V(\check{p}) > 0$, a contradiction. Suppose next that there is $p < \pi_0$ such that $V(p) > 0$. Then, we must have some $\check{p} \in (p, \pi_0)$ such that $V(\check{p}) > 0$ and $V'(\check{p}) < 0$. Since $h(\check{p}) > 0$ and thus $h(\check{p})V'(\check{p}) < 0$, (17) implies $V(\check{p}) < 0$, a contradiction. The argument so far establishes $V(p) = 0$ for all $p \leq \pi_0$. An analogous argument can be used to establish the same result for the range $[\pi_0, \hat{p}]$. □

Lemma 10. *If $\hat{p} \leq \pi_1$ (Case 2), then $V'(p) = \kappa$ for all $p \in [\hat{p}, 1]$.*

Proof. Let us differentiate both sides of (18) to obtain

$$rV'(p) = g'(p)V'(p) + g(p)V''(p) - \lambda x V(p) - p\lambda x V'(p) + \lambda x(1 + V(1)).$$

We can combine this equation and (18) to remove $V(p)$ and obtain

$$V'(p) = \frac{\lambda x(c + r + rV(1)) + (r + p\lambda x)g(p)V''(p)}{r^2 + r(\lambda x + \rho_L + \rho_H) + \lambda x \rho_L}. \tag{SA.11}$$

Let us denote $a = \frac{\lambda x(c+r+rV(1))}{r^2+r(\lambda x+\rho_L+\rho_H)+\lambda x\rho_L}$. We show that $V'(p) = a, \forall p \geq \hat{p}$. Note first that this is true for $p = \pi_1$ since $g(\pi_1) = 0$ and (SA.11) imply $V'(\pi_1) = a$. To first prove the linearity of V over $(\pi_1, 1]$, suppose not, i.e., $V'(p) \neq a$ for some $p \in (\pi_1, 1]$. Consider first the case $V'(p) > a$. We must then have some $\check{p} \in (\pi_1, p)$ such that $V''(\check{p}) > 0$ and $V'(\check{p}) > a$. Since $g(\check{p}) < 0$ and thus $g(\check{p})V''(\check{p}) < 0$, the equation (SA.11) implies $V'(\check{p}) < \frac{\lambda x(c+r+rV(1))}{r^2+r(\lambda x+\rho_L+\rho_H)+\lambda x\rho_L} = a$, a contradiction. Consider next the case $V'(p) < a$. Then, we must have some $\check{p} \in (\pi_1, p)$ such that $V''(\check{p}) < 0$ and $V'(\check{p}) < a$. Since $g(\check{p}) < 0$ and thus $g(\check{p})V''(\check{p}) > 0$, the equation (SA.11) implies $V'(\check{p}) > \frac{\lambda x(c+r+rV(1))}{r^2+r(\lambda x+\rho_L+\rho_H)+\lambda x\rho_L} = a$, a contradiction. The argument so far establishes that $V'(p) = a, \forall p \in [\pi_1, 1]$. An analogous argument can be used to show that $V(p)$ is linear on $[\hat{p}, \pi_1]$ as well.

According to Lemma 3(iv), V can be linear only if $K = \underline{K}$, in which case $V'(p) = \kappa, \forall p$ according to Lemma 3(ii), as desired. □

Lemma 11. *If $\hat{p} \in (\pi_1, \pi_0)$ (Case 3), then $V'(\hat{p}) = \sigma(\hat{p})$—equivalently $\Delta'(\hat{p}) = 0$.*

Proof. Given the cutoff class, the HJB equation (5) requires $y = 0(1)$ to be optimal for $p < (>)\hat{p}$. Thus, $\frac{\partial W(p,y)}{\partial y} \leq 0$ if $p < \hat{p}$ and $\frac{\partial W(p,y)}{\partial y} \geq 0$ if $p > \hat{p}$. Since $\Delta(p) = \frac{1}{x}\frac{\partial W(p,y)}{\partial y} = 0$ at $p = \hat{p}$, this condition implies $\Delta'(\hat{p}_-) \geq 0$ and $\Delta'(\hat{p}_+) \geq 0$.

Let us differentiate both sides of (17) to obtain $rV'(p) = h'(p)V'(p) + h(p)V''(p)$ or

$$V''(p) = \frac{(r+\rho_L+\rho_H)V'(p)}{h(p)}. \tag{SA.12}$$

Substituting this into the differentiation of $\Delta(p)$ in (19), we obtain for $p < \hat{p}$,

$$\begin{aligned}
\Delta'(p) &= \frac{\Delta(p)}{p} + \frac{c}{px} - p(1-p)\lambda V''(p) \\
&= \frac{\Delta(p)}{p} + \frac{c}{px} - \frac{p(1-p)\lambda(r+\rho_L+\rho_H)}{h(p)}V'(p), \\
&= \frac{\Delta(p)}{p} + \frac{p(1-p)\lambda(r+\rho_L+\rho_H)}{h(p)}\left[\sigma(p) - V'(p)\right],
\end{aligned} \tag{SA.13}$$

Since $\Delta(\hat{p}) = 0$, this equation and $\Delta'(\hat{p}_-) \geq 0$ imply

$$\Delta'(\hat{p}_-) = \frac{\hat{p}(1-\hat{p})\lambda(r+\rho_L+\rho_H)}{h(\hat{p})}\left[\sigma(\hat{p}) - V'(\hat{p})\right] \geq 0. \tag{SA.14}$$

Let us next differentiate both sides of (18) to obtain

$$rV'(p) = \lambda x + \lambda x[V(1) - V(p)] - p\lambda x V'(p) + g'(p)V'(p) + g(p)V''(p),$$

which can be rewritten as

$$\begin{aligned} g(p)V''(p) &= (r - g'(p) + p\lambda x)V'(p) + \lambda x[V(p) - (1 + V(1))] \\ &= (r + \rho_L + \rho_H)V'(p) + (1-p)\lambda x V'(p) + \lambda x[V(p) - (1 + V(1))] \\ &= (r + \rho_L + \rho_H)V'(p) - \frac{c}{p} - \frac{\Delta(p)x}{p}. \end{aligned} \tag{SA.15}$$

Substituting this into the differentiation of $\Delta(p)$, we obtain for $p \geq \hat{p}$

$$\begin{aligned} \Delta'(p) &= \frac{\Delta(p)}{p} + \frac{c}{px} - p(1-p)\lambda V''(p) \\ &= \frac{\Delta(p)}{p} + \frac{c}{px} - p(1-p)\lambda \frac{(r + \rho_L + \rho_H)V'(p) - \frac{c}{p} - \frac{\Delta(p)x}{p}}{g(p)} \\ &= \left(1 + \frac{p(1-p)\lambda x}{g(p)}\right)\left(\frac{\Delta(p)}{p} + \frac{c}{px}\right) - \frac{p(1-p)\lambda(r + \rho_L + \rho_H)}{g(p)}V'(p) \\ &= \frac{h(p)}{g(p)}\left(\frac{\Delta(p)}{p} + \frac{c}{px}\right) - \frac{p(1-p)\lambda(r + \rho_L + \rho_H)}{g(p)}V'(p) \\ &= \frac{h(p)}{g(p)}\left(\frac{\Delta(p)}{p}\right) + \frac{p(1-p)\lambda(r + \rho_L + \rho_H)}{g(p)}\left[\sigma(p) - V'(p)\right]. \end{aligned} \tag{SA.16}$$

Since $\Delta(\hat{p}) = 0$, this equation and $\Delta'(\hat{p}_+) \geq 0$ imply

$$\Delta'(\hat{p}_+) = \frac{\hat{p}(1-\hat{p})\lambda(r + \rho_L + \rho_H)}{g(\hat{p})}\left[\sigma(\hat{p}) - V'(\hat{p})\right] \geq 0. \tag{SA.17}$$

For the interior cutoff case where $\hat{p} \in (\pi_1, \pi_0)$, the natural drift of the hidden state implies $h(\hat{p}) > 0$ (the belief drifts up without monitoring), while $g(\hat{p}) < 0$ (the belief drifts down under full monitoring).

Applying these opposing signs to our single-crossing requirements yields a binding condition. First, since $h(\hat{p}) > 0$, (SA.14) requires that $\sigma(\hat{p}) \geq V'(\hat{p})$ for $\Delta'(\hat{p}_-) \geq 0$ to hold. Second, since $g(\hat{p}) < 0$, equation (SA.17) requires that $\sigma(\hat{p}) \leq V'(\hat{p})$ for $\Delta'(\hat{p}_+) \geq 0$ to hold. These two conditions can hold simultaneously if and only if $V'(\hat{p}) = \sigma(\hat{p})$. Thus, the single-crossing property must strictly bind, equivalently yielding $\Delta'(\hat{p}) = 0$. □

D.4 Proof of Lemma 5

The *only if* direction directly follows from Lemma 4 which shows that $\hat{p}_{\mathrm{M}} \leq \pi_1$ is equivalent to $\lambda x \geq \overline{\lambda}$.

To prove the *if* direction, suppose that $\lambda x \geq \overline{\lambda}$ and hence $\pi_1 \geq \max\{\hat{p}_M, \underline{p}\}$ by Lemma 4. Suppose for contradiction that $\hat{p} > \pi_1$ and hence $\hat{p} > \max\{\hat{p}_{\text{M}}, \underline{p}\}$. We consider two cases depending on $\hat{p} \in (\pi_1, \pi_0)$ or $\hat{p} \geq \pi_0$.

Consider first the case $\hat{p} \in (\pi_1, \pi_0)$. Recall from Lemma 3(ii) that for any $\varepsilon > 0$, $V'(p; \underline{K}) = \frac{\lambda x}{r+\rho_L+\rho_H} = \sigma(\underline{p}), \forall p \in [\pi_1 + \varepsilon, 1]$. Using this, we argue that $V(1) \geq \underline{K}$. Else if $V(1) < \underline{K}$, then we would have for all $p \geq \hat{p}$

$$V'(\hat{p}) = V'(\hat{p}; V(1)) > V'(\hat{p}; \underline{K}) = \sigma(\underline{p}) > \sigma(\hat{p}),$$

where the first inequality follows from Lemma 3(iii) while the second inequality holds since σ is decreasing. This inequality contradicts the condition in (SA.14) that requires $V'(\hat{p}) \leq \sigma(\hat{p})$ since $h(\hat{p}) > 0$.

That $V(1) \geq \underline{K}$ implies V is (weakly) convex over $[\hat{p}, 1]$ due to Lemma 3(iv). Then, we have a contradiction since, by (19),

$$\Delta(\hat{p}) = \hat{p}\lambda - \frac{c}{x} + \hat{p}\lambda \left[V(1) - V(\hat{p}) - (1-\hat{p})V'(\hat{p})\right] \geq \hat{p}\lambda - \frac{c}{x} > 0,$$

where the first inequality follows from the convexity of V while the second from $\hat{p} > \hat{p}_{\text{M}}$.

Consider next the case $\hat{p} \geq \pi_0$. By Lemma 9 (together with value-matching and smooth pasting), we have $V(\hat{p}) = V'(\hat{p}) = 0$. Then, by (19),

$$\Delta(\hat{p}) = \hat{p}\lambda - \frac{c}{x} + \hat{p}\lambda V(1) \geq \hat{p}\lambda - \frac{c}{x} > 0,$$

a contradiction, which establishes $\hat{p} \leq \pi_1$.

To lastly prove $\hat{p} = \hat{p}_{\text{M}}$, observe that if $\hat{p} \leq \pi_1$, then, by Lemma 10, $V'(p) = \kappa, \forall p \in [\hat{p}, 1]$. This linearity of V over $[\hat{p}, 1]$ implies $V(1) - V(p) - V'(p)(1-p) = 0$ for all $p \geq \hat{p}$, which implies by (19) that $\Delta(p) = p\lambda - \frac{c}{x}$ and hence $\hat{p} = \frac{c}{\lambda x} = \hat{p}_{\text{M}}$ due to (20).

D.5 Proof of Proposition 10

D.5.1 Proof of Lemma 6

The first statement follows directly from Lemma 3(ii) and the definition of $\underline{p}$.

To prove the second statement, note that by Lemma 4, $\lambda x < \overline{\lambda}$ is equivalent to $\pi_1 < \underline{p} < \hat{p}_{\text{M}}$. Thus, $\underline{p} \in (\pi_1, \min\{\hat{p}_{\text{M}}, \pi_0\})$. The monotonicity of $\underline{p}$ with respect to λ follows from the fact that σ is decreasing in p and λx and increasing in c while $\kappa = \frac{\lambda x}{r+\rho_L+\rho_H}$ is increasing in λx.

D.5.2 Proof of Lemma 7

Observe first that $\underline{p}_2 < \underline{p}_1$ since $\underline{p}$ is decreasing in λx and increasing in c as shown in Lemma 6.

To simplify notations, let $V_i(\cdot)$ denote $V_i(\cdot; K)$. Let us first establish $V_1'(p) < V_2'(p), \forall p \geq \underline{p}_1$. To do so, observe first that $V_1'(1) < V_2'(1)$ due to (27), $\lambda_1 x_1 \leq \lambda_2 x_2$, and $c_1 \geq c_2$ (with at least one inequality being strict). Suppose for contradiction that there is some $p \in [\underline{p}_1, 1]$ such that $V_1'(p) \geq V_2'(p)$. Then, there must exist $\check{p} \in [p, 1]$ such that $V_1'(\check{p}) = V_2'(\check{p})$ and $V_1'(p) < V_2'(p), \forall p > \check{p}$. Given this and $V_1(1) = V_2(1) = K$,

$$V_1(\check{p}) = V_1(1) - \int_{\check{p}}^{1} V_1'(p)dp > V_2(1) - \int_{\check{p}}^{1} V_2'(p)dp = V_2(\check{p}). \tag{SA.18}$$

Using (24) and the fact that $V_1'(\check{p}) = V_2'(\check{p})$, we can write

$$\begin{aligned}
&r(V_1(\check{p}) - V_2(\check{p})) \\
&= (\lambda_2 x_2 - \lambda_1 x_1)\check{p}(1-\check{p})V_1'(\check{p}) - (c_1 - c_2) + \check{p}(\lambda_1 x_1 - \lambda_2 x_2) + \check{p}\lambda_1 x_1 \left[K - V_1(\check{p})\right] - \check{p}\lambda_2 x_2 \left[K - V_2(\check{p})\right] \\
&\leq (\lambda_2 x_2 - \lambda_1 x_1)\check{p}(1-\check{p})V_1'(\check{p}) + \check{p}(\lambda_1 x_1 - \lambda_2 x_2)\left[V_1(1) - V_1(\check{p})\right] \\
&= \check{p}(\lambda_1 x_1 - \lambda_2 x_2)\left[V_1(1) - V_1(\check{p}) - (1-\check{p})V_1'(\check{p})\right] \leq 0,
\end{aligned}$$

where the first inequality follows from the fact that $K = V_1(1)$ and $V_1(\check{p}) > V_2(\check{p})$ while the second inequality from the convexity of V_1 for $K \geq \underline{K}_1$ as established in Lemma 3(iv). This inequality contradicts the inequality in (SA.18).

Then, $V_1(p) > V_2(p), \forall p \in [\underline{p}_1, 1)$ follows from the same argument as in (SA.18), using the fact that $V_1'(p) < V_2'(p), \forall p \in [\underline{p}_1, 1]$.

D.5.3 Proof of Step 8

Conside fist the case $\lambda x \searrow \underline{\lambda}$. Let (p^-, K^-) denote the limit of $(\hat{p}, K)$ satisfying (26) as λx decreases to $\underline{\lambda}$. By the continuity of V and V' with respect to p, K, and (λ, x, c), the pair (p^-, K^-) is a unique solution at $\lambda x = \underline{\lambda}$.[25]

To show that $p^- = \pi_0$, suppose for a contradiction that $p^- < \pi_0$.[26] Then, we must have $K^- < K_0$ since $p^- = p(K^-)$ and $\pi_0 = p(K_0)$ while $p(\cdot)$ is increasing. This implies $\Delta(p(K_0); K_0) > 0 = \Delta(p(K^-); K^-)$ since $\Delta(p(\cdot); \cdot)$ is increasing. Now let us choose some (λ_1, x_1, c) such that $\lambda_1 x_1 = \underline{\lambda} - \varepsilon$ with small $\varepsilon > 0$. Let $V_1(\cdot;\cdot)$, $\Delta_1(\cdot;\cdot)$, $\sigma_1(\cdot)$, and $p_1(\cdot)$ denote the corresponding functions. By the continuity of these functions with respect to

[25] The uniqueness follows from Step 1 and 4.

[26] Note that we cannot have $p^- > \pi_0$ since p^- is the limit of $\hat{p} < \pi_0$.

(λ, x, c), we can make $\varepsilon > 0$ sufficiently small that $\Delta_1(p_1(K_0); K_0) > 0$. Let K' be such that $\sigma_1(p^-) = V_1'(p^-; K')$, i.e., $p_1(K') = p^-$. Then, we have $K' < K^-$ since $\sigma_1(p^-) > \sigma(p^-) = V'(p^-; K^-) > V_1'(p^-; K^-)$ (where the second inequality holds due to Lemma 7) and since $V_1'(p^-; \cdot)$ is decreasing. Also, using $\lambda_1 x_1 < \underline{\lambda}$ and derivations analogous to those in Step 5, we can show that $\Delta_1(p^-; K') < \Delta(p^-; K^-) = 0$. Given that $\Delta_1(p_1(K_0); K_0) > 0 > \Delta_1(p_1(K'); K')$, we can find some $K_1 \in (K', K_0)$ such that $p_1(K_1) < \pi_0$ and $\Delta_1(p_1(K_1); K_1) = 0$ while $\sigma_1(p_1(K_1)) = V_1'(p_1(K_1); K_1)$, which contradicts the definition of $\underline{\lambda}$ (since $\lambda_1 x_1 < \underline{\lambda}$).

Consider next the case $\lambda x \nearrow \overline{\lambda}$. To prove the convergence, let (p^+, K^+) denote the limit of $(\hat{p}, \hat{K})$ satisfying (26) as λx increases to $\overline{\lambda}$. Clearly, $p^+ \geq \pi_1$. Suppose $p^+ > \pi_1$ for a contradiction. Since, at $\lambda x = \overline{\lambda}$, we have $\pi_1 = \underline{p} = \frac{c}{\lambda x}$ by Lemma 4, we can choose λx sufficiently close to $\overline{\lambda}$ (or sufficiently large if $\overline{\lambda} = \infty$) that π_1 is close to $\frac{c}{\lambda x}$. Letting $(\hat{p}, K)$ denote a pair satisfying (26) for such λx, we have $\hat{p} > \pi_1$, which implies that $\hat{p} > \frac{c}{\lambda x}$ and $\hat{p} > \underline{p}$ since $\hat{p}$ is close to $p^+ > \pi_1$. Then, $\Delta(\hat{p}; K) = 0$ requires $V(1; K) - V(\hat{p}; K) - (1-\hat{p})V'(\hat{p}; K) < 0$, which implies $K < \underline{K}$ by Lemma 3(iv). Then, by Lemma 3(iii) and Claim 5, $V'(\hat{p}; K) > V'(\hat{p}; \underline{K}) = \kappa = \sigma(\underline{p}) > \sigma(\hat{p})$, a contradiction.

D.6 Proof of Proposition 12

The proof proceeds in a few steps and will employ the notations and results from the proof of Proposition 11. For instance, we continue to use the notations K_0 and $\overline{K}$.

Step 1: *No pair $(\hat{p}, \hat{K})$ with $\hat{K} < K_0$ or $\hat{K} \geq \overline{K}$ can satisfy (32).*

Recall first that $V'(\pi_0; K_0) = 0$ by the definition of K_0. We can then observe that if $\hat{K} < K_0$, then, for all $p \geq \pi_0$,

$$V'(p; \hat{K}) \geq V'(\pi_0; \hat{K}) > V'(\pi_0; K_0) = 0$$

since V' is decreasing in K and increasing in p. Observe also that if $\hat{K} \geq \overline{K}$, then $V'(p; \hat{K}) < V'(1; \hat{K}) \leq V'(1; \overline{K}) = 0, \forall p < 1$.

Step 2: *For any $K \in [K_0, \overline{K})$, there is a unique $p(K) \in [\pi_0, 1)$ that satisfies $V'(p(K); K) = 0$ and is continuously increasing in K.*

For any $K \in [K_0, \overline{K})$,

$$V'(\pi_0; K) \leq V'(\pi_0; K_0) = 0 < V'(1; K),$$

where the second inequality holds since $V'(1; K) = \frac{\lambda x - c - rK}{\rho_H} > 0$ for $K < \overline{K} = \frac{\lambda x - c}{r}$. Thus,

there exists a unique $p(K) \in (\pi_0, 1)$ that satisfies the desired property, since $V'(\cdot; K)$ is strictly increasing.

The monotonicity of $p(K)$ follows easily from the fact that V' is increasing in p and decreasing in K.

Step 3: *For each $\lambda x \in (c, \underline{\lambda}]$, there exists a unique pair $(\hat{p}, \hat{K})$ satisfying (32) with $\hat{p} \in [\pi_0, 1)$ and $\hat{K} \in [K_0, \overline{K})$.*

Consider any $\lambda_1 x_1 \in (c, \underline{\lambda}]$ and $\lambda_2 x_2 = \underline{\lambda}$. Let $V_i(p; K)$, $\sigma_i(p)$, $\Delta_i(p; K)$, and $p_i(K)$ denote the functions associated with (λ_i, x_i, c). Let K_i and $\overline{K}_i$ denote K_0 and $\overline{K}$ corresponding to (λ_i, x_i, c). Let us first observe that $K_2 > K_1$. By Lemma 7, $V_2'(\pi_0; K_1) > V_1'(\pi_0; K_1) = 0 = V_2'(\pi_0; K_2)$, which implies $K_1 < K_2$ since V_2' is decreasing in K by Lemma 3(iii).

By the definition of K_0 and $p(\cdot)$, $p_2(K_2) = \pi_0$. Then, by Step 8 in the proof of Proposition 11 and the continuity of Δ, we have $\Delta_2(p_2(K_2); K_2) = 0$. Thus, using (30), we obtain

$$\begin{aligned} 0 = \Delta_2(p_2(K_2); K_2) &= \frac{rV_2(\pi_0; K_2)}{x} \\ &\geq \frac{rV_1(\pi_0; K_2)}{x} > \frac{rV_1(\pi_0; K_1)}{x} = \Delta_1(p_1(K_1); K_1), \end{aligned} \tag{SA.19}$$

where the first inequality follows from Lemma 7 while the second inequality from Lemma 3(iv) and $K_2 > K_1$.

Observe next that $p_1(\overline{K}_1) = 1$ since $V_1'(1; \overline{K}_1) = 0$. Thus, using (19), we obtain

$$\Delta_1(p_1(\overline{K}_1); \overline{K}_1) = \Delta_1(1; \overline{K}_1) = \lambda_1 - \frac{c}{x} > 0.$$

Combining this with (SA.19) and using the strict monotonicity of $\Delta_1(p_1(\cdot); \cdot)$, we can conclude that there is a unique $\hat{K} \in [K_1, \overline{K}_1)$ such that $\Delta_1(p_1(\hat{K}); \hat{K}) = 0$ and $p_1(\hat{K}) \in [\pi_0, 1)$.

Step 4: *For the pair $(\hat{p}, \hat{K})$ satisfying (32), $\hat{p}$ is decreasing in λx and increasing in c.*

The proof of this step is analogous to that of Step 5 in the proof of Proposition 11, and hence omitted.

D.7 Proofs of Claim 2 and Claim 4

Differentiating $\delta(\cdot)$ and letting $\alpha := \rho_L + \rho_H$, we obtain

$$\delta'(p) = (2p - 3p^2)\lambda(r + \alpha)V'(p) + \frac{c\alpha}{x} + p^2(1 - p)\lambda(r + \alpha)V''(p). \tag{SA.20}$$

To prove Claim 2, we have $V''(p) = \frac{(r+\alpha)V'(p) - \frac{c}{p} - \frac{\Delta(p)x}{p}}{g(p)}$ from (SA.15) and substitute this into (SA.20) to obtain

$$\begin{aligned}
\delta'(p) =&(2p-3p^2)\lambda(r+\alpha)V'(p) + \frac{c\alpha}{x} + p^2(1-p)\lambda(r+\alpha)\frac{(r+\alpha)V'(p) - \frac{c}{p} - \frac{\Delta(p)x}{p}}{g(p)} \\
=&(2p-3p^2)\lambda(r+\alpha)V'(p) + \frac{c\alpha}{x} - p(1-p)\lambda(r+\alpha)\frac{\Delta(p)x}{g(p)} - p(1-p)\lambda(r+\alpha)\frac{c}{g(p)} \\
&+ \frac{p^2(1-p)\lambda(r+\alpha)^2V'(p)}{g(p)} \\
=&(2p-3p^2)\lambda(r+\alpha)V'(p) + \frac{c\alpha}{x} - p(1-p)\lambda(r+\alpha)\frac{\Delta(p)x}{g(p)} - p(1-p)\lambda(r+\alpha)\frac{c}{g(p)} \\
&+ (r+\alpha)\frac{p^2(1-p)\lambda(r+\alpha)V'(p) - \frac{c}{x}h(p)}{g(p)} + (r+\alpha)\frac{ch(p)}{xg(p)} \\
=&(2p-3p^2)\lambda(r+\alpha)V'(p) + \frac{c\alpha}{x} - p(1-p)\lambda(r+\alpha)\frac{\Delta(p)x}{g(p)} + (r+\alpha)\frac{\delta(p)}{g(p)} + \frac{c(r+\alpha)}{x} \\
=&2p(1-p)\lambda(r+\alpha)V'(p) - p^2\lambda(r+\alpha)V'(p) \\
&+ \frac{c\alpha}{x} - p(1-p)\lambda(r+\alpha)\frac{\Delta(p)x}{g(p)} + (r+\alpha)\frac{\delta(p)}{g(p)} + \frac{c(r+\alpha)}{x} \\
=&2p(1-p)\lambda(r+\alpha)V'(p) - \frac{p^2(1-p)\lambda(r+\alpha)V'(p) - h(p)\frac{c}{x}}{1-p} - \frac{h(p)c}{(1-p)x} \\
&+ \frac{c\alpha}{x} - p(1-p)\lambda(r+\alpha)\frac{\Delta(p)x}{g(p)} + (r+\alpha)\frac{\delta(p)}{g(p)} + \frac{c(r+\alpha)}{x} \\
=&2p(1-p)\lambda(r+\alpha)V'(p) - \frac{\delta(p)}{1-p} - \frac{h(p)c}{(1-p)x} \\
&+ \frac{c\alpha}{x} - p(1-p)\lambda(r+\alpha)\frac{\Delta(p)x}{g(p)} + (r+\alpha)\frac{\delta(p)}{g(p)} + \frac{c(r+\alpha)}{x} \\
=&2p(1-p)\lambda(r+\alpha)V'(p) - \frac{\delta(p)}{1-p} - \frac{c\alpha}{x} + \frac{\rho_H c}{(1-p)x} \\
&+ \frac{c\alpha}{x} - p(1-p)\lambda(r+\alpha)\frac{\Delta(p)x}{g(p)} + (r+\alpha)\frac{\delta(p)}{g(p)} + \frac{c(r+\alpha)}{x} \\
=&2p(1-p)\lambda(r+\alpha)V'(p) - \delta(p)\left(\frac{1}{1-p} - \frac{r+\alpha}{g(p)}\right) + \frac{\rho_H c}{(1-p)x} \\
&- p(1-p)\lambda(r+\alpha)\frac{\Delta(p)x}{g(p)} + \frac{c(r+\alpha)}{x}.
\end{aligned}$$

Since $V'(p) \geq 0$ and $g(p) < 0$, we have $\delta'(p) > 0$ if $\Delta(p) \geq 0$ and $\delta(p) = 0$.

To prove Claim 4, let us substitute (SA.12) into (SA.20) to obtain

$$\begin{aligned}
\delta'(p) =&(2p-3p^2)\lambda(r+\alpha)V'(p)+\frac{c\alpha}{x}+p^2(1-p)\lambda(r+\alpha)^2\frac{V'(p)}{h(p)}\\
=&(2p-3p^2)\lambda(r+\alpha)V'(p)+\frac{(r+\alpha)}{h(p)}\left(p^2(1-p)(r+\alpha)\lambda V'(p)-\frac{c}{x}h(p)\right)+\frac{c(r+2\alpha)}{x}\\
=&2p(1-p)\lambda(r+\alpha)V'(p)-\frac{p^2(1-p)\lambda(r+\alpha)V'(p)-\frac{c}{x}h(p)}{1-p}-\frac{c}{(1-p)x}h(p)\\
&+\frac{(r+\alpha)}{h(p)}\delta(p)+\frac{c(r+2\alpha)}{x}\\
=&2p(1-p)\lambda(r+\alpha)V'(p)-\frac{\delta(p)}{1-p}-\frac{c\alpha}{x}+\frac{\rho_H c}{(1-p)x}+\frac{(r+\alpha)}{h(p)}\delta(p)+\frac{c(r+2\alpha)}{x}\\
=&2p(1-p)\lambda(r+\alpha)V'(p)-\frac{\delta(p)}{1-p}+\frac{\rho_H c}{(1-p)x}+\frac{(r+\alpha)}{h(p)}\delta(p)+\frac{c(r+\alpha)}{x}. \qquad \text{(SA.21)}
\end{aligned}$$

Since $V'(p)\geq 0$, we have $\delta'(p)>0$ if $\delta(p)=0$.

D.8 Proof of Proposition 13

The first statement follows from Proposition 10 in the case $\hat{p}\leq\pi_1$, where $\hat{p}=\frac{c}{\lambda x}$. To deal with the case $\hat{p}>\pi_1$, rewrite $\Delta(\hat{p})=0$ as

$$\hat{p}\lambda-\tfrac{c}{x}=-\lambda\hat{p}\left[V(1)-V(\hat{p})-(1-\hat{p})V'(\hat{p})\right]. \qquad \text{(SA.22)}$$

From the proofs of Proposition 11 and Proposition 12, we know that for $p\geq\hat{p}$, $V(p)=V(p;K)$ for some $K>\underline{K}$. Thus, by Lemma 3(iv), V is strictly convex, which implies that the expression inside the square bracket in (SA.22) is strictly positive, so $\hat{p}<\frac{c}{\lambda x}$.

To prove the second statement, let us focus on establishing the comparative statics with λx since doing so with c is analogous. To do so, consider any $\lambda_2 x_2>\lambda_1 x_1$. Let π_{0i}, π_{1i} and $\hat{p}_i$ denote the values of π_0, π_1 and $\hat{p}$, respectively, under $\lambda_i x_i, i=1,2$. Note that $\pi_{01}=\pi_{02}$ and $\pi_{11}>\pi_{12}$.

The desired result, $\hat{p}_1>\hat{p}_2$, follows immediately from Proposition 10 to Proposition 12 if $\lambda_2 x_2>\lambda_1 x_1\geq\overline{\lambda}$ or if $\overline{\lambda}>\lambda_2 x_2>\lambda_1 x_1>\underline{\lambda}$ or if $\lambda_1<\lambda_2\leq\underline{\lambda}$. In the case $\lambda_1 x_1<\overline{\lambda}\leq\lambda_2 x_2$, we have $\hat{p}_1>\pi_{11}>\pi_{12}\geq\hat{p}_2$ as desired. In the case $\lambda_1 x_1\leq\underline{\lambda}<\lambda_2 x_2<\overline{\lambda}$, we have $\hat{p}_1\geq\pi_{01}=\pi_{02}>\hat{p}_2$.

E Passive Learning

In this section, we consider a situation in which the detection can occur without monitoring, which we call *passive learning*: for instance, there is a community reporting by the victim of an infraction. Let $w \in (0,1)$ denote the probability of such detection (upon an infraction being committed). Assuming that the passive learning occurs independently of the principle's monitoring, an infraction is detected with the probability $y + w - wy$, so the rate at which the principal learns $\omega = H$ is $p_t \lambda x(y + w - yw)$. We assume that passive learning, which occurs without any monitoring, has no effect on reducing harm. Thus, by choosing y units of enforcement, the principal with posterior belief p_t mitigates the harm by the same rate as before, that is, $p_t \lambda x y$. As a result, the principal's flow loss and threshold belief for the myopically optimal policy are unchanged. Therefore, passive learning affects the principal's discounted total loss only through her belief updating.

The belief updating rule (2) changes to

$$\dot{p} = \rho_L(1-p) - \rho_H p - p(1-p)\lambda x(y + w - yw) =: f_w(p, y). \tag{SA.23}$$

With passive learning, learning takes place even when the principal does not enforce. This means that the posterior belief may drift below π_0 even when $y = 0$. Indeed, one can find a unique $\pi_w \in (\pi_1, \pi_0)$ such that $f_w(\pi_w, 0) = 0$.[27]

Recall that, without passive learning, the long-term belief under MP either stays at π_0 or cycles over $[\hat{p}_{\mathrm{M}}, 1]$. In either case, the principal obtains the same payoff under MP as under SP, for information is never collected when monitoring is strictly suboptimal. This is no longer the case when there is passive learning. Suppose $\pi_w < \hat{p}_{\mathrm{M}}$. Then, learning occurs with positive probability even when the posterior falls below $\hat{p}_{\mathrm{M}}$ (where enforcement is strictly suboptimal) until it reaches π_w. In this case, MP prescribes the (myopically) correct action, i.e., no monitoring, unlike SP, so the former performs strictly better:

Proposition 4. *If $\hat{p}_M > \pi_w$, MP achieves a strictly lower long-run loss for the principal than SP. Otherwise, MP and SP achieve the same long-run loss.*

Proof. We first establish the following result:

Claim 7. $\mathbb{E}[p_t | p_0 = \pi_0] = \pi_0$.

[27] This is because $f_w(\pi_0, 0) = f(\pi_0, w) < f(\pi_0, 0) = 0 = f(\pi_1, 1) < f(\pi_1, w) = f_w(\pi_1, 0)$. Note also that the benchmark π_1 remains relevant: Since $f_w(p, 1) = f(p, 1)$, we have $f_w(\pi_1, 1) = f(\pi_1, 1) = 0$.

Proof. The posterior belief conditional on the initial belief $p_0 = \pi_0$ and any history of monitoring and detection through time t, denoted by h^t, is equal to

$$\Pr[\omega_t = H | p_0 = \pi_0, h^t] = \Pr[\omega_t = H | \omega_0 = H, h^t]\pi_0 + \Pr[\omega_t = H | \omega_0 = L, h^t](1 - \pi_0).$$

Thus,

$$\begin{aligned}
\mathbb{E}[p_t | p_0 = \pi_0] &= \mathbb{E}[\Pr[\omega_t = H | p_0 = \pi_0, h^t]] \\
&= \mathbb{E}_{h^t}\Big[\Pr[\omega_t = H | \omega_0 = H, h^t]\pi_0 + \Pr[\omega_t = H | \omega_0 = L, h^t](1 - \pi_0)\Big] \\
&= \Pr[\omega_t = H | \omega_0 = H]\pi_0 + \Pr[\omega_t = H | \omega_0 = L](1 - \pi_0) = \pi_0,
\end{aligned}$$

where the last equality follows from the fact that π_0 is the stationary belief that the Markov chain ω_t is at state H. □

The expected total loss under MP with the initial belief $p_0 = \pi_0$ is given by

$$\begin{aligned}
L_{MP}(\pi_0) =& \mathbb{E}\left[\int_0^\infty (p_t \lambda x (1 - y_{\text{MP}}(p_t)) + c y_{\text{MP}}(p_t)) e^{-rt} dt \,\middle|\, p_0 = \pi_0\right] \\
=& \mathbb{E}\left[\int_0^\infty \min\{p_t \lambda x, c\} e^{-rt} dt \,\middle|\, p_0 = \pi_0\right] \\
=& \int_0^\infty \mathbb{E}\Big[\min\{p_t \lambda x, c\} \Big| p_0 = \pi_0\Big] e^{-rt} dt \\
\leq& \int_0^\infty \min\Big\{\mathbb{E}\big[p_t \lambda x \big| p_0 = \pi_0\big], c\Big\} e^{-rt} dt \\
=& \int_0^\infty \min\big\{\pi_0 \lambda x, c\big\} e^{-rt} dt \\
=& \min\{\pi_0 \lambda x, c\}/r = L_{SP}(\pi_0),
\end{aligned}$$

where the inequality follows from Jensen's inequality and the equality right after that inequality follows from Claim 7.

The weak inequality becomes an equality if $\hat{p}_{\text{M}} \leq \pi_w$. This is because $\min\{p_t \lambda x, c\} = c$ for all $p_t \in [\pi_w, 1]$. However, if $\hat{p}_{\text{M}} > \pi_w$, then both $p_t \lambda x < c$ and $p_t \lambda x > c$ occurs with positive probability. Hence, in this case, the inequality becomes strict. □

F Analysis for Section 4

Proof of Lemma 1. To identify the stationary distribution of the posterior belief, we invoke a *balance equation*: for any subset $S \subset [0,1]$, the mass of beliefs flowing out of S must equal the mass of beliefs flowing into S. It suffices to consider two cases: $S = \{\hat{p}\}$ and $S = (p,1]$ for each $p > \hat{p}$.

For $S = (p,1]$ for $p > \hat{p}$, the balance condition requires

$$[\Phi(p+dp) - \Phi(p)](1 - p\lambda x dt) = \left(\int_{\hat{p}}^{p} \phi(s)\, s\lambda x\, ds + \Phi(\hat{p})\hat{p}\lambda x z(\hat{p}) \right) dt + o(dt), \tag{SA.24}$$

where the LHS represents the mass flowing out of the set $(p,1]$ while the RHS represents the mass flowing into it. Since the instantaneous rate of belief change is $f(p,1) < 0$ by (2), we have $dp = -f(p,1)\, dt + o(dt)$. Using this, (SA.24) simplifies to

$$-\phi(p) f(p,1) = \int_{\hat{p}}^{p} \phi(s)\, s\lambda x\, ds + \Phi(\hat{p})\hat{p}\lambda x z(\hat{p}).$$

Differentiating both sides with respect to p yields

$$\phi(p)\, p\lambda x = -\phi'(p) f(p,1) - \phi(p) f'(p,1),$$

where $f'(p,1) = \frac{\partial f(p,1)}{\partial p}$. Thus, we obtain the ODE

$$\phi'(p) f(p,1) = -\phi(p)\big(p\lambda x + f'(p,1)\big), \tag{SA.25}$$

whose solution is given in (8).

Turning to $S = \{\hat{p}\}$, the balance equation requires

$$\Phi(\hat{p})\, \hat{p}\lambda x z(\hat{p})\, dt = [\Phi(\hat{p}+dp) - \Phi(\hat{p})](1 - \hat{p}\lambda x dt) + o(dt). \tag{SA.26}$$

The LHS represents the outflow of mass, which occurs when the belief jumps from $\hat{p}$ to 1 with probability $\hat{p} z(\hat{p}) \lambda x\, dt$ during dt. The RHS represents the inflow of mass into $\hat{p}$. Substituting this and $dp = -f(\hat{p},1)\, dt + o(dt)$ into the RHS of (SA.26) yields (9).

Given (8), the mass $\Phi(\hat{p})$ must satisfy

$$1 - \Phi(\hat{p}) = \int_{\hat{p}}^{1} \phi(p)\, dp = \phi(\hat{p}) \int_{\hat{p}}^{1} \exp\Big(\int_{\hat{p}}^{p} r(s)\, ds \Big)\, dp. \tag{SA.27}$$

Using (9), we obtain $\Phi(\hat{p}) = -\frac{\phi(\hat{p})f(\hat{p},1)}{\hat{p}\lambda x z(\hat{p})}$, which can be substituted into (SA.27) to yield

$$\phi(\hat{p})\left[\int_{\hat{p}}^{1}\exp\Big(\int_{\hat{p}}^{p} r(s)\,ds\Big)\,dp - \frac{f(\hat{p},1)}{\hat{p}\lambda x z(\hat{p})}\right] = 1. \tag{SA.28}$$

Since the expression in brackets is positive, there exists $\phi(\hat{p}) > 0$ satisfying this equation. □

Lemma 12. *As $\hat{p}$ becomes lower, $\Phi(p)$ falls for all $p \geq \hat{p}$. Hence, Φ exhibits a mean-preserving spread as $\hat{p}$ becomes lower.*

Proof. For $p \geq \hat{p}$, we have $\Phi(p) = \Phi(\hat{p}) + \phi(\hat{p})\int_{\hat{p}}^{p}\exp\left(\int_{\hat{p}}^{t} r(s)ds\right)dt$ and thus

$$\begin{aligned}\frac{\partial\Phi(p)}{\partial\hat{p}} &= \Phi'(\hat{p}) + \phi'(\hat{p})\int_{\hat{p}}^{p}\exp\left(\int_{\hat{p}}^{t} r(s)ds\right)dt - \phi(\hat{p})\left[1 + r(\hat{p})\int_{\hat{p}}^{p}\exp\left(\int_{\hat{p}}^{t} r(s)ds\right)dt\right]\\ &= \Phi'(\hat{p}) - \phi(\hat{p}) + [\phi'(\hat{p}) - \phi(\hat{p})r(\hat{p})]\int_{\hat{p}}^{p}\exp\left(\int_{\hat{p}}^{t} r(s)ds\right)dt = \Phi'(\hat{p}) - \phi(\hat{p}) > 0,\end{aligned}$$

where the second equality follows from the definition of ϕ. To prove the inequality, observe that

$$\Phi(\hat{p}) = -\frac{\phi(\hat{p})f(\hat{p},1)}{\hat{p}\lambda x z(\hat{p})} = -\frac{\phi(\hat{p})f(\hat{p},1)(1-\hat{p})}{h(\hat{p})} \quad \text{or} \quad \Phi(\hat{p})h(\hat{p}) = -\phi(\hat{p})f(\hat{p},1)(1-\hat{p}),$$

where the second equality follows from the definitioin of $z(\cdot)$. By differentiating both sides, we obtain

$$\begin{aligned}\Phi'(\hat{p})h(\hat{p}) + \Phi(\hat{p})h'(\hat{p}) &= -\phi'(\hat{p})f(\hat{p},1)(1-\hat{p}) - \phi(\hat{p})(f'(\hat{p},1)(1-\hat{p}) - f(\hat{p},1))\\ &= \phi(\hat{p})(\lambda x\hat{p} + f'(\hat{p},1))(1-\hat{p}) - \phi(\hat{p})[f'(\hat{p},1)(1-\hat{p}) - f(\hat{p},1)]\\ &= \phi(\hat{p})(\lambda x\hat{p}(1-\hat{p}) + f(\hat{p},1)) = \phi(\hat{p})h(\hat{p}),\end{aligned}$$

where the second equality follows from (SA.25). Rearranging the resulting equation yields

$$\Phi'(\hat{p}) - \phi(\hat{p}) = -\frac{\Phi(\hat{p})h'(\hat{p})}{h(\hat{p})} = \frac{\Phi(\hat{p})(\rho_L + \rho_H)}{(1-\hat{p})\rho_L - \hat{p}\rho_H} > 0,$$

where the inequality holds since $\hat{p} < \pi_0 = \frac{\rho_L}{\rho_L+\rho_H}$. □

Proof of Lemma 2. First, the long-run distribution of posterior belief conditional on $\omega = H$

can be obtained as

$$\Psi(p) := \Pr\{p' \leq p | \omega = H\} = \frac{\Pr\{p' \leq p \text{ and } \omega = H\}}{\Pr\{\omega = H\}} = \begin{cases} 0 & \text{if } p < \hat{p} \\ \frac{\Phi(\hat{p})\hat{p} + \int_{\hat{p}}^{p} \phi(s) s ds}{\pi_0} & \text{if } p \geq \hat{p}. \end{cases}$$

The expected enforcement conditional on $\omega = H$ is thus

$$\mathbb{E}_{p\sim\Psi}[y_{\hat{p}}(p)|\omega = H, x] = \frac{\Phi(\hat{p})\hat{p}}{\pi_0} z(\hat{p}, x) + \left(1 - \frac{\Phi(\hat{p})\hat{p}}{\pi_0}\right) = 1 - \frac{1}{\pi_0}\Phi(\hat{p})\hat{p}(1 - z(\hat{p}, x)). \quad \text{(SA.29)}$$

We first show that this expression is decreasing in $\hat{p}$. To do so, it suffices to prove that $\Phi(\hat{p})$ is increasing since $\hat{p}(1 - z(\hat{p}, x))$ is increasing.

To this end, let

$$\eta(\hat{p}) := \int_{\hat{p}}^{1} \exp\left(\int_{\hat{p}}^{p} r(s) ds\right) dp.$$

Using this and (SA.28), we obtain

$$\phi(\hat{p}) = \frac{1}{\eta(\hat{p}) - \frac{f(\hat{p},1)}{\hat{p}\lambda x z(\hat{p},x)}},$$

which can be plugged into (SA.27) to yield

$$1 - \Phi(\hat{p}) = \frac{\eta(\hat{p})}{\eta(\hat{p}) - \frac{f(\hat{p},1)}{\hat{p}\lambda x z(\hat{p},x)}} = \frac{1}{1 + \left(\frac{1}{\hat{p}\lambda x z(\hat{p},x)}\right)\left(\frac{-f(\hat{p},1)}{\eta(\hat{p})}\right)},$$

which is continuous in $\hat{p} \in (\pi_1, \pi_0)$.

Observe first that $0 = f(\hat{p}, z(\hat{p}, x)) = h(\hat{p}) - \lambda x \hat{p}(1 - \hat{p}) z(\hat{p}, x)$ and thus

$$\hat{p}\lambda x z(\hat{p}, x) = \frac{h(\hat{p})}{1 - \hat{p}} = \rho_L - \frac{\hat{p}}{1 - \hat{p}}\rho_H,$$

which is decreasing in $\hat{p}$. So, it suffices to show that $\left(\frac{-f(\hat{p},1)}{\eta(\hat{p})}\right)$ is increasing in $\hat{p}$. For this, observe

$$\begin{aligned} \frac{d}{d\hat{p}}\left(\frac{-f(\hat{p},1)}{\eta(\hat{p})}\right) &= \frac{-f'(\hat{p},1)\eta(\hat{p}) + f(\hat{p},1)\eta'(\hat{p})}{\eta(\hat{p})^2} \\ &= \frac{-f'(\hat{p},1)\eta(\hat{p}) + f(\hat{p},1)(-r(\hat{p})\eta(\hat{p}) - 1)}{\eta(\hat{p})^2} \\ &= \frac{\hat{p}\lambda x \eta(\hat{p}) - f(\hat{p},1)}{\eta(\hat{p})^2} > 0, \end{aligned}$$

where the second equality follows from the fact that $\eta'(\hat{p}) = -r(\hat{p})\eta(\hat{p}) - 1$ (by the definition of η), the third equality from the definition of $r(\hat{p})$, and the inequality from $f(\hat{p}, 1) < 0$.

As $\hat{p} \to \pi_1$, we have $z(\hat{p}, x) \to 1$. Then, by (9), we must have $\Phi(\hat{p}) \to 0$ since $f(\hat{p}, 1) \to 0$, which, by (SA.29), implies $\mathbb{E}_{p\sim\Psi}[y_{\hat{p}}(p)|\omega = H] \to 1$ as $\hat{p} \to \pi_1$.

As $\hat{p} \to \pi_0$, we have $z(\hat{p}, x) \to 0$ by definition of $z(\hat{p}, x)$. Then, by (9), we have $\phi(\hat{p}) \to 0$, which implies $\Phi(\hat{p}) \to 1$ by (SA.27). Thus, by (SA.29), $\mathbb{E}_{p\sim\Psi}[y_{\hat{p}}(p)|\omega = H] \to 0$ as $\hat{p} \to \pi_0$.

Thus, there exists $\bar{p}(\lambda x) \in (\pi_1, \pi_0)$ that satisfies (11). The continuity of $\bar{p}(\lambda x)$ in λx follows from the continuity of $z(\cdot)$ and $\Phi(\cdot)$ in x.

Given that $\mathbb{E}_{p\sim\Psi}[y_{\hat{p}}(p)|\omega = H, x]$ is decreasing from 1 to 0 as $\hat{p}$ increases from π_1 to π_0, it is clear that $\bar{p}(\lambda x)$ is decreasing from π_0 to π_1 as $\bar{y}$ increases from 0 to 1. □