



# Should I Stalk or Should I Go? An Auditing Exploration/Exploitation Dilemma

Reda Aboutajdine, Pierre Picard

## ► To cite this version:

Reda Aboutajdine, Pierre Picard. Should I Stalk or Should I Go? An Auditing Exploration/Exploitation Dilemma. 2019. <hal-02373199>

**HAL Id: hal-02373199**

**<https://hal.science/hal-02373199v1>**

Preprint submitted on 20 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Should I Stalk or Should I Go?

## An Auditing Exploration/Exploitation Dilemma

Reda Aboutajdine <sup>\*†1</sup> and Pierre Picard <sup>‡1</sup>

<sup>1</sup>*CREST - École Polytechnique*<sup>§</sup>

November 20, 2019

### Abstract

We consider a fraud inspection problem where service providers are central to the fraud generating process, either as the main protagonists or as colluding third parties. Because interactions are repeated between the auditor (insurer, tax collector, environmental regulation agency, etc.) and auditees (doctors, tax preparers, waste management subcontractors, etc.), auditing behaves as a learning mechanism to separate the wheat (honest agents) from the chaff (defrauders). We analyze a Bayesian inspector's dynamic auditing problem in the face of fraud, and characterize its optimal strategy as a strategic exploration/one-armed bandit one. The insurer faces the well-known reinforcement learning exploration/exploitation trade-off between gathering information for higher future profits (exploration) and prioritizing immediate profits (exploitation). We then derive optimal auditing strategies with multiple auditees and capacity constraints as the solution to a k-armed bandit problem. We finally investigate the extents to which learning occurs under optimality in terms of how much information is obtained and how quickly it is obtained.

**Keywords:** Fraud, dynamic optimal auditing, information acquisition, armed bandit

**JEL:** D82, D83, G22, L21

---

<sup>\*</sup>reda.aboutajdine@polytechnique.edu

<sup>†</sup>Corresponding author: Département d'Economie, École Polytechnique, Route de Saclay, Palaiseau, FRANCE

<sup>‡</sup>pierre.picard@polytechnique.edu

<sup>§</sup>*Acknowledgements:* We are grateful to conference participants at APRIA, ARIA and ESEM annual meetings, as well as seminar participants at CREST for helpful discussions and comments.

# 1 Introduction

When an auditor can monitor an auditee's action at some cost, the auditing decision comes down to balancing between investigation-generated gains and losses. These gains and losses usually include the administrative costs of auditing, the proceeds from recovering undue payments in case of a fraud-revealing audit, and the profits drawn from a deterrence effect when defrauders are threatened to be spotted. But another non-immediate (and less scrutinized) motivation for monitoring may arise: information acquisition. When auditees are heterogeneous and the auditor's optimal action is not type-constant, the latter could benefit from learning more about the auditee. For example, in insurance, policyholders may differ in their propensity to defraud: it would be optimal for the insurer to audit "low morality" types more since they defraud frequently enough to amortize auditing costs, with significant deterrence effects. On the contrary, "high morality" types usually submit invalid claims by mistake, rendering them impervious to incentives, and unfrequently enough for net proceeds of auditing to be negative. While this morality is not directly observable, the repeated fraud patterns may convey information about the true type, provided an audit is conducted. This situation is obviously not restricted to insurance fraud and is relevant to multiple settings where fraud and auditing can take place. Tax evasion, environmental compliance and subcontractor fraud all fall in its scope and share a common feature, that is the repeated character (annually, monthly or even daily) of auditor-auditee interactions.<sup>1</sup> But an audit is costly, and gains from uncertainty reduction may not be profitable. When are they then?

To answer this question, we consider a dynamic programming problem where a decision maker (DM), the auditor, faces a non-strategic auditee who takes an action that may be invalid. We willingly restrict the occurrence of fraud to a simple non-strategic stochastic process in order to single out the information acquisition effect and exclude the deterrence one. The DM can choose to inspect the auditee at each period and reveal whether there was fraud. As a Bayesian, he acquires information by updating his belief

---

<sup>1</sup>The frequency of interactions can also be explained by the existence of intermediaries who provide services to several agents: e.g., car repairers or health service providers in insurance, tax preparers for tax evasion

about the auditee’s type, conditionally on the outcome of the auditing decision. Our goal is to understand if this learning opportunity alters the DM’s myopic optimal auditing policy and how. In particular, we are interested in the role of the time horizon (finite or infinite), the time preference and the efficiency of the information acquisition mechanism in reducing uncertainty.

Our main result states that this learning opportunity induces the DM to extend the inspection target set to some auditees whose immediate auditing proceeds are negative. In other words, the DM faces a trade-off between not auditing to avoid losing money at the expense of not learning anything, and auditing with a negative expected immediate proceed while acquiring knowledge for more accurate future audits. The DM’s intertemporal strategy takes the shape of a cut-off in the belief space, whereby he audits all agents above the threshold and ignores the ones below. With a finite-horizon, this cut-off is increasing in time, implying that more exploration takes place early on in the relationship: since the information acquired is about a time-invariant type, any information obtained at a given period is useful for all remaining periods, and, the further the horizon, the easier it is to amortize the one-time information acquisition cost. At the last period, the threshold is the one-period myopic one and excludes auditees with negative auditing proceeds. With an infinite horizon, the cut-off is unsurprisingly constant, lower than the myopic/no-learning threshold, and corresponds to the limit of the finite-horizon first period belief threshold when the total number of periods tends to infinity. As for the time preference, a more patient DM will be willing to explore more to enjoy the future information-generated additional profits. More importantly, when the DM is indifferent between the present and the future, i.e., when the discount rate tends to one, total exploration takes place in the infinite horizon setting: all beliefs are audited, as it takes a finite time and cost to reach a given level of belief precision, and an undiscounted infinite number of periods are left to take optimal actions and amortize the initial finite cost.

The intuitive concept underlying this set of results is the classical Exploration versus Exploitation dilemma: the DM must choose between playing his best option according to his current knowledge at the risk of being mistaken, or diversify his actions early on to

refine his beliefs and take more accurate actions later on. This dilemma arises particularly in Reinforcement Learning problems, in particular Bandit problems. We establish that our dynamic programming problem is nothing but a Strategic Exploration/One-Armed Bandit one, i.e. a bandit with a risky arm (auditing) and a safe arm (not auditing). The value function is then analogous to the well-known Gittins index, which allows us to extend our main result to the case of multiple auditees and capacity constraints. Another interpretation of our problem, in line with the One-Armed Bandit approach, is that of an optimal stopping problem: the cut-off threshold defines a stopping time as the moment when the belief process crosses the threshold. In our fraud inspection problem, auditing has the peculiarity of being both the learning action and the profit generating action. Therefore, the DM's cut-off decision comes down to a trade-off between stopping early enough in case of a honest type (low stopping time/large threshold), and not excluding a dishonest type in case of mistaken initial beliefs (large stopping time/low threshold). In contrast to our approach, the classical setting is one where the learning action (e.g., attention allocation, price revelation) is chosen first, then the profit generating action (buying, selling, etc.) is taken, and the DM's decision is the stopping time, which consequently never goes to infinity.

A second set of results elaborates on the specificity of the information acquisition process and its implications in terms of optimal auditing. We start by answering the question of when (in)complete/(in)adequate learning occurs, as defined by [Easley & Kiefer \(1988\)](#) and [Aghion et al. \(1991\)](#), and show that the answer lies in the distribution of optimal stopping times. Then, we focus on the efficiency of information gathering, i.e., the speed of convergence of the belief process to the true state. To do so, we first exploit the (sub/super)-martingale feature of the Bayesian belief process and focus on the Doob-Meyer predictable component.

These features relate to several strands of the academic literature. Our approach relies on the seminal work of [Townsend \(1979\)](#) and [Gale & Hellwig \(1985\)](#) on Costly State Verification where a principal can costly reveal some hidden information before

taking the action. We extend the literature on optimal auditing<sup>2</sup> by characterizing the scheme of optimal strategies under dynamic information acquisition in a multi-period setting.<sup>3</sup> One important aspect through which our conclusions differ from previous ones is that our cutoff is in a belief space rather than on monetary amounts associated to fraud. Our main result bears a striking resemblance with that of [Dionne et al. \(2008\)](#) as the auditor is induced to monitor auditees with negative expected payoff, although for different reasons. For us, this is motivated by the perspective of future profits due to sequential learning, while the motivation stems from a static deterrence effect in their case. This dynamic aspect with information gathering links our work to the literature on sequential experiments and optimal stopping, following the seminal papers by [Wald \(1945\)](#) and [Arrow et al. \(1949\)](#). These optimal stopping problems are intrinsically related to bandit ones<sup>4</sup> and our model is easily reinterpreted as a One-Armed Bandit problem, bringing our work close to the literature on Bandits in economics ([Bergemann & Välimäki \(2008\)](#)) and on Reinforcement Learning ([Sutton & Barto \(1998\)](#)). More generally, it relates to the literature on dynamic information acquisition by a decision maker, as in [Easley & Kiefer \(1988\)](#) and [Aghion et al. \(1991\)](#).

The rest of the paper is organized as follows: Section 2 introduces the model. Section 3 characterizes the inter-temporal optimal auditing strategy and its interpretations in terms of bandits and optimal stopping. Section 4 focuses on the extents to which learning by auditing happens and its informativeness as a function of the problem's parameters. Section 5 concludes. All proofs are in the Appendix.

## 2 The Model

### 2.1 Setting

We consider a world inhabited by two protagonists: an auditor and an auditee. The fundamental aspect of their interaction in this context is that it is repeated in time.

---

<sup>2</sup>see [Picard \(2013\)](#) for a comprehensive review.

<sup>3</sup>This work extends our previous analysis of the two-period case, see [Aboutajdine & Picard \(2018\)](#).

<sup>4</sup>see [Ferguson \(2004\)](#).

Practically, it relates to the fact that many interactions occur between an auditing party and a service provider (SP). The SP can be the direct entity contractually related to the auditor, such as subcontractors who are major suppliers of a client firm or public institution. The SP can instead act as a third party to perform a commercial act and/or certify actions, such as health service providers (dentists, opticians, etc.) or tax preparers. Third party SPs handle therefore the cases of several policyholders of the same insurer or of several taxpayers, thus the frequent interaction with the auditor.<sup>5</sup> Hereafter, we restrict the auditee to be an SP, and use, for the sake of simplicity and without loss of generality, the terminology of insurance fraud. The auditor becomes the insurer, the auditee is an SP, and the SP's action is to submit a claim.

The SP channels one and only one claim at each period with value 1, each claim being either valid or invalid. The invalidity of a claim may result from honest mistakes or actual ill-disposed voluntary intent to defraud. The probability of submitting invalid claims depends on the SP's intrinsic type, with the SP being either honest (type  $H$ , transmits invalid claims only involuntarily) or dishonest (type  $D$ , transmits invalid claims both voluntarily and involuntarily). Because of this, a  $D$  type is more prone to submitting invalid claims than an  $H$  type, and we consider that he does so with probability  $p_D$ , while this probability is  $p_H$  for an  $H$  type, with  $p_H < p_D$ . Hence, the SP is a non-strategic<sup>6</sup> entity and his actions are represented by an exogenous Bernoulli process of parameter  $p_i$ .

Types are time-invariant private information, unobservable to the insurer. But the latter, based on his experience and his observations, has a subjective belief  $\pi$  that the SP is of type  $D$ . In other words, the insurer attaches probability  $\pi$  to the fact that the SP is of type  $D$ . An SP to whom the auditor assigns a prior  $\pi$  will therefore, from the insurer's point of view, submit an invalid claim with probability  $\bar{p}(\pi) = (1 - \pi)p_H + \pi p_D$ . The insurer may audit a claim and reveal its true status, valid or invalid: when invalid, the auditor gets back the illegitimate amount of 1. This decision is represented by the

---

<sup>5</sup>We know for a fact that anti-fraud efforts in health insurance focus on service providers, as their collaboration is a necessary condition for fraud to occur, and they may even defraud without the insureds knowledge. The case of tax preparers is less obvious, but [Boning et al. \(2018\)](#) suggest that they can play an important role in anti-fraud efforts.

<sup>6</sup>Reminder: this is important to isolate the information effect from the deterrence effect.

choice of an auditing strategy  $x \in [0, 1]$ , which is the probability with which the auditor will investigate the claim submitted by the SP. Since the only information available to the auditor is the prior, the auditing strategy is a function of  $\pi$ .

It costs  $c \in (p_H, p_D)$  to perform an audit, inducing a net proceed of  $1 - c$  when a claim is found invalid, and  $-c$  when it is found valid. In expectation, the net proceed of auditing an SP with prior  $\pi$  is  $\bar{p}(\pi) - c$ . In particular, if the type was known, auditing a dishonest SP would yield an expected net proceed of  $p_D - c > 0$  while auditing an honest one would yield  $p_H - c < 0$ . This difference in the profitability of auditing between types of SPs is what may motivate the insurer to acquire information.

Finally, interactions take place over an arbitrary number of periods  $T \geq 1$ .<sup>7</sup> Each period is indexed by  $t \in \{1, 2, \dots, T\}$  and the initial period  $t = 1$  is the beginning of the insurer-SP relationship. The period 1 initial prior may be considered as initialized at some arbitrary value (e.g. 0.5 if no relevant information is available) or based on some other observable characteristics of the SP. The total number of periods over which the interactions take place may be interpreted as the “lifespan” of an SP. For example, if an insurer checks the channeled claims every month and a pharmacist works for 30 years (i.e., 360 months), then  $T = 360$ . Whenever necessary, the time-dependent variables of interest will be indexed by both the period at which they are considered and the total number of periods, i.e., by  $(t, T) \in \{1, \dots, T\} \times \mathbb{N}^*$ . Otherwise, we will restrain the indexation to  $t \in \{1, \dots, T\}$ .

## 2.2 Information Acquisition

Period  $t$  audits also allow the insurer to update his beliefs at the beginning of period  $t + 1$ . Depending on whether an audit has been performed and, if so, whether the claim was valid or invalid (*Val* and *Inv*, respectively), posterior beliefs  $\tilde{\pi}_{t+1}$  are deduced from

---

<sup>7</sup>The model is now a generalization of the two-period model considered by [Aboutajdine & Picard \(2018\)](#).



initial beliefs  $\pi_t$  through Bayes' Law:

$$\tilde{\pi}_{t+1} = \begin{cases} \mathbb{P}(D|audit, Inv) = \varphi_{AI}(\pi_t) = \frac{p_D \pi_t}{\bar{p}(\pi_t)}, \\ \mathbb{P}(D|audit, Val) = \varphi_{AV}(\pi_t) = \frac{(1-p_D)\pi_t}{1-\bar{p}(\pi_t)}, \\ \mathbb{P}(D|no\ audit) = \varphi_{NA}(\pi_t) = \pi_t, \end{cases} \quad (1)$$

with

$$\begin{aligned} \varphi_{AV}(\pi_t) &< \pi_t < \varphi_{AI}(\pi_t), \\ \varphi'_{AI} &> 0, \quad \varphi''_{AI} < 0, \\ \varphi'_{AV} &> 0, \quad \varphi''_{AV} > 0. \end{aligned} \quad (2)$$

The updating mechanism is such that an invalid claim increases the belief that an SP is

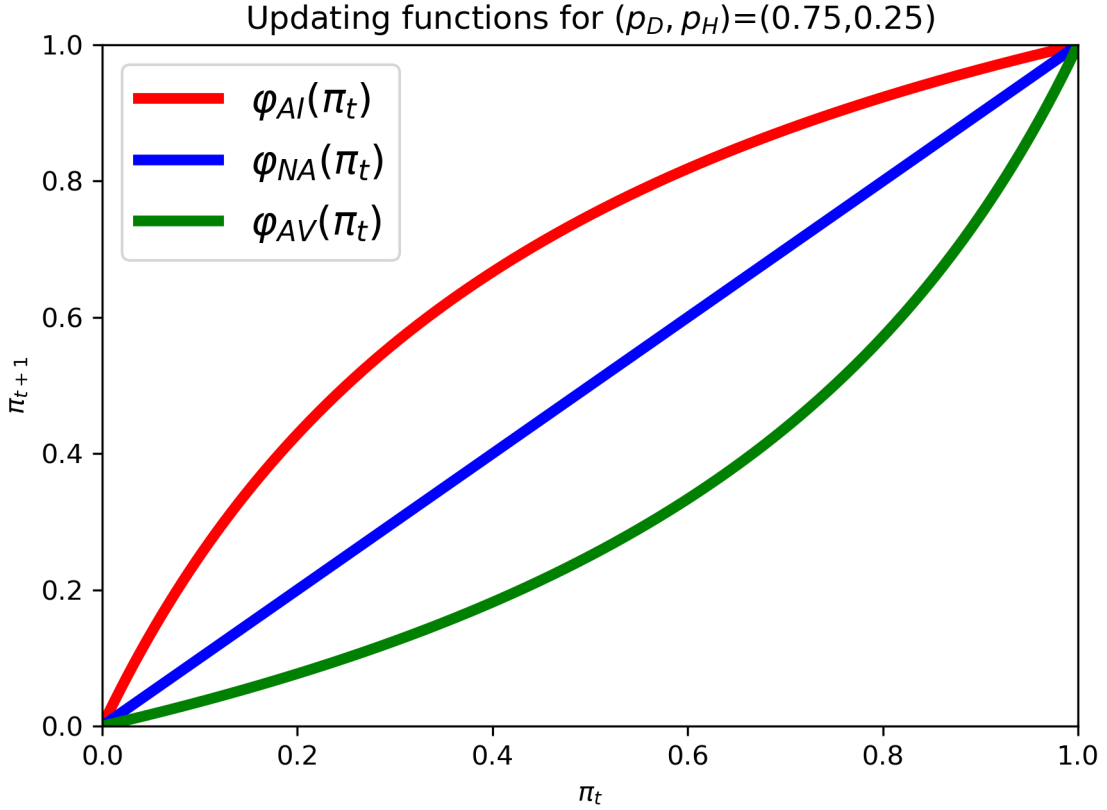


Figure 1: Updating Functions

dishonest ( $\pi_{t+1} = \varphi_{AI}(\pi_t) > \pi_t$ ) while a valid claim decreases it ( $\pi_{t+1} = \varphi_{AV}(\pi_t) < \pi_t$ ). Either way, both cases imply that the insurer acquires information about the auditee's true type. On the contrary, not auditing leaves the beliefs unchanged since no relevant

information is obtained.

As the setting is dynamic, the evolution of beliefs in times constitutes a stochastic belief process  $\langle \pi_t \rangle$  with a natural filtration  $\mathcal{F}_t$ , and, because it is governed by Bayes Law, the process behaves as a martingale.

**Lemma 1.** *The belief process  $\langle \pi_t \rangle$  is a martingale under the auditor's probability measure  $\mathbb{P}$*

$$\mathbb{E}_{\mathbb{P}}[\tilde{\pi}_{t+1}|\mathcal{F}_t] = \pi_t.$$

*In addition, with some auditing (i.e.  $x > 0$ ), under state of the world  $H$  (respectively  $D$ ) and the associated probability measures  $\mathbb{P}_H$  (resp.  $\mathbb{P}_D$ ), the belief process  $\langle \pi_t \rangle$  is a supermartingale (resp. a submartingale). With an infinite horizon, the process converges to the true belief about the state of the world*

$$i - \mathbb{E}_H[\tilde{\pi}_{t+1}|\mathcal{F}_t] \leq \pi_t \text{ and } \pi_t \xrightarrow{\mathbb{P}_H} 0.$$

$$ii - \mathbb{E}_D[\tilde{\pi}_{t+1}|\mathcal{F}_t] \geq \pi_t \text{ and } \pi_t \xrightarrow{\mathbb{P}_D} 1.$$

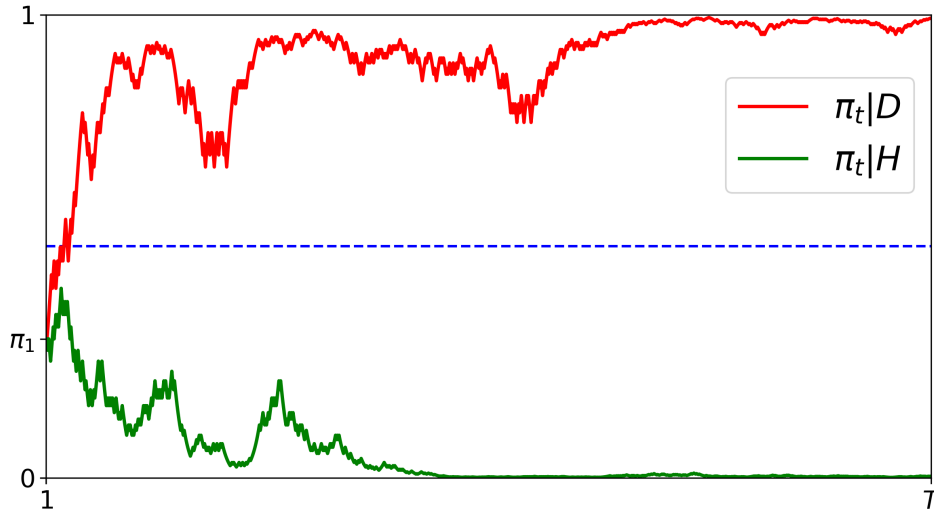


Figure 2: Simulated convergence of priors to the true beliefs

**Corollary 1.** *The posterior  $\tilde{\pi}_{t+1}$  is a mean-preserving spread of the prior  $\pi_t$ .*

Figure 2 illustrates the convergence of beliefs for each type, with the same initial belief  $\pi_1$ . This convergence of the Bayesian beliefs to the truth under systematic auditing means that auditing acts as a separating tool to isolate dishonest SPs from honest ones. Figure 3 simulates over  $T = 1000$  periods the trajectories of beliefs for a population of 2000 SPs, half honest half dishonest. At period 1, the population is uniformly distributed across priors, and a Bernoulli process is drawn for each SP's fraud behavior, resulting in a corresponding belief process realization under systematic auditing (i.e.,  $x = 1$ ). The progressive separation of types, as a convergence of beliefs to both ends of the spectrum, is such that, after enough time, dishonest SPs end up on the right, while honest ones are on the left.

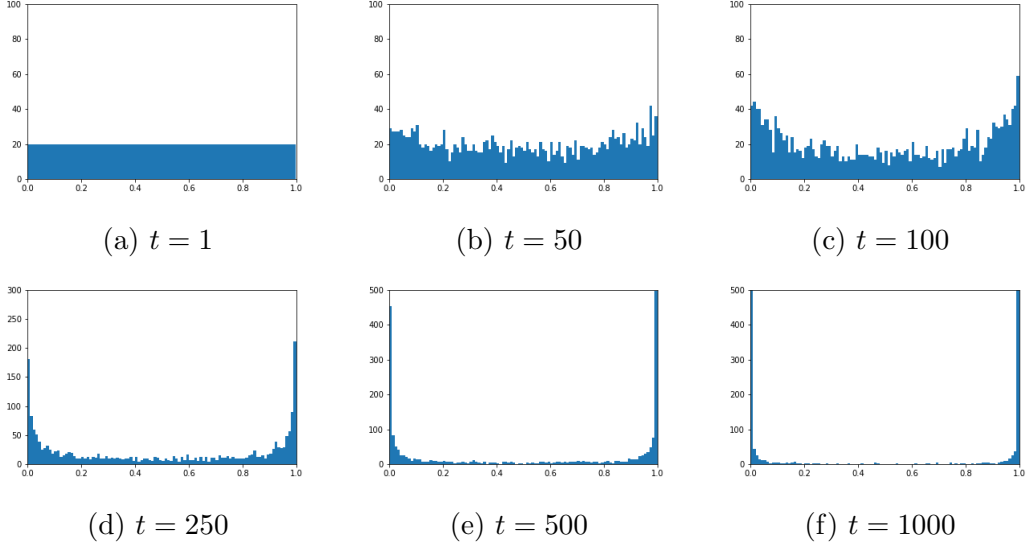


Figure 3: Distribution of priors after repeated audits ( $T = 1000$ ).

In the following, we go back to a situation where the insurer faces one single SP.

## 2.3 A Dynamic Programming Problem

### 2.3.1 The one-period/myopic problem

When  $T = 1$ , the problem is myopic, the auditor has no future profits to take into account, and information acquisition is not relevant. The auditor has objective function

$$\Omega_T(\pi, x) = u(\pi, x) = (\bar{p}(\pi) - c)x,$$

and solves for the value function

$$V_T(\pi) = \max_{x(\pi) \in [0,1]} \Omega_T(\pi, x).$$

Because of the linearity of the objective function in  $x$ , optimal auditing has a bang-bang solution characterized by a threshold  $\pi^+$  above which auditing proceeds are positive. In other words, the auditors targets only individually profitable claims. The value function is consequently piece-wise linear in  $\pi$ .

**Lemma 2.** *For  $T = 1$ , the optimal auditing strategy is bang-bang*

$$x_T^*(\pi) \begin{cases} = 0 & \text{if } \pi < \pi^+, \\ \in (0, 1) & \text{if } \pi = \pi^+, \\ = 1 & \text{if } \pi > \pi^+. \end{cases} \quad \text{where } \pi^+ = \frac{c-p_H}{p_D-p_H} \in (0, 1),$$

Figure 4 illustrates the myopic problem's objective function, optimal policy and value function.

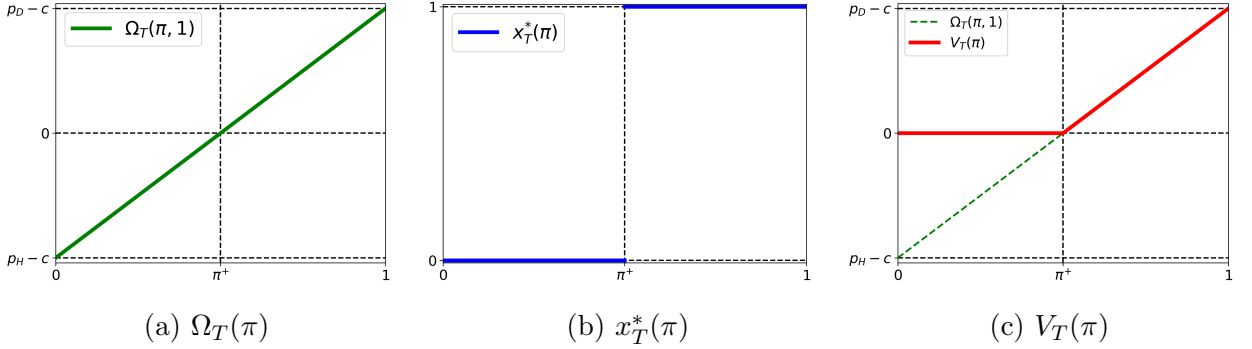


Figure 4: Myopic objective, policy and value functions.

### 2.3.2 The multi-period problem

With a multi-period setting, the existence of future periods and their interactions with the present through the learning process alters the objective function at any  $t < T$ ,

$$\Omega_t(\pi_t, x_t) = u(\pi_t, x_t) + \delta \mathbb{E}_t \left[ V_{t+1}(\tilde{\pi}_{t+1}) | x_t \right], \quad (3)$$

and

$$V_t(\pi_t) = \max_{x_t \in [0,1]} \Omega_t(\pi_t, x_t),$$

where  $\delta$  represents time-discounting.<sup>8</sup> Denote also with the superscript  $d$  the average discounted counterparts  $\Omega_t^d = \frac{1}{\sum_{j=t}^T \delta^{j-t}} \Omega_t$  and  $V_t^d = \frac{1}{\sum_{j=t}^T \delta^{j-t}} V_t$ . Our problem is now a dynamic programming one with parameters  $[\delta, u, \mathbb{P}(\tilde{\pi}_{t+1}), x]$ .

In a finite horizon setting, this problem can be solved by backward induction. Therefore, the total number of periods  $T$  is important and we will occasionally index functions by  $\{t, T\}$  rather than just  $\{t\}$ . The (piece-wise) linearity features of the objective/value functions persist through time and the corresponding optimal auditing strategies remain bang-bang. Lemma 3 characterizes all these features.

**Lemma 3.** *The objective function  $\Omega_{t,T}(\pi, x)$  is*

*i - linear in  $x$ , continuous and piece-wise linear in  $\pi$ ,*

*ii -  $\frac{\partial \Omega_{t,T}}{\partial x} \Big|_{\pi}$  is continuous and increasing in  $\pi$ ,  $\frac{\partial \Omega_{t,T}}{\partial x} \Big|_{\pi=1} > 0$  and  $\frac{\partial \Omega_{t,T}}{\partial x} \Big|_{\pi=0} < 0$ ,*

*iii - and there exists a threshold  $\pi_{t,T}^* \in (0, 1)$  defined uniquely by  $\frac{\partial \Omega_{t,T}}{\partial x} \Big|_{\pi_{t,T}^*} = 0$ .*

*The optimal auditing strategy  $x_{t,T}^*(\cdot)$  is bang-bang:  $x_{t,T}^*(\pi) = \mathbb{1}_{\{\pi \geq \pi_{t,T}^*\}}$ .*

*The value function  $V_{t,T}(\pi)$  verifies*

*i -  $V_{t,T}(\pi) = \Omega_{t,T}(\pi, x_{t,T}^*(\pi)) = \max(0, \Omega_{t,T}(\pi, 1))$ ,*

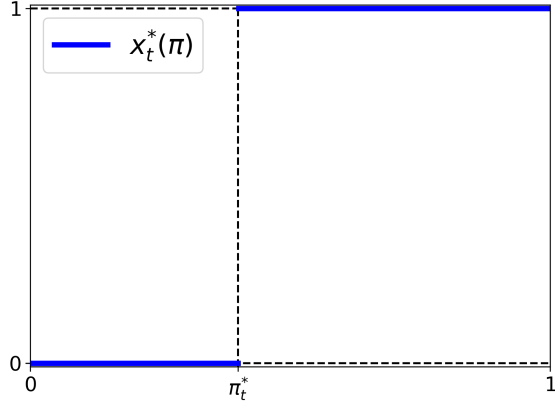
*ii - it is continuous and piece-wise linear in  $\pi$ ,*

*iii - it is convex in  $\pi$  (as the maximum of linear functions).*

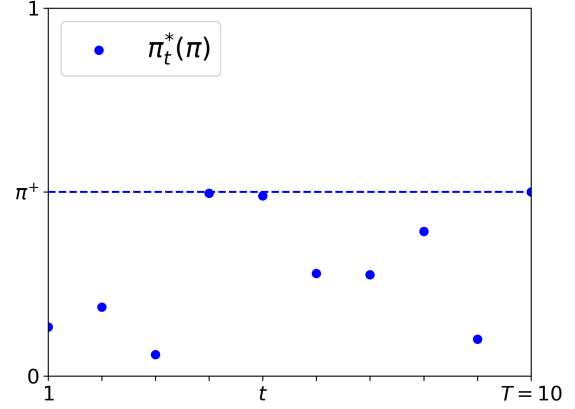
Figure 5 illustrates the Lemma 3 statements. Figure 5a illustrates the fact that optimal auditing is bang-bang. Figure 5b represents a possible shape of the intertemporal optimal auditing thresholds, with one threshold for each period. At this stage, we can't say anything more about these thresholds other than that they are lower than  $\pi^+$ . Figure

---

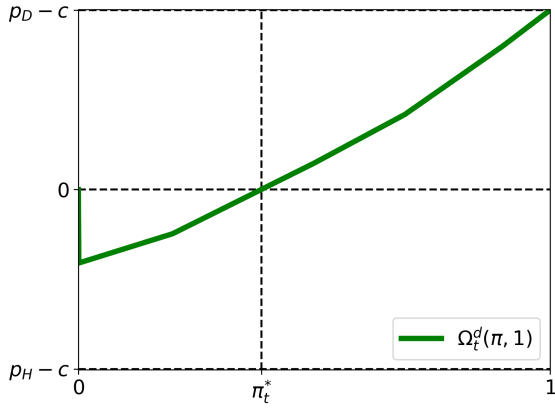
<sup>8</sup>Our model may be extended to the case where the horizon  $T$  is random, and  $\delta$  includes a termination probability.



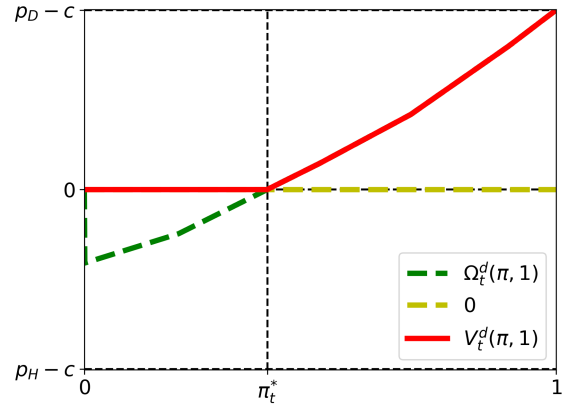
(a) Optimal Policy Function



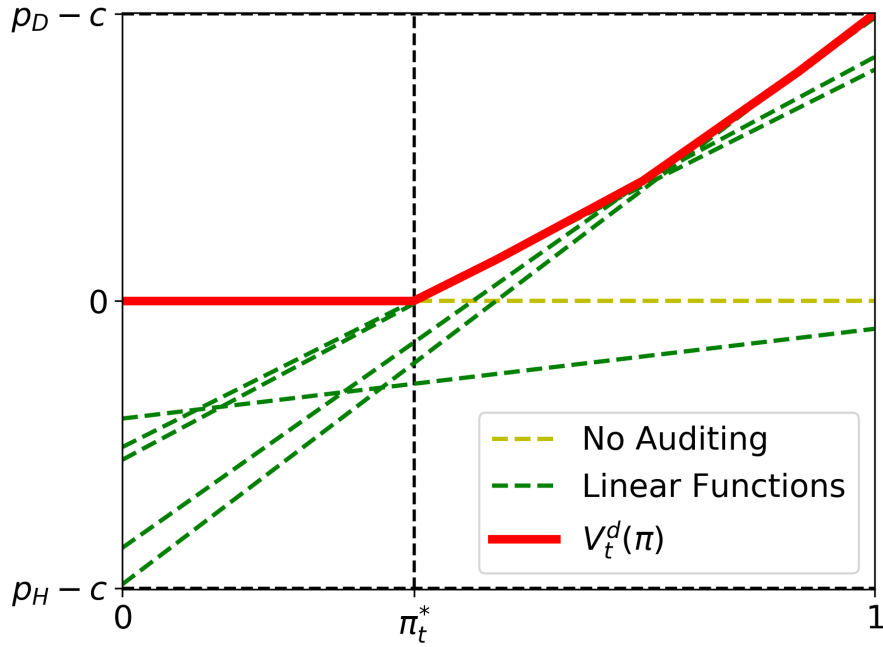
(b) Auditing thresholds



(c) Objective function



(d)  $V_t = \max(0, \Omega_t(\pi, 1))$



(e)  $V_t = \max(\text{Linear functions})$

Figure 5: Time  $t$  policy, objective and value functions

5c illustrates the fact that the objective function at time  $t$  where the insurer audits with probability one is null for the corresponding threshold  $\pi_t^*$ . Finally, Figures 5d and 5e show how the value function is obtained either as the maximum between 0 and the objective function under certain auditing, or as the maximum of linear functions. At this stage of the paper, these linear functions appear in the proof of Lemma 3, but we don't know more about their signification.

At any time  $t \in \{1, \dots, T-1\}$ , the objective function in equation (3) differs from the myopic one through the second expectation term, which represents the impact of current auditing decision  $x_t$  on expected future auditing proceeds through its effects on beliefs. Lemma 4 describes this component's dependence on  $x_t$ .

**Lemma 4.** *The component  $\mathbb{E}_t[V_{t+1}(\tilde{\pi}_{t+1})|x_t]$  is increasing in  $x_t$  and always positive.*

In other words, current auditing increases expected future proceeds because of an exploration effect: the information gathered at the current period allows more accurate audits in the future. The objective function can therefore be decomposed into an exploitation and an exploration component

$$\Omega_t(\pi_t, x_t) = \overbrace{u(\pi_t, x_t)}^{\text{Exploitation}} + \delta \overbrace{\mathbb{E}_t[V_{t+1}(\tilde{\pi}_{t+1})|x_t]}^{\text{Exploration}}.$$

		$(0, \pi^+)$	$(\pi^+, 1)$
Exploitation	$\frac{\partial}{\partial x} u$	-	+
Exploration	$\frac{\partial}{\partial x} \mathbb{E}_t[V_{t+1}]$	+	+
Objective	$\frac{\partial}{\partial x} \Omega_{t,T}(\pi)$	?	+

Table 1: Marginal effects of auditing

The myopic threshold  $\pi^+ = \pi_{T,T}^*$  divides the belief space into two segments where exploitation and exploration effects may go in different directions (see Table 1): above

$\pi^+$ , both components are increasing in  $x_t$ , while below  $\pi^+$ , the exploitation effect is decreasing in  $x_t$ , discouraging auditing, while the exploration one is increasing in  $x_t$ , encouraging it.

The exploration and exploitation effects are thus conflicting on  $(0, \pi^+)$ . The exploitation component is maximized for  $\pi_t^* = \pi^+$ . The exploration component is maximized for  $\pi_t^* = 0$ . We can therefore identify the sequence  $(\pi_t^*)$ , illustrated in Figure 5b, as a measure of the dynamic balance between exploitation and exploration. As illustrated in Figure 6, the closer  $\pi_t^*$  to 0, the more the auditor explores, and the closer  $\pi_t^*$  to  $\pi^+$ , the more he exploits.



Figure 6: Levels of Exploration/Exploitation

Our goal hereafter is to understand how this optimal balance between exploration and exploitation behaves through time, i.e., to characterize the sequence  $(\pi_t^*)$ .

### 3 Exploration vs Exploitation: Auditing to Separate the Wheat from the Chaff

In this section, we present our main results about the evolution of optimal auditing thresholds as functions of time and discounting, and we relate our problem to classical learning problems. In Section 3.1, we answer our question of interest for the finite horizon case, then take the limit when  $T$  tends to infinity to characterize the infinite horizon solution in Section 3.2. Finally we interpret our problem as an optimal stopping one in Section 3.3 and as a Bandit one in 3.4.



### 3.1 Finite horizon optimal auditing

The inclusion of the exploration component at early periods encourages more auditing by decreasing the optimal thresholds. In addition, the larger the discount rate  $\delta$ , the stronger the exploration effect. These results are stated in Proposition 1.

**Proposition 1** (Finite Horizon Optimal Thresholds). *The existence of the learning opportunity changes the auditing efforts as*

- i - the optimal thresholds sequence  $(\pi_{t,T}^*)_{t \in \{1, \dots, T\}}$  is strictly increasing, i.e., there is more exploration in the early stages of the relationship,*
- ii - the optimal level of exploration depends on the remaining number of periods  $(\pi_{t,T}^* = f(T - t))$  and for all  $i \in \{0, 1, \dots, t - 1\}$ ,  $\pi_{t,T}^* = \pi_{t-i, T-i}^* = \pi_{1, T-t+1}^*$ ,*
- iii - the sequence of first period thresholds  $(\pi_{1,T}^*)_{T \in \mathbb{N}^*}$  is strictly decreasing,*
- iv - the optimal levels of exploration are increasing with the DM's patience, i.e., for any  $t \in \{1, \dots, T - 1\}$ :*

$$\frac{\partial \pi_{t,T}^*}{\partial \delta} < 0 \quad \text{and} \quad \lim_{\delta \rightarrow 0} \pi_{t,T}^*(\delta) = \pi^+.$$

Statement i means that, for a given number of periods, the earlier the audit, the larger the auditing set  $(\pi_{t,T}^*, 1]$ . Figure 7a illustrates this through the fact that the optimal thresholds are increasing in time. This is due to the fact that interacting for a longer time with auditees induces the auditor to learn more about them, even when their claims are believed to be non profitable to audit at that point in time. Because information gained from the current period audit increases all subsequent periods' expected auditing proceeds, the generated gains cover the one-time preliminary expenses. In terms of exploration/exploitation, the auditor explores more at the beginning, hoping to identify dishonest types soon enough to exploit them later. Statement ii formalizes this dependence on the remaining number of periods. As shown in Figure 7b, starting at period 1 with a horizon  $T = 5$  yields the same optimal thresholds as starting at period 5 with a horizon

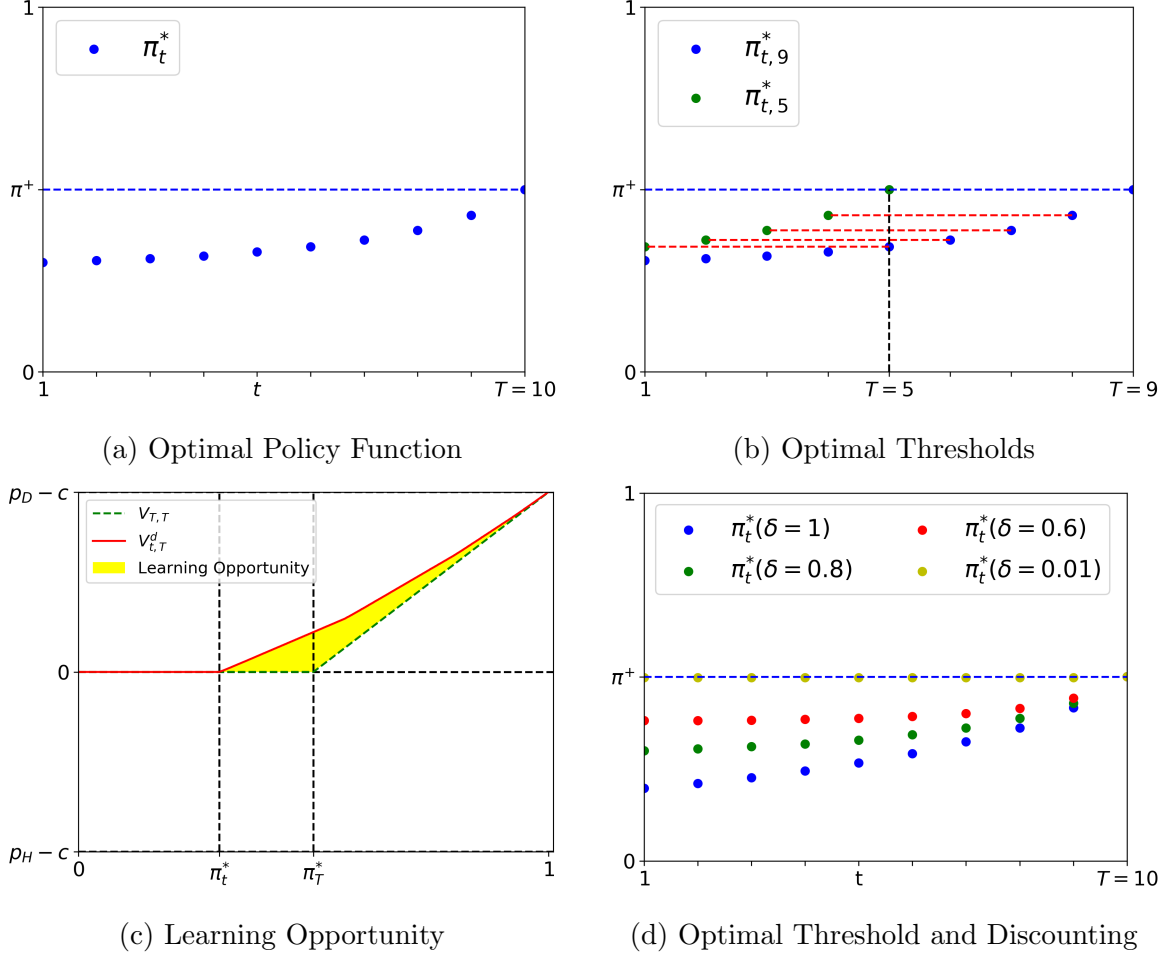


Figure 7: Finite Horizon Thresholds and Value Function

$T = 9$ : in both cases, there are 4 periods remaining, and  $(\pi_{t,5}^*)_{t \in \{1, \dots, 5\}} = (\pi_{t,9}^*)_{t \in \{5, \dots, 9\}}$ . As a consequence, the average discounted value function  $V_{t,T}^d$  is larger than the myopic value function  $V_{T,T}$ , and the difference represents the learning opportunity, as illustrated in Figure 7c. Statement iii is a direct consequence of statements i and ii because of the backward induction nature of the solution, and conveys the idea that, at the beginning of the relationship, more exploration takes place if there are more periods remaining (i.e.  $T$  increases). Statement iv tells us that the more patient the insurer, the more exploration takes place as more weight is put on the remaining periods. However, if the auditor is not patient, optimal auditing sequences tend to the myopic/one-period threshold  $\pi^+$ . Figure 7d illustrates this last statement.

Observe that the increasing optimal thresholds mean that there is more auditing at the beginning of the relationship. Mittone (2006) notes in a tax evasion experiment

that auditing may act as an “educating” mechanism when applied early. Participants audited at the beginning show less propensity to defraud in subsequent periods, even if they are no longer audited. On the contrary, subjects not audited before the second part of the experiments continue defrauding and are hardly deterred. Our result provides an alternative information based motivation for intense auditing earlier in an auditor-auditee relationship.

### 3.2 Infinite horizon optimal auditing

With an infinite horizon, the optimal auditing strategy is characterized in Proposition 2 below.

**Proposition 2** (Infinite Horizon Optimal Thresholds). *The infinite-horizon problem is obtained by taking the limit  $T \rightarrow \infty$*

- i - The optimal thresholds sequence  $(\pi_{t,T}^*)_{t \in \mathbb{N}^*}$  is time independent, i.e. constant and equal to  $\pi_\infty^* = \lim_{T \rightarrow \infty} \pi_{1,T}^* \geq 0$ .*
- ii - The infinite horizon objective function  $V_\infty(\pi)$  is convex in  $\pi$  as the supremum of an infinite countable family of linear functions.*
- iii - Total exploration occurs when the DM is patient enough*

$$\lim_{\delta \rightarrow 1} \pi_\infty^*(\delta) = 0.$$

Statements i is illustrated in Figure 8a and Figure 8b, while statement ii is illustrated in Figure 8c. According to statement iii, a necessary and sufficient condition for all priors to be audited at the initial period when  $T$  becomes arbitrarily large is for the auditor to be patient enough, i.e.,  $\delta$  close enough to 1. The underlying intuition is that, while additional auditing allows to have refined information and better targeted audits, the lower the initial period prior, the longer the insurer has to wait before he can derive positive proceeds for the corresponding SPs.

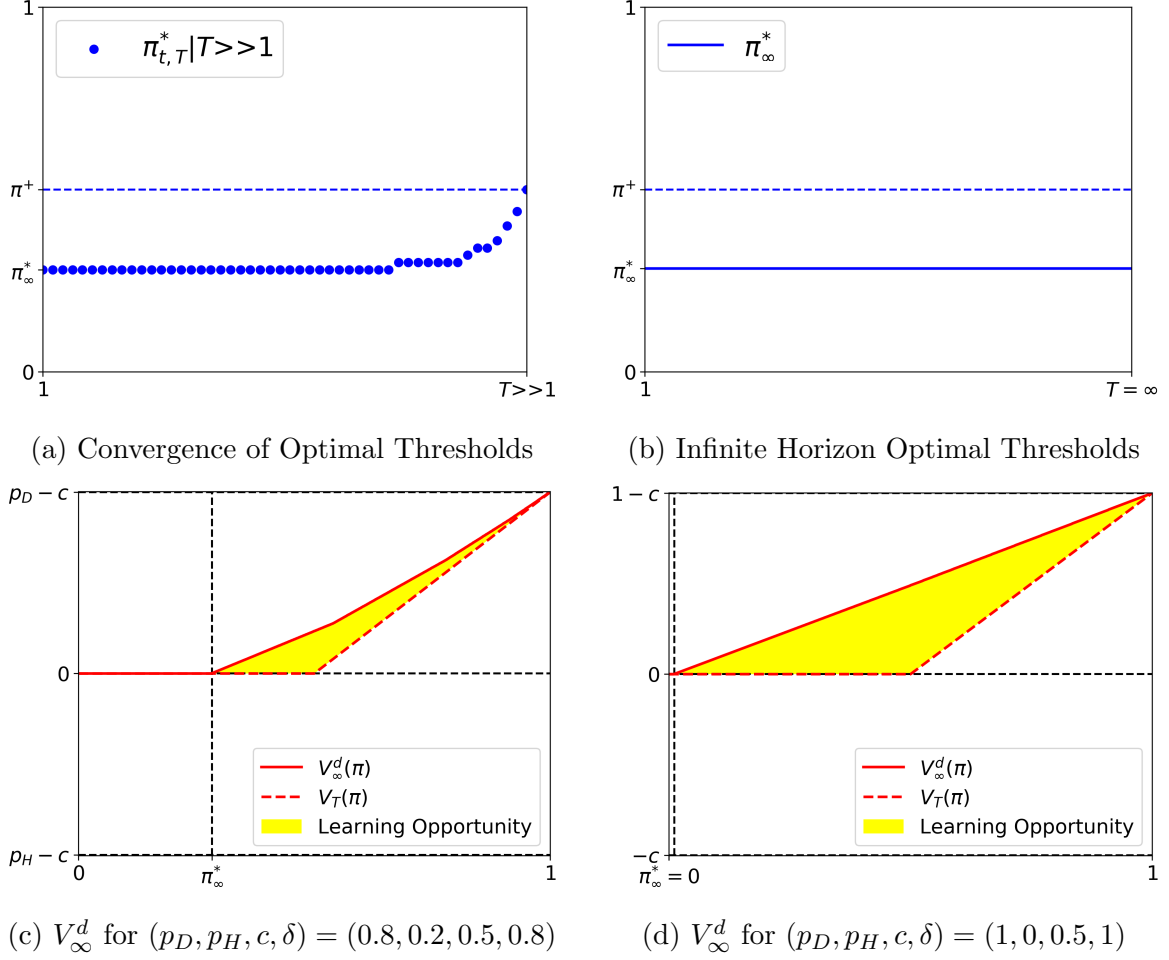


Figure 8: Infinite Horizon Thresholds and Value Functions

*Example 1.* Assume  $(p_D, p_H) = (1, 0)$ . In this context, types are revealed after only one period as  $\varphi_{AI} = 1$  and  $\varphi_{AV} = 0$ . Then, the objective function is given by

$$\begin{aligned}
 \Omega_\infty(\pi, x) &= \left( \pi - c + \pi \sum_{i \geq 1} \delta^i (1 - c) \right) x \\
 &= \left( \pi - c + \pi(1 - c) \frac{\delta}{1 - \delta} \right) x \\
 &= \left( \pi \left( \frac{1 - \delta c}{1 - \delta} \right) - c \right) x.
 \end{aligned}$$

Then  $\pi_\infty^*$  verifies  $\pi \left( \frac{1 - \delta c}{1 - \delta} \right) - c = 0 \Rightarrow \pi_\infty^* = \frac{(1 - \delta)c}{1 - \delta c}$  and

$$V_\infty(\pi) = \left( \pi \left( \frac{1 - \delta c}{1 - \delta} \right) - c \right) \mathbb{1}_{\{\pi > \pi_\infty^*\}}.$$

Figure 8d illustrates the average discounted value function  $V_\infty^d$  as a linear function when

$\delta \rightarrow 1$ , as linear combination with weights  $\pi$  and  $1 - \pi$  between per period payoffs  $1 - c$  and 0. Indeed, in this case, with probability  $\pi$ , the auditee is dishonest and the average future payoff per period is  $1 - c$ , and with probability  $1 - \pi$  the average future payoff per period is 0, as auditing is stopped. Current period proceeds do not appear as they are amortized over an infinite undiscounted number of periods.

### 3.3 Dynamic auditing as an optimal stopping problem

A consequence of 1 and 2 is that once it is optimal to not audit at a given period, not auditing is optimal at all subsequent periods. This is due to the fact that our optimal auditing problem can be formulated as an optimal stopping problem, as shown in Proposition 3 below.

**Proposition 3.** *Let  $\tau$  be a stopping time with respect to  $(\mathcal{F}_t)_{t \in \{1, \dots, T\}}$ , where  $T \in \mathbb{N} \cup \{\infty\}$ , and define  $W_{1,\tau}(\pi_1, \delta)$  as the expected proceeds of systematically auditing for  $\tau$  periods, starting from a prior  $\pi_1$  and with time discounting  $\delta$ ,*

$$W_{1,\tau}(\pi_1, \delta) = \mathbb{E} \left[ \sum_{j=1}^{\tau-1} \delta^{j-1} u(\tilde{\pi}_j, 1) \right], \quad \text{and} \quad W_{1,1}(\pi_1, \delta) = 0.$$

*Then, for  $\nu = \mathbb{1}\{\bigcup_{j \in \{1, \dots, T\}} \{\tilde{\pi}_j < \pi_j^*\}\}$ , the optimal stopping time  $\tau^*$ , defined as*

$$\tau^* = \nu \min \left\{ t \in \{1, \dots, T\} \mid \mathbb{1}\{\tilde{\pi}_t < \pi_t^*\} \right\} + (1 - \nu)(T + 1),$$

*verifies*

$$V_{1,T}(\pi_1, \delta) = W_{1,\tau^*}(\pi_1, \delta) = \max_{\tau} W_{1,\tau}(\pi_1, \delta).$$

Figure 9 illustrates the optimal stopping rule. When auditing is systematic, in Figure 9a, beliefs for both types converge towards the true belief. Under the optimal strategy, in Figure 9b, type  $H$  stops being audited after 89 periods, because the corresponding belief drops below the optimal threshold. Type  $D$ , on the contrary, continues being audited until the end. This optimal stopping rule defines an optimal stopping region in the belief

space at each period  $t$  as  $SR_t = \{\pi | \pi < \pi_t^*\}$ . As soon as the belief process  $\langle \pi_t \rangle$  generated under the optimal auditing strategy is lower than  $\pi_t^*$ , auditing and learning stop, and consequently the belief remains constant.

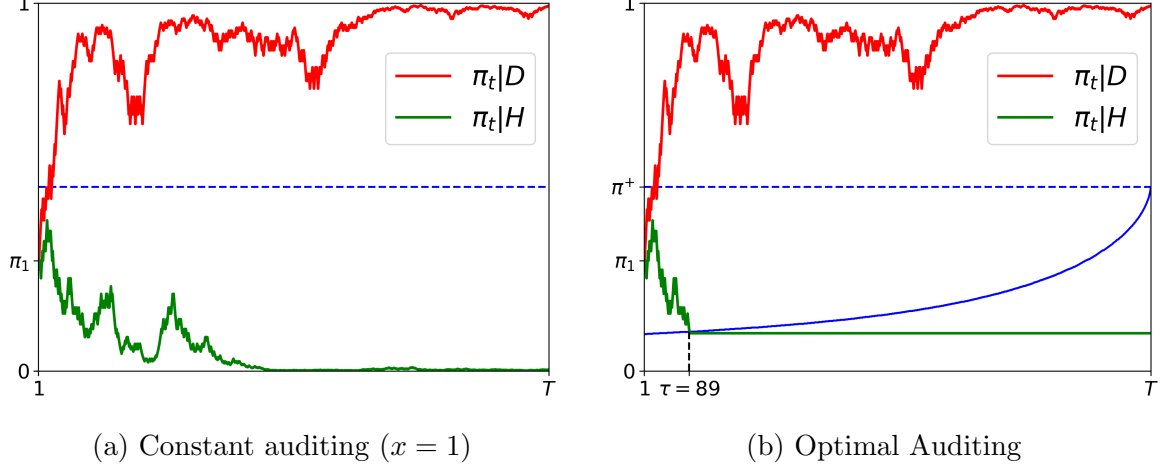


Figure 9: Optimal auditing as an optimal stopping rule

This formulation as an optimal stopping problem allows us to better understand what are the linear functions of which  $V_{1,T}$  is the maximum (Figure 5e). proposition 4 formalizes this idea.

**Proposition 4.** *There exists an integer  $M \in \mathbb{N} \cup \{\infty\}$  and a partition  $\bigcup_{m \in \{1, \dots, M\}} \Psi_m$  of  $(0, 1)$  defined as*

$$(\pi, \pi') \in \Phi_m^2 \iff \forall \omega = (\omega_1, \dots, \omega_T), \tau^*(\omega, \pi) = \tau^*(\omega, \pi'),$$

where  $\omega$  is a trajectory of the fraud generating process. Then,  $V_{1,T}(\pi)$  is linear on each  $\Psi_m$ , and the family of linear functions of which  $V_{1,T}$  is the maximum corresponds to the linear functions on  $\Psi_m$ .

Proposition 4 means that the value function is linear on intervals of beliefs for which auditing stops at the same time for all priors of the same interval. In other words, the trajectories of the belief process for a realization  $\omega$  are such that the time  $t$  belief is either higher or lower than  $\pi_t^*$ , for all the initial beliefs belonging to the same interval.

In particular, with a finite horizon, one such interval is defined as  $\{\pi | \pi > \varphi_{AV}^{t-T}(\pi_T^*)\}$ . This is the set of beliefs that are so large that they are always audited, because updating

will never bring them below an optimal threshold. In this case, the corresponding optimal stopping time is always equal to  $T + 1$ , and

$$\begin{aligned}
V_{1,T}((\pi_1, \delta)) &= W_{1,T+1}(\pi_1, \delta) = \sum_{j=1}^T \delta^{j-1} \mathbb{E}[u(\tilde{\pi}_j, 1)] \\
&= \sum_{j=1}^T \delta^{j-1} u(\mathbb{E}[\tilde{\pi}_j], 1) \\
&= \sum_{j=1}^T \delta^{j-1} u(\pi_1, 1), \quad \text{from the martingale property.}
\end{aligned}$$

This shows that for the highest beliefs, the average discounted value function is equal to the myopic value function.

### 3.4 Dynamic auditing as an armed bandit problem

Optimal stopping problems are also intimately related to Bandit problems. Bandit problems are settings where a slot machine, referred to as a "bandit", has  $k + 1$  arms that yield different payoffs. All arms but one have payoffs governed by an unknown random variable, the last arm representing an outside option with known deterministic payoff that may be normalized to 0. A gambler must choose a limited number of arms to play at each period, and can update his beliefs about the payoffs after observing the outcome of his play. He thus faces an exploration/exploitation dilemma between choosing once and for all the best arm at the initial period, or explore by updating his choice given the accumulated experience. Hereafter, we place ourselves in an infinite horizon setting, but results are valid in the finite horizon problem.

#### 3.4.1 One-armed bandit

We start with the most simple setting for a Bandit problem: the One-Armed Bandit. In this setting the player only faces one risky arm with returns  $(X_1, X_2, \dots, X_n, \dots)$ , and one safe arm with deterministic return  $\lambda$  at each period. Arms are characterized by their

Gittins Index.

**Definition 1** (Gittins Index for an arm). For an arm characterized by the returns  $(X_1, X_2, \dots, X_n, \dots)$ , and  $N$  a stopping time, its Gittins Index is given by

$$\Lambda(\delta) = \sup_{N \geq 1} \left\{ \frac{\mathbb{E}(\sum_1^N \delta^{j-1} X_j)}{\mathbb{E}(\sum_1^N \delta^{j-1})} \right\}.$$

Thus, the safe arm has a Gittins Index equal to  $\lambda/(1 - \delta)$ . The optimal policy is given by the Gittins Index Theorem for a One-Armed Bandit, whereby it is optimal to play the risky arm if  $\Lambda(\delta) > \lambda/(1 - \delta)$ . Otherwise, it is optimal to use the safe arm once and for all.

In our setting, the risky arm is auditing and the associated return are a function of the belief  $\pi$ . Its has a Gittins Index  $\Lambda(\pi, \delta)$  that depends on  $\pi$ . The risky returns are given by  $X_j = u(\tilde{\pi}_j, 1)$ . The safe arm corresponds to not auditing, with  $\lambda = 0$ . When the Gittins Index of the risky arm is larger than 0, it is equivalent to the condition  $\pi > \pi_\infty^*$ .

**Proposition 5.**

$$\Lambda(\pi_1, \delta) > 0 \iff \pi_1 > \pi_\infty^*.$$

This is a direct consequence of the fact that  $V_\infty(\pi_1) = \mathbb{E}[\sum_1^{\tau^*} \delta_{j-1} u(\tilde{\pi}_j, 1)]$  and  $V_\infty(\pi_1) > 0$  for  $\pi_1 > \pi_\infty^*$ .

### 3.4.2 k-armed bandit

This characterization of our problem as a Bandit one allows us to answer the question of what happens with multiple auditees and capacity constrained auditing. Multi-Armed Bandits optimal strategies consist in Gittins Index Rules.

**Definition 2** (Gittins Index Rule). A decision rule that at each stage chooses an arm that has the highest Gittins index is called a **Gittins index rule**.

The Gittins Index Theorem shows that such rules are optimal.



**Theorem 1** (The Gittins Index Theorem). *For a  $k$ -armed bandit problem with independent arms and geometric discounting, any Gittins index rule is optimal.*

We show in Proposition 6 that, in the context of our auditing problem, the Gittins Index Rule is equivalent to auditing arms by order of highest prior.

**Proposition 6.** *Given an initial prior  $\pi_1$ , using a Gittins Index Rule is equivalent to auditing the highest priors in priority, conditional on the priors being larger than  $\pi_\infty^*$ .*

$$\Lambda(\pi_1, \delta) > \Lambda(\pi'_1, \delta) \iff \pi_1 > \pi'_1 \quad (> \pi_\infty^*).$$

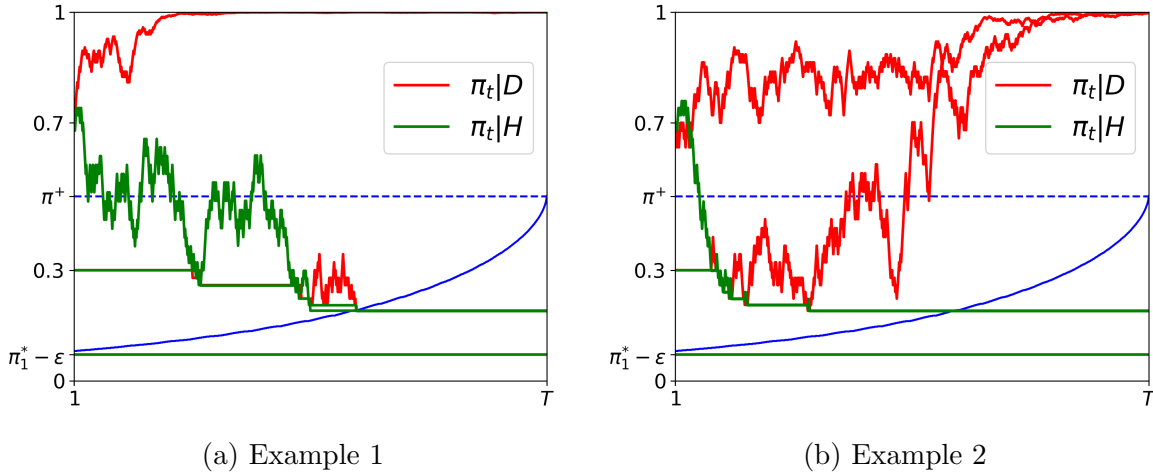


Figure 10: 6-armed bandit with a capacity constrained audit ( $k = 2$ )

Therefore, we can derive the optimal auditing strategy when there are multiple auditees and a capacity constrained audit. Let there be a population of auditees of size  $n$ . Let there also be a capacity constraint on auditing:

$$\sum_{i=0}^n x_{i,t} \leq k.$$

Then an optimal strategy is to audit by decreasing prior until the constraint is binding or the optimal threshold  $\pi_\infty^*$  is reached. Figure 10 shows two examples of a 6-armed bandit when  $k = 2$ . The two auditees, one of each type, with the lowest initial prior at  $\pi_1^* - \varepsilon$  are never audited, because the outside option is more valuable since the beginning. In both examples, the type  $D$  starting at a prior of 0.7 is kept in the auditing pool. The

other type  $D$ , starting at a lower prior of 0.3 is excluded from auditing in Figure 10a after a certain number of periods where it competes with the other two auditees of type  $H$ . In Figure 10b, this other type  $D$ , is kept in the auditing pool even if its prior decreases sharply at first. The two type  $H$  auditees are both excluded after a finite time.

## 4 The Extents of Learning

After showing how the possibility of learning through audits alters optimal auditing strategies, we now focus on different aspects of the learning process. First, in Section 4.1, we examine how much information the DM gets on average, i.e., how often he gets to know the truth, or at least how often he learns enough to take the optimal action. Second, in Section 4.2, we focus on the speed at which information is gathered, i.e., how fast we approach the truth.

### 4.1 How much can the DM learn?

We are hereafter interested in the question of whether the DM gets to learn the truth, i.e., how often the belief process converges to the true belief. In the Optimal Learning academic literature,<sup>9</sup> this question is addressed through the concepts of complete and adequate learning (e.g., Easley & Kiefer (1988), Aghion et al. (1991)). Complete learning occurs when, with probability one, the agent acquires the true information about the state of the world. Adequate learning is a weaker version of complete learning, whereby it occurs when, with probability one, the agent acquires enough information to allow him to obtain the true maximum payoff” (i.e., ”learning everything worth knowing”).

Aghion et al. (1991) address the question of characterizing situations where adequate learning obtains or does not obtain. More specifically, they ask the question of whether adequate learning is the generic outcome. One situation where they find adequate learning to obtain is the undiscounted case, i.e.,  $\delta \rightarrow 1$ .<sup>10</sup> This is exactly analogous to

---

<sup>9</sup>In this literature, the horizon is usually infinite.

<sup>10</sup>Other exhibited situations where adequate learning occurs include smoothness and quasi-concavity, or analyticality of the payoff function

Statement [iii](#) in [proposition 2](#): because of the absence of discounting, the optimal threshold tends to 0, which means systematic auditing, and therefore convergence of the belief process to its true value ([Lemma 1](#)). However, they argue that adequate learning is not the generic outcome, and define *partial learning* as occurring when the agent acquires adequate knowledge with probability strictly between 0 and 1. They suggest, through an informal discussion, that partial learning is the generic outcome.

Hereafter, we show that we can formally answer this question in our setting. With an infinite horizon, complete, adequate, and partial learning can be formulated through the distribution of the optimal stopping times. With a finite horizon, the comparison is not possible as we cannot talk about the convergence of beliefs.

**Infinite horizon** In the infinite horizon setting, adequate learning occurs if, in the case of a type  $H$ , the belief converges to a limit for which there is no auditing, or if, in the case of a type  $D$ , the belief converges to a limit for which there is auditing. In other words, conditional on type  $H$ , adequate learning occurs when  $\tau^* < \infty$ , and the belief process drops to a value  $\pi^{lim}$  below  $\pi_\infty^*$ . The belief therefore converges to  $\pi^{lim}$  since auditing stops. Conditional on type  $D$ , adequate learning occurs when  $\tau^* = \infty$ , auditing never stops, and the belief process converges to 1. [Definition 3](#) formalizes this idea and [Proposition 7](#) answers the question of genericity.

**Definition 3** (Adequate learning in infinite horizon). Let  $\tau^*$  be the stopping time  $\tau^* = \inf(t \in \mathbb{N}^* | \pi_t < \pi_\infty^*)$ . For an initial prior  $\pi_1$ , adequate learning occurs with probability

$$\begin{aligned} \mathbb{P}(AL|\pi_1) &= \pi_1 \mathbb{P}(\tau^* = \infty | D, \pi_1) + (1 - \pi_1)(1 - \mathbb{P}(\tau^* = \infty | H, \pi_1)), \\ &= \pi_1 \mathbb{P}(\tau^* = \infty | D, \pi_1) + (1 - \pi_1). \end{aligned} \tag{4}$$

**Proposition 7.** *In this setting, partial learning is the generic outcome for all initial priors  $\pi_1$*

$$0 < \mathbb{P}(AL|\pi_1) < 1, \quad \forall \pi_1 \in (0, 1). \tag{5}$$

In particular

$$\mathbb{P}(AL|\pi_1 < \pi_\infty^*) = (1 - \pi_1). \quad (6)$$

Adequate learning is never obtained because it is too strong a condition. Adequacy is always obtained conditional on type  $H$ , but never for type  $D$ . When the initial prior is below the optimal threshold, adequate learning occurs if the auditee is of type  $H$ , i.e., with probability  $(1 - \pi_1)$ , but not for type  $D$ , as he is never audited. This means  $\mathbb{P}(\tau^* = \infty|D, \pi_1) = 0$  and explains equation (6). When the initial prior is larger than  $\pi_\infty^*$ , there is a non null probability, conditional on type  $D$ , for the belief process to hit the optimal threshold, and  $\mathbb{P}(\tau^* = \infty|D, \pi_1) < 1$ . Example 2 below provides a special case where  $\mathbb{P}(\tau^* = \infty|D, \pi_1)$  is computable.

*Example 2* (Infinite horizon Gambler' Ruin). When  $p_D + p_H = 1$ , the updating functions are symmetrical ( $\varphi_{AI} = \varphi_{AV}^{-1}$ ). Given an infinite horizon threshold  $\pi_\infty^*$  and an initial prior  $\pi_1$ , the updating functions define a partition of  $(0, 1) = \bigcup_{k \in \mathbb{Z}} \Phi_k = \bigcup_{k \in \mathbb{Z}} (\varphi_{AI}^k(\pi_1), \varphi_{AI}^{k+1}(\pi_1)]$ , and there exists a  $k^*$  such that  $\pi_\infty^* \in \Phi_{k^*}$ . Then the problem is analogous to a Gambler's Ruin problem with a biased toss, where the player wins with probability  $p_D > \frac{1}{2}$  and has initial fortune  $k^*$ . Thus

$$\mathbb{P}(\tau^* = \infty|D, \pi_1) = \begin{cases} 0 & \text{if } k^* \geq 0, \\ 1 - (\frac{p_D}{1-p_D})^{k^*} & \text{if } k^* < 0, \end{cases}$$

and, when  $\pi_1 > \pi_\infty^*$ , partial learning occurs, with adequate learning happening with probability

$$\mathbb{P}(AL|\pi_1) = \pi_1(1 - (\frac{p_D}{1-p_D})^{k^*}) + (1 - \pi_1).$$

## 4.2 How fast can the DM learn?

Lemma 1 shows that the belief process is a martingale and converges to the true belief. Now, we are interested in how fast information is accumulated, and how this speed of

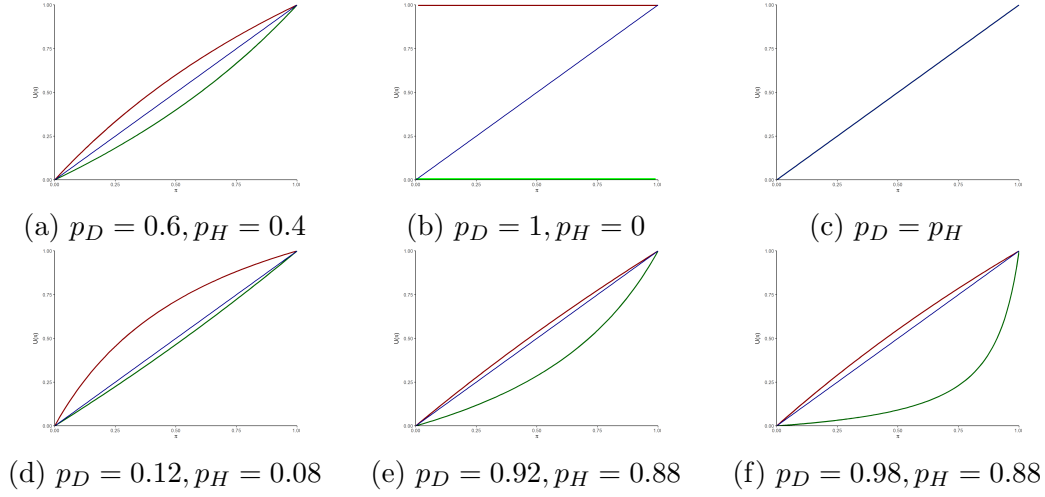


Figure 11: Updating functions for different parameters

information acquisition is related to the parameters of the model.

Figure 11 shows the updating functions  $\{\varphi_{AI}, \varphi_{AV}\}$  for different values of the parameters  $(p_D, p_H)$ . The first intuition is that the more polarized the behaviors of types  $H$  and  $D$ , i.e., the larger  $p_D - p_H$ , the more informative the audit. The extreme case  $p_D - p_H = 1$  is illustrated in Figure 11b. In this case, auditing is perfectly informative, as observing an invalid claim implies that the auditee is type  $D$ , while observing a valid claim implies he is of type  $H$ . On the contrary, when  $p_D = p_H$  (Figure 11c), types behave the same way and auditing reveals nothing about the type. The second intuition is related to the informativeness of extreme behaviors. i.e., when for a fixed difference  $p_D - p_H$ , either  $p_D$  is close to 1 or  $p_H$  is close to 0. If  $p_D \approx 1$ , observing a valid claim is a strong sign that the auditee is of type  $H$ , as shown in Figure 11f. If  $p_H \approx 0$ , observing an invalid claim is a strong sign that the auditee is of type  $D$ .

#### 4.2.1 Doob-Meyer decomposition

In proposition 8, we use the Doob-Meyer decomposition for submartingales and supermartingales to decompose the belief processes under both states of the world  $H$  and  $D$ . This decomposition allows us to identify the magnitude of information acquisition.

**Proposition 8.** *[Information acquisition] Under states of the world  $H$  and  $D$ , we can decompose the stochastic process of updated beliefs thanks to the Doob-Meyer decomposi-*

tion.

*i Under  $\mathbb{P}_H$ ,  $\tilde{\pi}_t = M_t^H + P_t^H$  where  $M_t^H$  is a martingale and  $P_t^H$  is a decreasing and predictable process.*

$$ii \ P_t^H = -(p_D - p_H) \sum_{i=1}^{t-1} \tilde{\pi}_i \left[ \varphi_{AI}(\tilde{\pi}_i) - \varphi_{AV}(\tilde{\pi}_i) \right] < 0$$

*iii Under  $\mathbb{P}_D$ ,  $\tilde{\pi}_t = M_t^D + P_t^D$  where  $M_t^D$  is a martingale and  $P_t^D$  is an increasing and predictable process.*

$$iv \ P_t^D = (p_D - p_H) \sum_{i=1}^{t-1} (1 - \tilde{\pi}_i) \left[ \varphi_{AI}(\tilde{\pi}_i) - \varphi_{AV}(\tilde{\pi}_i) \right] > 0$$

The predictable process  $P_t^i$  represents the average gain in information under the true state  $i \in \{H, D\}$

$$\mathbb{E}_i[\tilde{\pi}_t | \pi_1] = \pi_1 + \mathbb{E}_i[P_t^i | \pi_1].$$

#### 4.2.2 Informativeness and exploration

The decomposition in Proposition 8 allows us to quantify the amount of information obtained at each stage and to understand how it affects the optimal level of exploration. From the expression of  $P_t$ , we can see that the difference  $\beta = p_D - p_H$  plays an important role. This is rather intuitive in that a larger difference means more polarized behavior between the Honest and Dishonest behaviors. Proposition 9 formalizes the idea that the more polarized the behavior, the larger  $|P_t^H|$  and  $|P_t^D|$ , and the more information is gathered on average.

**Proposition 9.** *At any time  $t$ , the predictable process  $P_t^H$  is increasing in  $p_H$  and decreasing in  $p_D$*

$$\frac{\partial P_t^H}{\partial p_H} > 0 \quad \text{and} \quad \frac{\partial P_t^H}{\partial p_D} < 0,$$

*and the predictable process  $P_t^D$  is increasing in  $p_D$  and decreasing in  $p_H$*

$$\frac{\partial P_t^D}{\partial p_D} > 0 \quad \text{and} \quad \frac{\partial P_t^D}{\partial p_H} < 0.$$

Behaviors are more polarized when  $p_D$  increases and/or  $p_H$  decreases. When it is the case,  $P_t^H$  is smaller, and since  $P_t^H < 0$ , the average decrease in the belief  $|P_t^H|$  is larger, and beliefs converge towards the true belief 0 faster. The reasoning is analogous for type  $D$  as, when behaviors are more polarized,  $P_t^D$  is larger. As  $P_t^D > 0$ , the average increase in the belief  $P_t^D$  is larger and beliefs converge towards the true belief 1 faster.

The impact of a more efficient learning, in the sense that belief processes converge faster, is such that optimal thresholds are lower, and more exploration takes place under optimal auditing at each period. Figure 13 illustrates different optimal threshold sequences for different levels of polarization of behaviors  $\beta = p_D - p_H$ .

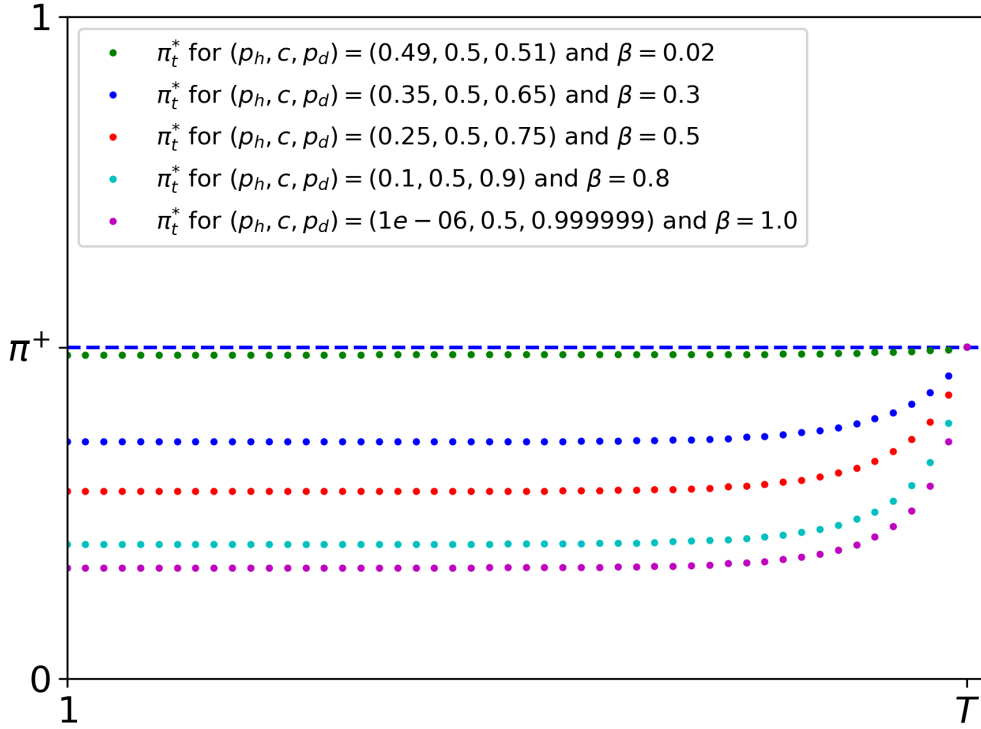


Figure 12: Optimal thresholds depend on  $\beta$

Figure 13: Polarized and extreme fraud behaviors induce more exploration

## 5 Conclusion

This article investigates the role of learning in the context of fraud inspection when inspector/inspectee interactions are repeated. Auditing acts as a learning mechanism that isolates the dishonest (worth auditing) types from the honest (unprofitable to audit)

types.

Our first set of results shows how future auditing proceeds are always increasing in the current auditing efforts, but current proceeds may be negative. Optimal auditing relies then on an exploration/exploitation trade-off consisting in balancing auditing-diminished present proceeds and auditing-enhanced discounted future proceeds. Because learning-generated advantages reverberate through all future periods, the number of remaining periods positively impact auditing efforts. As a weighting of future proceeds, low discounting encourages learning through auditing and complete exploration occurs in an infinite horizon when the auditor is indifferent between present and future proceeds.

Our second set of results examines the particularities of our learning process: we first show that, in our setting, partial learning is the generic outcome. Second, we identify the average change in the belief process as the predictable component of a Doob-Meyer decomposition of the belief process in each state of the world. This average change is larger when dishonest and honest types behave very differently, as the separating power of auditing is then stronger.



## References

- Aboutajdine, R. & Picard, P. (2018), ‘Preliminary Investigations for Better Monitoring: Learning in Repeated Insurance Audits’, *Risks* **6**(1), 15.
- Aghion, P., Bolton, P., Harris, C. & Jullien, B. (1991), ‘Optimal Learning by Experimentation’, *The Review of Economic Studies* **58**(4), 621–654.
- Arrow, K. J., Blackwell, D. & Girshick, M. A. (1949), ‘Bayes and Minimax Solutions of Sequential Decision Problems’, *Econometrica* **17**(3/4), 213–244.
- Artís, M., Ayuso, M. & Guillén, M. (2002), ‘Detection of Automobile Insurance Fraud With Discrete Choice Models and Misclassified Claims’, *Journal of Risk and Insurance* **69**(3), 325–340.
- Bergemann, D. & Välimäki, J. (2008), Bandit Problems, *in* S. N. Durlauf & L. E. Blume, eds, ‘The New Palgrave Dictionary of Economics: Volume 1 – 8’, Palgrave Macmillan UK, London, pp. 336–340.
- Boning, W. C., Guyton, J., Ronald H. Hodge, I., Slemrod, J. & Troiano, U. (2018), Heard it Through the Grapevine: Direct and Network Effects of a Tax Enforcement Field Experiment, Working Paper 24305, National Bureau of Economic Research.
- Brockett, P. L., Derrig, R. A., Golden, L. L., Levine, A. & Alpert, M. (2002), ‘Fraud Classification Using Principal Component Analysis of RIDITs’, *Journal of Risk and Insurance* **69**(3), 341–371.
- Derrig, R. A. (2002), ‘Insurance Fraud’, *Journal of Risk and Insurance* **69**(3), 271–287.
- Dionne, G., Giuliano, F. & Picard, P. (2008), ‘Optimal Auditing with Scoring: Theory and Application to Insurance Fraud’, *Management Science* **55**(1), 58–70.
- Easley, D. & Kiefer, N. M. (1988), ‘Controlling a Stochastic Process with Unknown Parameters’, *Econometrica* **56**(5), 1045–1064.

- Ferguson, T. (2004), *Optimal Stopping and Applications*, available at <https://www.math.ucla.edu/~tom/Stopping/Contents.html>.
- Gale, D. & Hellwig, M. (1985), ‘Incentive-Compatible Debt Contracts: The One-Period Problem’, *The Review of Economic Studies* **52**(4), 647–663.
- Kirchler, E. (2007), *The Economic Psychology of Tax Behaviour*, Cambridge University Press.
- Major, J. A. & Riedinger, D. R. (2002), ‘EFD: A Hybrid Knowledge/Statistical-Based System for the Detection of Fraud’, *Journal of Risk and Insurance* **69**(3), 309–324.
- Mittone, L. (2006), ‘Dynamic behaviour in tax evasion: An experimental approach’, *The Journal of Socio-Economics* **35**(5), 813–835.
- Mookherjee, D. & Png, I. (1989), ‘Optimal Auditing, Insurance, and Redistribution’, *The Quarterly Journal of Economics* **104**(2), 399–415.
- Picard, P. (2013), Economic Analysis of Insurance Fraud, in G. Dionne, ed., ‘Handbook of Insurance’, Springer New York, pp. 349–395.
- Sutton, R. S. & Barto, A. G. (1998), *Reinforcement Learning: An Introduction*, MIT Press.
- Tennyson, S. & Salsas-Forn, P. (2002), ‘Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objectives’, *Journal of Risk and Insurance* **69**(3), 289–308.
- Townsend, R. M. (1979), ‘Optimal contracts and competitive markets with costly state verification’, *Journal of Economic Theory* **21**(2), 265–293.
- Viaene, S., Derrig, R. A., Baesens, B. & Dedene, G. (2002), ‘A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection’, *Journal of Risk and Insurance* **69**(3), 373–421.
- Wald, A. (1945), ‘Sequential Tests of Statistical Hypotheses’, *The Annals of Mathematical Statistics* **16**(2), 117–186.

# Appendix

## 5.1 Proofs

### 5.1.1 Proof of Lemma 1

*Proof.* At period  $t + 1$ , with prior  $\pi_t$  and auditing decision  $x_t$

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}[\tilde{\pi}_{t+1} | \mathcal{F}_t, x_t] &= x_t[\bar{p}(\pi)\varphi_{AI}(\pi_t) + (1 - \bar{p}(\pi))\varphi_{AV}(\pi_t)] + (1 - x_t)[\varphi_{NA}(\pi_t)] \\
&= x_t[p_D\pi_t + (1 - p_D)\pi_t] + (1 - x_t)\pi_t \\
&= \pi_t
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_D[\tilde{\pi}_{t+1} | \mathcal{F}_t, x_t] &= x_t[p_D\varphi_{AI}(\pi_t) + (1 - p_D)\varphi_{AV}(\pi_t)] + (1 - x_t)[\varphi_{NA}(\pi_t)] \\
&= x_t[\bar{p}(\pi)\varphi_{AI}(\pi_t) + (1 - \bar{p}(\pi))\varphi_{AV}(\pi_t) + (p_D - \bar{p}(\pi))(\varphi_{AI}(\pi_t) - \varphi_{AV}(\pi_t))] + (1 - x_t)\pi_t \\
&= \pi_t + x_t[\underbrace{(p_D - \bar{p}(\pi))}_{\geq 0} \underbrace{(\varphi_{AI}(\pi_t) - \varphi_{AV}(\pi_t))}_{\geq 0}] \\
&\geq \pi_t
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_H[\tilde{\pi}_{t+1} | \mathcal{F}_t, x_t] &= x_t[p_H\varphi_{AI}(\pi_t) + (1 - p_H)\varphi_{AV}(\pi_t)] + (1 - x_t)[\varphi_{NA}(\pi_t)] \\
&= x_t[\bar{p}(\pi)\varphi_{AI}(\pi_t) + (1 - \bar{p}(\pi))\varphi_{AV}(\pi_t) + (p_H - \bar{p}(\pi))(\varphi_{AI}(\pi_t) - \varphi_{AV}(\pi_t))] + (1 - x_t)\pi_t \\
&= \pi_t + x_t[\underbrace{(p_H - \bar{p}(\pi))}_{\leq 0} \underbrace{(\varphi_{AI}(\pi_t) - \varphi_{AV}(\pi_t))}_{\geq 0}] \\
&\leq \pi_t
\end{aligned}$$

Finally, the convergence of the belief process to the true belief is a direct consequence of the Martingale Convergence Theorem for bounded martingales.  $\square$

### 5.1.2 Proof of Corollary 1

*Proof.* This is a direct consequence of the martingale property.  $\square$

### 5.1.3 Proof of Lemma 2

This is straightforward as

- $\bar{p}(\pi) - c$  is strictly increasing in  $\pi$ , strictly negative for  $\pi = 0$  and strictly positive for  $\pi = 1$ ,
- $\pi^+$  is then uniquely defined,
- $\text{sign}\left(\frac{\partial \Omega_T}{\partial x}\right) = \text{sign}(\pi - \pi^+)$ ,
- then  $x_T^* = \mathbb{1}\{\pi \geq \pi^+\}$ .

### 5.1.4 Proof of Lemma 3

*Proof.* We proceed by backward induction for  $t \in \{1, \dots, T\}$ .

**At the final stage  $t = T$ :** The objective function

$$\Omega_{T,T}(\pi, x) = u(\pi, x) = (\bar{p}(\pi) - c)x,$$

is trivially (piece-wise) linear in  $\pi$ , linear in  $x$ . Its derivative with regards to  $x$

$$\left. \frac{\partial \Omega_{T,T}}{\partial x} \right|_{\pi} = \bar{p}(\pi) - c = (p_D - p_H)\pi + p_H - c,$$

is increasing in  $\pi$ , is equal to  $p_D - c > 0$  for  $\pi = 1$  and to  $p_H - c < 0$  for  $\pi = 0$ . It implies the existence of a threshold  $\pi_{T,T}^* \in (0, 1)$  such that

$$\left. \frac{\partial \Omega_{T,T}}{\partial x} \right|_{\pi_{T,T}^*} = 0,$$

and

$$x_{T,T}^*(\pi) \begin{cases} = 0 & \text{if } \pi < \pi_{T,T}^*, \\ \in (0, 1) & \text{if } \pi = \pi_{T,T}^*, \\ = 1 & \text{if } \pi > \pi_{T,T}^*. \end{cases}$$

Then, the value function

$$V_{T,T}(\pi) = \max\left(0, \bar{p}(\pi) - c\right),$$

is (piece-wise) linear in  $\pi$  and convex as the maximum of the finite family of linear functions indexed by  $\mathcal{I}_T$

$$\mathcal{I}_T = \{0, \bar{p}(\pi) - c\}.$$

**At any stage  $t \in \{1, \dots, T-1\}$ :** The objective function

$$\begin{aligned} \Omega_{t,T}(\pi, x) &= u(\pi, x) + \delta \mathbb{E}_t \left[ V_{t+1,T}(\tilde{\pi}_{t+1}) \right], \\ &= (\bar{p}(\pi) - c)x + \delta \left[ x \left( \bar{p}(\pi) V_{t+1,T}(\varphi_{AI}(\pi)) + (1 - \bar{p}(\pi)) V_{t+1,T}(\varphi_{AV}(\pi)) \right) + (1 - x) V_{t+1,T}(\pi) \right], \end{aligned}$$

is trivially linear in  $x$  as  $x$  appears only in the transition probabilities, while  $V_{t+1,T}$  is independent of  $x$ . In addition, by induction,  $V_{t+1,T}$  is piece-wise linear in  $\pi$ . It implies, for every  $\pi \in (0, 1)$ , the existence of a couple  $(\alpha, \beta) \in \mathbb{R}^2$  such that

$$V_{t+1,T}(\pi) = \alpha\pi + \beta.$$

Therefore,

$$\begin{aligned} \bar{p}(\pi) V_{t+1,T}(\varphi_{AI}(\pi)) &= \bar{p}(\pi) \left( \alpha \varphi_{AI}(\pi) + \beta \right) = \alpha p_D \pi + \bar{p}(\pi) \beta, \\ \bar{p}(\pi) V_{t+1,T}(\varphi_{AV}(\pi)) &= (1 - \bar{p}(\pi)) \left( \alpha' \varphi_{AV}(\pi) + \beta' \right) = \alpha' (1 - p_D) \pi + (1 - \bar{p}(\pi)) \beta', \end{aligned}$$

and  $\Omega_{t,T}$  is piece-wise linear in  $\pi$ . It is also continuous in  $\pi$  as the sum of functions continuous in  $\pi$ . Its derivative with regards to  $x$  verifies

$$\begin{aligned}\left.\frac{\partial \Omega_{t,T}}{\partial x}\right|_{\pi} &= \bar{p}(\pi) - c + \delta \left[ \bar{p}(\pi) V_{t+1,T}(\varphi_{AI}(\pi)) + (1 - \bar{p}(\pi)) V_{t+1,T}(\varphi_{AV}(\pi)) - V_{t+1,T}(\pi) \right], \\ \left.\frac{\partial \Omega_{t,T}}{\partial x}\right|_{\pi=0} &= p_H - c < 0, \\ \left.\frac{\partial \Omega_{t,T}}{\partial x}\right|_{\pi=1} &= p_D - c > 0.\end{aligned}$$

In addition, a second derivative with regards to  $\pi$

$$\begin{aligned}\left.\frac{\partial \Omega_{t,T}}{\partial x \partial \pi}\right|_{\pi} &= p_D - p_H + \delta \left[ \alpha(\varphi_{AI}(\pi)) p_D + \alpha(\varphi_{AV}(\pi)) (1 - p_D) - \alpha(\pi) \right], \\ &> 0,\end{aligned}$$

is strictly positive,<sup>11</sup> thus defining the threshold  $\pi_{t,T}^*$ . Since the optimal action is either to audit or not to audit, the value function can be expressed as

$$\begin{aligned}V_{t,T}(\pi) &= \max \left( \Omega_{t,T}(\pi, 0), \Omega_{t,T}(\pi, 1) \right) \\ &= \max \left( 0, \bar{p}(\pi) - c + \delta V_{t+1}(\tilde{\pi}_{t+1}) \right) \\ &= \max \left( 0, \bar{p}(\pi) - c + \delta [\bar{p}(\pi) V_{t+1}(\varphi_{AI}(\pi)) + (1 - \bar{p}(\pi)) V_{t+1}(\varphi_{AV}(\pi))] \right).\end{aligned}$$

Then, for any element  $fl(\pi)$  of the linear functions family  $\mathcal{I}_{t+1}$  such that  $V_{t+1}(\pi) = \max_{fl \in \mathcal{I}_{t+1}} fl(\pi)$ , reasoning as previously,  $\bar{p}(\pi) fl_i(\varphi_{AI}(\pi))$  and  $(1 - \bar{p}(\pi)) fl_i(\varphi_{AV}(\pi))$  are linear in  $\pi$ , therefore  $\bar{p}(\pi) V_{t+1}(\varphi_{AI}(\pi))$  and  $(1 - \bar{p}(\pi)) V_{t+1}(\varphi_{AV}(\pi))$  are maximums of families of linear functions. Finally, using usual properties of sums or maximums, scalar multiplication of maximums and maximums of maximums, there exists a family  $\mathcal{I}_t$  of linear functions such that

$$V_t(\pi) = \max_{fl \in \mathcal{I}_t} fl(\pi).$$

---

<sup>11</sup>Because  $p_D - p_H = \max \alpha$

As the maximum of a finite family of linear functions,  $V_t(\pi)$  is continuous, piece-wise linear and convex in  $\pi$ .  $\square$

### 5.1.5 Proof of Lemma 4

*Proof.* This is a direct consequence of the martingale property, the convexity of  $V_{t+1}$  in  $\pi$  and Jensen's inequality:

$$\begin{aligned} \frac{\partial}{\partial x_t} \mathbb{E}_t[V_{t+1}(\tilde{\pi}_{t+1})|x_t] &= \bar{p}(\pi_t)V_{t+1}(\varphi_{AI}(\pi_t)) + (1 - \bar{p}(\pi_t))V_{t+1}(\varphi_{AV}(\pi_t)) - V_{t+1}(\pi_t) \\ &> V_{t+1}\left(\bar{p}(\pi_t)\varphi_{AI}(\pi_t) + (1 - \bar{p}(\pi_t))\varphi_{AV}(\pi_t) - \pi_t\right) \\ &\geq 0. \end{aligned}$$

The positivity is straightforward as

$$V_t(\pi) \geq \Omega_t(\pi, x_t = 0) = \delta V_{t+1}(\pi) \geq \dots \geq \delta^{T-t} V_T(\pi) \geq 0.$$

$\square$

### 5.1.6 Proof of Proposition 1

*Proof.* We proceed by backward induction.

**Initialization Stage  $t = T - 1$ :** The objective function verifies

$$\left. \frac{\partial \Omega_{T-1,T}}{\partial x} \right|_{\pi} = \bar{p}(\pi) - c + \delta \left[ \bar{p}(\pi)V_{T,T}(\varphi_{AI}(\pi)) + (1 - \bar{p}(\pi))V_{T,T}(\varphi_{AV}(\pi)) - V_{T,T}(\pi) \right],$$

and, since  $V_{T,T}(\pi) = 0$  below  $\pi_{T,T}^*$  and  $\varphi_{AV}(\pi) < \pi < \varphi_{AI}(\pi)$ ,

$$\left. \frac{\partial \Omega_{T-1,T}}{\partial x} \right|_{\pi_{T,T}^*} = \delta \bar{p}(\pi_{T,T}^*) V_{T,T}(\varphi_{AI}(\pi_{T,T}^*)) > 0.$$

As  $\pi_{T-1,T}^*$  is defined by  $\frac{\partial}{\partial x}\Omega_{T-1,T}\Big|_{\pi_{T-1,T}^*} = 0$  and  $\frac{\partial}{\partial x}\Omega_{T-1,T}$  is strictly increasing in  $\pi$ , it implies

$$\pi_{T-1,T}^* < \pi_{T,T}^*.$$

**At any stage  $t \in \{1, \dots, T-2\}$ :** The objective function verifies

$$\frac{\partial \Omega_{t,T}}{\partial x}\Big|_{\pi} = \bar{p}(\pi) - c + \delta \left[ \bar{p}(\pi) V_{t+1,T}(\varphi_{AI}(\pi)) + (1 - \bar{p}(\pi)) V_{t+1,T}(\varphi_{AV}(\pi)) - V_{t+1,T}(\pi) \right],$$

and, since  $V_{t+1,T}(\pi) = 0$  below  $\pi_{t+1,T}^*$  and  $\varphi_{AV}(\pi) < \pi < \varphi_{AI}(\pi)$ ,

$$\frac{\partial \Omega_{t+1,T}}{\partial x}\Big|_{\pi_{t+1,T}^*} = 0 = \bar{p}(\pi_{t+1,T}^*) - c + \delta \left[ \bar{p}(\pi_{t+1,T}^*) V_{t+1,T}(\varphi_{AI}(\pi_{t+1,T}^*)) \right],$$

thus, by subtraction,

$$\frac{\partial \Omega_{t,T}}{\partial x}\Big|_{\pi_{t+1,T}^*} = \delta \left[ \bar{p}(\pi_{t+1,T}^*) (V_{t+1,T} - V_{t+2,T})(\varphi_{AI}(\pi_{t+1,T}^*)) \right],$$

Since at prior  $\pi$  and time  $t+1$ , playing the sequence  $(x_{t+1,T}^*, x_{t+2,T}^*, \dots, x_{T,T}^*)$  is optimal and yields  $V_{t+1,T}(\pi)$ , it dominates any other sequence, in particular  $(x_{t+2,T}^*, x_{t+3,T}^*, \dots, x_{T,T}^*, 0)$ , which yields  $V_{t+2,T}(\pi)$ . Hence

$$V_{t+2,T}(\varphi_{AI}(\pi_{t+1,T}^*)) \leq V_{t+1,T}(\varphi_{AI}(\pi_{t+1,T}^*)).$$

In addition, since  $\varphi_{AI}(\pi_{t+1,T}^*) > \pi_{t+2,T}^*$ , then  $A^k(\pi_{t+1,T}^*) > \pi_{t+k+1,T}^*$  for  $k \in \{1, T-t-1\}$ , and the corresponding belief transitions happen with strictly positive probability  $\in (p_H^{T-t}, p_D^{T-t})$ . It implies that the sequence  $(x_{t+2,T}^*, x_{t+3,T}^*, \dots, x_{T,T}^*, x_{T,T}^*)$  strictly dominates  $(x_{t+2,T}^*, x_{t+3,T}^*, \dots, x_{T,T}^*, 0)$ . Therefore

$$V_{t+2,T}(\varphi_{AI}(\pi_{t+1,T}^*)) < V_{t+1,T}(\varphi_{AI}(\pi_{t+1,T}^*)).$$



In the end

$$\left. \frac{\partial \Omega_{t,T}}{\partial x} \right|_{\pi_{t+1,T}^*} > 0 = \left. \frac{\partial \Omega_{t,T}}{\partial x} \right|_{\pi_{t,T}^*}.$$

Since  $\frac{\partial \Omega_{t,T}}{\partial x}$  is a strictly increasing function of  $\pi$ ,

$$\pi_{t,T}^* < \pi_{t+1,T}^*.$$

This proves statement **i**. Statement **ii** is a direct consequence by backward induction.

Statement **iii** is a result of both these statements as

$$\pi_{1,T}^* = \pi_{2,T+1}^* > \pi_{1,T+1}^*.$$

As for the impact of  $\delta$ , when  $\delta \rightarrow 0$ , the result is straightforward as the problem comes down to the myopic one. A simple backward induction from stage  $t = T - 1$  shows that

$$\frac{\partial}{\partial \delta} V_{t,T}(\pi, \delta) > 0, \quad \frac{\partial}{\partial x \partial \delta} \Omega_{t,T} > 0,$$

implying that, since  $\pi_{t,T}^*(\delta)$  solves for  $\left. \frac{\partial}{\partial x} \Omega_{t,T} \right|_{\pi, \delta} = 0$  and  $\frac{\partial \Omega}{\partial x}$  is decreasing in  $\pi$ ,

$$\frac{\partial \pi_{t,T}^*}{\partial \delta} < 0.$$

□

### 5.1.7 Proof of Proposition 2

*Proof.* Statement **i** is a direct consequence of  $(\pi_{t,T}^*)_{T/in\mathbb{N}^*}$  being decreasing and bounded below by 0. Statement **ii** comes from the fact that at each stage the value function is the maximum of a finite family of linear functions.

As for Statement **iii**, assume that it is not true, i.e.,  $\lim_{\delta \rightarrow 1} \pi_{\infty}^*(\delta) = \underline{\pi} > 0$ . Then, for  $\pi < \underline{\pi}$ ,  $V_{\infty}(\pi) = 0$ . Consider the profits of the auditor under strategy  $(x_t = 1)_{1 \leq t \leq \bar{i}}$ .

Because of the linearity of  $u$  and the martingale property of beliefs

$$\mathbb{E}_1 \left[ \sum_{i=1}^{\bar{t}} \delta^{i-1} u(\tilde{\pi}_i, 1) \right] = \frac{1 - \delta^{\bar{t}}}{1 - \delta} u(\pi, 1).$$

With probability  $\pi$ , the auditee is initially believed to be of type  $D$ , and because of the convergence of beliefs to the true state, there exists  $r \in (0, 1)$  and a time  $t_r$  such that

$$\mathbb{P} \left( \tilde{\pi}_{t_r} > \pi^+ + \varepsilon \mid D, (x_t)_{1 \leq t \leq t_r} = 1 \right) > r.$$

Then, consider the alternative strategy  $(x_t^{al})_t$  defined as

$$x_t^{al} = \mathbb{1}_{\{t \leq t_r\}} + \mathbb{1}_{\{t > t_r\}} \mathbb{1}_{\{\pi_{t_r} > \pi^+ + \varepsilon\}},$$

with corresponding expected profits higher than

$$\frac{1 - \delta^{t_r}}{1 - \delta} \underbrace{u(\pi, 1)}_{<0} + \pi r \frac{\delta^{t_r}}{1 - \delta} \underbrace{u(\pi^+ + \varepsilon, 1)}_{>0}.$$

When  $\delta$  tends to one, the left hand side term is finite, while the right hand side term goes to infinity. This last result contradicts the fact that  $V_\infty(\pi) = 0$ .  $\square$

### 5.1.8 Proof of Proposition 3

*Proof.* First,  $\tau^*$  is indeed a stopping time as it is defined by the process  $\langle \pi_t \rangle$  that is by definition adapted to the filtration  $(\mathcal{F}_t)_{t \in \{1, \dots, T\}}$ . In addition,

$$\forall t < \tau^*, \quad \tilde{\pi}_t \geq \pi_t^*, \tag{7}$$

$$\forall t \geq \tau^*, \quad \tilde{\pi}_t = \tilde{\pi}_{\tau^*} < \pi_{\tau^*}^* < \pi_t^*. \tag{8}$$

Equation (7) is true because otherwise, by definition of  $\tau^*$  as a minimum of times at which the previous inequality is verified,  $\tau^* \leq t$ . In equation (8), the first equality relies on the fact that once audit stops, i.e., at  $\tau^*$ , priors remain constant, the following inequality stems from the definition of  $\tau^*$ , and finally the last inequality stems from the

fact that the sequence  $(\pi_t^*)_t$  is increasing, from Proposition 1.

Then rewriting the value function yields

$$\begin{aligned}
V_{1,T}(\pi_1, \delta) &= \mathbb{E} \left[ \sum_{j=1}^T \delta^{j-1} u(\tilde{\pi}_j, \mathbb{1}\{\tilde{\pi}_j \geq \pi_j^*\}) \right] \\
&= \mathbb{E} \left[ \sum_{j=1}^T \delta^{j-1} u(\tilde{\pi}_j, 1) \mathbb{1}\{\tilde{\pi}_j \geq \pi_j^*\} \right] \\
&= \mathbb{E} \left[ \sum_{j=1}^{\tau^*} \delta^{j-1} u(\tilde{\pi}_j, 1) \right], \quad \text{from Eq. (7) and (8)} \\
&= W_{1,\tau^*}(\pi_1, \delta).
\end{aligned}$$

Finally,  $\tau^*$  maximizes  $W$  because otherwise, there would be another stopping time based auditing strategy that would maximize  $W$  and yield auditing proceeds strictly higher than  $V_{1,T}(\pi_1, \delta)$ , which would be contradictory.  $\square$

### 5.1.9 Proof of Proposition 4

*Proof.* For a given  $\pi_1$  and the corresponding optimal stopping rule  $\tau^*$

$$\begin{aligned}
W_{1,\tau^*}(\pi_1, \delta) &= \mathbb{E} \left[ \sum_{j=1}^{\tau^*-1} \delta^{j-1} u(\tilde{\pi}_j, 1) \right] \\
&= \sum_{\omega} \mathbb{P}(\omega) \sum_{j=1}^{\tau^*(\omega)-1} \delta^{j-1} u(\pi_j(\omega), 1).
\end{aligned}$$

Using the same reasoning as in the proof of Lemma 3,  $u(\pi_j(\omega), 1)$  is a linear function of the initial belief  $\pi_1$ ,  $u(\pi_j(\omega), 1) = \alpha_j(\omega)\pi_1 + \beta_j(\omega)$ . Thus

$$\frac{\partial W_{1,\tau^*}}{\partial \pi_1} = \sum_{\omega} \mathbb{P}(\omega) \sum_{j=1}^{\tau^*(\omega)-1} \delta^{j-1} \alpha_j(\omega).$$

This is the slope of the value function at  $\pi_1$ . In addition, notice that for  $\pi'_1 < \pi_1$  such that they are not in the same  $\Psi_m$

$$\tau^*(\pi'_1, \omega) \leq \tau^*(\pi_1, \omega)$$

and there exists at least one trajectory  $\omega'$  such that the inequality is strict. Otherwise, by definition of the intervals  $\Psi$ ,  $\pi'_1$  and  $\pi_1$  would be in the same  $\Psi_m$ . Then

$$\left. \frac{\partial W_{1,\tau^*}}{\partial \pi_1} \right|_{\pi'_1} < \left. \frac{\partial W_{1,\tau^*}}{\partial \pi_1} \right|_{\pi_1}.$$

This last inequality shows that the linear functions slopes are increasing in  $\pi$ , as shown in Figure 5e. □

### 5.1.10 Proof of Proposition 6

*Proof.* We first show the direct implication of the equivalence. Denoting  $\tau^*(\pi)$  and  $\tau^*(\pi')$  the optimal stopping times for each prior

$$\begin{aligned} \Lambda(\pi, \delta) > \Lambda(\pi', \delta) &\Rightarrow \frac{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi)} \delta^{j-1} u(\tilde{\pi}_j, 1)]}{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi)} \delta^{j-1}]} > \frac{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi')} \delta^{j-1} u(\tilde{\pi}'_j, 1)]}{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi')} \delta^{j-1}]} \\ &\Rightarrow \frac{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi)} \delta^{j-1} u(\tilde{\pi}_j, 1)]}{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi)} \delta^{j-1}]} > \frac{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi)} \delta^{j-1} u(\tilde{\pi}'_j, 1)]}{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi)} \delta^{j-1}]} \\ &\Rightarrow \mathbb{E}\left[\sum_{j=1}^{\tau^*(\pi)} \delta^{j-1} u(\tilde{\pi}_j, 1)\right] > \mathbb{E}\left[\sum_{j=1}^{\tau^*(\pi)} \delta^{j-1} u(\tilde{\pi}'_j, 1)\right] \\ &\Rightarrow \pi > \pi'. \end{aligned}$$

Then, for the indirect implication

$$\begin{aligned} \pi > \pi' &\Rightarrow \forall \omega, u(\pi_j(\omega), 1) > u(\pi'_j(\omega), 1) \\ &\Rightarrow \frac{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi')} \delta^{j-1} u(\tilde{\pi}_j, 1)]}{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi')} \delta^{j-1}]} > \frac{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi')} \delta^{j-1} u(\tilde{\pi}'_j, 1)]}{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi')} \delta^{j-1}]} = \Lambda(\pi', \delta) \\ &\Rightarrow \Lambda(\pi, \delta) \geq \frac{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi')} \delta^{j-1} u(\tilde{\pi}_j, 1)]}{\mathbb{E}[\sum_{j=1}^{\tau^*(\pi')} \delta^{j-1}]} > \Lambda(\pi', \delta). \end{aligned}$$

□

### 5.1.11 Proof of Proposition 7

*Proof.* Because of the  $(1 - \pi_1)$  component, it is sufficient to show that  $\mathbb{P}(\tau^* = \infty | D, \pi_1) < 1$ . This is straightforward as there exists some  $l \in \mathbb{N}$  such that  $\varphi_{AV}^l(\pi_1) < \pi_\infty^* \leq \varphi_{AV}^{l-1}(\pi_1)$ .<sup>12</sup> This is the prior obtained from auditing as long as  $\pi_1 \geq \pi_\infty^*$  and observing valid claims every time, which happens with probability  $(1 - p_D)^l$ . Therefore,  $\mathbb{P}(\tau^* = \infty | D, \pi_1) \leq 1 - (1 - p_D)^l$ .  $\square$

### 5.1.12 Proof of Proposition 8

*Proof.* Let  $\langle I_t \rangle$  be the Bernoulli process representing the fraud outcome at period. Then  $\mathbb{P}_H(I_t = 1) = p_H$  and  $\mathbb{P}_D(I_t = 1) = p_D$ . The updated prior under a given state of the world  $H$  or  $D$  can be written as

$$\begin{aligned}\tilde{\pi}_{t+1} &= I_{t+1}\varphi_{AI}(\tilde{\pi}_t) + (1 - I_{t+1})\varphi_{AV}(\tilde{\pi}_t) \\ &= \tilde{\pi}_t + \left(I_{t+1} - \bar{p}(\tilde{\pi}_t)\right) \left[\varphi_{AI}(\tilde{\pi}_t) - \varphi_{AV}(\tilde{\pi}_t)\right]\end{aligned}$$

In particular

$$\begin{aligned}\mathbb{E}_D[\tilde{\pi}_{t+1} | \mathcal{F}_t] &= \tilde{\pi}_t + \left(p_D - \bar{p}(\tilde{\pi}_t)\right) \left[\varphi_{AI}(\tilde{\pi}_t) - \varphi_{AV}(\tilde{\pi}_t)\right] \\ &= \tilde{\pi}_t + \left(1 - \tilde{\pi}_t\right) \left(p_D - p_H\right) \left[\varphi_{AI}(\tilde{\pi}_t) - \varphi_{AV}(\tilde{\pi}_t)\right]\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_H[\tilde{\pi}_{t+1} | \mathcal{F}_t] &= \tilde{\pi}_t + \left(p_H - \bar{p}(\tilde{\pi}_t)\right) \left[\varphi_{AI}(\tilde{\pi}_t) - \varphi_{AV}(\tilde{\pi}_t)\right] \\ &= \tilde{\pi}_t - \tilde{\pi}_t \left(p_D - p_H\right) \left[\varphi_{AI}(\tilde{\pi}_t) - \varphi_{AV}(\tilde{\pi}_t)\right]\end{aligned}$$

---

<sup>12</sup>If  $\pi_1 < \pi_\infty^*$ , then  $l = 0$ .

One decomposition into a martingale and a predictable process is given, under state of the world  $D$ , by

$$\tilde{\pi}_t = M_t^D + P_t^D$$

$M_t^D$  is a martingale and  $P_t^D$  is predictable and increasing

$$P_t^D = (p_D - p_H) \sum_{i=1}^{t-1} (1 - \tilde{\pi}_i) \left[ \varphi_{AI}(\tilde{\pi}_i) - \varphi_{AV}(\tilde{\pi}_i) \right]$$

$$M_1^D = \tilde{\pi}_1$$

$$P_1^D = 0$$

We can verify that  $M_t^D$  is indeed a Martingale by induction (it is verified at  $t = 1$ )

$$\begin{aligned} \mathbb{E}_D[M_{t+1}^D | \mathcal{F}_t] &= \mathbb{E}_D[\tilde{\pi}_{t+1} - P_{t+1}^D | \mathcal{F}_t] \\ &= \mathbb{E}_D[\tilde{\pi}_{t+1} | \mathcal{F}_t] - P_{t+1}^D \\ &= \mathbb{E}_D[I_{t+1} \varphi_{AI}(\tilde{\pi}_t) + (1 - I_{t+1}) \varphi_{AV}(\tilde{\pi}_t) | \mathcal{F}_t] - P_{t+1}^D \\ &= \tilde{\pi}_t + \underbrace{\left( (1 - \tilde{\pi}_t) (p_D - p_H) \left[ \varphi_{AI}(\tilde{\pi}_t) - \varphi_{AV}(\tilde{\pi}_t) \right] \right)}_{=-P_t^D} - P_{t+1}^D \\ &= \tilde{\pi}_t - P_t^D \\ &= M_t^D \end{aligned}$$

Finally, because of the unicity of the Doob-Meyer decomposition, the above decomposition is the only one.

The same reasoning yields the decomposition under state of the world  $H$

$$\tilde{\pi}_t = M_t^H + P_t^H$$

$M_t^H$  is a martingale and  $P_t^H$  is predictable and increasing

$$P_t^H = (p_D - p_H) \sum_{i=1}^{t-1} (1 - \tilde{\pi}_i) \left[ \varphi_{AI}(\tilde{\pi}_i) - \varphi_{AV}(\tilde{\pi}_i) \right]$$

$$M_1^H = \tilde{\pi}_1$$

$$P_1^H = 0.$$

Positivity of  $P_t^D$  and negativity of  $P_t^H$  are direct consequences of inequalities (2).  $\square$

### 5.1.13 Proof of Proposition 9

*Proof.* From the formulas for the updating functions

$$\varphi_{AI} = \frac{p_D \pi}{\bar{p}(\pi)} \quad \text{and} \quad \varphi_{AV} = \frac{(1 - p_D) \pi}{(1 - \bar{p}(\pi))},$$

we derive the derivatives with respect to  $p_D$  and  $p_H$

$$\begin{aligned} \frac{\partial \varphi_{AI}}{\partial p_D} &= \frac{p_H \pi (1 - \pi)}{\bar{p}(\pi)^2} > 0 & \text{and} & \quad \frac{\partial \varphi_{AV}}{\partial p_D} = \frac{-(1 - p_H) \pi (1 - \pi)}{(1 - \bar{p}(\pi))^2} < 0, \\ \frac{\partial \varphi_{AI}}{\partial p_H} &= \frac{-p_D \pi (1 - \pi)}{\bar{p}(\pi)^2} < 0 & \text{and} & \quad \frac{\partial \varphi_{AV}}{\partial p_H} = \frac{(1 - p_D) \pi (1 - \pi)}{(1 - \bar{p}(\pi))^2} > 0. \end{aligned}$$

Finally, taking the derivatives of  $P_t^H$  and  $P_t^D$  and using the above derivatives yields the final result.  $\square$