Learning with endogenous information*

Subir Bose† University of Leicester
Tassos Magdalinos ‡ University of Southampton
Miltiadis Makris § University of Kent

May 2021

Abstract

We study the problem of learning when the decision maker (DM) is uncertain about the impact of current decisions on *future* data. In an environment where optimal experimentation is monotonically related to beliefs and the DM finds it optimal to experiment when past experimentation has been sufficiently low for given beliefs, experimentation takes place if past experimentation is below a history-dependent threshold. Since data are dependent and heterogeneously distributed random variables with sample moments that do not necessarily concentrate around their time-varying population counterparts, the usual techniques for establishing asymptotic learning based on laws of large numbers cannot be used. We overcome this difficulty by establishing a novel *a.s.* positive asymptotic lower bound for the sample mean of the data. We can thus establish that complete learning occurs asymptotically, with convergence of posterior beliefs taking place at an exponential rate.

Keywords: Bayesian learning, non-stationary dependent process, restless bandit, learning-by-experimentation

JEL: C11, D01, D80, D83

*We thank Daniel Fershtman, Emmanuel Guerre, Antonella Ianni, Alessandro Pavan, Katerina Petrova and Peter Phillips for helpful comments.

†School of Business, University of Leicester, Leicester LE1 7RH, sb345(at)le.ac.uk.

‡Department of Economics, University of Southampton, SO17 1BJ, A.Magdalinos(at)gmail.com.

§School of Economics, University of Kent, CT2 7FS, mmakris.econ(at)gmail.com.

1 Introduction

Economic agents often face sequential decision problems. For example, a firm may face innovation opportunities due to consecutive technology improvements, a research team often faces sequential opportunities to pursue new projects on related research themes, medical research breakthroughs create opportunities over time for the development of (or improvement of existing) cures and treatments of related medical conditions, an individual may face "urges" over time to consume fatty foods. However, when making their decisions, economic agents may have imperfect information about the impact of their actions on future information and, thereby, on future trade-offs and decisions. Some examples are:

- When contemplating whether to adopt a new technology or not, a firm may not know the impact a current innovation will have on the arrival of *profitable* opportunities to innovate in the future.
- When deciding whether to pursue or not a novel idea/project, a researcher may have imperfect information about the impact of pursuing the novel research on the arrival of future ideas/projects with *high* potential.
- Following a medical breakthrough, the development of a medical cure/treatment often takes place under uncertainty about how such a development can influence the emergence of future *profitable* breakthroughs.
- Faced with a technological opportunity to improve its products, a firm may have imperfect information how developing these improvements will influence the emergence of future *profitable* opportunities for product development.
- When contemplating to consume or not a certain good such as alcohol and fatty foods, a consumer may not know whether such consumption can make future *urges* to consume more likely.

Motivated by these important economic problems, we study optimal learning when (a) data that may carry pertinent information arrives over time *regardless* of past decisions, and (b) there is uncertainty about how current choices influence the process that generates future data/information. With reference to the above settings, examples of such data are: the arrival of innovation opportunities, the emergence of research ideas/projects, the arrival of medical or technological breakthroughs, and the occurrence of urges to consume.

The learning problem in hand is not standard because the data our decision maker (DM) faces sequentially are dependent and heterogeneously distributed random variables, and hence their mean does not necessarily concentrate around a number (as the population means are time-varying). As a result, we cannot appeal to a law of large numbers, and so the usual techniques for establishing asymptotic learning cannot be used. We overcome this difficulty by establishing a novel *a.s.* positive asymptotic lower bound for the data sample mean, which

enables us to show that our Bayesian agent learns the truth asymptotically, with the posterior probability converging *a.s.* at an exponential rate.

Our asymptotic bound approach is more generally applicable to problems that study the asymptotics of a sample mean of a function of a nonstationary process when the function is either convex in $\mathbb{R}$ and strictly increasing in $\mathbb{R}_+$ or concave in $\mathbb{R}$ and strictly decreasing in $\mathbb{R}_+$. Many Bayesian problems, including our Bayesian learning problem, involve ratios of conditional probabilities/likelihoods and posterior distributions whose asymptotic analysis naturally reduces to the above framework.

In the next section we discuss related literature. In Section 3 we develop the model, and in Section 4 we discuss optimal experimentation. Section 5 is devoted to learning, and Section 6 concludes.

2 Related Literature

The work in this paper is related to the statistical literature on sequential analysis (De Groot (1970)) and "learning-by-experimentation". For some of the early contributions in the economics literature on learning-by-experimentation see Aghion et. al. (1991), Easley et. al. (1988), Khilstrom et. al. (1984). An important general lesson from this literature is that while a rational decision-maker does not leave profitable learning opportunities unexploited in the limit, the cost of experimentation may force the decision-maker to stop experimenting and, hence, not learn the truth even in infinite horizon models. The stochastic process that generates the data in our model differs from the process assumed in that literature because, in our paper, the information that the signals carry depend on the extent of past experimentation.

The learning-by-experimentation literature is naturally closely related to the multi-arm bandit (MAB) problem (Robbins (1952); see also Gittins (1989), Berry and Fristedt (1985), Mahajan and Teneketzis (2007)). The main focus of this literature is studying how a rational Bayesian decision-maker (DM) balances the trade-off between exploration and exploitation. The primary interest of the multi-arm bandit literature is on characterizing the optimal decision rule.

The process that generates the observed data and beliefs in this paper makes our setting a "restless" (RB) bandit environment, where the "state of an arm" (beliefs in our case) changes even if the arm is not processed.[1] In such problems, it is known that optimal policies are in general not indexable (see Whittle (1998), Niño-Mora (2000), references therein, and the excellent survey of Mahajan and Teneketzis (2007)). The RB (as most of the MAB) literature has mainly considered models with finite state space possibly because RB problems are known to become very complicated otherwise. In contrast, the learning model when viewed as an

[1] Brezzi and Lai (2000) analyse a general MAB-type learning-by-experimentation model, where the process that generates the data is independent of the extent of past experimentation and beliefs that relate to an arm do not change if the arm is not processed. In that work, the optimal strategy is given by the Gittins index and the discounting of future rewards is shown to lead to incomplete learning.

exploration-vs-exploitation model would involve an infinite state space since the posterior beliefs would be a state variable. Fryer and Harms (2018) study an RB problem where, as in our setup, the information that the signals carry depend on past levels of experimentation. Nevertheless, in contrast to our model, an informative signal is not produced when their agent does not experiment. This enables them to show that in their setup a stopping strategy is optimal, as in the regular bandit problems, and that the optimal strategy is indexable and similar to the Gittins index.[2] Safronov (2017) also studies experimentation in an RB multi-arm bandit framework. There, the DM not only receives information about uncertain payoffs by using an arm, but also becomes experienced by using it. Experience provides an additional payoff to the agent. In that paper as well, the optimal policy turns out to be indexable. Crucially, the arrival of signals about uncertain payoffs is modelled as an independent process from the arrival of experience. In our model, instead, the arrival of signals does depend on the extent of past experimentation.

In the economics literature, Rothschild (1974) is the first to use the framework of the bandit problem to study a monopolist who tries to maximize profit in a dynamic setup when faced with a market with unknown demand. Subsequently, there have been many other important applications of the multi-arm bandit model - for example on search and matching, choice of research projects, R&D investments to name a few. See Bergmann and Valimaki (2006) for a survey of applications of the bandit problem in the economics literature. In all this work, the focus is on single-agent problems, but, in contrast to our case, a signal is generated only if experimentation has taken place in the immediate past, and the information the signal carries does not depend on the extent of past expeerimentation.

Much of the interest of the more recent learning models in the economics literature has gravitated towards models with multiple agents - models of "strategic experimentation". In that literature, the main focus is on the issue of agents/players learning from each other and the interesting informational externalities that these create. See Bolton and Harris (1999) for an early contribution. See also Keller et al (2005), Keller and Rady (2010) Klein and Rady (2011), Rosenberg, Solan Veille (2007)).[3] Horner and Skrzypacz (2017) is a superb survey of the literature on strategic experimentation. This literature has considered several interesting bandit environments (Poisson and exponential bandits for example, and both discrete and continuous time models). As in our paper, the focus has mainly been on a bandit setup with a binary action *per time instance*, and one safe and a single risky arm of two types.[4] While the framework studied in this literature is not explicitly that of a restless bandit, in these models,

[2] Urgun (2019) also show that the optimal strategy is indexable in an RB-type economic problem with finite state space where, however, learning is not a concern.

[3] There is also the related literature on social learning and informational cascades (Banerjee (1992), Bikhchandani etal (1992), Chamley (2004) and the references cited therein; see also Smith and Sorensen (2011)), Acemoglou et. al. (2011), Hann-Caruthers et. al. (2018) and references therein. The main interest of this literature is on what agents learn from each other's actions, when agents possess a privates signal about (typically) a binary state of nature, whereas we focus on the single-agent scenario and what the agent learns over time from her own signal.

[4] For a model where new arms can endogenously be generated during experimentation, see Fershtman and Pavan (2019).

the projects of the different agents are highly correlated (often, these are in fact common value models) so that even when an agent is not experimenting but is using the safe arm, the agent can learn about the value of their risky arm as long as there is some other agent experimenting. In other words, there is a *flavour* of our restless bandit framework in these models. These models, however, do not involve an environment like the one we study, where the signals may be generated by a history-dependent and non-stationary stochastic process. It would be interesting to extend the work in this paper to the setting of multiple agents and study the implications on social learning with the current Markovian underlying stochastic process.

The closest work to our paper is Bose and Makris (2018), which studies a restless bandit problem with a risky and a safe arm, where the informative signal can emerge with a positive probability in any period. The probability in question, depending on the state of nature, can be either constant or an increasing function of the stock of past experimentation, which is a special case of our current stochastic process framework. The main focus of interest in that paper is on studying the properties of the optimal experimentation rule. In fact, the section "Asymptotic Learning" in that paper shows only, by means of two simple build-for-purpose examples, the *possibility* of *both* complete and incomplete learning (depending on the example). The example where incomplete learning is the outcome asymptotically does not fit the framework here, while the example where complete learning takes place asymptotically is a special case of the current framework and such that the techniques developed here are not used.

3 Model

To fix ideas we describe the model in terms of a decision maker (DM) who contemplates whether to adopt a new technology or not. Appropriate changes in the narrative would make the following framework applicable to the other decision problems discussed in the Introduction.

Opportunities to innovate emerge sequentially *regardless* of the innovation history. A technology that becomes available and is not adopted is assumed for simplicity to no longer be available in the future.[5] A "period" $t \in \{1, 2, ...\}$ refers to the number of instances an opportunity to innovate has arisen up to (and including) the current instance.

Let $d_t = 0$ denote the action "adopt/innovate" in period t and $d_t = 1$ denote the action "do not adopt/innovate" in period t. The flow payoff from not adopting the new technology is normalized to zero. The reward/profit from innovating in period t is an independent draw from a time-invariant distribution over $\mathbb{R}$ which depends on the type $x_t \in \{0, 1\}$ of the new technology and a statistic h_t which is related to the innovation up to and including period t. If the new technology is "profitable" (denoted by $x_t = 0$) then the mean reward from innovating

[5] One can think of technologies that are not adopted to depreciate at an infinite rate or that the entrepreneurs responsible for the development of these technologies to go bankrupt or exit the market to avoid excessive losses.

in period t is equal to $b(h_t) > 0$, whereas if the the new technology is "loss-making" (denoted by $x_t = 1$) the mean reward from innovating in period t is equal to $b(h_t) - c(h_t) < 0$, where $b(.)$ and $c(.)$ are continuously differentiable functions.

We postulate that

$$h_{t+1} = g(h_t, d_t), \quad t \geq 1 \tag{3.1}$$

for some pre-determined $h_1 \in \mathbb{R}$ *a.s.*, where g is a differentiable function satisfying the following:

$$g(0,0) = 0, \;\; g(0,1) > 0, \;\; \inf_{z \geq 0} \frac{\partial g(z,1)}{\partial z} \geq 0, \;\; \inf_{z \geq 0} \frac{\partial g(z,0)}{\partial z} > 0 \tag{3.2}$$

Note that (3.2) implies that $g(\cdot, 1)$ is increasing in $\mathbb{R}_+$ and that $g(\cdot, 0)$ is strictly increasing in $\mathbb{R}_+$. We emphasize here that, conditional on its Markovian structure, the law-of-motion (3.1) is quite general as it allows for non-linearities in $g(\cdot, y)$. We also assume that

$$\sup_{z \geq 0} \frac{\partial g(z,1)}{\partial z} < 1, \tag{3.3}$$

which is a stability condition implying that there is $H > 0$ such that

$$H = g(H, 1).$$

The probability the period-t innovation is profitable is given by

$$\mathbb{P}(x_t = 1 | h_t, \theta) = p + \theta h_t, \quad t \geq 1 \tag{3.4}$$

where $p \in (0,1)$ and $\theta \in \{\underline{\theta}, \overline{\theta}\}$ with $\overline{\theta} > \underline{\theta} \geq 0$ such that

$$\underline{\theta} h_1 + p > 0, \quad \overline{\theta} H = p < 1, \tag{3.5}$$

which ensure that either profitability type is possible.[6]

The law-of-motion (3.1) for the statistic of past innovation, as well as the stochastic process $\mathbb{P}(x_t | h_t, \theta)$ are known by our agent. Crucially, our DM observes the profitability type of the current innovation opportunity before deciding whether to adopt it or not.

Observe, however, that knowing the profitability of the period-t innovation, x_t. does not reveal the profitability of the innovations that will become available in the future. In fact, the profitability type of an innovation is a (Bernoulli) random variable whose distribution depends on the extend of past innovations and the variable θ, which captures the impact of past decisions on the distribution of the reward from innovating. Notice that the case of "low impact", $\theta = \underline{\theta}$, is allowed to represent an environment where past decisions have no impact on current profits from innovation, i.e. $\underline{\theta} = 0$.

The agent does not know the true state of nature θ and would like to discover it. Denote

[6] In Bose and Makris (1998) it is assumed that $\underline{\theta} = 0 and \overline{\theta} = 1$.

with μ_0 the agent's "prior" belief that $\theta = \overline{\theta}$ (i.e. the belief before the realisation of the first sample/signal). Assume that $0 < \mu_0 < 1$.

The profitability of the new technology in period t (i.e. x_t) can serve, in effect, as a signal about θ. We assume that our agent is Bayesian, and so she uses the observed history $x^t = (x_1, ..., x_t)$ for inference purposes. Denote with ξ_t the (posterior) belief of the agent that $\theta = \overline{\theta}$, given that $x_t = 1$ and the statistic of past innovations h_t. This is given by Bayes rule as:

$$m(1, \mu_{t-1}, h_t) \equiv \frac{\mu_{t-1}[p + \overline{\theta} h_t]}{\mu_{t-1}[p + \overline{\theta} h_t] + (1 - \mu_{t-1})[p + \underline{\theta} h_t]}. \tag{3.6}$$

where μ_{t-1} is the belief of the agent at the end of period $t-1$ that $\theta = \overline{\theta}$ (i.e. the period-$\{t\}$ "prior"). Similarly, denote with $\hat{\xi}_t$ the (posterior) belief of the agent that $\theta = \overline{\theta}$, given that $x_t = 0$ and the statistic of past innovations h_t. This is given by Bayes rule as:

$$m(0, \mu_{t-1}, h_t) \equiv \frac{\mu_{t-1}[1 - p - \overline{\theta} h_t]}{\mu_{t-1}[1 - p - \overline{\theta} h_t] + (1 - \mu_{t-1})[1 - p - \underline{\theta} h_t]}. \tag{3.7}$$

Observe thus that

$$\mu_t = m(x_t, \mu_{t-1}, h_t) \equiv x_t m(1, \mu_{t-1}, h_t) + (1 - x_t) m(0, \mu_{t-1}, h_t) \tag{3.8}$$

Note that μ_t is a random variable, which depends on the realisation of x^t. Given these beliefs the subjective probability at the end of period t that the period-$\{t+1\}$ new technology is profitable is equal to

$$\pi_{t+1} = \pi(\mu_t, h_{t+1}) \equiv p + [\underline{\theta} + \mu_t \Delta\theta] h_{t+1} \tag{3.9}$$

where $\Delta\theta \equiv \overline{\theta} - \underline{\theta}$.

Before we state the agent's decision problem, it is worth clarifying the main differences between our setting and the literature we mentioned in Section 2 on (one-player or strategic) learning-by-experimentation with bandits that are not restless. In terms of our narrative, in the received literature, all innovation opportunities have the same profitability which is drawn from a given distribution before the first opportunity arrives, and agents have imperfect information on the profitability of new technologies. Here, instead, the profitability of new technologies may change over time as it is drawn anew each period, and our agent observes the realised profitability before making her decision. Furthermore, in the existing literature, the distribution of the reward from experimenting depends on the constant and unknown profitability and/or on the signal that is generated as a result of the said experimentation, while in our model it depends on the signal that has been generated as a result of past experimentation.[7]. Finally, the distribution of the signal (that depends on the unknown state-of-the-world and is generated as a result of the current action) is independent of past experimentation, in contrast to the distribution of the signal in our model.[8]

[7] In our model, the distribution of the reward from experimentation is also allowed to depend on the extent of past experimentation

[8] in Aghion et al. (1991), the unknown state-of-the-world affects directly both the reward from experi-

We can now state the decision problem of our agent. The DM chooses at each period t whether to adopt or not the available technology in order to maximize her intertemporal payoff:

$$\mathbb{E}\sum_{t=1}^{\infty}\delta^{t-1}[1-d_t]\{b(h_t)-x_t c(h_t)\}$$

where $\delta \in (0,1)$ is the discount rate, subject to (3.1) and (3.6)-(3.9).

We note that the only connection betweens periods occurs through the "stock" of experimentation as described by the statistic h, and the (posterior) beliefs μ. The current action affects directly the current mean reward and indirectly the value of the problem at the end of the current period through its effect on the evolution of the stock of experimentation and next period's beliefs. Notice also that the mean reward depends on the stock of experimentation and the profitability of the current innovation opportunity. Therefore, the decision problem can be formulated as a Markovian stochastic dynamic programming problem with the "state" in each period t given by (x_t, h_t, μ_t).

Define

$$V(x,h,\mu) = \max_{d\in\{0,1\}}\Big\{[1-d_t][b(h_t)-x_t c(h_t)]+ \tag{3.10}$$

$$\delta\Big[\pi(\mu,g(h,d))V(1,g(h,d),m(1,\mu,g(h,d)))+[1-\pi(\mu,g(h,d))]V(0,g(h,d),m(0,\mu,g(h,d)))\Big]\Big\}$$

Existence of the function V follows from standard results.[9] $V(x,h,\mu)$ is the value function of the DM's problem given the state (x,h,μ) at the beginning of a period.

In the next section, we characterise optimal experimentation. We restrict attention to environments that satisfy the following assumptions:

Assumption A : $b(h)$ and $b(h)-c(h)$ are nonincreasing in h.

This assumption states that the mean reward from innovating is nondecreasing in the extent of past innovation regardless of its profitability. It captures an environment with learning-by-doing in innovation, i.e. where more past innovations make the agent better in terms of taking advantage of and integrating in existing processes new technologies.

Assumption B : $\pi(m(0,\mu,h),g(h,0))$ is increasing in h.

This assumption states that the belief of the DM that next-period's innovation opportunity is loss-making, given that the current innovation opportunity has been profitable and the DM has chosen to adopt the new technology, in increasing in the stock of past experimentation. It ensures that an increase in the stock of past experimentation increases the belief of the DM that next-period's innovation opportunity is loss-making regardless of the profitability of the current new technology. This, in turn, guarantees that the DM always chooses $d_t = 0$ (to

mentation and the process that generates the data, while here that reward is not affected directly from the unknown state-of-the-world.

[9] Details are available by the authors upon request.

innovate) when the innovation opportunity is profitable. As we show in the Online Appendix A, this assumption is satisfied, for example, if $g(h,d) = q[\gamma h + d]$, with $\gamma > 0, 0 < q < \frac{1-p}{1+\gamma(1-p)}$ (and hence $q\gamma < 1$ and $H \equiv \frac{q}{1-q\gamma} < 1-p$), such that $H \leq (1-p)\frac{1-\sqrt{1-\frac{\underline{\theta}}{\overline{\theta}}}}{\underline{\theta}}$. The latter is a stronger condition than $H < (1-p)$ if $\underline{\theta}$ is close to zero, while it is a weaker condition than $H < (1-p)$ if $\underline{\theta}$ is close to $\overline{\theta}$ and $\overline{\theta} \leq 1$.

For the rest of the paper, we also make the tie-breaking assumption that an indifferent agent will choose $d_t = 0$. It can easily be checked that no qualitative result is affected if one were to break such a tie in the other way.

4 Optimal Experimentation

Let

$$D(h_t, \mu_t) \equiv V(0, g(h_t,1), m(0, \mu_t, g(h_t,1))) - V(1, g(h_t,1), m(1, \mu_t, g(h_t,1)))$$

That is, for any h_t and μ_t, and conditionally on the agent having chosen not to innovate in period t, the term $D(h_t, \mu_t)$ measures the net gain – evaluated in $t+1$ utils – from having a profitable innovation opportunity in period $t+1$. Observe that for a myopic or infinitely impatient agent we have that $d_t = x_t$ for any beliefs and history (of innovations), and thereby $D(h_t, \mu_t) = b(h_t) > 0$.

Let us define also:

$$\Gamma(h_t, \mu_t, x) \equiv V(x, g(h_t,0), m(x, \mu_t, g(h_t,0))) - V(x, g(h_t,1), m(x, \mu_t, g(h_t,1))$$

and

$$\Psi(h_t, \mu_t) \equiv \pi(\mu_t, g(h_t,0))\Gamma(h_t, \mu_t, 1) + [1 - \pi(\mu_t, g(h_t,0))]\Gamma(h_t, \mu_t, 0)$$

$\Gamma(h_t, \mu_t, x)$ represents the net gain - evaluated at $t+1$ - from adopting the new technology at period t, conditionally on next period's profitability-state being x. The interpretation of the term $\Psi(h_t, \mu_t)$ is more subtle. The profitability-state in period $t+1$ is not known in period t. Innovating at period t (as opposed to not innovating) has two effects: it changes the likelihood of a loss-making innovation opportunity next period, and it changes the next-period's value function conditional on next period's prrofittability-state. The latter is given by $\Gamma(h_t, \mu_t, x)$. Measuring the expected gain from innovating at period t, while maintaining the likelihood that the period$-\{t+1\}$ new technology is loss-making at the level that corresponds to innovation taking place in period t, gives rise to $\Psi(h_t, \mu_t)$. To understand these better, observe that for a myopic or infinitely impatient agent we have $\Gamma(h_t, \mu_t, 1) = \Gamma(h_t, \mu_t, 0) = \Psi(h_t, \mu_t) = 0$.

Given these definitions, we can now discuss optimal decisions and proceed in proving our first theorem. To this end, note (after a straightforward algebraic manipulation) that the net benefit from innovating at period t is:

$$NB_t \equiv \delta\{[\underline{\theta} + \mu_t \Delta\theta]\,[g(h_t,1) - g(h_t,0)]\,D(h_t, \mu_t) + \Psi(h_t, \mu_t)\} - [x_t c(h_t) - b(h_t)] \qquad (4.1)$$

This net benefit is bounded and consists of two parts. The term inside the second square brackets represents the "short-run cost" from innovation. This is negative if the new technology in hand is profitable, and positive if the the new technology in hand is loss-making. The "long-run benefit" from innovation evaluated in period-$\{t+1\}$ utils is captured by the term inside the curly brackets. This long-run benefit is twofold. First, innovation reduces the perceived frequency of difficult tasks in the future. This is captured by the first term inside the curly brackets. Second, innovation will influence future beliefs and histories, and thereby future decisions. This is captured by the second term inside the curly brackets. The agent will innovate at period t if and only if the long-run benefit from doing so is at least as high as the short-run cost, i.e. $NB_t \geq 0$.

The above decision rule is not very helpful in understanding optimal experimmentation, unless one signs NB_t for any given period and state. This is the main challenge in proving (in Appendix A) the following result:

Theorem 1. *Suppose that, conditional on x_t, beliefs in period-t are μ_t. Under $\mu_t < 1$ and Assumptions A and B, we have, for any $t \geq 1$, that $\Psi(h_t, \mu_t) \geq 0$ and $D(h_t, \mu_t) \geq b(h_t)$. Moreover, $d_t = 1$ is optimal if and only if $x_t = 1$ and*

$$c(h_t) - b(h_t) > \delta\{[\underline{\theta} + \mu_t \Delta\theta]\,[g(h_t, 1) - g(h_t, 0)]\, D(h_t, \mu_t) + \Psi(h_t, \mu_t)\} \tag{4.2}$$

When $\mu_t = 1$, Assumption B is not needed for these results.

Proof. Define first the following properties:

(A) $V(x_{t+1}, h_{t+1}, \mu_{t+1})$ is non-increasing in x_{t+1}, h_{t+1} and μ_{t+1},
(B) $V(0, h_{t+1}, m(0, \mu_t, h_{t+1}))$ is non-increasing in h_{t+1}.
(C) $V(0, h_{t+1}, m(0, \mu_t, h_{t+1})) - V(1, h_{t+1}, m(1, \mu_t, h_{t+1})) \geq b(h_{t+1})$,
(D) $d(0, h_{t+1}, m(0, \mu_t, h_{t+1})) = 0$,

Note that if properties (A) and (B) hold, then we clearly have that $\Psi(h_t, \mu_t) \geq 0$ as desired for the theorem. Property (C), in turn, implies $D(h_t, \mu_t) \geq b(g(h_t, 1))$ as also desired. Furthermore, property (D) is directly part of the theorem. Note, then, that the theorem follows directly from these properties and the definition of the net benefit from innovation (4.1). Therefore, proving Properties (A)-(D) are sufficient for the theorem to hold. We prove these four properties in Appendix A. □

A corollary of the above theorem is that if $\underline{\theta} = 0$, $c(.)$ and $b(.)$ are constant, and $\mu_t = 0$, then the DM chooses $d_t = x_t$ regardless of history of innovations. The reason is that in this case, actions do not affect future beliefs and rewards, and so $\Gamma(h_t, 1, x_t) = \Psi(h, 1) = 0$ for any history of experimentation h_t and any profitability-type x_t.

It may be instructive to discuss briefly here why we need Assumption B to prove the above results when $\mu_t < 1$. An increase in innovations up to and including period-t has two opposite effects on the (perceived) frequency of loss-making innovation opporttunities *after* the *next*

period (i.e. of $x_{t+2} = 1$), conditional on a profitable innovation opportunity occurring in the next period (i.e. on $x_{t+1} = 0$). On the one hand, treating period-$\{t+1\}$ posterior and action as fixed, having more innovations up to and including period t implies, due to the monotonicity properties of $g(.)$ and $\pi(.)$, that it is more likely that a loss-making innovation opportunity will occur in period $t+2$. On the other hand, treating as given period-$\{t+1\}$ history and action, under $x_{t+1} = 0$, having more innovations up to and including period t implies, due to the monotonicity properties of $m(.)$ and $\pi(.)$, a lower period-$t+1$ posterior and thereby a lower likelihood that a loss-making innovation opportunity will occur in period $t+2$. As we show in the proof, when the first effect (weakly) dominates the second effect, the long-run benefit from innovating at period t is non-negative. The first effect would be the dominant one if the posterior in period $t+1$ was independent of past actions, which would be the case if the agent was convinced about the state of the world (i.e. if $\mu_t \in \{0,1\}$). Assumption B is a sufficient condition for the first effect to (weakly) dominate the second effect.

We now turn to the study of asymptotic learning.

5 Learning

In each period $t \in \{1,2,...\}$, there may exist an opportunity to experiment ($x_t = 1$) or not ($x_t = 0$). The arrival of experimentation opportunities over time is governed by the "state of nature" $\theta \in \Theta \equiv \{\theta_0, \theta_1\}$. The state of nature is a random variable whose realisation is initially unknown to an agent, who observes the history of arrivals of experimentation opportunities and would like to discover the true state of nature. We are interested in whether the agent will learn asymptotically the true state of nature. The environment we restrict attention to is described next.

Let the vector $x^t = (x_1, ..., x_t)$ be a sample of length t of (observable) experimentation opportunities from the initial period up to and including period t, with x^0 denoting a scalar history initialisation. Denote by $X^t = \{0,1\} \times ... \times \{0,1\}$ a t-fold Cartesian product, and $X^0 = \{x^0\}$ is a singleton set representing the "null" history where no sample has been drawn.

The two parameters in Θ represent very different data-generating processes. The data-generating process is i.i.d. under θ_0 whereas under θ_1 it is (non-trivially) history-dependent. In particular, we postulate that the conditional probability of an experimentation opportunity arising in period t is given by two alternative generating mechanisms: for each $t \geq 1$

$$\mathbb{P}(x_t = 1 | x^{t-1}, \theta_0) = p + \phi_0 f_t(x^{t-1}) \tag{5.1}$$

and

$$\mathbb{P}(x_t = 1 | x^{t-1}, \theta_1) = p + \phi_1 f_t(x^{t-1}), \tag{5.2}$$

where $p \in (0,1)$, $\phi_1 > \phi_0 \geq 0$ and $f_t : X^{t-1} \to \mathbb{R}$ is a non-stochastic coordinate-wise increasing function satisfying $\inf_{t \geq 2} f_t(0_{t-1}) > -p/\phi_1$ and $\sup_{t \geq 2} f_t(1_{t-1}) < (1-p)/\phi_1$, where 0_t and

1_t denote a $t \times 1$ vector of zeros and ones respectively, and $f_1(x^0) < (1-p)/\phi_1$. For most economic applications, a non-negativity requirement applies to the initial condition: $f_1(x^0) \geq 0$. As we show in Lemma 1 in Appendix A, this together with the law-of-motion (5.6) below imply that $f_t(x^{t-1}) \geq 0$ for all $t \geq 1$. With this in mind, the image of f_t, $\mathbf{Im}(f_t)$, satisfies the following restriction according to whether the initial condition $f_1(x^0)$ can take negative values: there exists $\delta \in (0, (1-p)/\phi_1)$ such that

$$\mathbf{Im}(f_t) = \begin{cases} [-p/\phi_1 + \delta, (1-p)/\phi_1 - \delta], & \text{if } f_1(x^0) > -p/\phi_1 \\ [0, (1-p)/\phi_1 - \delta], & \text{if } f_1(x^0) \geq 0 \end{cases} \quad \text{for all } t > 1. \tag{5.3}$$

In both data-generating processes in (5.1) and (5.2), an experimentation opportunity arises at time t as a result of the dynamic evolution of the history x^{t-1} of past experimentation opportunities. The classical case of constant probability p of an experimentation opportunity ($x_t = 1$) at each period t (irrespective of past experimentation opportunities) may only arise under θ_0 when $\phi_0 = 0$. Henceforth, we will refer to the process $\mathbb{P}(\cdot|x^{t-1}, \theta)$ as the θ-process, $\theta \in \{\theta_0, \theta_1\}$.

Experimentation in period t takes place ($d_t = 1$) when an opportunity for experimentation arises and the extent of past experimentation is sufficiently low. Specifically, we postulate that experimentation in period t takes place according to the rule

$$d_t = x_t \mathbf{1}\{f_t(x^{t-1}) < \overline{f}_t(x^{t-1})\}, \tag{5.4}$$

where, for each $t > 1$, $\overline{f}_t$ is a non-stochastic function of the history x^{t-1} satisfying

$$\inf_{t \in \mathbb{N}} \overline{f}_t(x^{t-1}) > 0 \quad a.s. \tag{5.5}$$

and $\overline{f}_1(x^0)$ a pre-determined positive constant.

Note that this rule allows for the case when experimentation takes place (for reasons outside the model) whenever the opportunity arises, $d_t = x_t$, by selecting a sufficiently high threshold, e.g. $\inf_{t \in \mathbb{N}} \overline{f}_t(x^{t-1}) \geq (1-p)/\phi_0$. As the focus of this paper is on the issue of learning, we remain agnostic about the reason behind the experimentation rule (5.4). It could simply be exogenously given (as an experimentation heuristic or because experimentation needs to satisfy certain exogenously set criteria to be given the "go ahead" or not), or the result of some form of myopic reasoning on the part of an agent who is responsible for the experimentation (e.g. use of a "rule of thumb"). But it could also be the result of optimal decision making by a Bayesian agent faced with a learning-by-experimentation problem where the occurrence of an opportunity to experiment generates a short-run net benefit from experimentation but, depending on the state of nature, could also give rise to (additional) long-run costs. Examples of such a problem are when "experimentation" describes harmful consumption under stochastic habit formation, or refraining from investment when the current opportunity cost is high and there is stochastic learning-by-doing in the sense that future costs are more likely to be high the lower past investment has been. In Appendix

B we provide such an example and show the optimality of an experimentation rule such as (5.4). Finally, we note that the threshold type of experimentation we focus on here seems to fit naturally the applications we have in mind. This is especially true since both the beliefs (of the agent or the budget-holder or the approval committee) and the measurement of past experimentation are, in principle, continuous variables (see below) whereas the set of actions is a finite set.

Next, we provide some more structure to the data generating process.

Assumption 1. We assume that

$$f_t(x^{t-1}) = g(f_{t-1}(x^{t-2}), d_{t-1}), \quad t \geq 2 \tag{5.6}$$

for some pre-determined $f_1(x_0) \in \mathbb{R}$ *a.s.*, where d_{t-1} is defined in (5.4) with threshold sequence $\overline{f}_t$ satisfying (5.5), and $g : \mathbf{Im}(f_t) \times \{0,1\} \to \mathbf{Im}(f_t)$, with $\mathbf{Im}(f_t)$ defined in (5.3), is a differentiable function satisfying the following:

(a) when $\mathbf{Im}(f_t) = [0, (1-p)/\phi_1 - \delta]$, the function g in (5.6) satisfies

$$g(0,0) = 0, \;\; g(0,1) > 0, \;\; \inf_{z \geq 0} \frac{\partial g(z,y)}{\partial z} \geq 0, \;\; \inf_{z \geq 0} \frac{\partial g(z,0)}{\partial z} > 0 \tag{5.7}$$

(b) when $\mathbf{Im}(f_t) = [-p/\phi_1 + \delta, (1-p)/\phi_1 - \delta]$, in addition to (5.7), g satisfies

$$\gamma := \sup_{z<0} \sup_{y \geq 0} \frac{\partial g(z,y)}{\partial z} \in [0,1). \tag{5.8}$$

Note that (5.7) implies $g(\cdot, y)$ that is increasing and that both $g(\cdot, 0)$ and $g(0, \cdot)$ are strictly increasing in $\mathbb{R}_+$. The condition (5.8) is a stability requirement that allows to asymptotically eliminate negative initial condition effects. We emphasize here that, conditional on its Markovian structure, the law-of-motion (5.6) for the conditional probability of the occurrence of an experimentation opportunity under the state of nature $\theta = \theta_1$ is quite general as it allows for non-linearities in $g(\cdot, y)$.

The experimentation rule (5.4) and the law-of-motion (5.6) for the conditional probability of the occurrence of an experimentation opportunity under the state of nature $\theta = \theta_1$, as well as the stochastic process $\mathbb{P}(x^t | x^{t-1}, \theta)$ are known by our agent. The agent, however, does not know the true state of nature and would like to discover it. Denote with μ_0 the agent's "prior" belief that $\theta = \theta_1$ (i.e. the belief given the null history x_0 before the realisation of the first sample). Assume that $0 < \mu_0 < 1$. The agent does observe the history of arrivals of experimentation opportunities x^t. We assume that our agent is Bayesian, and so she uses the observed history for inference purposes. Denote with $\mu(x^t)$ the (posterior) belief of the agent that $\theta = \theta_1$, given the realisation of period-t sample x^t. This is given by Bayes rule as:

$$\mu(x^t) = \frac{\mu(x^{t-1})\mathbb{P}(x^t | x^{t-1}, \theta_1)}{\mu(x^{t-1})\mathbb{P}(x^t | x^{t-1}, \theta_1) + (1 - \mu(x^{t-1}))\mathbb{P}(x^t | x^{t-1}, \theta_0)}. \tag{5.9}$$

Note that $\mu(x^t)$ is a random variable, which depends on the realisation of $x^t = (x_1, ..., x_t)$.

We leave this section by noting that the above model can also capture an alternative model where the conditional probability of the occurrence of an experimentation opportunity in period t under the state of nature $\theta = \theta_1$ is equal to p plus a strictly increasing function $q(.)$, with $q(0) = 0$ and $q'(.) \leq 1$, of the "stock" of past experimentation, with the latter now be denoted by $f_t(x^{t-1})$ with f_t be a function as defined above. In this alternative specification, (5.6) would then be the law-of-motion for the "stock" of past experimentation. Such a Markovian law-of-motion is consistent with the laws-of-motion for "stock" variables used in most economic applications.

6 Asymptotic Learning

The question we address in this paper is whether the agent eventually learns the true state of nature. We show in what follows that under Bayesian updating in our environment, $\mu(x^t)$ indeed converges almost surely to one when $\theta = \theta_1$ and to zero when $\theta = \theta_0$.

Because the opportunities for experimentation are dependent and heterogeneously distributed random variables under $\theta = \theta_1$ and under $\theta = \theta_0$ with $\phi_0 > 0$, experimentation opportunities do not have a constant expectation across time and the sample mean of experimentation opportunities does not necessarily concentrate to a constant value. As a result, the usual techniques for establishing asymptotic learning based on laws of large numbers do not apply. We propose a novel approach to establishing asymptotic learning in the absence of stationarity by proving that all accumulation points of the sample mean of a suitable class of functions of experimentation history have the same sign. The above set of accumulation points need not be a singleton (as is the case under a law of large numbers) so the approach is applicable in non-stationary learning environments.

In what follows, we provide an exposition of the key elements of our approach, leaving the details of proofs to the Appendix A. We employ the shorthand notation

$$h_j := f_j(x^{j-1}) \text{ and } \overline{h}_j := \overline{f}_j(x^{j-1}). \tag{6.1}$$

Backward iteration of the law-of-motion (5.9) for the posterior $\mu(x^t)$ combined with the binary property of x_i and the formulae for the conditional distributions $\mathbb{P}\left(\cdot|x^{j-1}, \theta_1\right)$ and $\mathbb{P}\left(\cdot|x^{j-1}, \theta_0\right)$ in (5.2) and (5.1) give

$$\mu(x^t) = \left(1 + \frac{1-\mu_0}{\mu_0} \exp\left\{-t\tilde{\pi}_t\right\}\right)^{-1}, \tag{6.2}$$

where

$$\tilde{\pi}_t = \frac{1}{t}\sum_{j=1}^{t} \psi_i\left(h_j\right) + o_{a.s.}\left(1\right) \text{ as } t \to \infty, \quad i \in \{0,1\} \tag{6.3}$$

with the function ψ_1 (given by (8.22) in Appendix A) being convex in $\mathbb{R}$ and strictly increasing on $\mathbb{R}_+$, the function ψ_0 (given by (8.23) in Appendix A) being concave in $\mathbb{R}$ and

strictly decreasing on $\mathbb{R}_+$, and $\psi_1(0) = \psi_0(0) = 0$. It is clear from the expression in (6.2) that, when $\theta = \theta_1$, $\liminf_{t\to\infty} \tilde{\pi}_t > 0$ *a.s.* is sufficient for $\mu(x^t) \to_{a.s.} 1$ and, when $\theta = \theta_0$, $\limsup_{t\to\infty} \tilde{\pi}_t < 0$ *a.s.* is sufficient for $\mu(x^t) \to_{a.s.} 0$. Thus, asymptotic learning may occur irrespective of the existence of a limit of the sequence $(\tilde{\pi}_t)_{t\in\mathbb{N}}$. By (6.3), the sufficient conditions for asymptotic learning are equivalent to $\liminf_{t\to\infty} t^{-1}\sum_{j=1}^{t}\psi_1(h_j) > 0$ *a.s.* when $\theta = \theta_1$ and $\limsup_{t\to\infty} t^{-1}\sum_{j=1}^{t}\psi_1(h_j) < 0$ *a.s.* when $\theta = \theta_0$.

The above sufficient conditions on the accumulation points of $t^{-1}\sum_{j=1}^{t}\psi_i(h_j)$ suggest[10] to investigate asymptotic bounds for $\mathbb{E}_{\theta_i}\left(\psi_i(h_j)|\, x^{j-2}\right)$ as $j \to \infty$, but such bounds are difficult to obtain due to the nonlinearity of the functions ψ_1 and ψ_0. We overcome this difficulty by using the convexity and monotonicity properties of the functions ψ_1 and ψ_0 in order to reduce the above problem to the more tractable problem of deriving an asymptotic bound for $\mathbb{E}_{\theta_i}\left(h_j|\, x^{j-2}\right)$. This reduction is possible in view of the following proposition, the key requirement of which consists of an asymptotic lower bound for the conditional expectation of the original data Y_j rather than $\psi(Y_j)$; see (6.4) below.

Proposition 1. *Let $\{\mathcal{G}_j : j \in \mathbb{N}\}$ be an increasing sequence of σ-algebras, $(Y_j)_{j\in\mathbb{N}}$ be a sequence of random variables such that Y_j is $\mathcal{G}_j$-measurable for each $j \in \mathbb{N}$, and $\psi : \mathbb{R} \to \mathbb{R}$ be a function with the following properties:*

(a) *there exist $\kappa > 0$ and a sequence of random variables $\left(\xi_j\right)_{j\in\mathbb{N}}$ such that, for all $j \in \mathbb{N}$, ξ_j is $\mathcal{G}_{j-1}$-measurable,*

$$\mathbb{E}\left(Y_j|\,\mathcal{G}_{j-1}\right) \geq \kappa + \xi_j \ \ a.s. \tag{6.4}$$

and $\xi_n \to_{a.s.} 0$ as $n \to \infty$.

(b) *ψ is convex on $\mathbb{R}$, strictly increasing on $\mathbb{R}_+$ with $\psi(0) = 0$, continuous at κ and satisfies $\sup_{j\in\mathbb{N}} \|\psi(Y_j)\|_{L_2} < \infty$.*

Under conditions (a) and (b),

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n}\psi(Y_j) > 0 \ \ a.s.$$

Note that if the convexity and monotonicity assumptions in (b) are replaced by the assumption that ψ is concave and strictly decreasing on $\mathbb{R}_+$, then the conclusion of the proposition becomes

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n}\psi(Y_j) < 0 \ \ a.s.$$

simply by applying Proposition 1 to the function $-\psi$. In the context of the current paper, *a.s.* positivity of the limit infimum obtains when $\theta = \theta_1$ and *a.s.* negativity of the limit supremum

[10] Note that $t^{-1}\sum_{j=1}^{t}\psi_i(h_j)$ may always be approximated by $t^{-1}\sum_{j=1}^{t}\mathbb{E}_{\theta_i}\left(\psi_i(h_j)|\, x^{j-2}\right)$ as $t \to \infty$ under sufficient integrability (but not stationarity) of the sequence $\{\psi_i(h_j) : j \geq 1\}$, see Lemma 2 in Appendix A.

obtains when $\theta = \theta_0$. Proposition 1 can be used to derive positive or negative asymptotic bounds for the sample mean of a function of a non-stationary stochastic process, for types of functions that are commonly used in Economics, such as $\cup$-shaped and $\cap$-shaped functions around 0 and is of independent interest.

Returning to the main result on asymptotic learning, an asymptotic lower bound of the type (6.4) for $\mathbb{E}_{\theta_i}\left(h_j|\, x^{j-2}\right)$ is derived by employing the law of motion (5.6) of Assumption 1 together with the expressions for $\mathbb{P}\left(\cdot|x^{j-1},\theta_1\right)$ and $\mathbb{P}\left(\cdot|x^{j-1},\theta_0\right)$ in (5.2) and (5.1); see (8.31) in Appendix A.

Theorem 1. *Under Bayesian updating, the posterior belief that $\theta = \theta_i$ converges almost surely to unity with exponential rate when θ_i, $i \in \{0,1\}$, is the true data generating process. In particular, there exists $\rho > 0$ such that, as $t \to \infty$:*

(i) $1 - \mu(x^t) = O\left(e^{-\rho t}\right)$ *a.s. when* $\theta = \theta_1$.

(ii) $\mu(x^t) = O\left(e^{-\rho t}\right)$ *a.s. when* $\theta = \theta_0$.

Theorem 1 proves that asymptotic learning occurs with probability one in the non-stationary experimentation environment of the current paper. The main mathematical complication arising from non-stationarity in this Bayesian learning context is the lack of applicability of laws of large numbers. We propose an approach that replaces the role of a law of large numbers in establishing Bayesian consistency by *a.s.* bounds on the set of accumulation points of stochastic sequences that would satisfy a law of large numbers under stationarity but may diverge in our more general nonstationary setting. These bounds are established by using information from the law of motion governing experimentation (5.6), which is part of an economic model. The approach is flexible in that it employs information from the (non-stationary) experimentation time paths to replace the mathematically convenient but restrictive assumptions of stationarity and ergodicity in establishing Bayesian learning.

The proof of both Proposition 1 and Theorem 2 are included in Appendix B.

7 Appendix A: Properties of V needed in proof of Theorem 1

Let hereafter $u(d,x,h) = (1-d)(b(h) - xc(h))$, and $\overline{f}(h,d) = p + \overline{\theta}g(h,d)$.

7.1 Proof of Properties (A) and (B) in proof of Theorem 1

Let γ be an element of the set S of continuous (and bounded) functions of x, h and μ, with $x \in \{0,1\}$, $h \in [0,H]$ and $\mu \in [0,1]$. Define then the function of x, h and μ

$$(N\gamma)(x,h,\mu) = \vartheta(h)\times$$

$$\max_{d\in\{0,1\}}\left\{u(d,x,h) + \delta\left\{\begin{array}{c}\pi(\mu,g(h,d))\gamma\left(1,g(h,d),\frac{\mu\overline{f}(h,d)}{\pi(\mu,g(h,d))}\right)\\ +[1-\pi(\mu,g(h,d))]\gamma\left(0,g(h,d),\frac{\mu[1-\overline{f}(h,d)]}{\pi(\mu,g(h,d))}\right)\end{array}\right\}\right\}$$

In Online Appendix B we can show, using standard dynamic programming techniques, that the Bellman equation is well-defined and that the associated value function V is given by the mapping N, which is unique and uniformly contracting on the set S.

Let S^* be the subset of S such that all functions γ in S^* are non-increasing in x, h, μ and satisfy that $\gamma(0, h, m(0, \mu, h))$ is non-increasing in h for any μ. Observe that S^* is nonempty (as it includes all constant functions) and, notably, all functions in S^* satisfy properties (A) and (B).

Given that N is uniformly contracting on the complete normed vector space S (the subset of all continuous functions $\gamma(x, h, \mu)$ from $\{0, 1\} \times [0, H] \times [0, 1]$ to $\mathbb{R}$ endowed with the sup norm), we have that if $N : S^* \to S^*$ and S^* is closed, then the unique value function defined by the Bellman equation lies in S^*; see for instance Stockey and Lucas (1989), Ch. 3.2, Corollary 1. We show these two properties in turn (and thereby properties (A) and (B) in proof of Theorem 1).

7.1.1 Proof that N maps S^* into S^*

Let $\mathbb{E}_\pi v(X) \equiv \pi v(1) + (1 - \pi)v(0)$. We prove first three useful Lemmata:

Lemma 1. *Fix a continuous function $\gamma(x, h, \mu)$. If, for any $h'' \geq h'$, $\mu'' \geq \mu'$ and $d'' \geq d'$, we have*

$$v_1' \equiv \gamma(1, g(h', d'), m(1, \mu', g(h', d')) \geq v_1'' \equiv \gamma(1, g(h'', d''), m(1, \mu'', g(h'', d'')))$$

$$v_0' \equiv \gamma(0, g(h', d'), m(0, \mu', g(h', d')) \geq v_0'' \equiv \gamma(0, g(h'', d''), m(0, \mu'', g(h'', d'')))$$

$$v_0' \geq v_1'$$

then

$$\mathbb{E}_{\pi(\mu', g(h', d'))}\gamma(X, g(h', d'), m(X, \mu', g(h', d'))) \geq \mathbb{E}_{\pi(\mu'', g(h'', d''))}\gamma(X, g(h'', d''), m(X, \mu'', g(h'', d'')))$$

Proof. Let $\pi' \equiv \pi(\mu', g(h', d'))$ and $\pi'' \equiv \pi(\mu'', g(h'', d''))$ and note that $\pi' \leq \pi''$. Now, take the difference $\pi' v_1' + (1 - \pi')v_0' - \pi'' v_1'' - (1 - \pi'')v_0''$, add and subtract $\pi'' v_1' + (1 - \pi'')v_0'$ and rearrange terms to obtain $(1 - \pi'')(v_0' - v_0'') + (\pi' - \pi'')(v_1' - v_0') + \pi''(v_1' - v_1'') \geq 0$ □

Lemma 2. *Consider $\gamma \in S^*$ and let $(x'', h'', \mu'') \geq (x', h', \mu')$ with at least one inequality being strict. Let also $d'' \geq d'$, and set $m(0) = m(0, \mu, h)$, $m'(0) = m(0, \mu, g')$ and $m''(0) = m(0, \mu, g'')$, and $g'', d')$ and $g'''', d'')$. We have that:*

$$\begin{aligned}
\mathbb{E}_{\pi(\mu', g(h', d))}[\gamma(X, g(h', d), m(X, \mu', g(h', d)))] &\geq \mathbb{E}_{\pi(\mu'', g(h'', d))}[\gamma(X, g(h'', d), m(X, \mu'', g(h'', d)))] \\
\mathbb{E}_{\pi(\mu, g(h, 0))}[\gamma(X, g(h, 0), m(X, \mu, g(h, 0)))] &\geq \mathbb{E}_{\pi(\mu, g(h, d''))}[\gamma(X, g(h, d''), m(X, \mu, g(h, d'')))]
\end{aligned}$$

and

$$\mathbb{E}_{\pi(m'(0),g(g(h',d'),0))}[\gamma(X,g(g(h',d'),0),m(X,m'(0),g(g(h',d'),0)))] \geq$$
$$\mathbb{E}_{\pi(m''(0),g(g(h'',d''),0))}[\gamma(X,g(g(h'',d''),0),m(X,m''(0),g(g(h'',d''),0)))]$$

Proof. The first inequality follows from Lemma 1 and the following: (a) $\pi(\mu',g(h',d)) \leq \pi(\mu'',g(h'',d))$, (b) $\gamma(0,g(h',d),m(0,\mu',g(h',d))) \geq \gamma(1,g(h',d),m(1,\mu',g(h',d)))$ due to $\gamma \in S^*$ and the monotonicity of $m(x,\mu',g(h',d))$ with respect to x, (c) $\gamma(1,g(h',d),m(1,\mu',g(h',d)))$ $\geq \gamma(1,g(h'',d),m(1,\mu'',g(h'',d)))$ due to $\gamma \in S^*$ and the monotonicity of $m(1,\mu,h)$ with respect to μ and h, and finally (d) $\gamma(0,g(h',d),m(0,\mu',g(h',d))) \geq \gamma(0,g(h'',d),m(0,\mu'',g(h'',d)))$ due to $m(0,\mu',g(h'',d)) \leq m(0,\mu'',g(h'',d))$ and that $\gamma \in S^*$.

The second inequality follows from Lemma 1 and the following: (a) $\pi(m(0),g(h,0)) \leq \pi(m(0),g(h,1))$, (b) $\gamma(1,g(h,0),m(1,m(0),g(h,0))) \leq \gamma(0,g(h,0),m(0,m(0),g(h,0)))$ due to $\gamma \in S^*$ and $m(1,m(0),g(h,d)) > m(0,m(0),g(h,d))$, (c) $\gamma(1,g(h,0),m(1,m(0),g(h,0))) \geq \gamma(1,g(h,1),m(1,m(0),g(h,1)))$ due to $\gamma \in S^*$ and the monotonicity of $m(1,m(0),g(h,d))$ with respect to d, and (d) $\gamma(0,g(h,0),m(0,m(0),g(h,0))) \geq \gamma(0,g(h,1),m(0,m(0),g(h,1)))$ due to $\gamma \in S^*$.

The last inequality follows from Lemma 1 and the following: (a) $\pi(m'(0),g(g'''(0),g(g'',0))$ from Assumption B, (b) $\gamma(0,g(g'',g(g',0))) \geq \gamma(1,g(g'',g(g',0)))$ due to $\gamma \in S^*$ and the monotonicity of $m(x,\mu',g(g'0))$ with respect to x, (c) $\gamma(1,g(g'',g(g',0)) \geq \gamma(1,g(g'''',g(g'',0)))$ due to $\gamma \in S^*$ and the monotonicity of $m(1,\mu,h)$ with respect to μ and h, and finally (d) $\gamma(0,g(g'',g(g',0))) \geq \gamma(0,g(g'''',g(g'',0)))$ due to $m(0,\mu',g(g'',0)) \leq m(0,\mu'',g(g'',0))$ and that $\gamma \in S^*$. □

Note then that the definitions of maximum and the decision rule $d(.)$, and the properties of $\vartheta(h)$ and $u(.)$, and the above properties imply, after setting now, with some abuse of notation, $d' = d(x',h',\mu')$ and $d'' = d(x'',h'',\mu'')$, that, when $x' = x'' = x$,

$$\begin{aligned}
& (N\gamma)(x,h'') \\
= \;& \vartheta(h')\left\{u(d',x,h') + \delta\left\{\begin{array}{c}\pi(\mu',g(h''))\gamma\left(1,g(h''),\frac{\mu'\overline{f}(h'')}{\pi(\mu',g(h''))}\right) \\ +[1-\pi(\mu',g(h''))]\gamma\left(0,g(h''),\frac{\mu'[1-\overline{f}(h'')]}{\pi(\mu',g(h''))}\right)\end{array}\right\}\right\} \\
\geq \;& \vartheta(h')\left\{u(d'',x,h') + \delta\left\{\begin{array}{c}\pi(\mu',g(h'''))\gamma\left(1,g(h'''),\frac{\mu'\overline{f}(h''')}{\pi(\mu',g(h'''))}\right) \\ +[1-\pi(\mu',g(h'''))]\gamma\left(0,g(h'''),\frac{\mu'[1-\overline{f}(h''')]}{\pi(\mu',g(h'''))}\right)\end{array}\right\}\right\} \\
\geq \;& \vartheta(h'')\left\{u(d'',x,h'') + \delta\left\{\begin{array}{c}\pi(\mu'',g(h''''))\gamma\left(1,g(h''''),\frac{\mu''\overline{f}(h'''')}{\pi(\mu'',g(h''''))}\right) \\ +[1-\pi(\mu'',g(h''''))]\gamma\left(0,g(h''''),\frac{\mu''(1-\overline{f}(h''''))}{\pi(\mu'',g(h''''))}\right)\end{array}\right\}\right\} \\
= \;& (N\gamma)(x,h'''')
\end{aligned}$$

Furthermore, setting $x' = 0$ and $x'' = 1$, and $h' = h'' = h$ and $\mu' = \mu'' = \mu$, we have:

$$
\begin{aligned}
(N\gamma)(0,h,\mu) &= \vartheta(h)\left\{u(d',0,h)+\delta\left\{\begin{array}{c}\pi(\mu,g(h,d'))\gamma(1,g(h,d'),\frac{\mu\overline{f}(h,d')}{\pi(\mu,g(h,d'))})\\ +[1-\pi(\mu,g(h,d'))]\gamma(0,g(h,d'),\frac{\mu[1-\overline{f}(h,d')]}{\pi(\mu,g(h,d'))})\end{array}\right\}\right\}\\
&\geq \vartheta(h)\left\{u(0,0,h)+\delta\left\{\begin{array}{c}\pi(\mu,g(h,0))\gamma(1,g(h,0),\frac{\mu\overline{f}(h,0)}{\pi(\mu,g(h,0))})\\ +[1-\pi(\mu,g(h,0))]\gamma(0,g(h,0),\frac{\mu[1-\overline{f}(h,0)]}{\pi(\mu,g(h,0))})\end{array}\right\}\right\}\\
&\geq \vartheta(h)\left\{u(d'',1,h)+\delta\left\{\begin{array}{c}\pi(\mu,g(h,d''))\gamma(1,g(h,d''),\frac{\mu\overline{f}(h,d'')}{\pi(\mu,g(h,d''))})\\ +[1-\pi(\mu,g(h,d''))]\gamma(0,g(h,d''),\frac{\mu[1-\overline{f}(h,d'')]}{\pi(\mu,g(h,d''))})\end{array}\right\}\right\}\\
&= (N\gamma)(1,h,\mu)
\end{aligned}
$$

with the last inequality using also $d'' \in \{0,1\}$.

Therefore, $N\gamma$ is non-increasing as well.

Let $m'(0) = m(0,\mu,g(h',d'))$ and $m''(0) = m(0,\mu,g(h'',d''))$.To show that $(N\gamma)(0,g(h',d'),m'(0)) \geq (N\gamma)(0,g(h'',d''),m''(0))$, we show first that $d(0,h,\mu) = 0$ for any h,μ.
To see this, let us compare

$$u(0,0,h) + \delta\mathbb{E}_{\pi(\mu,h')}V(X,h',m(X,\mu,h'))$$

where $h' = g(h,0)$, with

$$u(1,0,h) + \delta\mathbb{E}_{\pi(\mu,h'')}V(X,h'',m(X,\mu,h''))$$

where $h'' = g(h,1)$. Observe that $\pi(\mu,h') \leq \pi(\mu,h'')$. Using that $\gamma \in S^*$ and the monotonicity properties of $m(x,\mu,g(h,d))$ with respect to x and d, we get $\gamma(1,h',m(1,\mu,h')) \leq \gamma(0,h',m(0,\mu,h'))$ and $\gamma(1,h',m(1,\mu,h')) \geq \gamma(1,h'',m(1,\mu,h''))$. From $\gamma \in S^*$ we also have $\gamma(0,h',m(0,\mu,h'))$ $\geq \gamma(0,h'',m(0,\mu,h'')))$. From these properties and Lemma 1 we have $E_{\pi(\mu,h'')}\gamma(X,h'',m(X,\mu,h'')) \leq E_{\pi(\mu,h')}\gamma(X,h',m(X,\mu,h'))$. This alongside $u(0,0,h) > u(1,0,h)$ (from Assumption A) gives us the desired result that $d(0,h,\mu) = 0$.

Equipped with this result, set now, with some abuse of notation, $d'' \geq d'$. Using Assumption B as well as the monotonicity properties of $m(x,m,g)$ $\vartheta(h)$ and $u(.)$, and Lemma 2) that:

$$
\begin{aligned}
&(N\gamma)(0, g(h', d'), m'(0)) \\
&= \vartheta(g(h', d'', d')) \\
&+\delta\vartheta(g(h', d'))\left\{\begin{array}{c} \pi(m'(0), g(g(h', d'), 0))\gamma(1, g(g(h', d'), 0), m(1, m'(0), g(g(h', d'), 0))) \\ +[1-\pi(m'(0), g(g(h', d'), 0))]\gamma(0, g(g(h', d'), 0), m(0, m'(0), g(g(h', d'), 0))) \end{array}\right\} \\
&\geq \vartheta(g(h'', d'''', d'')) \\
&+\delta\vartheta(g(h'', d''))\left\{\begin{array}{c} \pi(m''(0), g(g(h'', d''), 0))\gamma(1, g(g(h'', d''), 0), m(1, m''(0), g(g(h'', d''), 0))) \\ +[1-\pi(m''(0), g(g(h'', d''), 0))]\gamma(0, g(g(h'', d''), 0), m(0, m''(0), g(g(h'', d''), 0))) \end{array}\right\} \\
&= (N\gamma)(0, g(h'', d''), m''(0))
\end{aligned}
$$

Thus, we have that N maps S^* into itself.
****

Finally, we prove property (B) for $t = n-1$ After using Lemma 1, we have that

$$
\begin{aligned}
&V(0, g(h_n, a_n), m(0, \mu_n, h_n, a_n)) = \vartheta(g(h_n, a_n))v(0, 0, g(h_n, a_n)) \\
&+\delta\vartheta(g(h_n, a_n))E_{\pi(0, g(h_n, a_n), m(0, \mu_n, h_n, a_n))}V(X, g(g(h_n, a_n), 0), m(X, m(0, \mu_n, h_n, a_n), g(h_n, a_n), 0)).
\end{aligned}
$$

Using the monotonicity properties of $m(.,.,.)$, $g(.,.)$ and properties (A) and (B) for $t = n$ we also have, for any $(a_n'', h_n'') \geq (a_n', h_n')$, that
(a) $V(1, g(h_n', a_n'), m(1, \mu, h_n', a_n')) \geq V(1, g(h_n'', a_n''), m(1, \mu, h_n'', a_n''))$,
(b) $V(0, g(h_n', a_n'), m(0, \mu, h_n', a_n')) \geq V(0, g(h_n'', a_n''), m(0, \mu, h_n'', a_n''))$, and
(c) $V(0, g(h_n', a_n'), m(0, \mu, h_n', a_n')) \geq V(1, g(h_n'', a_n''), m(1, \mu, h_n'', a_n''))$. The proof is complete after noting that $\pi(0, g(h_n, a_n), m(0, \mu_n, h_n, a_n))$ is nondecreasing in h_n, a_n, by Assumption **??**, and recalling Assumption **??** and $\vartheta(h_n'') \leq \vartheta(h_n')$.
****

7.1.2 Proof that S^* is closed

In what follows we endow all the function spaces with the sup norm. Let the subset of S consisting of all non-increasing functions be denoted by $S^{0,d}$. Note that $S^{0,d}$ is a closed subset of S,[11] and hence a $S^{0,d}$ endowed with the sup norm is a complete normed vector space space itself.

Take now any function $\gamma \in S^{0,d}$. For any admissible (according to Assumption B) g and f, consider the function $\widetilde{\gamma}$ that maps h, d, μ to $\mathbb{R}$ as:

$$
\widetilde{\gamma}(h, d, \mu) \equiv \gamma(0, g(h, d), m(0, \mu, g(h, d))) \tag{7.1}
$$

Define then the set of functions, $S^{1,d} \subseteq S^{0,d}$ such that $\gamma \in S^{1,d}$ if and only if $\widetilde{\gamma}$ is non-increasing in h and d. Note that $S^* = S^{1,d}$. We therefore need to show that $S^{1,d}$ is closed.

[11] See proof of Theorem 4.7 in Stockey and Lucas (1989).

To do so, denote with γ^1 a generic element of $S^{1,d}$. Moreover, we use an upper-bar to denote a limit function. Thus, for instance, $\overline{\gamma}$ denotes the limit function of a sequence of functions $\{\gamma_n\}_{n=1}^{\infty}$.

Consider now a sequence of functions $\{\gamma_n^1\}_{n=1}^{\infty}$, where $\gamma_n^1 \in S^{1,d}$, such that this sequence has a limit function $\overline{\gamma}^1$. We need to show that $\overline{\gamma}^1 \in S^{1,d}$. Since $S^{1,d}$ is a subset of $S^{0,d}$ and since $S^{0,d}$ is closed, the limit function $\overline{\gamma}^1$ must be in $S^{0,d}$ so the only way $\overline{\gamma}^1$ can not be in $S^{1,d}$ is because its corresponding $\widetilde{\overline{\gamma}}^1$ is increasing in h and d. Thus, suppose that $\widetilde{\overline{\gamma}}^1$ is increasing in h and d. Fix a μ and consider $h'' \geq h'$ and $d'' \geq d'$, with at least one inequality strict, such that $\widetilde{\overline{\gamma}}^1(h'', d'', \mu) - \widetilde{\overline{\gamma}}^1(h', d', \mu) = \varepsilon$, for some $\varepsilon > 0$. Let $g' = g(h', d')$ and $g'' = g(h'', d'')$. Similarly, $\mu' = m(0, \mu, g(h', d'))$ and $\mu'' = m(0, \mu, g(h'', d''))$. From (7.1), we have, $\gamma_n^1(0, g', \mu') = \widetilde{\gamma}_n^1(h', d', \mu)$ and $\gamma_n^1(0, g'', \mu'') = \widetilde{\gamma}_n^1(h'', d'', \mu)$. Moreover, we have that $\bar{\gamma}^1(0, g', \mu') = \widetilde{\overline{\gamma}}^1(h', a', \mu)$ and $\bar{\gamma}^1(0, g'', \mu'') = \widetilde{\overline{\gamma}}^1(h'', a'', \mu)$. Since γ_n^1 converges to $\overline{\gamma}^1$ uniformly, there exists a integer $\widehat{n} > 0$, such that for all $n > \widehat{n}$,

$$|\gamma_n^1(0, g'', \mu'') - \overline{\gamma}^1(0, g'', \mu'')| < \frac{\varepsilon}{2} \quad \text{and} \quad |\gamma_n^1(0, g', \mu') - \overline{\gamma}^1(0, g', \mu')| < \frac{\varepsilon}{2}$$

From the above two inequalities we have,

$$\gamma_n^1(0, g'', \mu'') > \overline{\gamma}^1(0, g'', \mu'') - \frac{\varepsilon}{2} \quad \text{and} \quad \gamma_n^1(0, g', \mu') < \overline{\gamma}(0, g', \mu') + \frac{\varepsilon}{2}$$

We thus get $\gamma_n^1(0, g'', \mu'') - \gamma_n^1(0, g', \mu') > \overline{\gamma}^1(0, g'', \mu'') - \overline{\gamma}^1(0, g', \mu') - \varepsilon = \widetilde{\overline{\gamma}}^1(h'', d'', \mu) - \widetilde{\overline{\gamma}}^1(h', d', \mu) - \varepsilon = 0$ which implies that $\widetilde{\gamma}_n^1(h'', d'', \mu) - \widetilde{\gamma}_n^1(h', d', \mu) > 0$. This contradicts that $\gamma_n^1 \in S^{1,d}$, and so $S^{1,d}$ is closed.

7.2 Proof of properties (C) and (D) in proof of Theorem 1

To reduce notation we drop without risk of confusion the subscript $n+1$ from the functions $V_{n+1}(.,.,.)$ and $a_n(.,.,.)$ in what follows.

We start with proving property (D) for $t = n-1$ For any given μ_{n-1}, h_{n-1} and a_{n-1} and corresponding history $h_n = g(h_{n-1}, a_{n-1})$ and posteriors $m_n(x_n) \equiv m(x_n, \mu_{n-1}, h_{n-1}, a_{n-1})$, let $x_n = 0$. Then, conditional on facing a task, if $a_n = 0$, the expected discounted payoff in period n is

$$v(0, 0, h_n) + \delta \mathbb{E}_{\pi(0, h_n, m_n(0))} V(X, h', m(X, m_n(0), h_n, 0))$$

where $h' = g(h_n, 0)$. On the other hand, if $a_n = 1$, the payoff is

$$v(1, 0, h_n) + \delta \mathbb{E}_{\pi(1, h_n, m_n(0))} V(X, h'', m(X, m_n(0), h_n, 1))$$

where $h'' = g(h_n, 1)$.

Note now that $\pi(0, h_n, m_n(0)) \leq \pi(1, h_n, m_n(0))$. Using property (A) for $t = n$ and the monotonicity properties of $m(x, m_n(0), h_n, a)$ with respect to x and a, we get $V(1, h', m(1, m_n(0), h_n, 0)) \leq V(0, h', m(0, m_n(0), h_n, 0))$ and $V(1, h', m(1, m_n(0), h_n, 0)) \geq V(1, h'', m(1, m_n(0), h_n, 1))$. From property (B) for $t = n$ we have $V(0, h', m(0, m_n(0), h_n, 0)) \geq V(0, h'', m(0, m_n(0), h_n, 1))$.

These imply that $E_{\pi(1,h_n,m_n(0))}V(X,h'',m(X,m_n(0),h_n,1)) \leq E_{\pi(0,h_n,m_n(0))}V(X,h',m(X,m_n(0),h_n,0))$. This alongside $v(0,0,h_n) > v(1,0,h_n)$ completes the proof that $a_n(h_n,0,m_n(0)) = 0$.

We now prove property (C) for $t = n-1$ Recall that $m_n(1) \geq m_n(0)$. Given the previous result, we have, for any given μ_{n-1}, h_{n-1} and a_{n-1}, and corresponding period n history h_n and posteriors $m_n(x_n)$,

$$V(0,h_n,m_n(0)) = \vartheta(h_n)\{v(0,0,h_n) + \delta E_{\pi(0,h_n,m_n(0))}V(X,h',m(X,m_n(0),h_n,0))\}$$

and similarly

$$V(1,h_n,m_n(1)) = \vartheta(h_n)\{\max_a\{v(a,1,h_n)+\delta E_{\pi(a,h_n,m_n(1))}V(X,g(h_n,a),m(X,m_n(1),h_n,a))\}\}$$

Note $\tilde{a} \equiv a(h_n,1,m_n(1)) \geq 0$. Due to $v(0,0,h_n) - v(0,1,h_n) > v(0,0,h_n) - v(1,1,h_n) = D(h_n)$, if $E_{\pi(\tilde{a},h_n,m_n(1))}V(X,g(h_n,\tilde{a}),m(X,m_n(1),h_n,\tilde{a})) \leq E_{\pi(0,h_n,m_n(0))}V(X,h',m(X,m_n(0),h_n,0))$, the proof is complete. This follows from (a) $V(0,h',m(0,m_n(0),h_n,0)) \geq V(1,h',m(1,m_n(0),h_n,0))$ due to property (A) for $t = n$ and the monotonicity of $m(x,m_n(0),h_n,0)$ with respect to x, (b) $V(1,h',m(1,m_n(0),h_n,0)) \geq V(1,g(h_n,\tilde{a}),m(1,m_n(1),h_n,\tilde{a}))$ due to property (A) for $t = n$, that $g(h_n,\tilde{a}) \geq h'$, and the monotonicity properties of $m(1,\mu,h_n,a)$ with respect to μ and a, (c) $V(0,h',m(0,m_n(0),h_n,0)) \geq V(0,g(h_n,\tilde{a}),m(0,m_n(1),h_n,\tilde{a}))$ due to properties (A) and (B) for $t = n$, and $m(0,m_n(1),h_n,\tilde{a}) \geq m(0,m_n(0),h_n,\tilde{a})$, and (d) $\pi(0,h_n,m_n(0))$ $\leq \pi(\tilde{a},h_n,m_n(1))$.

8 Appendix B

In the notation of (6.1), (5.6) becomes

$$h_i = g\left(h_{i-1}, x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}\right), \quad i \geq 2, \tag{8.1}$$

with initial condition $h_1 := f_1(x_0)$ satisfying

$$\mathbb{P}\left(|h_1| < \infty\right) = 1, \tag{8.2}$$

and threshold process satisfying

$$\overline{h}_i \text{ is } \sigma\left(x_{i-1},...,x_1\right) \text{ -measurable with } \inf_{i\in\mathbb{N}} \overline{h}_i > 0 \ a.s. \tag{8.3}$$

in view of (5.5) and the fact that each $\overline{f}_t$ is a non-stochastic function.

We start by establishing two preliminary results.

Lemma 1. *Consider the sequence $(h_i)_{i\geq 1}$ defined by (8.1) and h_1 satisfying (8.2). The following hold for each $i \geq 2$:*

(i) *$h_{i-1} \geq 0$ a.s. implies that $h_i \geq 0$ a.s.*

(ii) $h_i \mathbf{1}\{h_i < 0\} \geq \gamma^{i-1} h_1 \mathbf{1}\{h_1 < 0\}$ *a.s*, for some $\gamma \in [0,1)$.

Proof. We employ the mean value theorem

$$g(z,y) = g(0,y) + z \left.\frac{\partial g(z,y)}{\partial z}\right|_{z=z_0} \quad \text{for } |z_0| < |z|, \tag{8.4}$$

with $z = h_{i-1}$ and $y = x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}$ to the right side of (8.1): when $h_{i-1} \geq 0$ *a.s.*,

$$\begin{aligned} h_i &\geq g\left(0, x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}\right) + h_{i-1} \inf_{z \geq 0} \frac{\partial g\left(z, x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}\right)}{\partial z} \\ &\geq g\left(0, x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}\right) \geq g(0,0) = 0 \end{aligned}$$

a.s. by (5.7), since $g(0,\cdot)$ is strictly increasing and $x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\} \geq 0$. For part (ii), write

$$h_i \mathbf{1}\{h_i < 0\} = h_i \mathbf{1}\{h_i < 0\} \mathbf{1}\{h_{i-1} < 0\} + h_i \mathbf{1}\{h_i < 0\} \mathbf{1}\{h_{i-1} \geq 0\}. \tag{8.5}$$

For the first term of (8.5), using the expression for h_i in (8.1), applying (8.4) with $z = h_{i-1}$ and $y = x_{i-1}$, and using Assumption 1 yields

$$\begin{aligned} h_i \mathbf{1}\{h_i < 0\} \mathbf{1}\{h_{i-1} < 0\} &= g(h_{i-1}, x_{i-1}) \mathbf{1}\{h_{i-1} < 0\} \mathbf{1}\{h_i < 0\} \\ &\geq \left(g(0, x_{i-1}) + h_{i-1} \sup_{z<0} \frac{\partial g(z, x_{i-1})}{\partial z}\right) \mathbf{1}\{h_{i-1} < 0\} \mathbf{1}\{h_i < 0\} \\ &\geq h_{i-1} \sup_{z<0} \frac{\partial g(z, x_{i-1})}{\partial z} \mathbf{1}\{h_{i-1} < 0\} \mathbf{1}\{h_i < 0\} \\ &\geq h_{i-1} \sup_{z<0} \sup_{y \geq 0} \frac{\partial g(z,y)}{\partial z} \mathbf{1}\{h_{i-1} < 0\} \end{aligned} \tag{8.6}$$

where the last inequality follow from the fact that $h_{i-1} < 0$ and (5.8) implies that the term above (8.6) is minimized when $\mathbf{1}\{h_i < 0\} = 1$. The MVT (8.4) with $z = h_{i-1}$ and $y = x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}$ yields

$$\begin{aligned} h_i \mathbf{1}\{h_i < 0\} \mathbf{1}\{h_{i-1} \geq 0\} &= g\left(h_{i-1}, x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}\right) \mathbf{1}\{h_{i-1} \geq 0\} \mathbf{1}\{h_i < 0\} \\ &\geq \left(g\left(0, x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}\right) + h_{i-1} \inf_{z \geq 0} \frac{\partial g\left(z, x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}\right)}{\partial z}\right) \\ &\quad \times \mathbf{1}\{h_{i-1} \geq 0\} \mathbf{1}\{h_i < 0\} \\ &\geq g\left(0, x_{i-1}\mathbf{1}\left\{h_{i-1} < \overline{h}_{i-1}\right\}\right) \mathbf{1}\{h_{i-1} \geq 0\} \mathbf{1}\{h_i < 0\} \\ &\geq 0 \end{aligned}$$

for all i, similarly to above. By substituting the last bound and (8.6) into (8.5), we conclude that

$$h_i \mathbf{1}\{h_i < 0\} \geq \gamma h_{i-1} \mathbf{1}\{h_{i-1} < 0\}$$

for all i, with $\gamma = \sup_{z<0} \sup_{y \geq 0} \partial g(z,y)/\partial z \in [0,1)$ defined in (5.8). Backward recursion of the final inequality yields the result. ■

Lemma 2. *Let $\{\mathcal{G}_j : j \in \mathbb{N}\}$ be an increasing sequence of σ-algebras and Y_j be $\mathcal{G}_j$-measurable random variables for each $j \in \mathbb{N}$. If $\sup_{j\in\mathbb{N}} \|Y_j\|_{L_2} < \infty$, then*

$$\frac{1}{n}\sum_{j=1}^{n} [Y_j - \mathbb{E}(Y_j|\mathcal{G}_{j-1})] \to_{a.s.} 0 \quad (n \to \infty).$$

Proof. Letting $Z_j := Y_j - \mathbb{E}(Y_j|\mathcal{G}_{j-1})$ and $M_n := \sum_{j=1}^{n} Z_j$, note that $(Z_j, \mathcal{G}_j)$ is a martingale difference sequence and $(M_n, \mathcal{G}_n)$ is a martingale with

$$\sup_{j\in\mathbb{N}} \|Z_j\|_{L_2}^2 \leq \max\left\{\sup_{j\in\mathbb{N}} \|Y_j\|_{L_2}^2, \sup_{j\in\mathbb{N}} \|\mathbb{E}(Y_j|\mathcal{G}_{j-1})\|_{L_2}^2\right\} \leq \sup_{j\in\mathbb{N}} \|Y_j\|_{L_2}^2 < \infty \tag{8.7}$$

where the second inequality is obtained by the Jensen inequality for conditional expectations applied to the convex function $x \mapsto x^2$ followed by the law of iterated expectations, so M_n is a square integrable martingale. By the strong law of large numbers for square integrable martingales (see e.g. Theorem 4 in chapter VII.5 of Shiryayev (1995) with $A_n = n$), the condition

$$\sum_{j=1}^{\infty} \frac{1}{j^2}\mathbb{E}\left((\Delta M_j)^2 \middle| \mathcal{G}_{j-1}\right) < \infty \ \ a.s. \tag{8.8}$$

is sufficient for $M_n/n \to_{a.s.} 0$. In our notation, the monotone convergence theorem and the law of iterated expectations yield

$$\left\|\sum_{j=1}^{\infty} \frac{1}{j^2}\mathbb{E}\left((\Delta M_j)^2 \middle| \mathcal{G}_{j-1}\right)\right\|_{L_1} = \sum_{j=1}^{\infty} \frac{1}{j^2}\mathbb{E}(Z_j^2) \leq \sup_{j\in\mathbb{N}} \|Z_j\|_{L_2}^2 \sum_{j=1}^{\infty} \frac{1}{j^2} < \infty,$$

by (8.7), which implies the sufficient condition (8.8), since $\|X\|_{L_1} < \infty$ implies that $|X| < \infty$ $a.s.$ for any random variable X. The validity of (8.8) ensures that $M_n/n \to_{a.s.} 0$, as required. ■

Proof of Proposition 1. We derive some preliminary bounds. First, since $\xi_j \to_{a.s.} 0$, there exists a random integer J_0 such that $\xi_j > -\kappa/2$, and thereby $\kappa + \xi_j > \kappa/2 > 0$ for all $j \geq J_0$ $a.s.$. Given that ψ is increasing on $\mathbb{R}_+$, the latter implies that

$$\psi(\mathbb{E}(Y_j|\mathcal{G}_{j-1})) \geq \psi(\kappa + \xi_j) \ \text{ for all } j \geq J_0 \ \ a.s. \tag{8.9}$$

by (6.4). Second, convexity of ψ yields

$$E(\psi(Y_j)|\mathcal{G}_{j-1}) \geq \psi(\mathbb{E}(Y_j|\mathcal{G}_{j-1})) \geq \psi(\kappa + \xi_j) \tag{8.10}$$

for all $j \geq J_0$ *a.s.*, where the first inequality follows by the Jensen inequality for conditional expectations and the second inequality follows from (8.9). Third, letting $M_n := \sum_{j=1}^{n} [\psi(Y_j) - \mathbb{E}(\psi(Y_j)|\mathcal{G}_{j-1})]$, $\sup_{j\in\mathbb{N}} \|\psi(Y_j)\|_{L_2} < \infty$ and Lemma 2 (stated and proved in Appendix A) imply that $M_n/n \to_{a.s.} 0$. Denoting

$$S_{J_0} = \sum_{j=1}^{J_0-1} \mathbb{E}(\psi(Y_j)|\mathcal{G}_{j-1})$$

for brevity, we can then write *a.s.*

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n}\psi(Y_j) &= \frac{1}{n}\sum_{j=J_0}^{n}\mathbb{E}(\psi(Y_j)|\mathcal{G}_{j-1}) + \frac{1}{n}S_{J_0} + \frac{1}{n}M_n \\ &\geq \frac{1}{n}\sum_{j=J_0}^{n}\psi(\kappa+\xi_j) + \frac{1}{n}S_{J_0} + \frac{1}{n}M_n \end{aligned} \tag{8.11}$$

by (8.10).

To prove the desired result, note first that the fact that ψ is strictly increasing with $\psi(0) = 0$ implies that $\psi(y) > 0$ for all $y > 0$. Thus, since $\kappa > 0$ by assumption, $\psi(\kappa) > 0$. Also, we have that $\xi_j \to_{a.s.} 0$ and continuity of ψ at κ imply that $\psi(\kappa+\xi_j) \to_{a.s.} \psi(\kappa)$. The Cesàro lemma (e.g. 12.6 in Williams (1991)) then gives

$$\frac{1}{n}\sum_{j=J_0}^{n}\psi(\kappa+\xi_j) \to_{a.s.} \psi(\kappa) > 0.$$

Furthermore, the law of iterated expectations and $\sup_{j\in\mathbb{N}} \|\psi(Y_j)\|_{L_2} < \infty$ imply that $\|\mathbb{E}(\psi(Y_j)|\mathcal{G}_{j-1})\|_{L_1} = \|\psi(Y_j)\|_{L_1} < \infty$, which, in turn, implies that $\mathbb{E}(\psi(Y_j)|\mathcal{G}_{j-1}) < \infty$ for each j *a.s.* and hence

$$\frac{1}{n}S_{J_0} \leq \frac{1}{n}(J_0-1)\max_{1\leq j<J_0}|\mathbb{E}(\psi(Y_j)|\mathcal{G}_{j-1})| \to_{a.s.} 0.$$

Finally, recall that we have shown that $M_n/n \to_{a.s.} 0$. We conclude that all but the first term on the right side of (8.11) tend to 0 *a.s.*, and so taking the limit inferior in both sides of (8.11) yields

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n}\psi(Y_j) \geq \psi(\kappa) > 0 \ \ a.s.$$

as required. ■

Proof of Theorem 1. We start by proving (6.2) and (6.3). Backward iteration of the law-of-motion (5.9) for the posterior $\mu(x^t)$ gives

$$\begin{aligned}\mu(x^t) &= \frac{\mu_0 \prod_{j=1}^t \mathbb{P}\left(x^j|x^{j-1},\theta_1\right)}{\mu_0 \prod_{j=1}^t \mathbb{P}\left(x^j|x^{j-1},\theta_1\right) + (1-\mu_0)\prod_{j=1}^t \mathbb{P}\left(x^j|x^{j-1},\theta_0\right)} \\ &= \left(1 + \frac{1-\mu_0}{\mu_0}\frac{\prod_{i=1}^t \mathbb{P}\left(x^i|x^{i-1},\theta_0\right)}{\prod_{j=1}^t \mathbb{P}\left(x^j|x^{j-1},\theta_1\right)}\right)^{-1} \end{aligned} \tag{8.12}$$

where, using the binary property of x_i and the formulae for $\mathbb{P}\left(\cdot|x^{j-1},\theta_1\right)$ and $\mathbb{P}\left(\cdot|x^{j-1},\theta_0\right)$ in (5.2) and (5.1), we may write

$$\begin{aligned}\frac{\prod_{i=1}^t \mathbb{P}\left(x^i|x^{i-1},\theta_0\right)}{\prod_{j=1}^t \mathbb{P}\left(x^j|x^{j-1},\theta_1\right)} &= \frac{\prod_{i=1}^t \left(p+\phi_0 h_i\right)^{x_i}\left(1-p-\phi_0 h_i\right)^{1-x_i}}{\prod_{j=1}^t \left(p+\phi_0 h_j+\left(\phi_1-\phi_0\right)h_j\right)^{x_j}\left(1-p-\phi_0 h_j-\left(\phi_1-\phi_0\right)h_j\right)^{1-x_j}} \\ &= \frac{1}{\prod_{j=1}^t \left(1+\frac{\left(\phi_1-\phi_0\right)h_j}{p+\phi_0 h_j}\right)^{x_j}\left(1-\frac{\left(\phi_1-\phi_0\right)h_j}{1-p-\phi_0 h_j}\right)^{1-x_j}} = \frac{1}{\pi_t}\end{aligned}$$

where

$$\pi_t = \prod_{j=1}^t \left(1+\frac{\left(\phi_1-\phi_0\right)h_j}{p+\phi_0 h_j}\right)^{x_j}\left(1-\frac{\left(\phi_1-\phi_0\right)h_j}{1-p-\phi_0 h_j}\right)^{1-x_j}. \tag{8.13}$$

We conclude that the expression for the posterior in (8.12) can be simplified to

$$\mu(x^t) = \left(1+\frac{1-\mu_0}{\mu_0}\frac{1}{\pi_t}\right)^{-1} \tag{8.14}$$

with π_t given in (8.13).

We will show that there exists $\rho > 0$ such that

$$\pi_t \geq e^{\rho t} \text{ for all but finitely many } t \text{ } a.s., \text{ when } \theta = \theta_1 \tag{8.15}$$

and that

$$\pi_t \leq e^{-\rho t} \text{ for all but finitely many } t \text{ } a.s., \text{ when } \theta = \theta_0. \tag{8.16}$$

We may express π_t conveniently it terms of the function

$$q\left(x,y\right) = x\log\left(\frac{p+\phi_1 y}{p+\phi_0 y}\right) + \left(1-x\right)\log\left(\frac{1-p-\phi_1 y}{1-p-\phi_0 y}\right) \tag{8.17}$$

by employing the identities

$$\begin{aligned}\pi_t &= e^{\log \pi_t} = \exp\left\{\sum_{j=1}^t \left[x_j \log\left(1+\frac{\left(\phi_1-\phi_0\right)h_j}{p+\phi_0 h_j}\right) + \left(1-x_j\right)\log\left(1-\frac{\left(\phi_1-\phi_0\right)h_j}{1-p-\phi_0 h_j}\right)\right]\right\} \\ &= \exp\left\{\sum_{j=1}^t q\left(x_j,h_j\right)\right\} = \exp\left\{t\tilde{\pi}_t\right\}\end{aligned} \tag{8.18}$$

where

$$\tilde{\pi}_t = \frac{1}{t}\sum_{j=1}^{t} q\left(x_j, h_j\right) = \frac{1}{t}\sum_{j=1}^{t} \mathbb{E}_{\theta_i}\left(q\left(x_j, h_j\right)\middle| x^{j-1}\right) + \frac{1}{t}M_t \tag{8.19}$$

with $\mathbb{E}_{\theta_i}\left(\cdot\middle| x^{j-1}\right)$ denoting the conditional expectation when $\theta = \theta_i$ for $i \in \{0,1\}$ and

$$M_t = \sum_{j=1}^{t}\left[q\left(x_j, h_j\right) - \mathbb{E}_{\theta_i}\left(q\left(x_j, h_j\right)\middle| x^{j-1}\right)\right]. \tag{8.20}$$

Observe that $M_t/t \to_{a.s.} 0$ by Lemma 2 since the monotonicity of the functions $y \mapsto \log\left(\frac{p+\phi_1 y}{p+\phi_0 y}\right)$ and $y \mapsto \log\left(\frac{1-p-\phi_1 y}{1-p-\phi_0 y}\right)$ and the binary property of x_i imply that

$$\sup_{j\geq 1}\left|q\left(x_j, h_j\right)\right| \leq \max\left\{\left|q\left(1, -p/\phi_1 + \delta\right)\right|, \left|q\left(1, \left(1-p\right)/\phi_1 - \delta\right)\right|, \left|q\left(0, -p/\phi_1 + \delta\right)\right|, \left|q\left(0, \left(1-p\right)/\phi_1 - \delta\right)\right|\right\} \quad a.s.$$

by (5.3), since $\mathbf{Im}\left(h_j\right) = \mathbf{Im}\left(f_j\right)$ by definition of the h_j sequence in (6.1). Substituting $M_t/t = o_{a.s.}\left(1\right)$ to (8.19) shows (6.3) and (6.2) follows by combining the identities (8.12)-(8.14) with (8.18) and (8.19).

Since each h_j is $\sigma\left(x_{j-1}, ..., x_1\right)$-measurable, the conditional expectation in (8.19) can be evaluated for $i \in \{0,1\}$ as

$$\begin{aligned}\mathbb{E}_{\theta_i}\left(q\left(x_j, h_j\right)\middle| x^{j-1}\right) &= \mathbb{P}\left(x_j = 1|x^{j-1}, \theta_i\right)\log\left(\frac{p+\phi_1 h_j}{p+\phi_0 h_j}\right) + \mathbb{P}\left(x_j = 0|x^{j-1}, \theta_i\right)\log\left(\frac{1-p-\phi_1 h_j}{1-p-\phi_0 h_j}\right) \\ &= \psi_1\left(h_j\right)\mathbf{1}\left\{i=1\right\} + \psi_0\left(h_j\right)\mathbf{1}\left\{i=0\right\} \end{aligned} \tag{8.21}$$

where, using the formulae for $\mathbb{P}\left(\cdot|x^{j-1}, \theta_1\right)$ and $\mathbb{P}\left(\cdot|x^{j-1}, \theta_0\right)$ in (5.2) and (5.1), the functions $\psi_0, \psi_1 : \mathbf{Im}\left(f_j\right) \to \mathbb{R}$ are given by

$$\psi_1\left(y\right) = \left(p+\phi_1 y\right)\log\left(\frac{p+\phi_1 y}{p+\phi_0 y}\right) + \left(1-p-\phi_1 y\right)\log\left(\frac{1-p-\phi_1 y}{1-p-\phi_0 y}\right) \tag{8.22}$$

and

$$\psi_0\left(y\right) = \left(p+\phi_0 y\right)\log\left(\frac{p+\phi_1 y}{p+\phi_0 y}\right) + \left(1-p-\phi_0 y\right)\log\left(\frac{1-p-\phi_1 y}{1-p-\phi_0 y}\right), \tag{8.23}$$

with $\mathbf{Im}\left(f_j\right)$ defined in (5.3).

It is easy to see that, when $\theta = \theta_1$, the function ψ_1 in (8.22) satisfies the assumption (b) of Proposition 1: $\psi_1\left(0\right) = 0$ and by computing the derivatives

$$\psi_1'\left(y\right) = \phi_1\log\left(\frac{p+\phi_1 y}{p+\phi_0 y}\right) - \phi_1\log\left(\frac{1-p-\phi_1 y}{1-p-\phi_0 y}\right) + \frac{p\left(\phi_1-\phi_0\right)}{p+\phi_0 y} - \frac{\left(1-p\right)\left(\phi_1-\phi_0\right)}{1-p-\phi_0 y}$$

$$\begin{aligned}\psi_1''(y) &= (\phi_1-\phi_0)\left[\frac{p}{p+\phi_0 y}\left(\frac{\phi_1}{p+\phi_1 y}-\frac{\phi_0}{p+\phi_0 y}\right)+\frac{1-p}{1-p-\phi_0 y}\left(\frac{\phi_1}{1-p-\phi_1 y}-\frac{\phi_0}{1-p-\phi_0 y}\right)\right] \\ &= (\phi_1-\phi_0)^2\left[\frac{p^2}{(p+\phi_1 y)(p+\phi_0 y)^2}+\frac{(1-p)^2}{(1-p-\phi_1 y)(1-p-\phi_0 y)^2}\right]\end{aligned}$$

it is readily checked that $\psi_1''(y) > 0$ for $y \in \mathbf{Im}(f_j)$, establishing the convexity of ψ_1. Also, since $\psi_1'(0) = 0$ and ψ_1' is increasing (by convexity of ψ_1), we deduce that $\psi_1'(y) > 0$ for all $y > 0$, i.e. ψ_1 is strictly increasing on $\mathbb{R}_+$.

Similary, when $\theta = \theta_0$, the function $-\psi_0$ can be seen to satisfy assumption (b) of Proposition 1: $-\psi_0(0) = 0$,

$$\psi_0'(y) = \phi_0 \log\left(\frac{p+\phi_1 y}{p+\phi_0 y}\right)+\frac{(\phi_1-\phi_0)p}{p+\phi_1 y}-\phi_0\log\left(\frac{1-p-\phi_1 y}{1-p-\phi_0 y}\right)-\frac{(\phi_1-\phi_0)(1-p)}{1-p-\phi_1 y}$$

and

$$\psi_0''(y) = -(\phi_1-\phi_0)^2\left(\frac{p^2}{(p+\phi_0 y)(p+\phi_1 y)^2}+\frac{(1-p)^2}{(1-p-\phi_0 y)(1-p-\phi_1 y)^2}\right)$$

imply that $\psi_0''(y) < 0$ for all $y \in \mathbf{Im}(f_j)$ which, combined with $\psi_0'(y)$, implies that $\psi_0'(y) < 0$ for all $y > 0$. We conclude that ψ_0 is strictly decreasing on $\mathbb{R}_+$ and concave on $\mathbf{Im}(f_j)$, completing the proof that $-\psi_0$ satisfies assumption (b) of Proposition 1.

The above properties of the functions ψ_1 and $-\psi_0$ suggest that Proposition 1 may be applied to the sample average of the sequences $\psi_1(h_j)$ and $-\psi_0(h_j)$ provided that a suitable lower bound of the form (6.4) satisfying part (a) of Proposition 1 can be derived for the conditional expectation of h_j. To obtain such a lower bound, use (8.1) and the fact that $g(\cdot,0)$ is increasing to write

$$\begin{aligned}h_j &= h_j\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\}+h_j\mathbf{1}\left\{h_{j-1} \geq \overline{h}_{j-1}\right\} \\ &= g\left(h_{j-1}, x_{j-1}\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\}\right)\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\} \\ &\quad +g\left(h_{j-1}, x_{j-1}\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\}\right)\mathbf{1}\left\{h_{j-1} \geq \overline{h}_{j-1}\right\} \\ &= g(h_{j-1}, x_{j-1})\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\}+g(h_{j-1},0)\mathbf{1}\left\{h_{j-1} \geq \overline{h}_{j-1}\right\} \\ &\geq g(h_{j-1}, x_{j-1})\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\}+g(h_*,0)\mathbf{1}\left\{h_{j-1} \geq \overline{h}_{j-1}\right\} \qquad (8.24)\end{aligned}$$

for some $h_* > 0$, since $g(h_{j-1},0) \geq g\left(\inf_{j\in\mathbb{N}} \overline{h}_{j-1}, 0\right)$ on the set $\left\{h_{j-1} \geq \overline{h}_{j-1}\right\}$ and $\inf_{j\in\mathbb{N}} \overline{h}_{j-1} \geq h_*$ for some $h_* > 0$ by (8.3). Also, since $g(\cdot,0)$ is strictly increasing and $g(0,0) = 0$, $h_* > 0$ implies that

$$g(h_*, 0) > 0. \qquad (8.25)$$

We turn to finding a lower bound for the first term in (8.24). Using (5.7) and $x_{j-1} \in \{0,1\}$ we have

$$g(0, x_{j-1}) \geq \beta x_{j-1} \qquad (8.26)$$

for some $\beta \in (0, (1-p)/\phi_1 - \delta)$. Next, on the set $\{h_{j-1} \geq 0\}$, applying the mean value theorem and using (5.7) and (8.26) gives

$$\begin{aligned} g(h_{j-1}, x_{j-1}) &\geq g(0, x_{j-1}) + h_{j-1} \inf_{z \geq 0} \frac{\partial g(z, x_{j-1})}{\partial z} \geq g(0, x_{j-1}) \\ &\geq \beta x_{j-1}. \end{aligned} \tag{8.27}$$

Finally, on the set $\{h_{j-1} < 0\}$, applying the mean value theorem and using (5.7) and (8.26) gives

$$\begin{aligned} g(h_{j-1}, x_{j-1}) &\geq g(0, x_{j-1}) + h_{j-1} \sup_{z < 0} \frac{\partial g(z, x_{j-1})}{\partial z} \\ &\geq \beta x_{j-1} + \gamma h_{j-1} \end{aligned} \tag{8.28}$$

where γ, defined in (5.8), satisfies $\gamma \in [0, 1)$. Combining the bounds in (8.27) and (8.28) we obtain

$$\begin{aligned} g(h_{j-1}, x_{j-1}) \mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\} &= g(h_{j-1}, x_{j-1}) \mathbf{1}\left\{0 \leq h_{j-1} < \overline{h}_{j-1}\right\} + g(h_{j-1}, x_{j-1}) \mathbf{1}\{h_{j-1} < 0\} \\ &\geq \beta x_{j-1} \mathbf{1}\left\{0 \leq h_{j-1} < \overline{h}_{j-1}\right\} + (\beta x_{j-1} + \gamma h_{j-1}) \mathbf{1}\{h_{j-1} < 0\} \\ &= \beta x_{j-1} \mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\} + \gamma h_{j-1} \mathbf{1}\{h_{j-1} < 0\} \\ &\geq \beta x_{j-1} \mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\} + \gamma^j h_1 \mathbf{1}\{h_1 < 0\} \end{aligned}$$

where the last inequality follows by Lemma 1(ii). Substituting back to (8.24), we obtain

$$h_j \geq \beta x_{j-1} \mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\} + g(h_*, 0) \mathbf{1}\left\{h_{j-1} \geq \overline{h}_{j-1}\right\} + \gamma^j h_1 \mathbf{1}\{h_1 < 0\}. \tag{8.29}$$

To take the conditional expectation of h_j, note first that all random variables in (8.29) apart from x_{j-1} are $\sigma(x_{j-2}, ..., x_1)$-measurable, and so the conditional expectation of x_{j-1},for $i \in \{0, 1\}$, is given by

$$\begin{aligned} \mathbb{E}_{\theta_i}\left(x_{j-1} | \, x^{j-2}\right) &= \mathbb{P}\left(x_{j-1} = 1 | \, x^{j-2}, \theta_i\right) \\ &= (p + \phi_0 h_{j-1}) \mathbf{1}\{\theta = \theta_0\} + (p + \phi_1 h_{j-1}) \mathbf{1}\{\theta = \theta_1\} \\ &= p + h_{j-1}(\phi_0 \mathbf{1}\{\theta = \theta_0\} + \phi_1 \mathbf{1}\{\theta = \theta_1\}) \\ &\geq p + \phi_1 h_{j-1} \mathbf{1}\{h_{j-1} < 0\} \\ &\geq p \mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\} + \phi_1 \gamma^{j-1} h_1 \mathbf{1}\{h_1 < 0\} \end{aligned} \tag{8.30}$$

where the first equality follows from the binary property $x_j = \mathbf{1}\{x_j = 1\}$, the second equality follows from the form of the conditional distributions (5.1) and (5.2) and the last inequality follows by Lemma 1(ii). Taking conditional expectations in (8.29) and using the

$\mathcal{F}_{j-2}$-measurability of $\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\}$ and h_1 and the bound in (8.30) we obtain

$$\begin{aligned}\mathbb{E}_{\theta_i}\left(h_j|\, x^{j-2}\right) &\geq \beta\mathbb{E}_{\theta_j}\left(x_{j-1}|\, x^{j-2}\right)\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\} + g\left(h_*, 0\right)\mathbf{1}\left\{h_{j-1} \geq \overline{h}_{j-1}\right\} + \gamma^j h_1 \mathbf{1}\left\{h_1 < 0\right\} \\ &\geq \beta p\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\} + g\left(h_*, 0\right)\mathbf{1}\left\{h_{j-1} \geq \overline{h}_{j-1}\right\} + \gamma^{j-1} h_1 \mathbf{1}\left\{h_1 < 0\right\}\left(\phi_1\beta + \gamma\right).\end{aligned}$$

Since $h_1 < 0$ and $\mathbf{1}\left\{h_{j-1} < \overline{h}_{j-1}\right\} + \mathbf{1}\left\{h_{j-1} \geq \overline{h}_{j-1}\right\} = 1$, the last bound implies that

$$\mathbb{E}_{\theta_i}\left(h_j|\, x^{j-2}\right) \geq \min\left\{\beta p, g\left(h_*, 0\right)\right\} + \left(\phi_1\beta + \gamma\right)\gamma^{j-1} h_1 \mathbf{1}\left\{h_1 < 0\right\} \quad \forall\ j \ \ a.s. \tag{8.31}$$

for $i \in \{0, 1\}$ and constants $\beta \in \left(0, \left(1-p\right)/\phi_1 - \delta\right)$ and $\gamma \in [0, 1)$. The range of β and γ together with (8.25) imply that $\kappa := \min\left\{\beta p, g\left(h_*, 0\right)\right\} > 0$, $\xi_j := \left(\phi_1\beta + \gamma\right)\gamma^{j-1} h_1 \mathbf{1}\left\{h_1 < 0\right\} \to_{a.s.} 0$. Hence, (8.31) and the properties of the functions ψ_1 and ψ_0 implies that Proposition applies with $Y_j = h_j$ and $\psi \in \{\psi_1, -\psi_0\}$, giving:

$$\liminf_{t\to\infty} \frac{1}{t}\sum_{j=1}^{t} \psi_1\left(h_j\right) > 0 \ a.s. \text{ and } \limsup_{t\to\infty} \frac{1}{t}\sum_{j=1}^{t} \psi_0\left(h_j\right) < 0 \ a.s. \tag{8.32}$$

where the second statement is implied by $\liminf_{t\to\infty} t^{-1}\sum_{j=1}^{t}\left(-\psi_0\left(h_j\right)\right) > 0$ $a.s.$ and the standard relationship $\limsup_{t\to\infty} z_t = -\liminf_{t\to\infty}\left(-z_t\right)$. Combining (8.32) with (8.21), substituting into (8.19) and using the fact that $t^{-1}M_t \to_{a.s.} 0$ implies that

$$\liminf_{t\to\infty} \tilde{\pi}_t > 0 \ \ a.s. \ \text{ when } \theta = \theta_1 \tag{8.33}$$

and

$$\limsup_{t\to\infty} \tilde{\pi}_t < 0 \ \ a.s. \ \text{ when } \theta = \theta_0. \tag{8.34}$$

When $\theta = \theta_1$, (8.33) and (8.18) imply (8.15) which, in turn, implies that

$$1/\pi_t \leq e^{-\rho t} \ \text{ for all but finitely many } t \ \ a.s.$$

for some $\rho > 0$. Substituting the last bound into (8.14) and using the well-known approximation

$$\left(1 + cx\right)^{-1} = 1 + O\left(x\right) \ \text{ as } x \to 0 \tag{8.35}$$

for any scalar $c \neq 0$ to (8.14), we obtain, for all but finitely many t,

$$\mu(x^t) \geq \left(1 + \frac{1-\mu_0}{\mu_0} e^{-\rho t}\right)^{-1} = 1 + O\left(e^{-\rho t}\right) \ \ a.s.,$$

proving the first part of the theorem.

When $\theta = \theta_0$, (8.34) and (8.18) imply (8.16), which, in turn, implies that

$$1/\pi_t \geq e^{\rho t} \ \text{ for all but finitely many } t \ \ a.s.$$

Substituting the last bound into (8.14) and using and (8.35) yields, for all but finitely many t,

$$\begin{aligned}\mu(x^t) &\leq \left(1+\frac{1-\mu_0}{\mu_0}e^{\rho t}\right)^{-1} = \frac{\mu_0}{1-\mu_0}e^{-\rho t}\left(1+\frac{\mu_0}{1-\mu_0}e^{-\rho t}\right)^{-1} \\ &= \frac{\mu_0}{1-\mu_0}e^{-\rho t}\left[1+O\left(e^{-\rho t}\right)\right] = O\left(e^{-\rho t}\right) \quad a.s.,\end{aligned}$$

proving the second part of the theorem.

9 Appendix C: Proving rule (5.4)

Theorem 1 implies that optimal decision-making is such that:

$$d_t = x_t \mathbf{1}\{G(h_t, \xi_t) < 0\}$$

where

$$G(h_t,\xi_t) := \delta\{[\underline{\theta}+\xi_t\Delta\theta][g(h_t,1)-g(h_t,0)]D(h_t,\xi_t)+\Psi(h_t,\xi_t)\}-[c(h_t)-b(h_t)]$$

Consider the case when $G(h,\xi)$ is strictly increasing with $G(h_1,1)<0$ and $G(H,1)\geq 0$.[12] These two inequalities capture a decision-maker who, in the first period, would experiment (not invest in a high-cost project) if she was convinced that $\theta=1$ as long as the "stock" of past experimentation, h_1, is sufficiently low; specifically, if $h_t < h_*$ where h_* satisfies $h_1 < h_* \leq H$ and $G(h_*,1)=0$. Suppose also that $G(H,0)\geq 0$. This ensures that if the decision-maker was convinced that $\theta=0$, then she would choose to not experiment (invest in a high-cost project) only if the stock h_t was high enough, i.e. if $G(h_t,0)\geq 0$. In such an environment, a decision-maker who has imperfect information about the state of nature chooses to experiment only when an experimentation opportunity arises ($x_t=1$), and the extent of past experimentation, h_t, and the posterior, ξ_t, are sufficiently low. Observe also that the monotonicity properties of $G(\cdot)$ and that $\lambda_1\in(0,1)$ and $G(h_1,1)<0$ imply that $G(h_1,\xi(h_1,\lambda_1))<0$, which guarantees that experimentation starts at the first opportunity.

Note now that $\xi_t=\xi(h_t,\lambda_t)$ is a (weakly) increasing function of the measure of past experimentation h_t and the prior belief λ_t that $\theta=1$. This, together with $G(h_1,\xi_1)<0$ and the monotonicity properties of $G(\cdot)$, implies that $G(h,\xi(h,\lambda_t))=0$ defines implicitly a threshold stock $h_*(\lambda_t)$ that depends (weakly) negatively on λ_t, and is such that $G(h_t,\xi(h_t,\lambda_t))<0$ for all period-t measures $h_t < h_*(\lambda_t)$. We thus can rewrite the above decision rule as

$$d_t = x_t \mathbf{1}\{h_t < h_*(\lambda_t)\} \tag{9.1}$$

We note that because of our assumption that $G(h_1,1)<0$, we have that if $\lambda_t\in(0,1)$ (and

[12] $G(h,\xi)$ will be strictly increasing if, for instance, $B(h)$ is strictly decreasing, and the direct effects of varying h and ξ (i.e. while keeping $D(h,\xi)$ and $\Psi(h,\xi)$ constant) are the dominant ones.

hence $\xi(h,\lambda_t) < 1$ for all $h \geq 0$) then $h_*(\lambda_t) > h_* := h_*(1) > h_1$. The figure below depicts the various definitions and relationships in the context of an example for the functions G and ξ. In the figure, $\widetilde{h}(h_t,\lambda_t)$ is given implicitly by the solution with respect to h of $G(h,\xi(h_t,\lambda_t)) = 0$. Note that in this figure, $h_t < \widetilde{h}(h_t,\lambda_t)$ if and only if $h_t < h_*(\lambda_t)$.

We should emphasise here that the decision rule (9.1) can also be obtained, in a similar manner to the one above, from the optimal decision rule found in Bose and Makris (1998) and described above, if $G(h,\xi)$ is increasing in h and decreasing in ξ with $G(0,0) \leq 0$ and $G(h_1,\xi(h_1,\lambda_1)) < 0$. Therefore, the decision rule (9.1) may be encountered in many applications that fit the above model.

Let now, for any given sample $x = (x_1, x_2...)$, the functions $\lambda_t^*(x^{t-1})$ and $h_t^*(x^{t-1})$ be the solution with respect to λ_t and h_t of the system of difference equations

$$\lambda_t = x_{t-1}\xi(h_{t-1},\lambda_{t-1}) + [1 - x_{t-1}]\hat{\xi}(h_{t-1},\lambda_{t-1}) \tag{9.2}$$

and

$$h_t = q[x_{t-1}\mathbf{1}\{h_{t-1} < h_*(\lambda_{t-1})\} + \gamma h_{t-1}], \tag{9.3}$$

for $t \geq 2$, given the initial values $(\lambda_1, h_1) = (\mu_0, f_1(x^0))$. We thus have that $f_{t+1}(x^t) = g(f_t(x^{t-1}), d_t)$ as postulated in the main text, where $g(f,d) := q[d + \gamma f]$, and that $h_t = f_t(x^{t-1})$. It follows that the decision rule (9.1) reduces to (5.4) in the main text after also defining $\overline{f}_t(x^{t-1}) := h_*(\lambda_t^*(x^{t-1}))$ for all $t \geq 2$.

10 References

AGHION, P., P. BOLTON, C. HARRIS, and B. JULLIEN (1991), "Optimal Learning by Experimentation", *Review of Economic Studies.* 58, 621-654.

ACEMOGLOU, D., M. A. DAHLEH, I. LOBEL, and A. OZDAGLER (2011), "Bayesian learning in social networks", *Review of Economic Studies.* 78, 1201-1236.

BANERJEE, A. (1992), "A Simple Model of Herd Behavior". *Quarterly Journal of Economics.* 107, 797-817.

BERGEMANN, D., and J. VALIMAKI (2006), "Bandit Problems", in New Palgrave Dictionary of Economics, McMillan.

BERRY, D. A., and B. FRISTEDT, (1985), Bandit problems: sequential allocation of experiments. Chapman and Hall.

BIKHCHANDANI, S., D. HIRSHLEIFER, and I. WELCH (1992), "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades", *Journal of Political Economy.* 100, 992-1026.

BOSE, S., and M. MAKRIS (2018), "Investment under Unknown Learning-by-doing: Experimenting with a Restless Bandit", mimeo.

BREZZI, M.. LAI, T. L. (2000), "Incomplete Learning from Endogenous Data in Dynamic Allocation", *Econometrica*. 68, 1511-1516.

CHAMLEY, C. P. (2004), Rational Herds: Economic Models of Social Learning. Cambridge University Press.

DE GROOT, M. H. (1970), Optimal Statistical Decisions. New York: McGraw Hill.

EASLEY, D., and N. M. KIEFER, (1988), "Controlling a Stochastic Process with Unknown Parameters", *Econometrica*. 56, 1045-1064.

FERSHTMAN, D., and A. PAVAN (2019), "Searching for Arms", mimeo.

FRYER, R. G. Jr., and P. HARMS (2018), "Two-Armed Restless Bandits with Imperfect Information: Stochastic Control and Indexability", *Mathematics of Operations Research*.

GITTINS, J. C. (1989), Multi-Armed Bandit Allocation Indices, John Wiley and Sons, New York, NY.

HANN-CARUTHERS, W., V.V. MARTYNOV, and O. Tamuz (2018), "The Speed of Sequential Asymptotic Learning", *Journal of Economic Theory*. 173, 383-409.

HORNER, J., and A. SKRZYPACZ (2017), Learning, Experimentation, and Information Design. In B. Honore , A. Pakes, M. Piazzesi, and L. Samuelson (Eds.), Advances in Economics and Econometrics: Eleventh World Congress (Econometric Society Monographs, pp. 63-98). Cambridge: Cambridge University Press.

KELLER, G., and S. RADY (2010), "Strategic Experimentation with Poisson Bandits", *Theoretical Economics*. 5, 275-311.

KELLER, G., S. RADY, and M. CRIPPS (2005), "Strategic Experimentation with Exponential Bandits", *Econometrica*. 73, 39-68.

KHILSTROM, R., L. MIRMAN, and A. POSTLEWAITE (1984), "Experimental Consumption and the ' Rothschild Effect ' " in Bayesian Models of Economic Theory. Edited by M. Boyer and R. Khilstrom. Amsterdam: Elsevier.

KLEIN, N., and S. RADY (2011), "Negatively Correlated Bandits", *Review of Economic Studies*. 78, 693-792.

MAHAJAN, A., and D. TENEKETZIS (2008), "Multi-armed Bandit Problems", in Foundations and Applications of Sensor Management, edited by A. O. Hero III, D. A. Castanon, D. Cochran and K. Kastella, Springer–Verlag.

NINO-MORA, J. (2001), "Restless bandits, partial conservation laws, and indexability", *Advances in Applied Probability.* 33, 76-98.

ROBBINS, H. (1952), "Some aspects of the sequential design of experiments'", *Bulletin of the American Mathematical Society.* 58, 527-535.

ROSENBERG, D., E. SOLAN, and N. VIEILLE (2007), "Social Learning in One-Armed Bandit Problems", *Econometrica.* 75, 1591-1611.

SAFRONOV, M. (2017), "Experimentation and Learning-by-Doing", mimeo.

SHIRYAYEV, A.N. (1995), Probability, Springer-Verlag.

SMITH, L., and P.N. SORENSEN (2011), "Observational Learning", New Palgrave Dictionary of Economics, McMillan.

URGUN, C., (2019), "Restless Contracting", mimeo.

WILLIAMS, D. (1991), Probability with Martingales, Cambridge University Press.

WHITTLE, P. (1988),"Restless bandits: Activity allocation in a changing world", *Journal of Applied Probability.* 25, 287-298.