Investment Timing with Incomplete Information and Multiple Means of Learning

J. Michael Harrison
Stanford Graduate School of Business, Stanford University, CA 94305, Harrison_Michael@gsb.stanford.edu

Nur Sunar
Kenan-Flagler Business School, The University of North Carolina at Chapel Hill, NC 27599, nursunar@alumni.gsb.stanford.edu

Motivated by applications in oil exploration and investment, we consider a firm that can use one of several costly learning modes to dynamically reduce uncertainty about the unknown value of a project. Each learning mode incurs cost at a particular rate and provides information of a particular quality. In addition to dynamic decisions about its learning mode, the firm must decide when to stop learning and either invest or abandon the project. Using a continuous-time Bayesian framework, and assuming a binary prior distribution for the project’s unknown value, we solve both the discounted and undiscounted versions of this problem. In the undiscounted case, the optimal learning policy is to choose the mode that has the smallest cost per signal quality. When the discount rate is strictly positive, we prove that an optimal learning and investment policy can be summarized by a small number of critical values, and the firm only uses learning modes that lie on a certain convex envelope in cost-rate-versus-signal-quality space. We extend our analysis to consider a firm that can choose multiple learning modes simultaneously, which requires the analysis of both investment timing and dynamic subset selection decisions. We solve both the discounted and undiscounted versions of this problem, and explicitly identify sets of learning modes that are used under the optimal policy.

Key words: optimal control, optimal stopping, costly learning, Bayesian sequential hypothesis testing, subset selection

1. Introduction

The problem we study involves the timing of an investment decision, possibly preceded by costly learning about the economic value of the project under consideration. The value of the project is modeled as a binary random variable, the value of which is initially unknown to the firm. The novel feature of our formulation is the dynamic choice of a learning mode (for example, an exploration technology) based on observations to date. Each learning mode incurs cost at a particular rate and provides information of a particular quality.

Our theoretical framework is related to exploration and investment decisions of oil (or more broadly, natural resources) exploration and production companies. An oil exploration and production (E&P) company usually does not know the economic value of a site that is under consideration
for drilling. Of course, the company can use alternative exploration technologies, including seismic surveys, geological surveys, and various types of exploratory drilling, to learn about the value of the site. However, using such technologies is costly: leading oil E&P companies spend billions of dollars on exploration activities each year. For instance, in 2010 alone, Exxon Mobil Corporation spent about 3 billion U.S. dollars on exploration activities whereas the corresponding figure for PetroBras, the Brazilian oil giant, was around 5 billion U.S. dollars, which was more than 25% of its annual profit (Ernst & Young 2011, BBC 2011). Incorrectly assessing the economic value of potential production sites, or exploring potential sites with inappropriate technologies, may drastically reduce the profits earned by E&P companies. For instance, in 2011, the Edinburgh-based energy company Cairn had to quit its oil exploration activities in Greenland after spending $600 million on fruitless exploration (Bloomberg 2011). So, there is significant practical value in creating guidelines for efficient natural resource exploration and investment decisions.

There does not exist a generally accepted conceptual framework for E&P technology management decisions, despite their importance; see Bickel and Smith (2006, 2008), Brown and Smith (2012) and references therein for more information on the oil exploration literature. The results in this paper provide insights about how an E&P company can use alternative technologies to optimally explore a specific site, and when it should invest in (or abandon) the site. Although our motivating application is oil (or more broadly, natural resources) exploration and investment, we provide a general theoretical framework that has potential relevance to problems arising in other contexts as well, ranging from research and development projects to alternative marketing strategies before a new product launch.

1.1. Main Results

We adopt a Bayesian model with the following binary structure: the project is either favorable or unfavorable, and the firm does not know initially which case prevails; if the project is favorable and the firm decides to invest, then the net present value of cash flows from the project is a known positive quantity; but if the project is unfavorable and the firm decides to invest, then the net present value of such cash flows is a known negative quantity. The state of the firm’s decision problem at any time before termination is summarized by the posterior belief that the project is favorable, hereafter referred to as the firm’s posterior belief, given exploratory information gathered up to that time.

The continuation region is the interval of values for the posterior belief over which continued learning is optimal; below the left endpoint of that interval the optimal action is to abandon the project, and above the right endpoint it is optimal to invest immediately. We call those two
endpoints the lower critical belief and the upper critical belief. Those critical beliefs are distinct, so the continuation region is always non-empty.

We consider two different problem formulations, one with discounting and one without. In the undiscounted case, we derive explicit formulas for the upper and lower critical beliefs, and show that the optimal policy uses a single learning mode throughout the continuation region. Roughly speaking, that best mode is the one that minimizes cost per signal quality. In the discounted formulation, we prove that the optimal policy is described by a small number of critical values: those critical values divide the continuation region into sub-intervals in which different learning modes are used. We prove that the firm only chooses modes that lie on a certain lower convex envelope in cost-rate-versus-signal-quality space. In the left-most of those sub-intervals (that is, when the firm’s posterior belief is lowest), the optimal learning mode is the one with the lowest cost per signal quality; as the posterior belief gets larger, the optimal policy chooses learning modes with successively higher signal quality (faster learning) and with successively higher cost per signal quality.

In the undiscounted problem, we explicitly quantify the effect of problem parameters on optimal critical beliefs. We prove that both critical beliefs decrease if either (i) the net reward from a favorable project is increased, or (ii) the net loss from an unfavorable project is decreased. Perhaps surprisingly, the width of the optimal continuation region does not change monotonically with the net reward; the widest continuation region is achieved with a moderate value of that parameter. In both discounted and undiscounted problems, we show that an increase in the volatility of observations results in a decrease (respectively, increase) in the upper (respectively, lower) critical belief.

We extend our analysis to a setting where the firm is allowed to choose multiple learning modes simultaneously. When the discount rate is zero, the firm chooses the set that has the smallest cost per signal quality. In the case where measurement errors under different learning modes are uncorrelated, that set has just one element, namely, the learning mode with the smallest cost per signal quality. When the discount rate is strictly positive, we prove that the firm only uses sets that lie on a lower convex envelope in cost-rate-versus-signal-quality space. Identifying sets on that convex envelope via exhaustive enumeration might be challenging, especially when the number of available learning modes is large. We explicitly identify sets that lie on the lower convex envelope when measurement errors under different learning modes are uncorrelated.

1.2. Literature Review

Our paper contributes to the statistics literature by introducing and analyzing a problem of sequential hypothesis testing with alternative learning modes. In very basic terms, sequential hypothesis
testing is sequential application of a statistical test that allows three decisions each time a new observation is made: 1) rejecting the null hypothesis, 2) accepting the null hypothesis, or 3) continuing the test by making another observation. This is a pure optimal stopping problem, with no decisions to be made about how sampling will be done if option 3 is chosen. \textcite{wald1945} pioneered sequential hypothesis testing theory, introducing the sequential probability ratio test (or likelihood ratio test) in a discrete time setting. \textcite{siegmund1985} analyzes sequential testing of alternative hypotheses about the drift of a Brownian Motion, where the drift can take two values. In both these studies, the optimal stopping rule is a two-sided threshold policy: the decision maker rejects the null hypothesis if the likelihood ratio exceeds the upper threshold, accepts it if the likelihood ratio is below a lower threshold, and continues experimenting if the likelihood ratio lies between those two thresholds. The two thresholds are determined from the desired size and power of the statistical test. \textcite{shiryaev1967} and \textcite{peskir2006} study optimal sequential hypothesis testing about the drift of a Brownian Motion, assuming that the drift can take two possible values, and using a Bayesian formulation. Their goal is to find a stopping rule that minimizes the expected sum of three undiscounted costs: sampling costs that continue until a decision is made, the cost of a type I error, and the cost of a type II error. \textcite{kwon2011} extend \textcite{shiryaev1967} by introducing discounting. Again, all of these models study pure optimal stopping problems. To the best of our knowledge, ours is the first paper that studies sequential hypothesis testing with alternative learning modes or learning sets.

There are a few papers in the real options literature that incorporate a learning aspect in their problem formulations. \textcite{decamps2005, decamps2009, klein2009} all analyze optimal investment timing for a firm when the project value evolves as a geometric Brownian Motion, and the drift of the Brownian motion is unknown to the firm. \textcite{grenadier2010} solve for the optimal investment rule when the firm can learn if received cash flow shocks are permanent. Again, all of these papers focus on pure optimal stopping problems. A more comprehensive literature review on real options and related fields can be found in \textcite{grenadier2010}.

Our paper also contributes to the literature in applied probability that studies mixed stochastic control and optimal stopping. \textcite{davis1994} solve a singular control problem with discretionary stopping. \textcite{karatzas1999} characterize the optimal stopping rule and optimal drift-variance control for a diffusion on a fixed interval, with absorption at the interval’s endpoints. \textcite{karatzas2000} study the optimal consumption-investment problem with discretionary stopping.

In the operations management literature, learning-and-earning tradeoffs have been studied in the context of dynamic pricing with demand model uncertainty. Because explicit characterization
of an optimal policy is usually impossible in that setting, the main focus of this literature is to propose and analyze simple and well-performing heuristics. Some noteworthy recent papers in that literature are by Aviv and Pazgal (2005), Araman and Caldentey (2009), Besbes and Zeevi (2009), Farias and Roy (2010), Harrison et al. (2012) and Keskin and Zeevi (2012). Unlike those papers, we derive a truly optimal policy for our problem, which differs in essential ways from dynamic pricing problems.

Our paper is also related to the economics literature that involves a tradeoff between learning and earning. Keller and Rady (1999) study the structural properties of an optimal experimentation plan for a monopolistic firm whose demand model changes in a Markovian fashion. Moscarini and Smith (2001) show the existence of and conduct the sensitivity analysis on an optimal experimentation plan in a setting where experimentation incurs a particular cost and affects the variance of signals in a particular way. Bolton and Harris (1999) analyze a two-armed bandit problem with multiple agents where each agent solves the same experimentation problem. Interested readers can find a comprehensive economics literature review on the learning-versus-earning tradeoff in Bergemann and Valimaki (2008).

The organization of the paper is as follows. Section 2 describes the model to be studied, while Section 3 states our main results and the insights they provide. Section 4 extends the analysis to firms that can choose multiple learning modes simultaneously. Section 5 discusses two particular extensions of the models analyzed in Sections 3 and 4. Section 6 summarizes our main results. Proofs of Propositions 3 through 8 appear in Appendices A–E. Proofs of all other propositions, and of various supplementary lemmas, can be found in the electronic companion. (The proofs relegated to the electronic companion are either straightforward analyses having no stochastic element or variants of standard arguments in stochastic differential equation theory and filtering theory.)

2. Model

In our formulation of the firm’s problem, a learning strategy and the information it generates will be modeled by processes that go on forever, but a complete strategy for the firm consists of a learning strategy plus a stopping time \( \tau \), and only learning decisions over the interval \([0, \tau]\) will ultimately be relevant. That is, the unlimited time horizon for the learning process is just a mathematical convenience.

We take as given a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) on which are defined a standard Brownian Motion \(W = \{W_t, t \geq 0\}\) and a binary random variable \(\theta\) that is independent of \(W\). For each \(t \geq 0\), let \(\mathcal{F}_t\) be the smallest sub-\(\sigma\)-algebra of \(\mathcal{F}\) with respect to which \(\theta\) and \(\{W_s, 0 \leq s \leq t\}\) are all measurable, augmented as usual to include all \(\mathbb{P}\)-null sets; we shall refer to \(\{\mathcal{F}_t, t \geq 0\}\) as our ambient filtration.
The event \( \{ \theta = 1 \} \) is interpreted to mean that the project under consideration is favorable. If \( \theta = 1 \) and the firm ultimately decides to invest, then a net reward of \( R > 0 \) will be received at the time of investment, but if \( \theta = 0 \) and the investment is made, a net loss of \( L > 0 \) will be incurred. (Of course, either \( R \) or \( L \) can be interpreted as the expected present value of a cash flow stream.)

In the development to follow we denote by \( k \in \{ 0, 1 \} \) a generic value of \( \theta \), and set \( \pi_0 = \mathbb{P}(\theta = 1) \), assuming \( 0 < \pi_0 < 1 \). For reasons that will become clear, \( \pi_0 \) is called the prior probability of a favorable project, or prior belief.

We denote by \( \mathcal{M} = \{1, 2, \ldots, N\} \) the finite set of potential learning modes available to the firm. There is associated with each mode \( j \in \mathcal{M} \) a cost rate \( c(j) > 0 \), and a drift rate \( \mu(j,k) \) for each \( k \in \{0, 1\} \) such that \( \mu(j,1) > \mu(j,0) \) for each \( j \in \mathcal{M} \). A learning strategy is defined as a right continuous process \( M = \{M_t, t \geq 0 \} \) on \((\Omega, \mathcal{F}, \mathbb{P})\) that takes values in \( \mathcal{M} \); in the obvious way, we interpret \( M_t \) as the learning mode chosen by the firm at time \( t \). A learning strategy \( M \) determines an observation process \( X = \{X_t, t \geq 0 \} \) via the relationship

\[
X_t = \int_0^t \mu(M_s, \theta) ds + \sigma W_t, \quad t \geq 0,
\]

where \( X_0 = 0 \) by convention. The integrals involved in (1) are well defined in the Lebesgue sense for almost all \( \omega \in \Omega \), and the observation process \( X \) is obviously continuous. Denoting by \( \{\mathcal{F}^X_t, t \geq 0 \} \) the filtration generated by \( X \) (that is, the smallest increasing family of sub-\( \sigma \)-algebras that is \( \mathbb{P} \)-complete and right-continuous, and to which \( X \) is adapted), we say that \( M \) is an admissible learning strategy if it is adapted to \( \{\mathcal{F}^X_t, t \geq 0 \} \). That is, a learning strategy is called admissible if it is adapted to the filtration that is generated by its own observation process, in which case we define the posterior probability assessments or posterior beliefs

\[
\pi_t = \mathbb{P}(\theta = 1|\mathcal{F}^X_t), \quad t \geq 0.
\]

The following is a standard result in filtering theory; see Liptser and Shiryayev (1977), Theorem 9.1.

**Proposition 1.** The posterior belief process \( \pi = \{\pi_t, t \geq 0 \} \) satisfies the stochastic differential equation

\[
d\pi_t = \frac{1}{\sigma}[\mu(M_t,1) - \mu(M_t,0)]\pi_t(1 - \pi_t) \, dZ_t, \quad t \geq 0,
\]

where \( Z = \{Z_t, t \geq 0 \} \) is a standard Brownian motion with respect to \( \{\mathcal{F}^X_t, t \geq 0 \} \). Specifically,

\[
Z_t = \frac{1}{\sigma}X_t - \frac{1}{\sigma} \int_0^t [\pi_s \mu(M_s,1) + (1 - \pi_s)\mu(M_s,0)] ds, \quad t \geq 0.
\]
Continuing with the model formulation, we define an admissible policy as a pair \((M, \tau)\) where \(M\) is an admissible learning strategy and \(\tau\) is a finite stopping time with respect to \(\{F^X_t, t \geq 0\}\). The expected net profit under an admissible policy \((M, \tau)\) is defined to be
\[
E \left[ e^{-\lambda \tau} g(\pi_{\tau}) - \int_0^\tau e^{-\lambda t} c(M_t) \, dt \right],
\]
where \(\pi\) is the posterior belief process corresponding to \(M\) via (3), \(\lambda \geq 0\) is a given discount rate, and \(g : [0, 1] \rightarrow \mathbb{R}\) is defined via
\[
g(y) = [Ry - L(1 - y)]^+, \quad 0 \leq y \leq 1.
\]
The rationale for this definition is as follows. Suppose the firm stops learning at time \(\tau\). If it were to invest at that time, then its conditional expected net return from the investment would be \(R\pi_{\tau} - L(1 - \pi_{\tau})\); if that conditional expectation is negative, then the firm will choose to abandon the project rather than invest. In the obvious way, an admissible policy for the firm is said to be optimal if it maximizes the expected net profit (5).

3. Analysis
Because the posterior belief \(\pi_t\) summarizes all that is relevant in history for purposes of decision making at time \(t\) and afterward, intuition suggests that attention can be restricted to policies where \(M_t\) is a deterministic function of \(\pi_t\), and where \(\tau\) is the first time at which \(\pi\) enters some stopping set. Even more specifically, we shall focus our attention on a particular parametric family of policies, eventually showing that there exists an optimal policy in this family. To define that parametric family of policies, we begin by defining the signal quality or information quality of a learning mode \(j \in \mathcal{M}\) as
\[
q(j) = (\mu(j, 1) - \mu(j, 0))^2 / \sigma^2;
\]
one may describe \(|\mu(j, 1) - \mu(j, 0)| / \sigma\) as the contrast-to-noise ratio of learning mode \(j \in \mathcal{M}\). Next, we observe the following: there exist an index \(n \leq N\), and a numbering of learning modes such that
\[
0 < q(1) < \ldots < q(n) = \max_{1 \leq j \leq N} q(j),
\]
\[
0 < \frac{c(1)}{q(1)} < \frac{c(2) - c(1)}{q(2) - q(1)} < \ldots < \frac{c(n) - c(n - 1)}{q(n) - q(n - 1)},
\]
\[
c(i) = \phi(q(i)) \quad \text{for all } i = 1, \ldots, n, \quad \text{and}
\]
\[
c(i) \geq \phi(q(i)) \quad \text{for all } i = 1, \ldots, N,
\]
where \(\phi\) is the strictly increasing, piecewise linear and convex function on \([0, T]\) that connects the points \((0, 0), (q(1), c(1)), \ldots, (q(n), c(n))\). In words, \(\phi(\cdot)\) defines the lower convex envelope pictured in Figure 1 and we denote by \(1, 2, \ldots, n\) the learning modes \(i \in \mathcal{M}\) such that \((q(i), c(i))\) is an
extreme point of that envelope (that is, it is either the right end point or a point on the envelope at which the slope of the envelope changes.)

\[ \phi \]

Figure 1  Cost Rate versus Signal Quality for Learning Modes. The dashed line depicts the function \( \phi \) that satisfies conditions (10) and (11).

An admissible learning strategy \( M \) is called an \textit{envelope strategy} (or envelope learning strategy) if

\[ M_t = m(\pi_t), \quad t \geq 0, \]  

(12)

where \( \pi \) is the posterior belief process under \( M \) and \( m(\cdot) \) has the following structure: there exist an index \( k \in \{1, \ldots, n\} \) and critical beliefs \( \beta_0, \beta_1, \ldots, \beta_k \) such that

\[ 0 = \beta_0 < \beta_1 < \beta_2 < \ldots < \beta_k = 1, \]  

(13)

\[ m(y) = i, \quad \text{if } y \in [\beta_{i-1}, \beta_i) \text{ for } i = 1, \ldots, k. \]  

(14)

(It is immaterial how \( m(0) \) and \( m(1) \) are defined, because \( \pi \) cannot achieve those values in finite time.) That is, an envelope strategy divides the unit interval (the state space of the posterior belief process) into \( k \) sub-intervals over which learning modes \( 1, \ldots, k \) are employed. Hereafter, we refer to \( m(\cdot) \) as the \textit{learning function} that generates the envelope strategy \( M \). The following proposition shows that each learning function \( m \) generates a unique envelope learning strategy. It is proved in the electronic companion by suitably combining standard results in stochastic differential equation theory.

PROPOSITION 2. Suppose that \( k \in \{1, \ldots, n\} \) and that \( m(\cdot) \) satisfies (13) and (14). There exists a triple of processes \((M, X, \hat{\pi})\), each adapted to the ambient filtration \( \{\mathcal{F}_t, t \geq 0\} \) and each unique up
to an equivalence, that jointly satisfy (1), (3), (4) and (12). Moreover, \( \hat{\pi} \) and \( M \) are adapted to \( \{ F_t^X, t \geq 0 \} \) and \( \hat{\pi} \) is equal almost surely to the posterior belief process \( \pi \) defined by (2).

A structured policy consists of an envelope learning strategy \( M \) with parameters \( k, \beta_1, \ldots, \beta_k \) plus a stopping time of the form

\[
\tau = \inf \{ t \geq 0 : \pi_t \in [0, \ell] \cup [u, 1] \},
\]

where \( \pi \) is the posterior belief process under learning strategy \( M \), and \( \ell \) and \( u \) satisfy

\[
\ell < u, \quad 0 < \ell < \beta_1 \quad \text{and} \quad \beta_{k-1} < u < 1.
\]

Hereafter we shall refer to \( \ell \) and \( u \) as a lower critical belief and upper critical belief, respectively, and to \( (\ell, u) \) as the associated continuation region.

**Proposition 3.** Let \( k, \beta_1, \ldots, \beta_k, \ell, u \) be the parameters of a structured policy such that \( \ell \in (0, L/(L + R)) \) and \( u \in (L/(L + R), 1) \). There exists a unique continuous function \( v : [0, 1] \to \mathbb{R} \) that satisfies the following conditions.

(i) \( v(y) = 0 \) for \( y \in [0, \ell] \) and \( v(y) = Ry - L(1 - y) \) for \( y \in [u, 1] \).

(ii) \( v(\cdot) \) is continuously differentiable at each of the critical beliefs \( \beta_1, \ldots, \beta_{k-1} \) and is twice continuously differentiable at all other points in \( (\ell, u) \).

(iii) \( v(\cdot) \) satisfies the following ODE at each point \( y \in (\ell, u) \) other than the critical beliefs \( \beta_1, \ldots, \beta_{k-1} \):

\[
q(m(y))y^2(1-y)^2v''(y)/2 - c(m(y)) - \lambda v(y) = 0,
\]

where \( m \) is the learning function as in (14) with the given parameters \( \beta_1, \ldots, \beta_k \).

Moreover, \( v(\pi_0) \) is the expected net profit \[5\] for the structured policy.

Hereafter, we call \( v(\cdot) \) the value function for the structured policy, that terminology being justified by the last sentence of Proposition 3. To characterize an optimal policy, we will treat the discounted case \( (\lambda > 0) \) and the undiscounted case \( (\lambda = 0) \) separately. It will be shown below that there exists a structured policy that is optimal. Intuition suggests, and our analysis will confirm, that the learning modes selected outside the optimal continuation region \( (\ell, u) \) are in fact irrelevant.

3.1. Discounted Problem

The following proposition says that there exists a structured policy whose associated value function satisfies certain smoothness conditions along with two inequalities.

**Proposition 4.** There exists a structured policy whose value function \( v(\cdot) \) is continuously differentiable at \( \ell \) and \( u \) (the structured policy's lower and upper critical beliefs), is twice continuously
differentiable at each of the intermediate critical beliefs $\beta_1, \ldots, \beta_{k-1}$ and further satisfies the following inequalities:

$$v(y) \geq g(y), \quad y \in (0, 1),$$

$$-c(j) + q(j)y^2(1-y)^2v''(y)/2 \leq \lambda v(y), \quad \text{for all } y \in (0, 1)/\{\ell, u\} \text{ and all } j \in \mathcal{M}. \quad (18)$$

$$-c(j) + q(j)y^2(1-y)^2v''(y)/2 \leq \lambda v(y), \quad \text{for all } y \in (0, 1)/\{\ell, u\} \text{ and all } j \in \mathcal{M}. \quad (19)$$

**Remark 1.** In the proof of Proposition 4 (see Appendix B) such a policy will be constructed.

The structured policy specified in Proposition 4 will be referred to hereafter as our candidate policy. In Proposition 5, we prove that the candidate policy is optimal by showing that the value function for that policy is an upper bound on the value function for any admissible policy $(M, \tau)$.

**Proposition 5.** The structured policy specified in Proposition 4 is optimal. Under this policy, the optimal action for the firm is to abandon the project when $\pi_t \in [0, \ell]$, to invest immediately when $\pi_t \in [u, 1]$, and to continue to learn when $\pi_t \in (\ell, u)$.

The proof of Proposition 5 establishes (18) and (19) as sufficient conditions for optimality of the candidate policy, but they are also rather obviously necessary, as follows. First, if (18) did not hold, then there would exist an initial state $y$ from which the firm could achieve a higher expected net reward by stopping immediately than by following the candidate policy. Second, if (19) did not hold, then there would exist a state $y$ and a learning mode $j$ such that, starting in state $y$, the firm could achieve a higher expected net reward by ignoring the candidate policy initially, continuing to learn in mode $j$ until the first exit time from a small interval $(y-\epsilon, y+\epsilon)$, and using the candidate policy thereafter; this follows directly from Itô’s formula and the dynamics of the posterior belief process $\pi$ (see Proposition 1).

Hereafter, the function $v(\cdot)$ referred to in Proposition 4 will be called the optimal value function, and parameters $\ell$ and $u$ of the structured policy specified in Proposition 4 will be called optimal lower and upper critical beliefs. Of course, some of the inequalities in (18) and (19) are known to hold with equality by Proposition 3. To be more specific, Propositions 3 and 4 together imply that $v(\cdot)$ satisfies the following Hamilton-Jacobi-Bellman (HJB) equation, but our analysis does not make any direct use of this relationship:

$$0 = \max \left\{ g(y) - v(y), \max_{j \in \mathcal{M}} \left\{ -c(j) + q(j)y^2(1-y)^2v''(y)/2 - \lambda v(y) \right\} \right\}, \quad y \in (0, 1)/\{\ell, u\}.$$

Figures 2 through 4 present optimal solutions for three numerical examples in which the discount rate $\lambda$ changes and all other data remain fixed. The optimal value function is shown in the left panel, and the optimal policy parameters (that is, the optimal stopping parameters $\ell$ and $u$, and the optimal learning function $m^*$ to use in between those limits) are shown in the right panel.
According to Proposition 5, critical beliefs divide the continuation region \((\ell, u)\) into \(k\) sub-intervals in which the optimal policy chooses different learning modes. As the firm’s posterior belief \(\pi_t\) increases, the optimal policy chooses learning modes with higher signal quality, and with higher cost per signal quality. (In the extreme case \(k = 1\), only mode 1 is used by the optimal policy.) In Figure 2, three modes are used by the optimal policy, starting with the least costly (and least informative) mode 1 when the posterior belief is lowest, and ending with the most costly (and most informative) mode 3 when it is highest. The key factor motivating that progression is the following: as the assessed probability of a favorable project increases, the expected reward to the firm from investing gets larger, which makes the firm less patient; learning modes with higher signal quality resolve uncertainty faster, and as the assessed probability increases, the firm is willing to incur greater cost per signal quality to speed things up. As we shall see later, the modes with higher signal quality and higher cost rate are dropped from the optimal policy as the discount rate \(\lambda\) decreases toward zero; it is the firm’s impatience for profit sooner rather than later that motivates the use of costly learning modes.

### 3.1.1. Dependence of optimal critical beliefs on \(\sigma\) and \(\lambda\)
Proposition 6. The optimal upper critical belief $u$ (respectively, the optimal lower critical belief $\ell$) is a decreasing function (respectively, increasing function) of the observational volatility $\sigma$. That is, $\partial u / \partial \sigma \leq 0$ and $\partial \ell / \partial \sigma \geq 0$.

Proposition 7. The optimal upper critical belief $u$ (respectively, the optimal lower critical belief $\ell$) is a decreasing function (respectively, increasing function) of the discount rate $\lambda$. That is, $\partial u / \partial \lambda \leq 0$ and $\partial \ell / \partial \lambda \geq 0$.

From these two propositions we see that the optimal continuation region shrinks as either $\sigma$ increases or $\lambda$ increases. Increasing $\sigma$ makes observations less informative, which slows down learning in every available mode. Thus learning becomes more expensive when $\sigma$ increases, and it is intuitively plausible that the optimal continuation region would shrink, but that conclusion is certainly not obvious. Similarly, increasing the discount rate $\lambda$ motivates faster resolution of the investment decision, so Proposition 7 is intuitively plausible but again not obvious.

To fully understand the effect of changing the discount rate, one must consider not just the boundaries of the continuation region, but also the learning modes that are selected within it. Theoretical analysis of that subject is difficult, but Figures 3 and 4 show the effect of changing $\lambda$. 
in the numerical example depicted earlier in Figure 2. As \( \lambda \) decreases from 5 to 1 (corresponding to Figures 2 and 3, respectively), the intervals in which the optimal policy chooses modes 1 and 2 grow wider, while the interval associated with mode 3 shrinks. That is, smaller discount rate shifts the optimal learning strategy toward modes with lower signal quality (slower learning) and lower cost per signal quality. Moreover, as \( \lambda \) decreases further from 1 to 0.1 (corresponding to Figures 3 and 4, respectively), the optimal policy no longer uses mode 3, and the interval associated with mode 1 grows still wider.

From Proposition 5 one sees that for every positive value of \( \lambda \) there exists an optimal policy which uses mode 1 over a non-empty interval \((\ell, \beta_1)\). That is, the optimal policy referred to in Proposition 5 uses the mode with minimum cost per signal quality whenever the posterior belief \( \pi_t \) is sufficiently close to the lower critical belief \( \ell \) (but above it). We conjecture that the number of learning modes used by an optimal policy decreases monotonically as \( \lambda \) decreases (that is, modes with higher cost per signal quality are dropped successively by the optimal policy as the discount rate declines), and that only mode 1 is used when \( \lambda \) is sufficiently close to zero. The latter conjecture is consistent with results reported immediately below for the undiscounted case.
3.2. Undiscounted Problem

Recall from Section 3 that learning mode 1 is by definition a mode with the minimum cost per signal quality. For purposes of this section, it will be convenient to define

\[ \kappa = \frac{c(1)}{q(1)} = \min_{j \in M} \frac{c(j)}{q(j)}. \]

**Proposition 8.** There exist unique values \( \ell \) and \( u \) such that \( 0 < \ell < u < 1 \) and the following non-linear equations are satisfied:

\[
\log \left( \frac{u}{1-u} \right) - \frac{1}{u} - \log \left( \frac{\ell}{1-\ell} \right) + \frac{1}{\ell} - \frac{R}{2\kappa} = 0 \tag{20}
\]

\[
\frac{(1-2\ell)^2}{\ell(1-\ell)} - \frac{(1-2u)^2}{u(1-u)} - \frac{(R-L)}{2\kappa} = 0. \tag{21}
\]

Moreover, the structured policy with \( k = 1 \) and \( \ell \) and \( u \) chosen as above is optimal.

Interested readers can find the explicit formula for the optimal value function \( v \) in the proof of Proposition 8.

Under the optimal policy described in Proposition 8, the firm uses the learning mode with the minimum cost per signal quality throughout the continuation region \((\ell, u)\). The intuition is that when there is no discounting, the rate at which the firm learns about the value of its potential investment becomes irrelevant, because delaying investment has no impact on the expected net present value of the termination payoff. If there are multiple learning modes with the same minimum \( c/q \) ratio, the firm is indifferent among them and that indifference can be resolved in favor of the mode with the highest signal quality.

Using the nonlinear equations (20) and (21), we now proceed to analyze how the optimal lower and upper critical beliefs \( \ell \) and \( u \) change with problem parameters.

**Corollary 1.** The optimal lower and upper critical beliefs \( \ell \) and \( u \) both decrease strictly with the net reward \( R \), and both increase strictly with the net loss \( L \). That is,

\[ \frac{\partial \ell}{\partial R} < 0 \quad \frac{\partial u}{\partial R} < 0, \quad \frac{\partial \ell}{\partial L} > 0 \quad \text{and} \quad \frac{\partial u}{\partial L} > 0. \]

Moreover, \( \ell \) strictly increases with \( \kappa \) and \( u \) decreases with \( \kappa \). That is,

\[ \frac{\partial \ell}{\partial \kappa} > 0 \quad \text{and} \quad \frac{\partial u}{\partial \kappa} \leq 0. \]

Intuitively, as net reward \( R \) increases, the firm becomes more inclined to invest, and less inclined to abandon the project. Parallel to this intuition, one might expect to see that an increase in the net reward \( R \) would result in a narrower continuation region, because the firm would be more inclined to invest and less inclined to continue learning about the value of its potential investment. Interestingly, this plausible intuition fails under some conditions. As net reward \( R \) increases, the
continuation region \((\ell, u)\) shrinks if and only if \(\ell(1 - \ell)^2 < u(1 - u)^2\). Therefore, contrary to the intuition above, increasing the net reward \(R\) results in a wider continuation region when the optimal lower critical belief \(\ell\) is strictly larger than \(1/3\), whereas increasing the net reward \(R\) results in a narrower continuation region when the optimal upper critical belief \(u\) is smaller than \(1/3\). In fact, there exists a threshold \(R^*\) such that \(R \in (0, \min\{R^*, L\})\) implies \(\ell > 1/3\). From this and the fact that \(\ell\) and \(u\) decrease with \(R\), we see that the width of the continuation region does not change monotonically with \(R\).

This non-monotonicity can be further explained as follows. When the net reward \(R\) is sufficiently small compared to the loss \(L\), the optimal lower and upper critical beliefs are both close to 1, and the continuation region is very narrow. As the net reward \(R\) increases, the optimal lower and upper critical beliefs both decrease, but of different rates. As a result, the continuation region expands. When the net reward \(R\) is sufficiently high compared to the loss \(L\), the optimal policy is to invest immediately except when the posterior belief lies in a small neighborhood of zero. Thus, we see that the widest continuation region occurs at a moderate level of net reward \(R\).

4. Dynamic Subset Selection

The analysis up to this point allows the firm to learn about the value of the project by using just one learning mode at each time \(t\). Let us now suppose instead that the firm can use several learning modes simultaneously, calling this the “subset selection problem”, versus the “single-mode selection problem” considered earlier. To be more specific, let us assume the following.

First, if the firm chooses to use all learning modes over some period of time, then it observes an \(N\)-dimensional Brownian motion \(X\) whose individual components are independent, each component \(j\) having variance parameter \(\sigma^2\) and a drift parameter \(\mu(j, \theta)\) that depends on the unknown project value \(\theta\). That is, the \(j^{th}\) component of the \(N\)-dimensional Brownian motion \(X\) is precisely the observational process associated with the \(j^{th}\) learning mode in our single-mode selection problem. (The assumption that \(X\) has independent components with identical variance parameters is imposed primarily for ease of exposition; the extension to an arbitrary variance-covariance structure will be discussed in Section 5.) Second, if the firm chooses to employ a strict subset of the learning modes \(\mathcal{M} = \{1, \ldots, N\}\), then it is only able to observe those components of the \(N\)-dimensional Brownian motion \(X\), and choosing a subset \(I \subseteq \mathcal{M}\) incurs cost at rate

\[
c(I) = \sum_{j \in I} c(j). \tag{22}
\]

Rather than jump directly to an abstract general formulation of the subset selection problem, first consider the following relatively simple problem of dynamic inference. Suppose that the firm chooses to observe a subset \(I\) of the components of \(X\), where \(I\) has \(d\) elements. The observed
process is then a $d$-dimensional Brownian motion with the drift vector $\mu(I, \theta) = (\mu(j, \theta), j \in I)$ and covariance matrix $\sigma^2 I$, where $I$ is the $d$-dimensional identity matrix. In the electronic companion, Appendix EC.2, we show that the firm’s posterior belief process $\{\pi_t, t \geq 0\}$ evolves according to the stochastic differential equation

$$d\pi_t = \frac{||\mu(I, 1) - \mu(I, 0)||}{\sigma} \pi_t (1 - \pi_t) d\tilde{Z}_t, \quad t \geq 0,$$

where $|| \cdot ||$ is the Euclidean norm and $\tilde{Z}$ is a certain standard Brownian motion. Comparing (23) with (3) and (7), we have the following: observing the $d$ components of $X$ in the subset $I$ is informationally equivalent to observing a one dimensional Brownian signal having quality

$$q(I) = \frac{||\mu(I, 1) - \mu(I, 0)||^2}{\sigma^2} = \sum_{j \in I} q(j).$$

Figure 5  Cost Rate versus Signal Quality Graph for All Subsets. Learning mode data are the same as in Figure 1. Cost rate and signal quality of each point in this figure are calculated via relations (22) and (24), respectively.

With this preliminary, we see that the subset selection problem is actually equivalent to the single-mode selection problem, except that now what plays the role of a “learning mode” is a non-empty subset $I$ of the index set $M$, and the cost rate and the signal quality for each such subset are obtained from more basic data via formulas (22) and (24). The analysis and results developed in Section 3 then apply without change. As a result, in the undiscounted problem the optimal policy uses a single subset throughout the continuation region, namely, the subset of $M$ with the
minimum cost per signal quality; in the discounted problem the optimal policy uses only subsets of $\mathcal{M}$ whose $(q,c)$ pair is an extreme point of the lower convex envelope.

In order to identify subsets of $\mathcal{M}$ that lie on the lower convex envelope, the most direct approach is exhaustive enumeration, listing all the subsets of $\mathcal{M}$ and calculating two parameters for each such subset (a signal quality and a cost rate) using (22) and (24). However, the following proposition says that we can directly identify subsets of $\mathcal{M}$ that lie on the lower convex envelope, without resorting to exhaustive enumeration. For ease of exposition, we restrict attention to the case where all the cost-per-signal-quality ratios $c(j)/q(j)$ are distinct; interested readers can easily work out the analogous result when some of the ratios are equal.

**Proposition 9.** Assume that the ratios $\{c(j)/q(j), j \in \mathcal{M}\}$ are distinct, and let the learning modes be (re)numbered so that $c(1)/q(1) < \ldots < c(N)/q(N)$. Also, define

$$I_i \equiv \{1, \ldots, i\}, \quad \text{for } i = 1, \ldots, N.$$  

The extreme points of the lower convex envelope for the subset selection problem, ordered from lowest to highest value of the ratio $c(I)/q(I)$, are the $(q,c)$ pairs corresponding to the subsets $I_1, \ldots, I_N$.

The intuition behind Proposition 9 is as follows: As the firm’s posterior belief $\pi_t$ increases, the expected value of the potential investment increases, and in a discounted problem that makes the firm increasingly eager to resolve its investment decision quickly; that eagerness translates as a
desire for higher signal quality (that is, a desire for faster learning), even if it must be achieved at a higher cost per signal quality; and the least costly way to increase signal quality is to add whatever learning mode has the next lowest $c/q$ ratio, relative to the modes already in use.

The optimal policy in the subset selection problem may use learning modes which do not lie on the lower convex envelope for the single-mode selection problem. This is illustrated in Figure 6, which shows the optimal solution of the subset selection problem when the underlying data are the same as in Figures 1 and 2. The optimal policy uses mode 4 and mode 5 when the firm’s posterior belief is sufficiently high, and neither of these modes lie on the lower convex envelope depicted in Figure 1. What makes these expensive learning modes attractive in the subset selection problem is the firm’s ability to use them in combination with other, more cost-effective modes, providing an added boost to the speed of learning.

Propositions 8 and 9 together imply that, when the discount rate is zero in the subset selection problem, the optimal policy uses subset $I_1 = \{1\}$ throughout the continuation region. That is, the optimal policies and optimal value functions in the subset and the single-mode selection problems are exactly the same in the undiscounted case. When the discount rate is strictly positive, however, the optimal value function will increase, or at least not decrease, as we go from single-mode selection to subset selection, because the set of controls available to the firm is thereby expanded. Moreover, that change widens the optimal continuation region, or at least does not narrow it, because the means of learning available to the firm are expanded. The following proposition formalizes the latter conclusion; here the words “higher” and “lower” are used in the weak sense.

**Proposition 10.** The optimal upper critical belief $u$ (respectively, the optimal lower critical belief $\ell$) in the subset selection problem is higher (respectively, lower) than the optimal upper critical belief $u$ (respectively, the optimal lower critical belief $\ell$) in the single-mode selection problem.

## 5. Extensions

In Section 3, we assumed the observational volatility $\sigma$ to be the same for all learning modes. However, the analysis in Section 3 easily extends to the case where the observation process for learning mode $j$ has volatility parameter $\sigma(j) > 0$, $j \in M$. Specifically, one then defines the signal quality for mode $j$ to be $q(j) = (\mu(j, 1) - \mu(j, 0))^2/\sigma^2(j)$, and all results remain valid as stated.

In Section 4, the covariance matrix for the full $N$-dimensional observation process $X$ was assumed to be $\sigma^2 I$. Suppose instead that we take it to be a general $N \times N$ non-singular covariance matrix $\Sigma$, and for each non-empty subset $I$ of the index set $M$ let $\Sigma(I)$ denote the principle sub-matrix of $\Sigma$ consisting of rows and columns $j \in I$. In this more general setting, we redefine signal quality as

$$q(I) = (\mu(I, 1) - \mu(I, 0))^T \Sigma(I)^{-1} (\mu(I, 1) - \mu(I, 0)), \quad (26)$$
and the conclusions preceding Proposition 9 remain valid exactly as stated. That is, the subset selection problem actually reduces to a single-mode selection problem, and all of the results developed in Section 3 apply. However, Proposition 9, which further characterizes subsets lying on the lower convex envelope, does not extend in any obvious way to the case with general variance-covariance structure.

6. Summary

We study optimal learning and investment strategies in a setting where the firm does not know the value of the project under consideration, but can reduce uncertainty about the project value by using any of several available learning modes. We adopt a continuous-time Bayesian framework with a binary prior distribution for the unknown project value. The “signal quality” for each learning mode is defined in terms of more primitive data, and we show that an optimal policy uses only learning modes that lie on a certain lower convex envelope, or efficient frontier, in cost-rate-versus-signal-quality space. The state of the firm’s decision problem at any point in time is expressed by a certain posterior probability assessment, or posterior belief, given information gathered up to that time, and an optimal policy is characterized by just a few critical numbers: a lower critical belief \( \ell \) below which the firm should simply abandon the project; an upper critical belief \( u \) above which the firm should invest immediately, without further learning; and intermediate critical beliefs that divide the interval \((\ell, u)\) into sub-intervals within which different learning modes on the efficient frontier are used. We extend the analysis to allow simultaneous use of several different learning modes, showing that the apparently more complex “subset selection problem” actually reduces, in a certain sense, to our original single-mode selection problem.

Appendix A: Proof of Proposition 3

The required function \( v(\cdot) \) is fully specified on \((0, \ell)\) and \([u, 1)\) by condition (i) of the proposition. To construct a continuous extension to all of \((0, 1)\) that further satisfies (ii) and (iii), we observe the following: the ODE (17) has the form \( \frac{1}{2} \gamma^2(y) v''(y) = \tilde{c}(y) + \lambda v(y), \ \ell < y < u, \) where \( \tilde{c}(\cdot) \) is piecewise constant and \( \gamma(\cdot) \) is piecewise continuous, bounded and bounded away from zero on \((\ell, u)\). As boundary conditions, both \( v(\ell) \) and \( v(u) \) are specified by condition (i) of the proposition. Thus, the first statement of the proposition follows from standard results in elementary ODE theory.

With regard to the second statement, we have the following considerations. First, by Propositions 1 and 2, the posterior belief process \( \pi \) under the structured policy satisfies the stochastic differential equation

\[
\frac{d\pi_t}{\pi_t} = \gamma(\pi_t) dZ_t,
\]

where \( Z \) is a standard Brownian motion and

\[
\gamma(y) = \frac{1}{\sigma} \left[ \mu(m(y), 1) - \mu(m(y), 0) \right] y(1 - y), \quad 0 < y < 1.
\]
That is, \( \pi \) is a diffusion process with zero drift and instantaneous variance function \( \gamma^2(\cdot) \). Second, the function \( v(\cdot) \) that we have constructed is \( C^1 \) and piecewise \( C^2 \) on \([\ell, u]\) and satisfies the ODE
\[
\frac{1}{2} \gamma^2(y)v''(y) - \lambda v(y) = c(m(y))
\]
(29)
at all points \( y \in (\ell, u) \) except possibly the intermediate critical beliefs \( \beta_1, \ldots, \beta_{k-1} \). Finally, assuming \( \ell < \pi_0 < u \), our stopping time \( \tau \) is the first time at which \( \pi \) hits either \( \ell \) or \( u \), and because \( \gamma(\cdot) \) is bounded away from zero on \([\ell, u]\) it is well known that
\[
\mathbb{E}(\tau) < \infty.
\]
Furthermore, \( v'(\cdot) \) is bounded on \([\ell, u]\), so a standard argument using Itô’s formula gives us the following from (27), (28) and (29):
\[
\mathbb{E} \left[ e^{-\lambda \tau} v(\pi_\tau) \right] = v(\pi_0) + \mathbb{E} \left[ \int_0^\tau e^{-\lambda t} c(m(\pi_t)) dt \right].
\]
(31)
Because \( v \) satisfies the boundary conditions \( v(\ell) = g(\ell) \) and \( v(u) = g(u) \), we have \( v(\pi_\tau) = g(\pi_\tau) \), so (31) establishes that \( v(\pi_0) \) equals the expected net profit \( \{3\} \) for the structured policy. This proves the second statement of Proposition \( \{3\} \) for the case \( \ell < \pi_0 < u \), and if \( \pi_0 \in (0, \ell) \cup [u, 1) \) the conclusion is trivially true. \( \square \)

**Appendix B: Proof of Proposition \( \{4\} \)**

To express compactly the differential equations that appear throughout this appendix, we define the differential operator
\[
\Gamma f(y) = y^2(1 - y)^2 f''(y)/2, \quad 0 < y < 1, \quad \text{for } f \in C^2(0, 1).
\]
(32)
The following result from elementary differential equation theory will be needed in our constructive proof of Proposition \( \{4\} \).

**Lemma 1.** Let \( \alpha \in (0, 1) \) and \( q, c, \lambda > 0 \) be given constants, and let \( f \) be the unique solution of the ODE
\[
-c + q \Gamma f(y) - \lambda f(y) = 0, \quad \alpha < y < 1,
\]
(33)
with boundary conditions
\[
f(\alpha) = a \geq 0 \quad \text{and} \quad f'(\alpha) = b \geq 0.
\]
(34)
Then \( f(\cdot) \) is strictly convex and strictly increasing, with \( f(y) \uparrow \infty \) as \( y \uparrow 1 \), and \( \Gamma f(\cdot) \) is also strictly increasing, with \( \Gamma f(y) \uparrow \infty \) as \( y \uparrow 1 \).

**Proof of Lemma \( \{1\} \).** We first rewrite (33) as
\[
f''(y) = 2(\lambda f(y) + c)/ (qy^2(1 - y)^2), \quad \text{for all } y \in (\alpha, 1).
\]
(35)
Thus \( f''(\alpha+) = 2(\lambda a + c)/ (q \alpha^2(1 - \alpha)^2) > 0 \). Then, the boundary conditions (34) and the form of the differential equation (35) ensure that the following hold on all of \((\alpha, 1)\): \( f''(\cdot) > 0 \), \( f'(\cdot) > 0 \) and
\( f(\cdot) > 0 \). Moreover, (35) gives \( f''(y) \geq \hat{\kappa}(1-y)^{-2} \) for some \( \hat{\kappa} > 0 \) and all \( y \) sufficiently close to 1, implying that \( f(y) \geq -\kappa \log(1-y) \) for some \( \kappa > 0 \) and all \( y \) sufficiently close to 1. Thus \( f(y) \uparrow \infty \) as \( y \uparrow 1 \). The last sentence of the lemma is then immediate from (33) and the established properties of \( f(\cdot) \). □

To begin the proof of Proposition 4 we fix a trial value of \( \ell \in (0,1) \) and construct a function \( v_\ell \in C^2(\ell,1) \) that satisfies

\[
\max_{j \in M} \{-c(j) + q(j)\Gamma v_\ell(y) - \lambda v_\ell(y)\} = 0, \quad y \in (\ell,1),
\]

with boundary conditions

\[
v_\ell(\ell) = v_\ell'(\ell) = 0.
\]

Later the value of \( \ell \) will be chosen to ensure that \( v_\ell(\cdot) \) is tangent to the upward-sloping portion of our payoff function \( g(\cdot) \) at some point \( u \in (\ell,1) \); those values \( \ell \) and \( u \) will be the lower and upper critical beliefs of our optimal structured policy, and \( v(\cdot) = v_\ell(\cdot) \) will be the policy’s value function on \([\ell,u]\).

With \( \ell \in (0,1) \) fixed, the construction proceeds as follows. First, let \( f_1 \) be the unique \( C^2 \) solution of the ODE

\[
-c(1) + q(1)\Gamma f_1(y) - \lambda f_1(y) = 0, \quad y \in (\ell,1)
\]

with boundary conditions

\[
f_1(\ell) = f_1'(\ell) = 0.
\]

If the number of extremal control modes on the lower convex envelope in Figure 1 is just \( n = 1 \) (that is, if the lower convex envelope is just a straight line segment), then we set \( v_\ell(y) = f_1(y) \) for all \( y \in [\ell,1) \) and the construction is complete. If \( n \geq 2 \), we now claim the following: there exists a unique \( \alpha_1 \in (\ell,1) \) such that

\[
\Gamma f_1(\alpha_1) = \frac{c(2) - c(1)}{q(2) - q(1)}.
\]

For a proof of this claim, we first rewrite for ease of reference the inequalities (40) that appeared earlier in the definition of the extremal control modes \( 1, \ldots, n \):

\[
\frac{c(1)}{q(1)} < \frac{c(2) - c(1)}{q(2) - q(1)} < \cdots < \frac{c(n) - c(n-1)}{q(n) - q(n-1)}.
\]

Also, \( \Gamma f_1(\ell+) = c(1)/q(1) \) by (38) and (39), so the claim follows from the last sentence of Lemma 1.

Continuing our construction of \( v_\ell(\cdot) \) when \( n \geq 2 \), let \( f_2 \) be the unique solution of the ODE

\[
-c(2) + q(2)\Gamma f_2(y) - \lambda f_2(y) = 0, \quad y \in (\alpha_1,1),
\]

with boundary conditions

\[
f_2(\alpha_1) = f_2'(\alpha_1) = 0.
\]
with boundary conditions

\[ f_2(\alpha_1) = f_1(\alpha_1) \quad \text{and} \quad f'_2(\alpha_1) = f'_1(\alpha_1). \tag{43} \]

If \( n = 2 \), we set \( v_\ell = f_1 \) on \([\ell, \alpha_1)\) and \( v_\ell = f_2 \) on \([\alpha_1, 1)\) and the construction is complete. From \( (38), \ (40), \ (42) \) and the first equation in \( (43) \), we have that \( \Gamma f_2(\alpha_1) = \Gamma f_1(\alpha_1) \), so \( v''_\ell(\cdot) \) is continuous at \( \alpha_1 \).

If \( n > 2 \) we continue in this same way, defining critical values \( \alpha_2, \ldots, \alpha_{n-1} \) that satisfy

\[ \alpha_1 < \alpha_2 < \ldots < \alpha_{n-1} < 1, \tag{44} \]

and defining functions \( f_3, \ldots, f_n \) via the ODEs

\[ -c(j) + q(j)\Gamma f_j(y) - \lambda v_j(y) = 0, \quad y \in (\alpha_{j-1}, 1), \tag{45} \]

with boundary conditions

\[ f_j(\alpha_{j-1}) = f_{j-1}(\alpha_{j-1}) \quad \text{and} \quad f'_j(\alpha_{j-1}) = f'_{j-1}(\alpha_{j-1}) \tag{46} \]

for \( j = 3, \ldots, n \). To be more specific, the critical values \( \alpha_2, \ldots, \alpha_{n-1} \) are chosen to satisfy

\[ \Gamma f_{j-1}(\alpha_{j-1}) = \frac{c(j) - c(j-1)}{q(j) - q(j-1)} \quad \text{for} \quad j = 3, \ldots, n. \tag{47} \]

The inequalities in \( (41) \) and the last sentence of Lemma 1 ensure that critical values \( \alpha_2, \ldots, \alpha_{n-1} \) are well defined by \( (47) \) and further satisfy \( (44) \). In the end, we define \( v_\ell = f_1 \) on \([\ell, \alpha_1)\), \ldots, and \( v_\ell = f_n \) on \([\alpha_{n-1}, 1)\). Finally, it follows from the general criterion \( (17) \) for choosing successively the critical values \( \alpha_2, \ldots, \alpha_{n-1} \) that \( \Gamma f_{j-1}(\alpha_{j-1}) = \Gamma f_j(\alpha_{j-1}) \) for all \( j = 2, \ldots, n \), so our constructed function \( v_\ell(\cdot) \) is \( C^2 \) everywhere in \((\ell, 1)\). This completes our construction of \( v_\ell(\cdot) \) for fixed \( \ell \in (0, 1) \), but to establish property \( (36) \) it still must be shown that

\[ -c(j) + q(j)\Gamma v_\ell(y) - \lambda v_\ell(y) \leq 0 \quad \text{for all} \quad y \in (\ell, 1) \quad \text{and} \quad j \in \mathcal{M}. \tag{48} \]

For that purpose, it will be useful to establish the following terminology: we say that \textit{mode} \( j \) \textit{dominates} \textit{mode} \( i \) \textit{in state} \( y \) \((i, j \in \mathcal{M}, \ell < y < 1)\) if

\[ -c(i) + q(i)\Gamma v_\ell(y) - \lambda v_\ell(y) \leq -c(j) + q(j)\Gamma v_\ell(y) - \lambda v_\ell(y). \tag{49} \]

Also, setting \( \alpha_0 = \ell \) and \( \alpha_n = 1 \) for notational convenience, let

\[ \mathcal{I}_j \doteq [\alpha_{j-1}, \alpha_j) \quad \text{for} \quad j = 1, \ldots, n, \tag{50} \]
which means that $\mathcal{I}_j$ is the interval of $y$ values over which we have set $v_\ell(y) \doteq f_j(y)$. That is, \eqref{assumption} holds with equality for all $j = 1, \ldots, n$ and all $y \in \mathcal{I}_j$. Because $\Gamma f_j$ is strictly increasing (see Lemma 1), it follows from the inequalities \eqref{inequality}, and criteria \eqref{condition} and \eqref{condition2} for setting the critical values \{\alpha_j\} that, for each $j = 1, \ldots, n$, mode $j$ dominates all other extremal modes $i \in \{1, \ldots, n\}/\{j\}$ over the entire interval $\mathcal{I}_j$. Moreover, if $i \in \mathcal{M}$ is not what we have called an extremal mode (that is, if $(q(i), c(i))$ is not an extreme point of the lower convex envelope $\phi$ identified by conditions \eqref{Gamma} through \eqref{inequality}), then readers can easily verify the following: for every state $y \in [\ell, 1)$, there exists at least one extremal mode $j \in \{1, \ldots, n\}$ such that mode $j$ dominates mode $i$ in state $y$. Thus, \eqref{assumption} is established and we arrive at the following.

**Lemma 2.** For each $\ell \in (0, 1)$, the constructed function $v_\ell(\cdot)$ is in $C^2(\ell, 1)$ and satisfies \eqref{inequality} and \eqref{condition}. Moreover, the constructed function $v_\ell(\cdot)$ is strictly increasing and strictly convex on $(\ell, 1)$.

To conclude the proof of Proposition \[1\] it will be convenient to define $\hat{y} \doteq L/(L + R)$ and

$$g(y) = \begin{cases} 0 & \text{if } y \in [0, \hat{y}], \\ \hat{y} & \text{if } y \in [\hat{y}, 1]. \end{cases}$$

(51)

Then our definition \eqref{payoff} of the payoff function, $g(\cdot)$ can be re-expressed as

$$g(y) = \begin{cases} 0 & \text{if } y \in [0, \hat{y}], \\ \hat{y} & \text{if } y \in [\hat{y}, 1]. \end{cases}$$

(52)

Also, for each $\ell \in (0, 1)$ let

$$\hat{v}_\ell(y) = \begin{cases} 0 & \text{if } y \in [0, \ell), \\ v_\ell(y) & \text{if } y \in [\ell, 1]. \end{cases}$$

(53)

The following lemma states that the functions $\hat{v}_\ell(\cdot)$ are ordered as in Figure 7 with a larger value of $\ell$ giving a uniformly smaller function $\hat{v}_\ell(\cdot)$. In addition, for each $\ell \in (0, 1)$, $\hat{v}_\ell(\cdot)$ is strictly convex and strictly increasing on $(\ell, 1)$.

**Lemma 3.** The functions \{$\hat{v}_\ell(\cdot), \ell \in (0, 1)$\} are ordered, with larger $\ell$ producing uniformly smaller function $\hat{v}_\ell(\cdot)$. That is, for each $y \in (0, 1)$, $\hat{v}_\ell(y)$ is a decreasing function of $\ell$. In particular, for any $a, b$ such that $0 < a < b < 1$, $\hat{v}_a(y) > \hat{v}_b(y)$ for all $y \in [b, 1)$.

**Proof of Lemma 3.** We first prove the last sentence of the lemma. Take any $a, b$ such that $0 < a < b < 1$. For each $\ell \in \{a, b\}$, construct function $v_\ell(\cdot)$ that satisfies \eqref{inequality} and \eqref{condition} by using \eqref{assumption}, \eqref{assumption2}, \eqref{assumption3}, \eqref{assumption4}, \eqref{assumption5}, and the general criteria \eqref{condition} and \eqref{condition2} for selecting critical values \{\alpha_j\}. We will first prove following three relations:

$$v_a (b) > v_b (b) = 0,$$

(54)

$$v'_a (b) > v'_b (b) = 0,$$

(55)

$$v''_a (b) > v''_b (b+).$$

(56)
The relation (54) follows from facts that \( v_a(\cdot) \) is strictly increasing on \((a, 1)\) by Lemma 2 and \( v_a(a) = 0 \) by (39). The relation (55) holds because \( v_a(\cdot) \) is strictly convex on \((a, 1)\) by Lemma 2 and \( v'_a(a) = 0 \) by (39). With regard to inequality (56), from (38), (42) and (45) it follows that there exists some extremal mode \( j \in \{1, \ldots, n\} \) such that
\[
v''_a(b) = \frac{2(c(j) + \lambda v_a(b))}{q(j)b^2(1-b)^2}.
\] (57)

We also know from (38) and (39) that
\[
v''_b(b+) = \frac{2c(1)}{q(1)b^2(1-b)^2}.
\] (58)

This, (57), (58) and the fact that \( v_a(b) > 0 \) imply (56).

We now claim and below prove that if there exists some \( \tilde{y} \in (b, 1) \) such that \( v_a(\tilde{y}) > v_b(\tilde{y}) > 0 \), then \( v''_a(\tilde{y}) > v''_b(\tilde{y}) \). If \( v''_a(\tilde{y}) > v''_b(\tilde{y}) \) also holds, then \( v_a(\cdot) > v_b(\cdot) \) on \([\tilde{y}, 1)\). Relations (54), (55) and (56) guarantee that for \( \tilde{y} \) sufficiently close to \((a, 1)\) and \( v_a(\tilde{y}) > v_b(\tilde{y}) \) and \( v'_a(\tilde{y}) > v'_b(\tilde{y}) \). Using this, our claim and (54), it follows that \( v_a(y) > v_b(y) \) for \( y \in [b, 1) \). Because \( v_a(a) = 0 \), \( v_b(y) = 0 \) for all \( y \in [a, b] \) and \( v_a(\cdot) \) is strictly increasing on \([a, 1)\), we have that \( v_a(y) > v_b(y) \) for \( y \in (a, 1) \). Therefore, \( v_a(y) > v_b(y) \) for all \( y \in (a, 1) \). Then, the lemma follows from the definitions of \( \hat{v}_\ell(\cdot) \) in (53) for \( \ell \in \{a, b\} \).

We now proceed to show our claim above. Take any \( \tilde{y} \in (b, 1) \) such that
\[
v_a(\tilde{y}) > v_b(\tilde{y}).
\] (59)
From (38), (42), (45), and the general criteria (40) and (47) for setting critical values \{\alpha_j\}, for each \(\ell \in \{a, b\}\), \(\alpha_i(\ell)\) is the unique belief that satisfies

\[
\lambda v_i(\alpha_i(\ell)) = -c(i) + q(i)[c(i+1) - c(i)]/[q(i+1) - q(i)], \quad i = 1, \ldots, n - 1.
\]  

(60)

Observe from (41) that the right hand side of (60) is increasing in \(\ell\) for each \(\ell \in \{a, b\}\), \(\ell > 7\), when \(y\) is sufficiently close to and strictly larger than \(\hat{y}\), implying that \(\hat{\lambda}\) is strictly increasing on \((\ell, 1)\) by Lemma 2 imply that

\[
\max\{i : i \in \{0, \ldots, n-1\}, \alpha_i(a) \leq \hat{y}\} \geq \max\{i : i \in \{0, \ldots, n-1\}, \alpha_i(b) \leq \hat{y}\}. \quad (61)
\]

Therefore, from (38), (42) and (45), there exist two extremal modes \(i, j \in \{1, \ldots, n\}\) such that following relations hold:

\[
c(i)/q(i) \geq c(j)/q(j),
\]

(62)

\[
\lambda v_a(\hat{y}) = -c(i) + q(i)\Gamma a(\hat{y}), \quad \lambda v_b(\hat{y}) = -c(j) + q(j)\Gamma b(\hat{y}).
\]  

(63)

If \(i = j\), then it immediately follows from (63) and (59) that \(v''_a(\hat{y}) > v''_b(\hat{y})\). If \(i > j\), then by using (62) and (63), we have

\[
v''_a(\hat{y}) = 2(\lambda v_a(\hat{y}) + c(i))/q(i)\hat{y}^2(1 - \hat{y})^2 > 2(\lambda v_b(\hat{y}) + c(i))/q(i)\hat{y}^2(1 - \hat{y})^2 = v''_b(\hat{y}) = 2(\lambda v_b(\hat{y}) + c(j))/q(j)\hat{y}^2(1 - \hat{y})^2.
\]  

(64)

The first inequality above is from (59) and we have the second inequality from inequalities (41) and the fact that \(\lambda v_b(\hat{y}) < \lambda v_b(\alpha_j(b)) = -c(j) + q(j)(c(j+1) - c(j))/(q(j+1) - q(j))\).

We now proceed to show that there exists a unique \(\ell^* \in (0, 1)\) such that the function \(\hat{v}_\ell(\cdot)\) is tangent to \(\hat{g}(\cdot)\) at some point \(u^* \in (\hat{y}, 1]\) and prove that such \(\ell^*\) lies in the interval \((0, \hat{y})\). We begin the proof of this claim with the following observation: \(\hat{v}_\ell(\cdot)\) is tangent to \(\hat{g}(\cdot)\) at some point \(u \in (\hat{y}, 1]\) if and only if there exists some \(u \in (\hat{y}, 1]\) such that

\[
\hat{v}_\ell(u) = uR - (1 - u)L, \quad \hat{v}_\ell'(u) = R + L.
\]  

(65)

From (58), (39) and the definition (53), we have that

\[
\hat{v}_\ell''(\ell+) = 2c(1)/(q(1)\ell^2(1 - \ell)^2).
\]  

(66)

Therefore, if we take \(\ell > 0\) sufficiently close to zero, then \(\hat{v}_\ell''(\ell+) \geq \hat{\kappa}/\ell^2\) for some constant \(\hat{\kappa} > 0\), implying that \(\hat{v}_\ell(y) > \hat{\kappa}(y - \ell)/\ell^2 > R + L\) and \(\hat{v}_\ell(y) > \hat{\kappa}(y - \ell)^2/(2\ell^2)\) for \(\ell\) sufficiently close to zero and \(y\) sufficiently close to and strictly larger than \(\ell\). As a result, as depicted by curve \(i\) in Figure 7 when \(\ell > 0\) is sufficiently close to zero, \(\hat{v}_\ell(\cdot)\) does not intersect with \(\hat{g}(\cdot)\), that is, \(\hat{v}_\ell(y) > \hat{g}(y)\) for \(y \in (0, 1]\). If we take \(\ell = \hat{y}\), \(\hat{v}_\ell(\cdot)\) intersects \(\hat{g}(\cdot)\) twice because \(\hat{v}_\ell(y) \uparrow \infty\) as \(y \uparrow 1\) and \(\hat{v}_\ell(\hat{y}) = \hat{g}(\hat{y})\).

Then, from Lemma 3 and the fact that for each \(y \in (\ell, 1)\) \(\hat{v}_\ell(y)\) is continuous in \(\ell\), it follows that
there exists a unique $\ell^* \in (0, \hat{y})$ that guarantees the existence of belief $u^* \in (\hat{y}, 1)$ such that $\hat{v}_{\ell^*}(\cdot)$ satisfies (65). By Lemma 2, the function $\hat{v}_{\ell^*}(\cdot)$ is in $C^2(\ell^*, 1)$, and satisfies (36) and (37) with $\ell = \ell^*$. The argument explained above to prove the existence of a unique $\ell^*$ is depicted in Figure 7.

Denoting by $\alpha_1(\ell^*), \ldots, \alpha_{n-1}(\ell^*)$ the critical values chosen by using boundary conditions (37) with $\ell = \ell^*$ and the general criterion (47) for $j = 2, \ldots, n$, we define the function $v(\cdot)$ and parameters $k, \beta_1, \ldots, \beta_k$ as following:

$$v(y) = \begin{cases} \hat{v}_{\ell^*}(y) & \text{if } y \in [0, u^*) \\ g(y) = yR - (1 - y)L & \text{if } y \in [u^*, 1] \end{cases}$$

$$k = \max\{j \in \{1, \ldots, n - 1\} : \alpha_j(\ell^*) \leq u^*\} + 1 \quad \text{if } \alpha_1(\ell^*) \in (\ell^*, u^*)$$

$$\beta_i = \alpha_i, \quad i = 1, \ldots, k - 1, \quad \text{and} \quad \beta_k = 1.$$  

Then, $v(\cdot)$ is continuously differentiable on $(0, 1)$ and twice continuously differentiable on $(\ell^*, u^*)$.

We already know from construction of $v_{\ell^*}(\cdot)$ that $v(\cdot)$ satisfies (17) on $(\ell^*, u^*)$ with learning function $m$ (as in (44)) that uses parameters $k, \beta_1, \ldots, \beta_k$ defined above. Therefore, from Proposition 8 it follows that $v(\cdot)$ is the value function (that is, the expected net profit of the firm) under the structured policy with parameters $\ell^*, u^*, k, \beta_1, \ldots, \beta_k$. By construction in (67) through (69), $v(\cdot)$ satisfies (18) and (19). \hfill \square

**Appendix C: Proof of Proposition 5**

**Proof of Proposition 5.** We want to first establish the following result. The proof of the following lemma is a trivial application of Itô’s lemma, and so its proof is omitted.

**Lemma 4.** Let $\{\pi_t, t \geq 0\}$ be the belief process under arbitrary learning strategy $\{M_t, t \geq 0\}$, and let $f : [0, 1] \to \mathbb{R}$ be $C^1$ and piecewise $C^2$. Finally, let $T$ be a stopping time such that $\mathbb{E}(T) < \infty$. Then,

$$\mathbb{E}\left[e^{-\lambda T} f(\pi_T)\right] = f(\pi_0) + \mathbb{E}\left[\int_0^T e^{-\lambda t} \left(q(M_t)\pi_t^2(1 - \pi_t)^2 f''(\pi_t)/2 - \lambda f(\pi_t)\right) dt\right].$$  

(70)

Now we let $f(\cdot)$ to be specifically the value function $v(\cdot)$ for our candidate structured policy. We know from (19) that $v(\cdot)$ satisfies

$$q(j)g^2(1 - y)^2v''(y)/2 - \lambda v(y) \leq c(j).$$  

(71)

for all $j \in M$ and all $y \in (0, 1)/\{\ell, u\}$. At belief $u$ and $\ell$, we can assign a value to $v''(\ell)$ and $v''(u)$ such that (71) holds for all $j \in M$ and all $y \in (0, 1)$. Therefore, (70) implies

$$\mathbb{E}\left[e^{-\lambda T} v(\pi_T)\right] \leq v(\pi_0) + \mathbb{E}\left[\int_0^T e^{-\lambda t} c(M_t) dt\right].$$  

(72)
We apply this with \( T = \tau \land t \) and let \( t \uparrow \infty \). Then, we get

\[
\mathbb{E} \left[ e^{-\lambda \tau} g(\pi_\tau) - \int_0^\tau e^{-\lambda t} c(M_t) dt \right] \leq v(\pi_0)
\]

(73)

where the expression on the left hand side of the above inequality is the expected net profit of the firm under an arbitrary admissible policy \((M, \tau)\). Because \( \ell < \hat{y} \equiv L/(L + R) \) and \( u > \hat{y} \), the last sentence of the proposition trivially follows. \( \square \)

**Remark 2.** The optimal value function \( v(\cdot) \), which is the value function for the optimal structured policy referred to in Proposition 5, is strictly increasing on \((\ell, 1)\) and satisfies following conditions:

\[
v(y) = g(y) = 0, \quad y \in [0, \ell],
\]

(74)

\[
v(y) > g(y), \quad y \in (\ell, u),
\]

(75)

\[
v(y) = g(y) = yR - (1 - y)L > 0, \quad y \in [u, 1],
\]

(76)

where \( \ell \) and \( u \) are the lower and upper critical beliefs of the optimal structured policy referred to in Proposition 5.

**Appendix D: Proofs of Propositions 6 and 7**

The following lemma will be used in the proofs of Propositions 6 and 7.

**Lemma 5.** Fix a function \( g(\cdot) \) and consider two different sets of problem parameters, each of which consists of parameters \( \sigma, \lambda, \{c(j), j \in M\} \) and \( \{\mu(j, k), j \in M, k \in \{0, 1\}\} \). Denote by \( \ell_i \) and \( u_i \), optimal lower and upper critical beliefs and let \( v_i \) be the optimal value function associated with the parameter set \( i \in \{1, 2\} \). Then we have the following: If \( v_i(y) = v_2(y) \) for \( y \in [0, 1] \), then \( \ell_1 \leq \ell_2 \) and \( u_1 \geq u_2 \).

**Proof of Lemma 5.** Because \( v_1(\ell_i) = 0 \) for \( i = 1, 2 \) by (74) and \( v_1(y) \geq v_2(y) \) for \( y \in [0, 1] \), it follows that \( v_1(\ell_2) \geq v_2(\ell_2) = 0 \). From this and the fact that \( v_1(\ell_i) = 0 \) for \( i = 1, 2 \), we have that

\[
v_1(\ell_1) = 0 \leq v_1(\ell_2).
\]

(77)

Since \( v_1(\cdot) \) is increasing by Remark 2, it follows from inequality (77) that \( \ell_1 \leq \ell_2 \). We now proceed to prove that \( u_1 \geq u_2 \) by obtaining contradiction under two cases. Suppose for a contradiction that \( u_1 \in (\ell_2, u_2) \). Then by (75), we have that \( v_2(u_1) = v_1(u_1) \). This contradicts with the fact that \( v_1(y) \geq v_2(y) \) for all \( y \in [0, 1] \). Suppose now that \( u_1 \in (\ell_1, \ell_2) \). Then, by (76), \( g(y) = v_1(y) > 0 \) and \( g(y) = v_2(y) = 0 \) for all \( y \in [u_1, \ell_2] \), which is a contradiction. Therefore, \( u_1 \geq u_2 \). \( \square \)
**Proof of Proposition 6.** Fix all problem parameters except the observational volatility. Take \(\sigma_1\) and \(\sigma_2\) such that \(0 < \sigma_1 < \sigma_2\). Denote by \((M_i, \tau_i), v_i(\cdot), \ell_i\) and \(u_i\), optimal structured policy referred to as in Proposition 5, optimal value function, optimal lower and upper critical beliefs, respectively, when the observational volatility is \(\sigma_i\), \(i = 1, 2\). Define \(\tau(M_2, \sigma_1) = \inf\{t \geq 0 : \pi_t \notin (\ell_2, u_2) ; M_2, \sigma_1\}\). In words, \(\tau(M_2, \sigma_1)\) is the first exit time from the belief interval \((\ell_2, u_2)\) under learning strategy \(M_2\) when the observational volatility is \(\sigma_1\). Because \(\sigma_1 < \sigma_2\), it follows that \(\tau_2 \geq \tau(M_2, \sigma_1)\) almost surely for \(\pi_0 \in (0, 1)\), which implies that

\[
E\left[e^{-\lambda \tau(M_2, \sigma_1)} g(\pi_{\tau(M_2, \sigma_1)}) - \int_0^{\tau(M_2, \sigma_1)} e^{-\lambda t} c(M_2, t) dt ; \sigma_1\right] \geq v_2(\pi_0), \quad \pi_0 \in (0, 1).
\]

(78)

Also, because \((M_2, \tau(M_2, \sigma_1))\) is an admissible but not necessarily an optimal policy when the observational volatility is \(\sigma_1\), we have that

\[
v_1(\pi_0) \geq E\left[e^{-\lambda \tau(M_2, \sigma_1)} g(\pi_{\tau(M_2, \sigma_1)}) - \int_0^{\tau(M_2, \sigma_1)} e^{-\lambda t} c(M_2, t) dt ; \sigma_1\right], \quad \pi_0 \in (0, 1).
\]

(79)

By (78) and (79), we obtain that \(v_1(\pi_0) \geq v_2(\pi_0)\) for all \(\pi_0 \in (0, 1)\). Then, from Lemma 5, our claim follows. \(\Box\)

**Proof of Proposition 7.** Fix all problem parameters except the discount rate. Take any two discount rates \(\lambda_1\) and \(\lambda_2\) such that \(0 < \lambda_1 < \lambda_2\). Denote by \((M_i, \tau_i), v_i(\cdot), \ell_i\) and \(u_i\), optimal structured policy identified in Proposition 5, optimal value function, optimal lower and upper critical beliefs, respectively, when the discount rate is \(\lambda_i\), \(i = 1, 2\). Then, for any \(\pi_0 \in (0, 1)\),

\[
v_1(\pi_0) \geq E\left[e^{-\lambda_1 \tau_2} g(\pi_{\tau_2}) - \int_0^{\tau_2} e^{-\lambda_1 t} c(M_2, t) dt\right]
\]

\[
\geq E\left[e^{-\lambda_2 \tau_2} g(\pi_{\tau_2}) - \int_0^{\tau_2} e^{-\lambda_2 t} c(M_2, t) dt\right] = v_2(\pi_0).
\]

(80)

(81)

The inequality in (80) is because the structured policy \((M_2, \tau_2)\) is an admissible policy, but not necessarily an optimal policy when the discount rate is \(\lambda_1\); the inequality in (81) is because \(z'(\xi) \leq 0\) for all \(\xi \geq 0\) where \(z(\xi) = E[e^{-\xi \tau_2} g(\pi_{\tau_2})] - E\left[\int_0^{\tau_2} e^{-\xi t} c(M_2, t) dt\right]\). Since \(v_1(y) \geq v_2(y)\) for \(y \in [0, 1]\), the claim follows from Lemma 5. \(\Box\)

**Appendix E: Proof of Proposition 8**

We first state the following lemma that proves the first statement of the proposition. The proof of this lemma can be found in the electronic companion.
Lemma 6. Define $\tilde{\kappa} = 2c(1)/q(1)$, $f(y) = \log(y/(1-y)) - 1/y$ and $r(y) = (1-2y)^2/(y(1-y))$ for $y \in (0,1)$. Then, there exists a unique pair $(\ell,u)$ that satisfies $0 < \ell < u < 1$ and following two nonlinear equations:

\[ f(u) - f(\ell) - R/\tilde{\kappa} = 0, \]  
\[ r(\ell) - r(u) - (R - L)/\tilde{\kappa} = 0. \]  

(82) \quad (83)

It remains to prove the last sentence of the proposition. We fix a trial value of $\ell \in (0,1)$. Recall the differential operator $\Gamma$ defined in (32). We now proceed to show that there exists a unique function $v_\ell(\cdot)$ that satisfies

\[ \max_{j \in M} \{-c(j) + q(j)\Gamma v_\ell''(y)\} = 0, \quad y \in (\ell,1), \]  

(84)

with boundary conditions

\[ v_\ell(\ell) = v_\ell'(\ell) = 0. \]  

(85)

Observe that if such $v_\ell(\cdot)$ exists, then mode 1 achieves the maximum in (84) for all $y \in (\ell,1)$. Therefore, (84) can be replaced with the following ODE:

\[ -c(1) + q(1)y^2(1-y)^2v_\ell''(y)/2 = 0, \quad y \in (\ell,1). \]  

(86)

Elementary ODE theory implies that for each $\ell \in (0,1)$ there exists a unique function $v_\ell(\cdot)$ that is in $C^2(\ell,1)$ and satisfies (86) with boundary conditions (85). Moreover, from (86) and (85), it follows that for each $\ell \in (0,1)$, $v_\ell(\cdot)$ is strictly positive, strictly increasing and strictly convex on $(\ell,1)$. Defining function $\hat{v}_\ell(\cdot)$ as in (53) for each $\ell \in (0,1)$ and applying the same arguments as in the proof of Proposition 4 it follows that there exists a unique pair $(\ell,u)$ such that $\ell \in (0,L/(L+R))$, $u \in (L/(L+R),1]$ and $\hat{v}_\ell(\cdot)$ satisfies

\[ \hat{v}_\ell(u) = uR - (1-u)L, \quad \hat{v}_\ell'(u) = R + L. \]  

(87)

Using this unique pair $(\ell,u)$ and unique function $\hat{v}_\ell(\cdot)$, define the function $v(\cdot)$ as following:

\[ v(y) = \begin{cases} \hat{v}_\ell(y) & \text{if } y \in [0,u) \\ yR - (1-y)L > 0 & \text{if } y \in [u,1] \end{cases}. \]  

(88)

Then, $v(\cdot)$ is continuously differentiable on $(0,1)$ and twice continuously differentiable on $(\ell,u)$. We already know from (86) that $v(\cdot)$ satisfies (17) with learning function $m$ (as in (14)) that uses parameters $k = 1$ and $\beta_1 = 1$. Therefore, from Proposition 3 $v(\cdot)$ is the value function for the structured policy with parameters $k, \beta_1$ and the unique pair $(\ell,u)$. By (88), $v(\cdot)$ further satisfies inequalities (18) and (19). Therefore, from the same arguments as in the proof of Proposition 5 the
structured policy with parameters $\beta_1 \equiv 1$, $k = 1$ and the unique pair $(\ell, u)$ is optimal. Hereafter, the value function $v(\cdot)$ for optimal structured policy will be called optimal value function, parameters $\ell$ and $u$ of the same policy will be called optimal lower and upper critical beliefs.

We now proceed to explicitly identify the optimal value function $v(\cdot)$ and optimal lower and upper critical beliefs $\ell$ and $u$. By (53) and (88), $v(y)$ is already explicitly characterized for $y \in [0, \ell] \cup [u, 1]$. Therefore, we shall only focus our attention on explicit characterization of $v(\cdot)$ on $(\ell, u)$ by considering following differential equation

$$v''(y) = \tilde{\kappa}/(y^2(1-y)^2)$$

subject to boundary conditions

$$v(\ell) = 0, \quad v'(\ell) = 0, \quad v(u) = uR - (1-u)L, \quad v'(u) = R + L.$$  

Define $\psi(y) \equiv \log(y/(1-y))$ for $y \in (0, 1)$. Then, general form of the solution to the differential equation (89) is

$$v(y) = (2y-1)\psi'(y)\tilde{\kappa} + Ay + B$$

where $A$ and $B$ are constants to be determined from boundary conditions at either $y = \ell$ or $y = u$ in (90). If we solve for $A$ and $B$ by considering boundary conditions at $y = \ell$ only, we have that

$$A(\ell) = -\tilde{\kappa} \left[ 2\psi(\ell) + (2\ell - 1)/(\ell(1-\ell)) \right] \quad \text{and} \quad B(\ell) = \tilde{\kappa} \left[ \psi(\ell) + (2\ell - 1)/(1-\ell) \right].$$

Similarly, if we consider boundary conditions at $y = u$ only, we have that

$$A(u) = R + L - \tilde{\kappa} \left[ 2\psi(u) + (2u - 1)/(u(1-u)) \right], \quad \text{and} \quad B(u) = -L + \tilde{\kappa} \left[ \psi(u) + (2u - 1)/(1-u) \right].$$

From elementary ODE theory, it follows that there exists at most one solution to equation (89) subject to (90). As a result, there exists a unique function that satisfies (89) subject to (90) if and only if the pair $(\ell, u)$ satisfies

$$A(\ell) = A(u) \quad \text{and} \quad B(\ell) = B(u).$$

We already proved that there exists a unique pair $(\ell, u)$ and unique function $v$ that solve both (89) and (90). This implies that the pair $(\ell, u)$ satisfies both equations in (93). Because solving for pairs $(\ell, u)$ in (93) is equivalent to solving for pairs $(\ell, u)$ that satisfy both (82) and (83), Lemma 6 shows that the unique pair $(\ell, u)$ is also the unique solution of (93). Then, from (91), we explicitly characterize the optimal value function on $(\ell, u)$ as follows:

$$v(y) = (2y-1)\psi(y)\tilde{\kappa} + A(\ell)y + B(\ell) = (2y-1)\psi(y)\tilde{\kappa} + A(u)y + B(u), \quad y \in (\ell, u),$$

where the pair $(\ell, u)$ uniquely satisfies (82) and (83). □
Acknowledgments
We are indebted to Ruth Williams for her expert help in structuring the proof of Proposition 2.

References


BBC. 2011. Brazil companies Petrobras and Vale post record profits  


*Decision Analysis* **3**(1) 16–32.


Bloomberg. 2011. Cairns $600 Million Greenland Oil Campaign Ends in Failure  

Bolton, Patrick, Christopher Harris. 1999. Strategic experimentation.  


Ernst&Young. 2011. Global E&P Benchmark Study  


Electronic Companion of “Investment Timing with Incomplete Information and Multiple Means of Learning”

Here, we provide proofs of lemmas stated in Appendices A–E and proofs of all propositions not treated in Appendices A–E.

Appendix EC.1: Proof of Proposition [2]

A key to our analysis is the following result on one-dimensional stochastic differential equations.

**Lemma EC.1.** Consider a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) on which are defined a filtration \(\{\mathcal{F}_t, t \geq 0\}\) that satisfies the usual conditions, and a process \(W = \{W_t, t \geq 0\}\) that is a Brownian motion with respect to \(\{\mathcal{F}_t, t \geq 0\}\). Also, let \(\xi_0\) be a given constant, and let \(a(\cdot)\) and \(b(\cdot)\) be bounded Borel functions \(\mathbb{R} \to \mathbb{R}\) such that \(a(\cdot)\) has finite variation on every compact interval and \(a(\cdot) \geq \epsilon > 0\).

There exists a process \(\xi = \{\xi_t, t \geq 0\}\), unique up to an equivalence, that is adapted to \(\{\mathcal{F}_t, t \geq 0\}\) and satisfies the following stochastic differential equation:

\[
\xi_t = \xi_0 + \int_0^t a(\xi_s) dW_s + \int_0^t b(\xi_s) ds, \quad t \geq 0. \tag{EC.1}
\]

Moreover, the unique solution \(\xi\) is adapted to the filtration \(\{\mathcal{F}_t^W, t \geq 0\}\) that is generated by the Brownian motion \(W\).

**Proof of Lemma [EC.1].** The lemma is implied by the following three results that are either proved or cited by Karatzas and Shreve [1991]. First, Theorem 5.15 (Section 5.5) establishes existence of a weak solution for the SDE (EC.1) in the distributional sense. Second, the result of Nakao [1972] that is discussed in Section 5.10 establishes that pathwise uniqueness holds for equation (EC.1); see also Section V.41 of Williams and Rogers [1987] for a development of Nakao’s pathwise-uniqueness theorem. Finally, the result by Yamada and Watanabe that appears as Corollary 3.23 (Section 5.3) in Karatzas and Shreve [1991] establishes existence of a strong solution \(\xi\), and in our setting, where \(\xi_0\) is deterministic, that strong solution is adapted to the filtration generated by \(W\). Pathwise uniqueness of equation (EC.1) implies that any other solution of (EC.1) that is adapted to the larger filtration \(\{\mathcal{F}_t, t \geq 0\}\) must be equal almost surely to the strong solution \(\xi\). □

For the application of Lemma [EC.1] to our problem, we take \((\Omega, \mathcal{F}, \mathbb{P}), \{\mathcal{F}_t, t \geq 0\}\) and \(W\) as in Section [2]. Fixing a piecewise constant learning function \(m(\cdot)\) of the form [(11)], equation (12) requires that \(M_t = m(\hat{\pi}_t)\), so equations (11), (3) and (4) can be rewritten as follows:

\[
X_t = \int_0^t [\alpha(\hat{\pi}_s) + \theta \sigma \delta(\hat{\pi}_s)] ds + \sigma W_t, \quad t \geq 0, \tag{EC.2}
\]

\[
\hat{\pi}_t = \hat{\pi}_0 + \int_0^t \delta(\hat{\pi}_s) \gamma(\hat{\pi}_s) dZ_s, \quad t \geq 0, \quad \text{and} \tag{EC.3}
\]

\[
Z_t = \frac{1}{\sigma} X_t - \frac{1}{\sigma} \int_0^t [\alpha(\hat{\pi}_s) + \hat{\pi}_s \sigma \delta(\hat{\pi}_s)] ds, \quad t \geq 0, \tag{EC.4}
\]
where \( \hat{\pi}_0 = \pi_0 \),
\[
\alpha(y) \doteq \mu(m(y), 0), \quad \gamma(y) \doteq y(1 - y), \quad \text{and} \quad \delta(y) \doteq \frac{1}{\sigma}[\mu(m(y), 1) - \mu(m(y), 0)], \quad 0 < y < 1.
\] (EC.5)

Substituting (EC.2) and (EC.4) into (EC.3), we see that \( \hat{\pi} \) must satisfy
\[
\hat{\pi}_t = \hat{\pi}_0 + \int_0^t \delta(\hat{\pi}_s) \gamma(\hat{\pi}_s) dW_s + \int_0^t \delta^2(\hat{\pi}_s) \gamma(\hat{\pi}_s)(\theta - \hat{\pi}_s) ds.
\] (EC.6)

**Lemma EC.2.** There exists a continuous process \( \hat{\pi} = \{\hat{\pi}_t, t \geq 0\} \) that is adapted to \( \{F_t, t \geq 0\} \) and uniquely (up to an equivalence) solves the SDE (EC.6).

**Proof of Lemma EC.2.** Consider the following stochastic differential equation:
\[
\xi_t = \xi_0 + \int_0^t a(\xi_s) dW_s + \int_0^t [\theta b_1(\xi_s) + (1 - \theta) b_0(\xi_s)] ds, \quad t \geq 0,
\] (EC.7)
where
\[
\xi_0 \doteq \log(\hat{\pi}_0/(1 - \hat{\pi}_0)),
\] (EC.8)
\[
a(x) \doteq [\mu(m(x), 1) - \mu(m(x), 0)]/\sigma, \quad \tilde{m}(x) \doteq m(\exp(x)/(1 + \exp(x))), \quad x \in \mathbb{R},
\] (EC.9)
\[
b_1(x) \doteq [\mu(m(x), 1) - \mu(m(x), 0)]^2/2\sigma^2 \quad \text{and} \quad b_0(x) = -b_1(x), \quad x \in \mathbb{R}.
\] (EC.10)

Observe that \( a(\cdot), b_0(\cdot) \) and \( b_1(\cdot) \) are bounded, piecewise continuous functions on \( \mathbb{R} \) with just finitely many pieces, and \( a(\cdot) \geq \epsilon > 0 \). Of course, (EC.7) comprises two separate SDEs, one of which determines \( \xi \) on the set \( \{\theta = 0\} \), and the other of which determines \( \xi \) on the set \( \{\theta = 1\} \). The hypotheses of Lemma EC.1 are satisfied by each of those SDEs individually, so each of them has a solution that is adapted to \( \{F_t^W, t \geq 0\} \) and is unique up to an equivalence. Define
\[
\hat{\pi}_t = h(\xi_t), \quad t \geq 0, \quad \text{where} \quad h(x) \doteq \exp(x)/(1 + \exp(x)), \quad x \in \mathbb{R},
\] (EC.11)
so that
\[
\xi_t = f(\hat{\pi}_t), \quad t \geq 0, \quad \text{where} \quad f(y) \doteq \log(y/(1 - y)), \quad y \in (0, 1).
\] (EC.12)

Now \( h'(x) = h(x)(1 - h(x)) \) and \( h''(x) = h(x)(1 - h(x))(1 - 2h(x)) \), so we can use (EC.11) and Itô’s formula to reexpress (EC.7) in the form
\[
\hat{\pi}_t = \hat{\pi}_0 + \int_0^t \delta(\hat{\pi}_s) \gamma(\hat{\pi}_s) dW_s + \int_0^t \delta^2(\hat{\pi}_s) \gamma(\hat{\pi}_s)(\theta - \hat{\pi}_s) ds, \quad t \geq 0.
\] (EC.13)

Therefore, we conclude that \( \hat{\pi} \) defined from \( \xi \) via (EC.11) is adapted to \( \{F_t, t \geq 0\} \) and uniquely (up to an equivalence) solves the SDE (EC.6). \( \square \)
Based on Lemma [EC.2] $X$ is uniquely defined from $\hat{\pi}$ via [EC.2] and $M$ is uniquely defined from $\hat{\pi}$ via equation $M_t = m(\hat{\pi}_t)$ for $t \geq 0$. To prove that $\hat{\pi}$ is adapted to $\{F^X_t, \ t \geq 0\}$, we begin with the following observation: [EC.3] and [EC.4] together imply that $\hat{\pi}$ satisfies the differential relationship

$$d\hat{\pi}_t = \frac{1}{\sigma} \delta(\hat{\pi}_t) \gamma(\hat{\pi}_t) dX_t - \frac{1}{\gamma(\hat{\pi}_t)} \nu(\hat{\pi}_t) dt,$$

where $\nu(y) = \sigma y \delta^2(y) + \delta(y) \alpha(y)$ for all $y \in (0, 1)$. Using the transformation [EC.12], [EC.14] and Itô’s formula, we can then reexpress (EC.7) in terms of $\xi$ and $X$, rather than $\xi$ and $W$, as follows:

$$\xi_t = f(\hat{\pi}_0) + \frac{1}{\sigma} \int_0^t a(\xi_s) dX_s + \frac{1}{\sigma} \int_0^t \hat{b}(\xi_s) ds, \ t \geq 0,$$

where $a(\cdot)$ is the same as in [EC.9] and $\hat{b}(x) = -(\mu^2(\bar{m}(x), 1) - \mu^2(\bar{m}(x), 0)) / 2\sigma$ for $x \in \mathbb{R}$ is a bounded, piecewise continuous function with just finitely many pieces. To explore the implications of (EC.15) let us view $\xi$ and $X$ as continuous processes on a finite time domain $[0, T]$. Given the representation [EC.2] of $X$, a routine application of the Girsanov’s theorem then gives us two equivalent measures $\mathbb{P}$ (that is, $\mathbb{P}$ and $\mathbb{P}^*$ have the same null sets) and such that $\{\frac{1}{\sigma} X_t, \ 0 \leq t \leq T\}$ is a standard Brownian motion (zero drift and unit variance) with respect to ambient filtration $\{F_t, \ 0 \leq t \leq T\}$ on $([0, T], F, \mathbb{P})$. Moreover, the coefficient functions $a(\cdot)$ and $\hat{b}(\cdot)$ satisfy the hypotheses of pathwise-uniqueness theorem of Nakao (1972). So, exactly as in the proof of Lemma [EC.1] we conclude that (EC.15) has a solution $\{\xi_t, \ 0 \leq t \leq T\}$ that is unique up to an equivalence and is adapted to $\{F^X_t, \ 0 \leq t \leq T\}$. As noted earlier, (EC.15) and (EC.7) are just two different ways of writing the same equation, and $\mathbb{P}$ and $\mathbb{P}^*$ are equivalent measures, so their solutions (each unique up to an equivalence) must agree almost surely on $[0, T]$. This establishes that the solution of (EC.7) is adapted to $\{F^X_t\}$ over the restricted domain $[0, T]$, and since $T > 0$ was arbitrary, we conclude that $\{\xi_t, \ t \geq 0\}$ is adapted to $\{F^X_t, \ t \geq 0\}$ and so $\{\hat{\pi}_t, \ t \geq 0\}$ is adapted to $\{F^X_t, \ t \geq 0\}$.

To recapitulate, we began by fixing a piecewise constant learning function $m(\cdot)$. A process $\hat{\pi}$ was defined as the unique solution of the SDE (EC.6), whose coefficient functions are determined by $m(\cdot)$, and we used $\hat{\pi}$ to define a learning strategy $M$ via $M_t = m(\hat{\pi}_t), \ t \geq 0$. Next, the observed process $X$ was determined by $M$ via [11]. By proving that $\hat{\pi}$ (and hence $M$) is adapted to $\{F^X_t, \ t \geq 0\}$ we have confirmed that $M$ is an admissible learning strategy. Let us now denote by $\pi$ the posterior belief process induced by the learning strategy $M$, meaning that $\pi_t = \mathbb{P}(\theta = 1 | F^X_t), \ t \geq 0$. Making the change of variable $\hat{\xi}_t = f(\pi_t)$ where $f(\cdot)$ is defined by (EC.12), and using Itô’s formula, we can restate Proposition [1] in the following form:

$$\hat{\xi}_t = f(\pi_0) + \int_0^t \Delta_s dY_s - \frac{1}{2} \int_0^t \Delta_s^2 ds, \ t \geq 0,$$

(EC.16)
where

\[ \Delta_t = \frac{1}{\sigma} [\mu(M_t, 1) - \mu(M_t, 0)], \quad t \geq 0, \]  

\[ dY_t = \frac{1}{\sigma} dX_t - \frac{1}{\sigma} \mu(M_t, 0) dt, \quad t \geq 0. \]  

This alternative expression of Proposition 1 shows that \( \pi_t \) uniquely satisfies the filtering equations (EC.13) and (EC.14) given \( M \) and \( X \). By rearranging terms in (EC.15) and using the fact that \( \pi_0 = \hat{\pi}_0 \), we conclude that the right hand side of equation (EC.15) is equal to the right hand side of equation (EC.16). Therefore, \( \xi_t \) and \( \hat{\xi}_t \) are almost surely identical. As a result, \( \pi_t \) and \( \hat{\pi}_t \) are almost surely identical, which concludes the proof of Proposition 2. \( \square \)

**Appendix EC.2: Proof of Equation (23)**

Take \( \mathcal{I} = \{1, 2, \ldots, d\} \) without loss of generality. Let \( \mathbb{W} = \{\mathbb{W}_t, t \geq 0\} \) be a standard \( d \)-dimensional Brownian Motion on probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Moreover, let \( \theta \) be a binary random variable defined on the probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and independent of \( \mathbb{W} \). Consider \( d \)-dimensional stochastic process \( X = \{X_t, t \geq 0\} \) such that

\[ X_t = \mu(I, \theta) t + \sigma \mathbb{W}_t, \quad t \geq 0, \]  

and denote by \( \mathcal{F}_t^X \), the filtration generated by \( X \). Define posterior belief process \( \pi_t \equiv \{\pi_t, t \geq 0\} \) such that

\[ \pi_t \equiv \mathbb{P}(\theta = 1|\mathcal{F}_t^X), \quad t \geq 0. \]  

For brevity, we define the drift difference as

\[ D(j) \equiv \mu(j, 1) - \mu(j, 0), \quad \text{for } j \in I. \]  

Denoting by \( X_{j,t} \) the \( j \)th component of \( X \) at time \( t \), we define

\[ z(X_t) \equiv \exp (\frac{1}{2\sigma^2} \sum_{j=1}^d D(j)(2X_{j,t} - \mu(j, 1)t - \mu(j, 0)t)), \quad t > 0, \]  

\[ h(z) \equiv \pi_0 z/(\pi_0 z + (1 - \pi_0)), \quad z > 0, \]  

\[ b_t \equiv h(z(X_t)), \quad t > 0. \]  

Using Bayes’ rule, we obtain the following result:

**Lemma EC.3.** \( b_t = \mathbb{P}(\theta = 1|X_t) \) for \( t \geq 0 \).
Proof of Lemma EC.3. For each $k \in \{0, 1\}$ and $t > 0$, define conditional density functions

$$
\varphi_{k,t}(x) \triangleq \mathbb{P}(X_t \in dx | \theta = k) = (2\pi \sigma^2 t)^{-d/2} \exp(-(x_j - \mu(j,k)t)^2/(2\sigma^2 t)), \quad x \in \mathbb{R}^d,
$$

where $x_j$ is the $j^{th}$ component of vector $x$. From Bayes’ rule,

$$
\mathbb{P}(\theta = 1|X_t) = \pi_0 \varphi_{1,t}(X_t) / [\pi_0 \varphi_{1,t}(X_t) + (1-\pi_0) \varphi_{0,t}(X_t)], \quad t > 0.
$$

(EC.26)

Plugging the expression for $\varphi_{k,t}(\cdot)$ in (EC.25) into (EC.24), and using (EC.24) and (EC.22), we have that $b_t = \mathbb{P}(\theta = 1|X_t)$ with probability 1 for $t > 0$.

Proof of Lemma EC.4. Fix time $t > 0$ and take any $A \in \mathcal{F}_t^X$. By definition of conditional expectation, it is sufficient to prove that

$$
\int_A b_t \ d\mathbb{P} = \int_A \mathbb{I}_{\{\theta = 1\}} \ d\mathbb{P}.
$$

(EC.27)

Define probability measure $Q$ as

$$
Q(\cdot) \triangleq \pi_0 \ Q^1(\cdot) + (1-\pi_0) \ Q^0(\cdot),
$$

(EC.28)

where for each $k \in \{0, 1\}$ $Q^k(\cdot)$ is the distribution of $d$-dimensional Brownian motion with drift vector $\mu(I, k)$ and covariance matrix $\Sigma = \sigma^2 I$. Here, $I$ is a $d \times d$ identity matrix. Notice that there exists a Borel set $B \in \mathcal{B}(\mathbb{C}^d([0,t]))$ (that is, the Borel $\sigma$-algebra generated by open sets of all real-valued $d$-dimensional continuous functions on $[0,t]$) such that $A = \{\omega \in \Omega : X \in B\}$. Therefore, the right hand side of (EC.27) can be written as

$$
\int_A \mathbb{I}_{\{\theta = 1\}} \ d\mathbb{P} = \mathbb{P}(\theta = 1, X \in B) = \pi_0 Q^1(B).
$$

(EC.29)

From (EC.24) and (EC.28), the left hand side of (EC.27) can be written as

$$
\int_A b_t \ d\mathbb{P} = \int_B h(z(X_t)) \ dQ = \int_B h(z(X_t)) \ (\pi_0 dQ^1 + (1-\pi_0) dQ^0).
$$

(EC.30)

In addition, by multidimensional Girsanov theorem, we have that $dQ^1 = \tilde{z}(X_t) \ dQ^0$ where

$$
\tilde{z}(X_t) \triangleq \exp \left( \frac{1}{\sigma^2} \sum_{j=1}^d D(j) X_{j,t} - \left[ \mu(j,0) D(j) + D(j)^2/2 \right] t \right).
$$

Observe that $z(X_t) = \tilde{z}(X_t)$ with probability 1, which implies that $dQ^1 = z(X_t) \ dQ^0$. From this and the definition of $h(\cdot)$, it follows that

$$
h(z(X_t)) \ (\pi_0 dQ^1 + (1-\pi_0) dQ^0) = h(z(X_t)) \ (\pi_0 z(X_t) + 1 - \pi_0) \ dQ^0 = \pi_0 z(X_t) \ dQ^0.
$$
This and the fact that \( dQ^1 = z(X_t) dQ^0 \) imply that (EC.31) can be written as
\[
\int_A b_t \, d\mathbb{P} = \int_B \pi_0 \, dQ^1 = \pi_0 Q^1(B). \tag{EC.31}
\]
Then, equations (EC.29) and (EC.31) conclude that \( b_t = \pi_t \) with probability 1. \( \square \)

It only remains to derive the SDE \((23)\), which the posterior belief process \( \pi = \{\pi_t, t \geq 0\} \) needs to satisfy. In this analysis, we will use the following result:

**Lemma EC.5.** The stochastic process \( Z_t = \{Z_t, t \geq 0\} \) where for all \( t \geq 0 \)
\[
Z_t = \frac{1}{\sigma} \left[ X_{1,t} - \int_0^t (\pi_s \mu(1, 1) + (1 - \pi_s) \mu(1, 0)) \, ds, \ldots, X_{d,t} - \int_0^t (\pi_s \mu(d, 1) + (1 - \pi_s) \mu(d, 0)) \, ds \right]^T,
\]
is a standard \(d\)-dimensional Brownian motion with respect to filtration \( \{\mathcal{F}_t^X, t \geq 0\} \).

**Proof of Lemma EC.5.** For each \( j \in \{1, \ldots, d\} \) and for all \( t \geq 0 \), define
\[
\psi_{j,t}(X) = \mathbb{E} [\mu(j, \theta)|\mathcal{F}_t^X] = [\pi_t \mu(j, 1) + (1 - \pi_t) \mu(j, 0)]. \tag{EC.33}
\]
Then, from (EC.19) and (EC.32), \( Z \) can be written as
\[
Z_t = W_t + \frac{1}{\sigma} \left[ \int_0^t (\mu(1, \theta) - \psi_{1,s}(X)) \, ds, \ldots, \int_0^t (\mu(d, \theta) - \psi_{d,s}(X)) \, ds \right]^T, \quad t \geq 0. \tag{EC.34}
\]
By (EC.33) and tower property of conditional expectation, it follows that
\[
\mathbb{E} \left[ \left[ \int_0^t (\mu(1, \theta) - \psi_{1,s}(X)) \, dz, \ldots, \int_0^t (\mu(d, \theta) - \psi_{d,s}(X)) \, dz \right]^T \bigg| \mathcal{F}_s^X \right] = 0, \quad 0 \leq s \leq t. \tag{EC.35}
\]
Also, because \( W \) is a standard \(d\)-dimensional Brownian motion, we have that
\[
\mathbb{E} \left[ W_t - W_s \bigg| \mathcal{F}_s^X \right] = 0, \quad 0 \leq s \leq t. \tag{EC.36}
\]
Then, from (EC.35) and (EC.36), we obtain that
\[
\mathbb{E} \left[ Z_t - Z_s \bigg| \mathcal{F}_s^X \right] = 0, \quad 0 \leq s \leq t. \tag{EC.37}
\]
Because \( Z \) is adapted to filtration \( \{\mathcal{F}_t^X, t \geq 0\} \) and is integrable, it follows from (EC.37) that \( Z \) is a continuous martingale adapted to filtration \( \{\mathcal{F}_t^X, t \geq 0\} \). Moreover, denoting by \( X_{j, \cdot} \) and \( Z_{j, \cdot} \) the \( j \)-th components of \( X \) and \( Z \), observe from (EC.19) and (EC.32) that \( \frac{1}{\sigma} [X_t, X_{j,t}] = [Z_t, Z_{j,t}] = t \) if \( i = j \), and \( [X_t, X_{i,t}] = [Z_t, Z_{i,t}] = 0 \) otherwise. Then, by Levy’s characterization theorem (c.f., page 337 of Duffie (2010)), it follows that \( Z \) is a standard \(d\)-dimensional Brownian motion adapted to filtration \( \{\mathcal{F}_t^X, t \geq 0\} \). \( \square \)
Defining $\xi_t = z(X_t)$ for $t > 0$, it follows from multidimensional Itô’s formula that

$$d\xi_t = \xi_t \left( \frac{1}{\sigma^2} \sum_{j=1}^d D(j) dX_{j,t} + \left[ D(j)^2 - \left( \mu(j, 1)^2 - \mu(j, 0)^2 \right) \right] dt \right).$$ (EC.38)

Using the fact that $\pi_t = h(\xi_t)$ for $t > 0$ and applying Itô’s formula, for each $t > 0$ we have that

$$d\pi_t = h'(\xi_t) d\xi_t + \frac{1}{2} h''(\xi_t) (d\xi_t)^2 = \frac{h(\xi_t)(1 - h(\xi_t))}{\xi_t} d\xi_t - \frac{h^2(\xi_t)(1 - h(\xi_t))}{\xi_t^2} (d\xi_t)^2$$

$$= \sum_{j=1}^d \frac{D(j)}{\sigma^2} \pi_t(1 - \pi_t) [dX_{j,t} - (\pi_t \mu(j, 1) + (1 - \pi_t) \mu(j, 0)) dt]$$ (EC.39)

Then, by Lemma [EC.5] (EC.39) can be written as

$$d\pi_t = \sum_{j=1}^d \frac{D(j)}{\sigma} \pi_t(1 - \pi_t) dZ_{j,t}, \quad t \geq 0.$$ (EC.40)

Define the stochastic process $\tilde{Z} = \{\tilde{Z}_t, t \geq 0\}$ such that

$$\tilde{Z}_t = \sum_{j=1}^d \frac{D(j)}{||\mu(I, 1) - \mu(I, 0)||} Z_{j,t}, \quad t \geq 0.$$ (EC.41)

Then, (EC.40) can be written as

$$d\pi_t = \frac{||\mu(I, 1) - \mu(I, 0)||}{\sigma} \pi_t(1 - \pi_t) d\tilde{Z}_t$$ (EC.42)

Because $\{Z_{1,t}, \ldots, Z_{d,t}, t \geq 0\}$ are independent standard Brownian motions with respect to filtration $\{\mathcal{F}_{t}^X, t \geq 0\}$ and variance of $\tilde{Z}_t$ is $t$ for all $t \geq 0$, clearly $\tilde{Z}$ is a standard Brownian motion with respect to filtration $\{\mathcal{F}_{t}^X, t \geq 0\}$. □

**Appendix EC.3: Proofs of Propositions 9 and 10**

**Proof of Proposition 9.** We begin our analysis with the following observation: there exist an index $z \leq 2^N - 1$ and a labelling scheme $B_1, \ldots, B_{2^N-1}$ for all subsets of $M$ such that:

$$0 < q(B_1) < \ldots < q(B_z) = \bar{q} \leq \max_{1 \leq j \leq 2^N-1} q(B_j),$$ (EC.43)

$$0 < \frac{c(B_1)}{q(B_1)} < \frac{c(B_2)}{q(B_2)} < \ldots < \frac{c(B_z)}{q(B_z)} < \frac{c(B_{z+1})}{q(B_{z+1})},$$ (EC.44)

$$c(B_i) = \phi(q(B_i)) \text{ for all } i = 1, \ldots, z,$$ (EC.45)

$$c(B_i) \geq \phi(q(B_i)) \text{ for all } i = 1, \ldots, 2^N - 1,$$ (EC.46)

$\phi$ is the nondecreasing concave function that is increasing on $[0, \bar{q}]$, decreasing on $[\bar{q}, +\infty)$, and $\phi(0) = 0$. □
where $q$ is as in [22], $c$ is as in (22) and $\phi$ is the strictly increasing, piecewise linear and convex function on $[0,7]$ that connects the points $(0,0)$, $(q(B_1),c(B_1))$, $\ldots$, $(q(B_z),c(B_z))$ as in Figure 5. Hereafter, we call $B_1,\ldots,B_z$ extremal subsets.

Let $B_0$ be an artificial set with $c(B_0) = 0$ and $q(B_0) = 0$. Then, because $c/q$ ratio of each mode in $M$ is distinct, using (EC.43) through (EC.46), we have the following lemma:

**Lemma EC.6.** For each $i = 1,\ldots,z$, $B_i$ is the unique subset of $M$ that has the minimum $(c(\cdot) - c(B_{i-1}))/q(\cdot)$ ratio over all subsets of $M$ with strictly higher cost rate than $c(B_{i-1})$ and strictly higher signal quality than $q(B_{i-1})$.

For each $i = 1,\ldots,z$, define $\nu_i = (c(B_i) - c(B_{i-1}))/q(B_i) - q(B_{i-1})$, that is, the minimum $(c(\cdot) - c(B_{i-1}))/q(\cdot) - q(B_{i-1})$ ratio over all subsets of $M$ with strictly higher cost rate than $c(B_{i-1})$ and strictly higher signal quality than $q(B_{i-1})$. We now proceed to prove that $B_i = I_i$ for $i = 1,\ldots,z$ by induction. Consider $i = 1$ for the base step. It follows from simple algebra that $\min\{c(S_1)/q(S_1),c(S_2)/q(S_2)\} < (c(S_1) + c(S_2))/(q(S_1) + q(S_2))$ for any two non-empty subsets $S_1$ and $S_2$ of $M$ with distinct $c/q$ ratios. This and the fact that cost rate and signal quality are additive (see [22] and [24]) imply that

$$c(1)/q(1) = \min_{j \in M} c(j)/q(j) < c(S)/q(S), \quad \text{(EC.47)}$$

for any non-empty subset $S$ of $M$ other than $\{1\}$. Therefore, from Lemma EC.6 it follows that $B_1 = I_1 = \{1\}$.

For the inductive step, suppose that $B_i = I_i$ for $i = 1,\ldots,j-1 \leq z-1$. Then, clearly, $\nu_j = c(j)/q(j)$ and so $B_j = I_j$, which completes the proof for the inductive step. Since there are $N$ modes with distinct $c/q$ ratio, we have that $z = N$. □

**Proof of Proposition 5** Consider the optimal structured policy (identified in Proposition 5) in the single mode selection problem. That policy is an admissible, but not necessarily optimal policy in the subset selection problem. Therefore, optimal value function in the subset selection problem is weakly larger than the optimal value function in the single mode selection problem for each state variable $y \in [0,1]$. From this and the fact that $g(\cdot)$ is the same in both problems, the claim immediately follows from Lemma 5 □

**Appendix EC.4: Proof of Lemma 6**

**Proof of Lemma 6.** We first focus our attention on the equation (22). Because $R/\tilde{\kappa} > 0$ and $f(\cdot)$ is continuous and strictly increasing on $(0,1)$, it follows that there exists a unique continuous function $\tilde{f}$ such that for each $u \in (0,1)$, the pair $(l,u) = (\tilde{f}(u),u)$ uniquely satisfies (22) and

$$0 < \tilde{f}(u) < u. \quad \text{(EC.48)}$$
Moreover, $\tilde{f}(\cdot)$ is strictly increasing and has following properties:

$$
\tilde{f}(u) \uparrow 1 \text{ as } u \uparrow 1 \quad \text{and} \quad \tilde{f}(u) \downarrow 0 \text{ as } u \downarrow 0.
$$

Using these facts, we now proceed to prove our lemma under following three cases: $R - L > 0$, $R - L = 0$, and $R - L < 0$.

Suppose that $R - L > 0$. We now analyze the structural properties of $(\ell, u)$ pairs that satisfy (S3). Observe from the form of function $r(\cdot)$ that for each $u \in (0, 1)$ there are two $(\ell, u)$ pairs that satisfy (S3). From this, structural properties of $r(\cdot)$ and the assumption of strictly positive $R - L$, it follows that there exists a correspondence that consists of two functions $\ell_1 : (0, 1) \mapsto (0, 1)$ and $\ell_2 : (0, 1) \mapsto (0, 1)$ with following structural properties: (i) Both $(\ell_1(u), u)$ and $(\ell_2(u), u)$ pairs satisfy (S3) at each $u \in (0, 1)$, (ii) $\ell_1(u) < 1/2$ and $\ell_2(u) > 1/2$ for $u \in (0, 1)$, (iii) $\ell_1(\cdot)$ is concave and $\ell_2(\cdot)$ is convex on $(0, 1)$, (iv) $\ell_1(\cdot)$ is strictly increasing on $(0, 1/2)$ and strictly decreasing on $(1/2, 1)$; $\ell_2(\cdot)$ is strictly decreasing on $(0, 1/2)$ and strictly increasing on $(1/2, 1)$. Since we look for $(\ell, u)$ pairs that satisfy both (S2) and (S3), (EC.48) implies that we need to focus our attention only on $(\ell, u)$ pairs that satisfy $0 < \ell < u < 1$. By simple algebra, we have that

$$
\lim_{u \uparrow 1} \ell_2(u) = 1 \quad \text{and} \quad \lim_{u \downarrow 0} \ell_2(u) = 1.
$$

By perturbation analysis, we conclude that there exists a constant $\epsilon > 0$ sufficiently close to 0 such that $\ell_2(1 - \epsilon) = 1 - \epsilon + (R - L)\epsilon^2/\kappa + o(\epsilon^2) > 1 - \epsilon$. From this, (EC.50) and structural properties of $\ell_2$ stated in (ii) through (iv), it follows that $\ell_2(u) > u$ for $u \in (0, 1)$. As a result, it is sufficient to focus our attention only on $(\ell_1(u), u)$ pairs that satisfy (S3). Because $\ell_1(u) \to 0$ as $u \uparrow 1$ or $u \downarrow 0$ by (S3) and $\tilde{f}$ is strictly increasing and satisfies (EC.49), by structural properties of $\ell_1$ stated in (ii) through (iv) we have the following: if $\tilde{f}(\epsilon) < \ell_1(\epsilon)$ for $\epsilon > 0$ sufficiently close to 0, there exists a unique pair $(\ell, u)$ that satisfies (S2) and (S3). By (EC.48), this unique pair $(\ell, u)$ also satisfies $0 < \ell < u < 1$. Using perturbation analysis when $u = \epsilon > 0$ is sufficiently close to 0, we have that $\tilde{f}(\epsilon) = \epsilon - Re^2/\kappa + o(\epsilon^2)$ and $\ell_1(\epsilon) = \epsilon - (R - L)e^2/\kappa + o(\epsilon^2) > \tilde{f}(\epsilon)$. This completes our argument for the case of strictly positive $R - L$.

Suppose now that $R - L = 0$. Then, the set of $(\ell, u)$ pairs that satisfy (S3) is $\mathcal{J} = \{(\ell, u) : \ell \in \{u, 1 - u\}, u \in (0, 1)\}$. Therefore, set of $(\ell, u)$ pairs that satisfy both (S2) and (S3) is $\tilde{\mathcal{J}} = \mathcal{J} \cap \{(\tilde{f}(u), u), u \in (0, 1)\}$. From (EC.48), (EC.49) and the fact $\tilde{f}(\cdot)$ is continuous and strictly increasing, it follows that $\tilde{\mathcal{J}}$ includes a unique $(\ell, u)$ pair satisfying $0 < \ell < u < 1$.

Lastly, suppose that $R - L < 0$. Similar to the case of strictly positive $R - L$, there are two functions $u_1 : (0, 1) \mapsto (0, 1)$ and $u_2 : (0, 1) \mapsto (0, 1)$ such that for each $\ell \in (0, 1)$, both $(\ell, u_1(\ell))$ and $(\ell, u_2(\ell))$ satisfy (S3). Structural properties of $r(\cdot)$, (S3) and $R - L < 0$ imply following structural
functions \( u_1(\cdot) \) and \( u_2(\cdot) \): \( u_1(\ell) < 1/2 \) and \( u_2(\ell) > 1/2 \) for \( \ell \in (0, 1) \); \( u_1(\cdot) \) is concave and \( u_2(\cdot) \) is convex on \((0, 1)\); \( u_1(\cdot) \) is strictly increasing on \((0, 1/2)\) and strictly decreasing on \((1/2, 1)\); \( u_2(\cdot) \) is strictly decreasing on \((0, 1/2)\) and strictly increasing on \((1/2, 1)\). Using similar arguments used in the case of strictly positive \( R - L \), we conclude that it is sufficient to focus on pairs \( \{(\ell, u_2(\ell)), \ell \in (0, 1)\} \). Because \( u_2(\ell) \to 1 \) as \( \ell \downarrow 0 \) or as \( \ell \uparrow 1 \), we have the following from (EC.49), structural properties of \( u_2(\cdot) \) and facts that \( \bar{f}(\cdot) \) is continuous and strictly increasing: if \( u(1 - \epsilon) > u_2(1 - \epsilon) \) for \( \epsilon > 0 \) sufficiently close to 0 where the pair \( (1 - \epsilon, u(1 - \epsilon)) \) satisfies (EC.49), then there exists a unique pair \((\ell, u)\) that satisfies (EC.49), and \( 0 < \ell < u < 1 \). From perturbation analysis, it follows that for \( \epsilon > 0 \) sufficiently close to 0 \( (\ell, u) = (1 - \epsilon, 1 - \exp(-R/\bar{\kappa})\epsilon + o(\epsilon)) \) satisfies (EC.49) and \( u_2(1 - \epsilon) = 1 - \epsilon + o(\epsilon) \), which concludes our argument for this case.

Combining all three cases explained above completes our proof. \( \Box \)

**Appendix EC.5: Proof of Corollary**

**Proof of Corollary**

Recall the notation introduced in Lemma 6. Define \( F(u, \ell) = f(u) - f(\ell) - R/\bar{\kappa} \) and \( G(u, \ell) = r(\ell) - r(u) - (R - L)/\bar{\kappa} \). Then, it follows from multidimensional implicit function theorem that

\[
\left( \begin{array}{ccc} \frac{\partial u}{\partial R} & \frac{\partial u}{\partial L} & \frac{\partial u}{\partial \kappa} \\ \frac{\partial G}{\partial u} & \frac{\partial G}{\partial L} & \frac{\partial G}{\partial \kappa} \end{array} \right) = \left( \begin{array}{ccc} \frac{\partial F}{\partial u} & \frac{\partial F}{\partial L} & \frac{\partial F}{\partial \kappa} \\ \frac{\partial G}{\partial u} & \frac{\partial G}{\partial L} & \frac{\partial G}{\partial \kappa} \end{array} \right)^{-1}
\]

(EC.51)

Define \( \Delta(u, \ell) \equiv \ell^2(1 - \ell)^2u^2(1 - u)^2/(u - \ell) \). Using the definition of \( F \) and \( G \) and by simple algebra, we have that

\[
\left( \begin{array}{cc} \frac{\partial F}{\partial u} & \frac{\partial F}{\partial \ell} \\ \frac{\partial G}{\partial u} & \frac{\partial G}{\partial \ell} \end{array} \right)^{-1} = -\Delta(u, \ell) \left( \begin{array}{cc} \frac{(2\ell - 1)}{u^2(1 - u)^2} & \frac{1}{u(1 - u)} \\ \frac{(2u - 1)}{u^2(1 - u)^2} & \frac{1}{u(1 - u)} \end{array} \right),
\]

(EC.52)

\[
\left( \begin{array}{ccc} \frac{\partial F}{\partial R} & \frac{\partial F}{\partial L} & \frac{\partial F}{\partial \kappa} \\ \frac{\partial G}{\partial R} & \frac{\partial G}{\partial L} & \frac{\partial G}{\partial \kappa} \end{array} \right) = \left( \begin{array}{ccc} -\bar{\kappa}^{-1} & 0 & \frac{R}{\bar{\kappa}} \\ -\bar{\kappa}^{-1} & -\bar{\kappa}^{-1} & \frac{(R-L)}{\bar{\kappa}^2} \end{array} \right).
\]

(EC.53)

Plugging expressions in (EC.52) and (EC.53) into (EC.51), we obtain the following: \( \partial \ell / \partial R = -u\ell^2(1 - \ell)^2/(\bar{\kappa}(u - \ell)) < 0 \), \( \partial u / \partial R = -u\ell^2(1 - u)^2/(\bar{\kappa}(u - \ell)) < 0 \), \( \partial \ell / \partial L = (1 - \ell)u^2(1 - \ell)^2/(\bar{\kappa}(u - \ell)) > 0 \), \( \partial u / \partial L = (1 - \ell)u^2(1 - u)^2/(\bar{\kappa}(u - \ell)) > 0 \), \( \partial \ell / \partial \bar{\kappa} = (uR - (1 - u)L)\ell^2(1 - \ell)^2/\bar{\kappa} > 0 \), and \( \partial u / \partial \bar{\kappa} = (\ell R - (1 - \ell)L)u^2(1 - u)^2/\bar{\kappa} \leq 0 \). \( \Box \)