Bandit problems with Lévy processes

Asaf Cohen

The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel, cohenasa@post.tau.ac.il,

Eilon Solan

The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel, eilons@post.tau.ac.il,
www.math.tau.ac.il/~eilons

Bandit problems model the trade-off between exploration and exploitation in various decision problems. We
study two-armed bandit problems in continuous time, where the risky arm can have two types: High or Low;
both types yield stochastic payoffs generated by a Lévy process. We show that the optimal strategy is a
cut-off strategy and we provide an explicit expression for the cut-off and for the optimal payoff.

Key words: Two-armed bandit, Lévy processes, cut-off strategies.
MSC2000 subject classification: Primary: 60G40; secondary: 60G51
OR/MS subject classification: Primary: Decision analysis: theory; secondary: Probability: Markov
processes

1. Introduction Bandit problems, first described in Robbins (1952) [23], are a mathematical
model for studying the trade between exploration and exploitation. In its simplest formulation,
a decision maker (DM) faces N slot machines (called *arms*) and has to choose one of them at
each time instance. Each slot machine delivers a reward when and only when chosen. The reward's
distribution of each slot machine is drawn according to an unknown distribution, which itself
is drawn according to a known probability distribution from a set of known distributions. The
DM's goal is to maximize his total discounted payoff. The trade-off that the DM faces at each
stage is between exploiting the information that he already has, that is, choosing the arm that
looks optimal according to his information, and exploring the arms, that is, choosing a suboptimal
arm to improve his information about its payoff distribution. A good strategy for the DM will
involve phases of exploration and phases of exploitation. In exploration phases the DM samples
the rewards of the various machines and learns their rewards' distributions. In exploitation phases
the DM samples the machine whose reward's distribution so far is best until evidence shows that
its reward's distribution is not as good as expected.

Bandit problems have been applied to various areas, like economics, control, statistics, and
learning; see, e.g., Chernoff (1972) [7], Rothschild (1974) [25], Weitzman (1979) [28], Roberts and
Weitzman (1981) [24], Lai and Robbins (1984) [16], Bolton and Harris (1999) [6], Moscarini and
Squintani (2010) [18], Keller, Rady, and Cripps (2005) [13], Bergemann and Välimäki (2006) [2],
Besanko and Wu (2008) [5], and Klein and Rady (2011) [14].

Gittins and Jones (1979) [10] proved that in discrete time the optimal strategy of the DM has
a particularly simple form: at every period the DM calculates for each arm an index, which is a
real number, based on past rewards of that arm, and chooses the arm with the highest index. It

turns out that to calculate the index of an arm it is sufficient to consider an auxiliary problem
with two arms: an arm for which the index is calculated and an arm that yields a constant payoff.
The former arm is termed the *risky* arm, because its payoff distribution is not known, while the
latter is termed the *safe* arm. The literature therefore focuses on such problems, called two-armed
bandit problems.[1]

Once the optimality of the index strategy is guaranteed, one looks for the relation between the
data of the problems and the index. Explicit formulas for the index when the payoff is one of two
distributions that have a simple form have been established in the literature. Berry and Friestedt
(1985) [3] provide the solution to the problem in several cases, e.g., in discrete time when the
payoff distribution is one of two Bernoulli distributions, and in continuous time when the payoff
distribution is one of two Brownian motions. By studying the dynamic programming equation that
describes the problem in continuous time, Keller, Rady, and Cripps (2005) [13] and Keller and
Rady (2010) [12] provided an explicit form for the index when time is continuous and the payoff's
distribution is Poisson.[2]

In the present paper we study two-armed bandit problems in continuous time and provide an
explicit solution when the payoff distribution of the risky arm is one of two Lévy processes. We
assume that one distribution, called *High*, dominates the other, called *Low*, in a strong sense (see
Assumption 3 below). To eliminate trivial cases, we assume that the expected payoff generated by
the safe arm is lower than the expected payoff generated by the High distribution, and higher than
the expected payoff generated by the Low distribution.

In discrete time, under these assumptions the optimal strategy is a cut-off strategy: the DM
keeps experimenting as long as the posterior belief that the distribution is High is higher than
some cut-off point, and, once the posterior probability that the distribution is High falls below the
cut-off point, the DM switches to the safe arm forever. We extend this result to our setup, and
prove that when the two payoff distributions are Lévy processes that satisfy several requirements,
the optimal strategy is a cut-off strategy. Moreover, we provide an explicit expression for the cut-off
point in terms of the data of the problem. When particularized to the models studied by Bolton
and Harris (1999) [6], Keller, Rady, and Cripps (2005) [13], and Keller and Rady (2010) [12] our
expression reduces to the expressions that they obtained.

Apart from unifying previous results, our characterization shows that the special form of the
optimal payoff derived by Bolton and Harris (1999) [6] and Keller, Rady, and Cripps (2005) [13] is
valid in a general setup: the optimal payoff is the sum of the expected payoff, if no information is
available, and an option value that measures the expected gain from the ability to experiment. It
also shows that the data of the problem can be divided into information-relevant parameters and
payoff-relevant parameters; the information-relevant parameters can be summarized in a single real
number, and the payoff-relevant parameters are the expectations of the processes that contribute
to the DM's payoff. Finally, the characterization allows one to derive comparative statics on the
optimal cut-off and payoff. For example, as the discount rate increases, or the signals become less
informative, the cut-off point increases and the DM's optimal payoff decreases.

The rest of the paper is organized as follows. In Section 2 we present the model, the types of
strategies that we allow, and the assumptions that the payoff process should satisfy. In Section 3
we define the process of posterior belief and we develop its infinitesimal generator. In Section 4
we present the value function, and in Section 5 we present the Hamilton–Jacobi–Bellman (HJB)
equation. The main result, which characterizes the optimal strategy and the optimal payoff of the
DM, is formulated and proved in Section 6. The Appendix contains the proofs of several results
that are needed in the paper.

---

[1] In the literature these problems are also called one-armed bandit problems.

[2] These authors also studied the strategic setup in which several DMs have the same set of arms and their arms'
payoff distributions are the same (and unknown), and they compared the cooperative solution to the non-cooperative
solution.

2. The model

2.1. Reminder about Lévy processes Lévy processes are the continuous-time analog of discrete-time random walks with i.i.d. increments. A *Lévy process* $X = (X(t))_{t \geq 0}$ is a continuous-time stochastic process that (a) starts at the origin: $X(0) = 0$, (b) admits càdlàg modification,[3] and (c) has stationary independent increments. Examples of Lévy processes are a Brownian motion, a Poisson process, and a compound Poisson process.

Let $(X(t))$ be a Lévy process. For every Borel measurable set $A \subseteq \mathbb{R} \backslash \{0\}$, and every $t \geq 0$, let the Poisson random measure $N(t, A)$ be the number of jumps of $(X(t))$ in the time interval $[0, t]$ with jump size in A:

$$N(t, A) = \sharp \{ 0 \leq s \leq t \,|\, \Delta X(t) := X(s) - X(s-) \in A \}.$$

By Applebaum (2004) [1], one can define a Borel measure ν on $\mathcal{B}(\mathbb{R} \backslash \{0\})$ by

$$\nu(A) := E[N(1, A)] = \int N(1, A)(\omega) dP(\omega),$$

where (Ω, P) is the underlying probability space. The measure $\nu(A)$ is called the *Lévy measure* of $(X(t))$, or the *intensity measure* associated with $(X(t))$.

We now present the Lévy–Itō decomposition of Lévy processes. Let $(X(t))$ be a Lévy process; then there exists a constant $b \in \mathbb{R}$, a Brownian motion $\sigma Z(t)$ with standard deviation σ, and an independent Poisson random measure $N_\nu(t, dh)$ with the associated Lévy measure ν such that, for each $t \geq 0$,

$$X(t) = bt + \sigma Z(t) + \int_{h > |1|} h N_\nu(t, dh) + \int_{h \leq |1|} h \widetilde{N}_\nu(t, dh),$$

where $\widetilde{N}_\nu(t, A) := N_\nu(t, A) - t\nu(A)$ is the *compensated Poisson random measure*. This representation is called the *Lévy–Itō decomposition* of the Lévy process $(X(t))$. Thus, a Lévy processes is characterized by the triplet $\langle b, \sigma, \nu \rangle$.

If the Lévy process has finite expectation for each t, that is, $E|X(t)| < \infty$ for all $t \geq 0$, then the Lévy process can be represented as

$$X(t) = \mu t + \sigma Z(t) + \int_{\mathbb{R} \backslash \{0\}} h \widetilde{N}_\nu(t, dh);$$

that is, $X(t)$ can be represented as the sum of a linear drift, a Brownian motion, and an independent purely discontinuous martingale[4] (see Sato (1999, Theorem 25.3) [26]).

REMARK 1. Even though the process $(X(t))$ has finite expectation, it is possible that

$$E \left[\int_{\mathbb{R} \backslash \{0\}} |h| N_\nu(t, dh) \right] = \infty,$$

which means that the expectation of the sum of the jumps of $X(t)$ in any time interval is infinite.

---

[3] That is, it is continuous from the right, and has limits from the left: for every t_0, the limit $X(t_0-) := \lim_{t \nearrow t_0} X(t)$ exists a.s. and $X(t_0) = \lim_{t \searrow t_0} X(t)$.

[4] A *purely discontinuous process* is a process that is orthogonal to all continuous local martingales. For details, see Jacod and Shiryaev (1987, Ch. I, Definition 4.11) [11].

2.2. Lévy bandits A DM operates a two-armed bandit machine in continuous time, with a safe arm that yields a constant payoff ϱ, and a risky arm that yields a stochastic payoff $(X(t))$ that depends on its type θ. The risky arm can be of two types, High or Low. With probability $p_0 = p$ the arm's type is High, and with probability $1 - p$ it is Low. If the type is High (resp. Low) we set $\theta = 1$ (resp. 0). The process $(X(t))$ is a Lévy process with the triplet $\langle \mu_\theta, \sigma, \nu_\theta \rangle$; that is, the Lévy–Itō decomposition of $(X(t))$ is $X(t) = \mu_\theta t + \sigma_\theta Z(t) + \int_{\mathbb{R} \setminus \{0\}} h \widetilde{N}_{\nu_\theta}(t, dh)$. Formally, for $\theta \in \{0, 1\}$, let $(X_\theta(t))$ be a Lévy process with triplet $\langle \mu_\theta, \sigma, \nu_\theta \rangle$ and let θ be an independent Bernoulli random variable with parameter p. The process $(X(t))$ is defined to be $(X_\theta(t))$. We denote by P_p the probability measure over the space of realized paths that corresponds to this description. From now on, unless mentioned otherwise, all the expectations are taken under the probability measure P_p.

2.3. Strategies We adopt the concept of continuous-time strategies first introduced by Mandelbaum, Shepp, and Vanderbei (1990) [17]. An *allocation strategy* $\mathbf{K} = \{\mathbf{K}(t) \mid t \in [0, \infty)\}$ is a nonnegative stochastic process $\mathbf{K}(t) = (K_R(t), K_S(t))$ that satisfies

$$K_R(0) = K_S(0) = 0, \text{ and } (K_R(t)) \text{ and } (K_S(t)) \text{ are nondecreasing processes,} \tag{K1}$$
$$K_R(t) + K_S(t) = t, \ t \in [0, \infty), and \tag{K2}$$
$$\{K_R(t) \leq s\} \in \mathcal{F}_s^X, \ t, s \in [0, \infty), \tag{K3}$$

where $\mathcal{F}_s^X$ is the sigma-algebra generated by the process $(X(t))_{t \leq s}$. The interpretation of an allocation process is that the quantity $K_R(t)$ (resp. $K_S(t)$) is the time that the DM devotes to the risky arm (resp. safe arm) during the time interval $[0, t)$. The process $(\mathbf{K}(t))$ is basically a two-parameter time change of the two-dimensional process $(X(t), \varrho t)$.

Below we will define a stochastic integral with respect to $(X_\theta(t))$, and therefore we assume throughout that both Lévy processes $(X_1(t))$ and $(X_0(t))$ have finite quadratic variation, that is, $E[X_\theta^2(t)] < \infty$ for every $t \geq 0$ and each $\theta \in \{0, 1\}$. It follows that the processes $(X_\theta(t))$ have finite expectation.

ASSUMPTION 1.
A1. $E[X_\theta^2(1)] = \mu_\theta^2 + \sigma^2 + \int h^2 \nu_\theta(dh) < \infty$.

For every $(t, p) \in [0, \infty) \times [0, 1]$, every real-valued function $S : \mathbb{R} \to \mathbb{R}$, and every pair of Markov processes $(H_1(t))$ and $(H_2(t))$ with respect to the filtration $(\mathcal{F}_t^X)_{t \geq 0}$ under both P_0 and P_1 such that $E\left[S(\int_t^\infty H_1(s) dH_2(s)) \mid \theta\right]$ are well defined for both $\theta \in \{0, 1\}$, we define the following expectation operator:

$$E^{t,p}\left[S\left(\int_t^\infty H_1(s) dH_2(s)\right)\right] := p E\left[S\left(\int_t^\infty H_1(s) dH_2(s)\right)\Big| \theta = 1\right]$$
$$+ (1-p) E\left[S\left(\int_t^\infty H_1(s) dH_2(s)\right)\Big| \theta = 0\right]. \tag{2.1}$$

Using this notation, the expected discounted payoff from time t onwards under allocation strategy $\mathbf{K}$ when the prior belief at time t is $p_t = p$ can be expressed as

$$V_{\mathbf{K}}(t, p) := E^{t,p}\left[\int_t^\infty r e^{-rs} dY(\mathbf{K}(s))\right], \tag{2.2}$$

where $Y(\mathbf{K}(s)) := X(K_R(s)) + \varrho K_S(s)$. The goal of the DM is to maximize $V_{\mathbf{K}}(0,p)$. Let

$$U(t,p) := \sup_{\mathbf{K}} V_{\mathbf{K}}(t,p) \tag{2.3}$$

be the maximal payoff the DM can achieve from time t onwards, given that the prior belief at time t is $p_t = p$. As we show in Theorem 1 below, under proper assumptions the DM has an optimal strategy, so in fact the supremum in Eq. (2.3) is achieved. Moreover, we give explicit expressions for both the optimal strategy and the optimal value function $U(t,p)$.

REMARK 2. By Conditions (K1) and (K2), K_R and K_S are Lipschitz and thus absolutely continuous. Therefore, there exists a two-dimensional stochastic process $\mathbf{K}'(t) = \dfrac{d\mathbf{K}}{dt}(t) = (K_R'(t), K_S'(t))$ such that $\mathbf{K}(t) = \int_0^t \mathbf{K}'(s)ds$. To simplify notation we denote $K(t) := K_R(t)$, and $k(t) := K_R'(t)$. Hence, $\mathbf{K}(t) = (K(t), t - K(t))$ and $\mathbf{K}'(t) = (k(t), 1 - k(t))$. The process $(k(t))$ may be interpreted as follows: At each time instance t, the DM chooses $k(t)$ (resp. $1 - k(t)$), the proportion of time in the interval $[t, t + dt)$ that is devoted to the risky arm (resp. the safe arm). The process $(k(t))$ will be treated as a stochastic control parameter of the process $(X(t))$. Denote by $\mathcal{F}_{K(t)}$ the sigma-algebra generated by $(X(K(s)))_{s \le t}$.

DEFINITION 1. An *admissible control strategy* $(k(t, \omega))$ is any predictable process such that $0 \le k \le 1$ with probability 1, and such that the process $K(t) = \int_0^t k(s)ds$ satisfies[5] Condition (K3). Denote by Υ the set of all admissible control strategies.

In the sequel we will not distinguish between the allocation strategy $(K(t))$ and the corresponding admissible control strategy $(k(t))$.

2.4. Assumptions If the DM could deduce the type of the risky arm by observing the payoff of the risky arm in an infinitesimal time interval, then an almost-optimal strategy is to start at time 0 with the risky arm, and switch at time δ to the safe arm if the type of the risky arm is Low, where $\delta > 0$ is a small real number. Throughout the paper we make the following assumption, which implies that the DM cannot distinguish between the two types in any infinitesimal time.

ASSUMPTION 2.
A2. $\sigma_1 = \sigma_0$.
A3. $|\nu_1(\mathbb{R} \setminus \{0\}) - \nu_0(\mathbb{R} \setminus \{0\})| < \infty$.
A4. $|\int h(\nu_1(dh) - \nu_0(dh))| < \infty$.

Assumption A2 states that the Brownian motion component of both the High type and the Low type have the same standard deviation. By Revuz and Yor (1999, Ch. I, Theorem 2.7) [22] the realized path reveals the standard deviation and therefore if Assumption A2 does not hold then the DM can distinguish between the arms in any infinitesimal time interval. Assumption A3 states that the difference between the Lévy measures is finite and Assumption A4 states that the difference between the expectation of the jump part of the processes is finite. Otherwise, by comparing the jump part of the processes, the DM could distinguish between the arms in any infinitesimal time interval.

We also need the following assumption, which states that the High type is better then the Low type in a strong sense.

ASSUMPTION 3.
A5. $\mu_0 < \varrho < \mu_1$.
A6. For every $A \in \mathcal{B}(\mathbb{R} \setminus \{0\})$, $\nu_0(A) \le \nu_1(A)$.

---

[5] Since $0 \le k \le 1$, it follows that $K(t)$ satisfies Conditions (K1) and (K2) as well.

Assumption A5 merely says that the High (resp. Low) type provides higher (resp. lower) expected payoff than the safe arm. Assumption A6 is less innocuous; it requires that the Lévy measure of the High type dominates the Lévy measure of the Low type in a strong sense. Roughly, jumps of any size h, both positive and negative, occur more often (or at the same rate) under the High type than under the Low type. A consequence of this assumption is that jumps always provide good news, and (weakly) increase the posterior probability of the High type.

REMARK 3. Although we require that the zeroth and first moments of $(\nu_1(dh) - \nu_0(dh))$ are finite (Assumptions A3 and A4), this requirement is not made for moments of higher order, since, by Assumption 1, $\int_{\mathbb{R}\setminus\{0\}} h^2 \nu_\theta(dh) \le \mu_\theta^2 + \sigma^2 + \int_{\mathbb{R}\setminus\{0\}} h^2 \nu_\theta(dh) = E[X_\theta^2(1)] < \infty$, for $\theta \in \{0,1\}$.

3. The posterior belief

3.1. Motivation At time $t = 0$ the type θ is chosen randomly with $P(\theta = 1) = 1 - P(\theta = 0) = p$. The DM does not observe θ, but he knows the prior p and observes the controlled process $(X(K(t)))$. Let $p_t := P(\theta = 1 \mid \mathcal{F}_{K(t)})$ be the posterior belief at time t that the risky arm's type is High under the allocation strategy $(\mathbf{K}(t))$. The following proposition asserts that the payoff $V_{\mathbf{K}}(t, p)$ can be expressed solely by the[6] posterior process (p_{t-}) and the allocation strategy $(\mathbf{K}(t))$. This representation motivates the investigation of the posterior process.

PROPOSITION 1. *For every allocation strategy $\mathbf{K}$,*

$$V_{\mathbf{K}}(t,p) = E^{t,p}\left[\int_t^\infty re^{-rs}[(\mu_1 p_{s-} + \mu_0(1-p_{s-}))k(s) + \varrho(1-k(s))]ds\right].$$

Proof. We will prove the following series of equations, which proves the claim:

$$E^{t,p}\left[\int_t^\infty re^{-rs}dY(\mathbf{K}(s))\right] = E^{t,p}\left[\lim_{x\to\infty}\int_t^x re^{-rs}dY(\mathbf{K}(s))\right] \tag{3.1}$$

$$= \lim_{x\to\infty} E^{t,p}\left[\int_t^x re^{-rs}dY(\mathbf{K}(s))\right] \tag{3.2}$$

$$= \lim_{x\to\infty} E^{t,p}\left[\int_t^x re^{-rs}[(\mu_1 p_{s-} + \mu_0(1-p_{s-}))k(s) + \varrho(1-k(s))]ds\right] \tag{3.3}$$

$$= E^{t,p}\left[\lim_{x\to\infty}\int_t^x re^{-rs}[(\mu_1 p_{s-} + \mu_0(1-p_{s-}))k(s) + \varrho(1-k(s))]ds\right] \tag{3.4}$$

$$= E^{t,p}\left[\int_t^\infty re^{-rs}[(\mu_1 p_{s-} + \mu_0(1-p_{s-}))k(s) + \varrho(1-k(s))]ds\right]. \tag{3.5}$$

Eqs. (3.1) and (3.5) hold by the definition of the improper integral. Let $[X(K(s))]$ be the quadratic variation of the time-changed process $(X(K(s)))$. From the Itō isometry and Kobayashi (2011, pages 797–799) [15], it follows that

$$E^{t,p}\left[\int_t^\infty re^{-rs}dX(K(s)) - \int_t^x re^{-rs}dX(K(s))\right]^2$$

$$= E^{t,p}\left[\int_x^\infty re^{-rs}dX(K(s))\right]^2 = E^{t,p}\left[\int_x^\infty (re^{-rs})^2 d[X(K(s))]\right]$$

$$= E\left[\int_x^\infty (re^{-rs})^2 d[X(K(s))]\Big|\theta=1\right]p + E\left[\int_x^\infty (re^{-rs})^2 d[X(K(s))]\Big|\theta=0\right](1-p)$$

$$= E\left[\int_x^\infty (re^{-rs})^2 c_1 dK(s)\Big|\theta=1\right]p + E\left[\int_x^\infty (re^{-rs})^2 c_0 dK(s)\Big|\theta=0\right](1-p),$$

---

[6] p_{t-} is the posterior belief at time $K(t)-$.

where $c_\theta = E[X_\theta^2(1)] = \mu_\theta^2 + \sigma^2 + \int_{\mathbb{R}\setminus\{0\}} h^2 \nu_\theta(dh)$, for $\theta \in \{0,1\}$. Hence, $\int_t^x re^{-rs} dX(K(s))$ convergence to $\int_t^\infty re^{-rs} dX(K(s))$ in L^2 and Eq. (3.2) follows. Eq. (3.3) follows from Corollary 1 (part C1) in the appendix. Eq. (3.4) follows from the dominated convergence theorem, since for every $x \geq t$,

$$\left| \int_t^x re^{-rs}[(\mu_1 p_{s-} + \mu_0(1 - p_{s-}))k(s) + \varrho(1 - k(s))]ds \right| \leq \max\{|\mu_0|, |\mu_1|\}.$$

$\blacksquare$

3.2. Formal definition of the posterior belief An elegant formulation of the Bayesian belief updating process was presented by Shiryaev (1978, Ch. 4.2) [27] to a model in which the observed process is a Brownian motion with unknown drift and extended later to a model in which the observed process is a Poisson process with unknown rate in Peskir and Shiryaev (2000) [20].[7] We follow this formulation and extend it to the time-changed Lévy process. For every $p \in [0,1]$, the probability measure P_p satisfies $P_p = pP_1 + (1-p)P_0$. An important auxiliary process is the Radon–Nikodym density, given by

$$\varphi_t := \frac{d(P_0 \mid \mathcal{F}_{K(t)})}{d(P_1 \mid \mathcal{F}_{K(t)})}, \ \ t \in [0, \infty).$$

LEMMA 1. *For every $t \in [0, \infty)$,*

$$p_t = \frac{p}{p + (1-p)\varphi_t}.$$

Proof. Define the following Radon–Nikodym density process

$$\pi_t = p\frac{d(P_1 \mid \mathcal{F}_{K(t)})}{d(P_p \mid \mathcal{F}_{K(t)})}, \ \ t \in [0, \infty),$$

where $P_p(\cdot \mid \mathcal{F}_{K(t)}) = pP_1(\cdot \mid \mathcal{F}_{K(t)}) + (1-p)P_0(\cdot \mid \mathcal{F}_{K(t)})$. From the definition of (φ_t) it follows that $\pi_t = \frac{p}{p+(1-p)\varphi_t}$. Therefore, it is left to prove that $p_t = \pi_t$ for every $t \in [0, \infty)$. Let $A \in \mathcal{F}_{K(s)}$ where $s \geq t$. The following series of equations yields that $p_t = \pi_t$ for every $t \in [0, \infty)$:

$$E^p[\chi_A p_s | \mathcal{F}_{K(t)}] = E^p[\chi_A E^p[\chi_{\{\theta=1\}} | \mathcal{F}_{K(s)}] | \mathcal{F}_{K(t)}] \tag{3.6}$$
$$= E^p[\chi_{A \cap \{\theta=1\}} | \mathcal{F}_{K(t)}] \tag{3.7}$$
$$= pE^1[\chi_A | \mathcal{F}_{K(t)}] \tag{3.8}$$
$$= E^p[\chi_A \pi_s | \mathcal{F}_{K(t)}], \tag{3.9}$$

where $\chi_A = 1$ if A is satisfied and zero otherwise. Eq. (3.6) follows from the definition of p_t. Eq. (3.7) follows since $s \geq t$, and, therefore, $\mathcal{F}_{K(s)} \supseteq \mathcal{F}_{K(t)}$. Eq. (3.8) follows from the definition of the probability measure P_p, and Eq. (3.9) follows from the property of the Radon–Nikodym density π_t. $\blacksquare$

By Jacod and Shiryaev (1987, Ch. III, Theorems 3.24 and 5.19) [11], the process (φ_t) admits the following representation:

$$\varphi_t = \exp\left\{ \beta\sigma Z(K(t)) + (\bar{\nu}_1 - \bar{\nu}_0 - \frac{1}{2}\beta^2\sigma^2)K(t) + \int_{\mathbb{R}\setminus\{0\}} \ln\left(\frac{\nu_0}{\nu_1}(h)\right) N(K(t), dh) \right\},$$

---

[7] Similar work has been done in the disorder problem; see, e.g., Shiryaev (1978) [27], Peskir and Shiryaev (2002) [21], and Gapeev (2005) [9].

where $\beta := \dfrac{\mu_0 - \mu_1 - \int\limits_{\mathbb{R}\setminus\{0\}} h(\nu_0 - \nu_1)(dh)}{\sigma^2}$ and $\bar{\nu}_1 - \bar{\nu}_0 := \int\limits_{\mathbb{R}\setminus\{0\}} (\nu_1(dh) - \nu_0(dh))$. By Assumption A6, $\bar{\nu}_1 - \bar{\nu}_0$

is finite and the Radon–Nikodym derivative $\frac{\nu_0}{\nu_1}(h)$ exists.[8]

REMARK 4. 1. Let $B_\infty \in \mathcal{B}(\mathbb{R} \setminus \{0\})$ be a maximal set (up to ν_1-measure zero) such that $\nu_1(B_\infty) \geq 0 = \nu_0(B_\infty)$. Occurrence of a jump from B_∞ indicates that the risky arm is High. By definition, $\varphi_t = 0$ after such a jump and therefore $p_t = 1$.

2. By ignoring jumps from B_∞, $(\ln(\varphi_t))$ is a Lévy process[9] with time change $(K(t))$, under both P_0 and P_1 with respect to the filtration generated by (φ_t), which coincides with $(\mathcal{F}_{K(t)})$. From the one-to-one correspondence between φ_t and p_t it follows that p_t is a Markov process. Therefore, our optimal control problem falls in the scope of optimal control of Markov processes. Hence, we can limit the allocation strategies to *Markovian allocation strategies*, or equivalently to *Markovian control strategies*, which we define as follows.

DEFINITION 2. A control strategy $(k(t, \omega))$ is *Markovian* if it depends solely on the Markovian process (t, p_{t-}). That is, $k(t, \omega) = k(t, p_{t-})$. Denote by Υ_M the set of all Markovian control strategies.

REMARK 5. A convenient way to understand the "Girsanov style" process (φ_t) is to examine the process $(X(t))$. We may assume that

$$X(t) = \mu_1 t + \sigma Z(t) + \int\limits_{\mathbb{R}\setminus\{0\}} h \widetilde{N}_{\nu_1}(t, dh),$$

where, under P_1, $(Z(t))$ is a Brownian motion and the last term is a purely discontinuous martingale. By the definition of β, the same process can be represented as

$$X(t) = \mu_0 t + \sigma(\beta \sigma t + Z(t)) + \int\limits_{\mathbb{R}\setminus\{0\}} h \widetilde{N}_{\nu_0}(t, dh).$$

Under P_0 the process $(\beta \sigma t + Z(t))$ is a Brownian motion and the last term is a purely discontinuous martingale. For details, see Jacod and Shiryaev (1987, Ch. III) [11].

3.3. The infinitesimal operator An important tool in the proofs is the *infinitesimal operator* of the process (t, p_t) with respect to the Markovian control strategy k, which we will calculate in this section. The infinitesimal operator (or infinitesimal generator) of a stochastic process is the stochastic analog of a partial derivative (see Øksendal, 2000). In this section we calculate the infinitesimal operator of the process (t, p_t) with respect to the Markovian control strategy k, which we will use in the proof of Theorem 1. By Itō's formula (see, e.g., Kobayashi (2011), pages 797–799) [15], the posterior process (p_t) solves the following stochastic differential equation:

$$
\begin{aligned}
dp_t = {} & [\beta^2 \sigma^2 (1 - p_{t-})^2 p_{t-} - (\bar{\nu}_1 - \bar{\nu}_0) p_{t-}(1 - p_{t-})] dK(t) \qquad\qquad (3.10)\\
& - p_{t-}(1 - p_{t-}) \beta \sigma \, dZ(K(t)) \\
& + p_{t-}(1 - p_{t-}) \int\limits_{h \in \mathbb{R}\setminus\{0\}} \frac{1 - \frac{\nu_0}{\nu_1}(h)}{p_{t-} + \frac{\nu_0}{\nu_1}(h)(1 - p_{t-})} N(dK(t), dh) \\
= {} & p_{t-}(1 - p_{t-}) \Bigg[-\beta dM(K(t)) - (\bar{\nu}_1 - \bar{\nu}_0) dK(t) \\
& + \int\limits_{h \in \mathbb{R}\setminus\{0\}} \frac{1 - \frac{\nu_0}{\nu_1}(h)}{p_{t-} + \frac{\nu_0}{\nu_1}(h)(1 - p_{t-})} N(dK(t), dh) \Bigg],
\end{aligned}
$$

---

[8] To ensure the existence of the Radon–Nikodym derivative one does not need the full power of Assumption A6. Its full power will be used for the proof of Theorem 1.

[9] However, it is not a Lévy process under P_p for $0 < p < 1$, since it is not time-homogeneous.

where

$$M(K(t)) = X(K(t)) - \int_{\mathbb{R}\setminus\{0\}} h\widetilde{N}_{\nu_0}(K(t), dh) - \mu_0 K(t) + \beta\sigma^2 \int_0^t p_{s-}dK(s)$$

is a martingale under P_p with respect to $\mathcal{F}_{K(t)}$; see Corollary 1 (part C2) in the appendix.

The first term on the right-hand side of (3.10), $-p_{t-}(1 - p_{t-})\beta dM(K(t))$, is the contribution of the continuous part of the payoff process to the change in the belief, while the second term, $-p_{t-}(1 - p_{t-})(\bar{\nu}_1 - \bar{\nu}_0)dK(t)$, is the contribution of the fact that no jump occurred. This latter contribution is negative due to Assumption A6. If a jump of size h occurs during the interval $[t, t+dt)$, then the contribution of the jump is $P_t^h - p_{t-}$, where, $P_t^h := \frac{p_{t-}\nu_1(dh)}{p_{t-}\nu_1(dh)+(1-p_{t-})\nu_0(dh)}$ is the Bayesian update of the probability that the risky arm is High given that a jump of size h occurs. By Assumption A6, for every $0 < p < 1$ we have $P_p(p_t < P_t^h) = 1$.

To calculate the infinitesimal operator of the process (t, p_t) with respect to the Markovian control strategy k we apply Itō's formula[10] for $f(t, p) \in C^{1,2}([0, \infty) \times [0, 1])$ and obtain

$$\begin{aligned}
f(t, p_t) =& f(0, p_0) + \int_0^t f_t(s, p_{s-})ds + \int_0^t f_p(s, p_{s-})dp_s \qquad (3.11)\\
&+ \frac{1}{2}\int_0^t f_{pp}(s, p_{s-})p_{s-}^2(1-p_{s-})^2\beta^2\sigma^2 dK(s)\\
&+ \sum_{s\leq t}[f(s, p_s) - f(s, p_{s-}) - f_p(s, p_{s-})\Delta(p_s)]\\
=& f(0, p_0) + \int_0^t f_t(s, p_{s-})ds\\
&- \int_0^t f_p(s, p_{s-})[(\bar{\nu}_1 - \bar{\nu}_0)p_{s-}(1-p_{s-})]dK(s)\\
&+ \frac{1}{2}\int_0^t f_{pp}(s, p_{s-})p_{s-}^2(1-p_{s-})^2\beta^2\sigma^2 dK(s)\\
&+ \int_{s=0}^t \int_{h\in\mathbb{R}\setminus\{0\}} \left(f\left(s, \frac{p_{s-}}{p_{s-} + (1-p_{s-})\frac{\nu_0}{\nu_1}(h)}\right) - f(s, p_{s-})\right)\\
&\cdot(p_{s-}\nu_1(dh) + (1-p_{s-})\nu_0(dh))dK(s)\\
&- \int_0^t f_p(s, p_{s-})p_{s-}(1-p_{s-})\beta dM(K(s))\\
&+ \int_{s=0}^t \int_{h\in\mathbb{R}\setminus\{0\}} \left(f\left(s, \frac{p_{s-}}{p_{s-} + (1-p_{s-})\frac{\nu_0}{\nu_1}(h)}\right) - f(s, p_{s-})\right)\\
&\cdot[N(dK(s), dh) - (p_{s-}\nu_1(dh) + (1-p_{s-})\nu_0(dh))dK(s)].
\end{aligned}$$

The fifth and sixth terms on the right-hand side of Eq. (3.11) are stochastic integrals with respect to martingales and therefore they are local martingales (see Jacod and Shiryaev (1987, Ch. I, Theorem 4.40) [11]). The seventh term is a stochastic integral with respect to a compensated random measure, as will be shown in Corollary 1 (parts C2 and C3) in the appendix. Therefore, it is a local martingale (see Jacod and Shiryaev (1987, Ch. II, Theorem 1.8) [11]). Hence, by taking

---

[10] $C^{1,2}$ is the set of all functions $f : [0, \infty) \times [0, 1] \to \mathbb{R}$, which are C^1 in their first coordinate, and C^2 in their second coordinate.

expectations of both sides it follows that the infinitesimal operator of the process (t, p_t) with respect to the Markovian control strategy $(k(t,p))$ is

$$(\mathbb{L}^k f)(t,p) = f_t(t,p) - (\bar{\nu}_1 - \bar{\nu}_0)p(1-p)f_p(t,p)k(t,p) + \frac{1}{2}\beta^2\sigma^2 f_{pp}(t,p)p^2(1-p)^2 k(t,p)$$
$$+ \int_{\mathbb{R}\setminus\{0\}} \left(f\left(t, \frac{p}{p + (1-p)\frac{\nu_0}{\nu_1}(h)}\right) - f(t,p) \right)(p\nu_1(dh) + (1-p)\nu_0(dh))k(t,p).$$

When f is a function of p only, we will use the same notation for the infinitesimal operator of the process (p_t) with respect to the time-homogeneous Markovian control strategy $k(p)$. Specifically,

$$(\mathbb{L}^k f)(p) = -(\bar{\nu}_1 - \bar{\nu}_0)p(1-p)f'(p)k(p) + \frac{1}{2}\beta^2\sigma^2 f''(p)p^2(1-p)^2 k(p) \tag{3.12}$$
$$+ \int_{\mathbb{R}\setminus\{0\}} \left(f\left(\frac{p}{p + (1-p)\frac{\nu_0}{\nu_1}(h)}\right) - f(p) \right)(p\nu_1(dh) + (1-p)\nu_0(dh))k(p).$$

4. The value function In the next section we will introduce the Hamilton–Jacobi–Bellman (HJB) for our problem. The value function $U(t,p)$ is not C^2 in its second coordinate, and therefore we need to formalize the optimal problem differently. Additionally to the Markovian control strategy $k(t,p)$, we will add an artificial stopping time τ to the new strategy space. This new form will help us solve the HJB although $U(t,p)$ is not C^2. We start with a few basic properties of the value function $U(t,p)$.

PROPOSITION 2. *For every fixed $t \geq 0$, the function $p \mapsto U(t,p)$ is monotone, nondecreasing, convex, and continuous.*

Proof. Fix for a moment an allocation strategy $\mathbf{K}$. By Definition 2.1 and Eq. (2.2) the expected discounted payoff from time t onwards under strategy $\mathbf{K}$ when $p_t = p$ is

$$V_{\mathbf{K}}(t,p) = E^{t,p}\left[\int_t^\infty re^{-rs}dY(\mathbf{K}(s)) \right]$$
$$= pE\left[\int_t^\infty re^{-rs}dY(\mathbf{K}(s)) \Big| \theta = 1 \right] + (1-p)E\left[\int_t^\infty re^{-rs}dY(\mathbf{K}(s)) \Big| \theta = 0 \right].$$

For every fixed $t \geq 0$ the function $p \mapsto V_{\mathbf{K}}(t,p)$ is linear. Therefore $U(t,p)$, as the supremum of linear functions, is convex. By always choosing the safe arm, the DM can achieve at least $e^{-rt}\varrho$, and by always choosing the risky arm the DM can achieve at least $e^{-rt}(p\mu_1 + (1-p)\mu_0)$. Since $U(t,0) = e^{-rt}\varrho$ and $U(t,1) = e^{-rt}\mu_1$, the convexity of $U(t,p)$ implies that the function $p \mapsto U(t,p)$ is continuous and nondecreasing in p. ∎

It follows from Proposition 2 that for every fixed $t \geq 0$ there is a time-dependent cut-off p_t^* in $[0,1]$ such that $U(t,p) = \varrho$ if $p \leq p_t^*$ and $U(t,p) > \varrho$ otherwise. It follows that for every fixed t the strategy $k(t,p) \equiv 0$ that always chooses the safe arm is optimal for prior beliefs in $[0, p_t^*]$. By this conclusion, Proposition 1, and Remark 2 we deduce that the optimal problem (2.3) can be reduced to a combined optimal stopping and stochastic control problem as follows:

$$U(t,p) = \sup_{t \leq \tau, \, k \in \Upsilon_M} E^{t,p}\left[\int_t^\tau re^{-rs}W(p_{s-}, k(s, p_{s-}))ds + \varrho e^{-r\tau} \right], \tag{4.1}$$

where $W(p,l) := (\mu_1 p + \mu_0(1-p))l + \varrho(1-l)$ is the instantaneous payoff given the posterior p, using the Markovian control l. This representation of the value function will help us solve the HJB equation. Denote the continuation region to be

$$D := \{(t,p) \mid U(t,p) > \varrho e^{-rt}\}.$$

This is the region where the optimal action of the DM is to continue (that is, $k(t,p) > 0$, and $\tau > t$). The next lemma shows that the region D is invariant with respect to t. This means that the optimal stopping time τ (whenever it exists) does not depend[11] on t.

LEMMA 2. *For every $t \geq 0$ and every $p \in [0,1]$ one has $U(t,p) = e^{-rt}U(0,p)$. In particular, $(t,p) \in D$ if and only if $(s,p) \in D$, for every $t, s \geq 0$ and every $p \in [0,1]$.*

Proof. The first claim follows from the following list of equalities:

$$
\begin{aligned}
U(t,p) &= \sup_{t \leq \tau,\, 0 \leq k \leq 1} E^{t,p} \left[\int_t^\tau re^{-rs}W(p_{s-}, k(s,p_{s-}))ds + \varrho e^{-r\tau} \right] \quad (4.2)\\
&= \sup_{0 \leq \tilde{\tau},\, 0 \leq k \leq 1} E \left[\int_0^{\tilde{\tau}} re^{-r(t+u)}W(p_{u-}^p, k(t+u, p_{u-}^p))du + \varrho e^{-r(t+\tilde{\tau})} \right]\\
&= e^{-rt} \sup_{0 \leq \tilde{\tau},\, 0 \leq k \leq 1} E^{0,p} \left[\int_0^{\tilde{\tau}} re^{-ru}W(p_{u-}, k(t+u, p_{u-}))du + \varrho e^{-r\tilde{\tau}} \right]\\
&= e^{-rt}U(0,p),
\end{aligned}
$$

where the second equality follows from the Markovian property of p_t (see Remark 4). ∎

This lemma yields that the cut-off p_t^* discussed earlier is independent of t. We therefore denote it by p^*.

5. The HJB equation The following proposition introduces the HJB equation for our problem.

PROPOSITION 3. *Let $F \in C^1[0,1]$ be a function that satisfies*

$$F(p) \geq \varrho \text{ for every } p \in [0,1]. \quad (5.1)$$

Define the continuation region of F by

$$C := \{p \in [0,1] \mid F(p) > \varrho\}. \quad (5.2)$$

Suppose that

$$[0,\infty) \times C = D. \quad (5.3)$$
$$F \in C^2([0,1]\setminus\partial C) \text{ with locally bounded derivatives near } \partial C. \quad (5.4)$$
$$\mathbb{L}^k F(p) + rW(p,k(p)) - rF(p) \leq 0 \text{ on } [0,1]\setminus\partial C \quad (5.5)$$
$$\text{for all } k \in \Upsilon_M, \text{ and all } p \in [0,1].$$
$$\text{There is a control } k^* \text{ for which the inequality in Eq. (5.5)} \quad (5.6)$$
$$\text{holds with equality.}$$

Then, k^* is the optimal control, $\tau_D := \inf\{t \geq 0 \mid F(p_t) \notin C\}$ is the optimal stopping time, and $U(t,p) = e^{-rt}F(p)$.

---

[11] In fact, we will show in Theorem 1 that an optimal stopping time and an optimal control strategy do exist and the optimal control k is also time-homogeneous; that is, k does not depend on t, and therefore the allocation strategy **K** does not depend on t either.

Conditions (5.5) and (5.6) represent the HJB equation in our model.

REMARK 6. The function F need not be C^2 at the boundary of C. This is due to the representation of $U(t,p)$ in Eq. (4.1) as a combined optimal stopping and stochastic control problem. This issue will be further discussed in the proof.

Proof. Define $J(t,p) := e^{-rt}F(p)$. Then for every $(t,p) \in [0,\infty) \times ([0,1] \setminus \partial C)$,

$$\mathbb{L}^k J(s,p) = -re^{-rt}F(p) + e^{-rt}\mathbb{L}^k F(p). \tag{5.7}$$

By Eq. (5.4), $J(t,p) \in C^{1,2}([0,\infty) \times ([0,1] \setminus \partial C))$ and $J_{pp}(s,p)$ is bounded near $[0,\infty) \times \partial C$. Therefore, there exists a sequence $\{J^n\}_{n \geq 1} \subseteq C^{1,2}(D)$ such that

$$J^n \to J, \ J_t^n \to J_t, \ J_p^n \to J_p, \ J_{pp}^n \to J_{pp}$$

uniformly on every compact subset of $[0,\infty) \times ([0,1] \setminus \partial C)$ as n goes to infinity (see Øksendal (2000, Theorem C.1) [19]). Denote by $L(t)$ the sum of the last three terms on the right-hand side of Eq. (3.11). The process $(L(t))$, as the sum of local martingales, is a local martingale. Let (δ_m) be a sequence of increasing (a.s.) stopping times that diverge (a.s.), such that $L(\delta_m \wedge t)$ is[12] a martingale for every m. Let τ be an arbitrary stopping time and define $\tau_m := \tau \wedge m \wedge \delta_m$. We will prove the following series of equations:

$$
\begin{aligned}
E^{0,p} &\left[e^{-r\tau_m}F(p_{\tau_m})\right] - F(p) \\
&= E^{0,p}\left[J(\tau_m, p_{\tau_m})\right] - J(0,p) \tag{5.8} \\
&= \lim_{n \to \infty} \left(E^{0,p}\left[J^n(\tau_m, p_{\tau_m})\right] - J^n(0,p)\right) \tag{5.9} \\
&= \lim_{n \to \infty} E^{0,p}\left[\int_0^{\tau_m} \mathbb{L}^k J^n(s, p_{s-})ds\right] \tag{5.10} \\
&= \lim_{n \to \infty} E^{0,p}\left[\int_0^{\tau_m} \mathbb{L}^k J^n(s, p_{s-})\chi_{\{p_{s-} \notin \partial C\}}ds\right] \tag{5.11} \\
&= E^{0,p}\left[\int_0^{\tau_m} \mathbb{L}^k J(s, p_{s-})\chi_{\{p_{s-} \notin \partial C\}}ds\right] \tag{5.12} \\
&= E^{0,p}\left[\int_0^{\tau_m} e^{-rs}\left(-rF(p_{s-}) + \mathbb{L}^k F(p_{s-})\right)\chi_{\{p_{s-} \notin \partial C\}}ds\right] \tag{5.13} \\
&\leq -E^{0,p}\left[\int_0^{\tau_m} e^{-rs}W(p_{s-}, k(p_{s-}))\chi_{\{p_{s-} \notin \partial C\}}ds\right] \tag{5.14} \\
&= -E^{0,p}\left[\int_0^{\tau_m} e^{-rs}W(p_{s-}, k(p_{s-}))ds\right], \tag{5.15}
\end{aligned}
$$

where, $\chi_A = 1$ if A is satisfied and zero otherwise. Eq. (5.8) follows from the definition of $J(t,p)$. Eqs. (5.9) and (5.12) follow from the choice of the sequence J^n. Eq. (5.10) follows from the definition of the infinitesimal operator. By condition (5.3), the boundary ∂C of C is a single point (specifically, it is the cut-off point p^*). Therefore, Eqs. (5.11) and (5.15) follow from Lemma 5 in the appendix. Eq. (5.13) follows from Eq. (5.7), and inequality (5.14) follows from (5.5).

By Eq. (5.1) and the series of Eqs. (5.8)–(5.15) we obtain

$$
\begin{aligned}
E^{0,p} &\left[\int_0^{\tau_m} e^{-rs}W(p_{s-}, k(p_{s-}))ds + \varrho e^{-r\tau_m}\right] \\
&\leq E^{0,p}\left[\int_0^{\tau_m} e^{-rs}W(p_{s-}, k(p_{s-}))ds + e^{-r\tau_m}F(p_{\tau_m})\right] \leq F(p).
\end{aligned}
$$

---

[12] $a \wedge b := \min\{a, b\}$.

By taking $m \to \infty$ we deduce that

$$E^{0,p}\left[\int_0^\tau e^{-rs}W(p_{s-}, k(p_{s-}))ds + \varrho e^{-r\tau}\right] \leq F(p). \tag{5.16}$$

The left-hand side of Eq. (5.16) is the payoff of the DM using the stopping time τ and the stationary Markovian control strategy k. By taking the supremum in Eq. (5.16) it follows that $U(0,p) \leq F(p)$ for every $0 \leq p \leq 1$. To prove the opposite inequality, apply the argument above to the stationary Markovian control strategy $k^* = k^*(p_{s-})$ and the stopping time τ_D, so that the inequality in Eq. (5.14) is replaced by an equality. By taking the limit $m \to \infty$ and by the definition of D we obtain

$$\begin{aligned}
U(0,p) &\geq E^{0,p}\left[\int_0^{\tau_D} e^{-rs}W(p_{s-}, k^*(p_{s-}))ds + e^{-r\tau_D}\varrho\right] \\
&= E^{0,p}\left[\int_0^{\tau_D} e^{-rs}W(p_{s-}, k^*(p_{s-}))ds + e^{-r\tau_D}F(p_{\tau_D})\right] = F(p).
\end{aligned}$$

∎

6. The optimal strategy In this section we present our main result that states that there is a unique optimal allocation strategy, and that it is a cut-off strategy. The theorem also provides the exact cut-off point and the corresponding expected payoff in terms of the data of the problem. Let α^* be the unique solution in $(0,\infty)$ of the equation $f(\alpha) = 0$, where

$$f(\alpha) := \int\left(\left(\frac{\nu_0}{\nu_1}(h)\right)^\alpha - 1\right)\nu_0(dh) + \alpha(\bar{\nu}_1 - \bar{\nu}_0) + \frac{1}{2}(\alpha + 1)\alpha\left(\frac{\mu_1 - \mu_0}{\sigma}\right)^2 - r = 0. \tag{6.1}$$

The existence and the uniqueness of such a solution are proved in Lemma 6 in the appendix.

THEOREM 1. *Denote $p^* := \frac{\alpha^*(\varrho - \mu_0)}{(\alpha^*+1)(\mu_1 - \varrho) + \alpha^*(\varrho - \mu_0)}$. Under Assumptions A1–A6 there is a unique optimal strategy k^* that is time-homogeneous and is given by*

$$k^* = \begin{cases} 0 & \text{if } p \leq p^*, \\ 1 & \text{if } p > p^*. \end{cases} \tag{6.2}$$

The expected payoff under k^* is

$$U(0,p) = V_{k^*}(0,p) = \begin{cases} \varrho & \text{if } p \leq p^*, \\ p\mu_1 + (1-p)\mu_0 + C_{\alpha^*}(1-p)(\frac{1-p}{p})^{\alpha^*} & \text{if } p > p^*, \end{cases} \tag{6.3}$$

where $C_{\alpha^*} = \frac{\varrho - \mu_0 - p^*(\mu_1 - \mu_0)}{(1-p^*)\left(\frac{1-p^*}{p^*}\right)^{\alpha^*}}$.

We now discuss the relation between Theorem 1 and the results of Bolton and Harris (1999) [6], Keller, Rady, and Cripps (2005) [13], and Keller and Rady (2010) [12]. The expected payoff from the risky arm if no information is available is

$$\int_0^\infty re^{-rs}E[X(s)]ds = \int_0^\infty re^{-rs}[p\mu_1 + (1-p)\mu_0]s\,ds = p\mu_1 + (1-p)\mu_0.$$

One can verify that the strategy $k^* \equiv 1$ and the function $F(p) = p\mu_1 + (1-p)\mu_0$ satisfy Condition (5.6), and following the results of Bolton and Harris (1999) [6] and Keller, Rady, and Cripps (2005) [13], one can "guess" that a function of the form $C(1-p)(\frac{1-p}{p})^\alpha$ satisfies Condition (5.6)

as well. This leads to the form of the optimal payoff that appears in Eq. (6.3). The function $C_{\alpha^*}(1-p)(\frac{1-p}{p})^{\alpha^*}$ is the option value for the ability to switch to the safe arm. The parameters of the payoff processes that determine the cut-off point p^* and the optimal payoff $U(0,p)$ are the expected payoffs μ_1 and μ_0. In Bolton and Harris (1999) [6] the only component in the risky arm is the Brownian motion with drift. Therefore, $\nu_i \equiv 0$, so that $\alpha^* = (-1 + \sqrt{1 + 8r\sigma^2/(\mu_1 - \mu_0)^2})/2$. In Keller, Rady, and Cripps (2005) [13], the risky arm is either the constant zero (Low type, so that $\nu_0 \equiv 0$), or it yields a payoff $\bar{h}$ according to a Poisson process of rate λ (High type). If the risky arm is High, then the only component in the Lévy–Itō decomposition is the purely discontinuous component, and $\nu_1(\bar{h}) = \lambda$ and zero otherwise. Therefore, $\mu_1 = \lambda\bar{h}$, $\mu_0 = 0$, and $\alpha^* = r/\lambda$. In Keller and Rady (2010) [12], the risky arm yields a payoff $\bar{h}$ according to a Poisson process. For the High type, the Poisson process rate is λ_1, and for the Low type the rate is λ_0, where $\lambda_0 < \lambda_1$. The only component in the Lévy–Itō decomposition is the purely discontinuous component, and $\nu_i(\bar{h}) = \lambda_i$ and zero otherwise. Therefore, $\mu_1 = \lambda_1\bar{h}$, $\mu_0 = \lambda_0\bar{h}$, and α^* is the unique solution of the equation

$$f(\alpha) := \lambda_0 \left(\frac{\lambda_0}{\lambda_1}\right)^{\alpha} + \alpha(\lambda_1 - \lambda_0) - \lambda_0 - r = 0.$$

Proof of Theorem 1. Let p^*, α^*, and C_α^* be the parameters that were defined in the theorem. Define the cut-off strategy k^* associated with the cut-off p^* by

$$k^* := \begin{cases} 0 & \text{if } p \leq p^*, \\ 1 & \text{if } p > p^*, \end{cases} \tag{6.4}$$

and the function F by

$$F(p) := \begin{cases} \varrho & \text{if } p \leq p^*, \\ F_1(p) & \text{if } p > p^*, \end{cases} \tag{6.5}$$

where $F_1(p) := p\mu_1 + (1-p)\mu_0 + C_{\alpha^*}(1-p)(\frac{1-p}{p})^{\alpha^*}$. We will show that the function $F(p)$ and the cut-off strategy k^* are the optimal payoff and the optimal Markovian control strategy respectively. To this end, we verify that conditions (5.1)–(5.6) are satisfied for $F(p)$ and k^*. Conditions (5.1)–(5.4) can be easily verified for the function $F(p)$. To verify conditions (5.5) and (5.6) we need the full power of Assumption A6. To prove that $F(p)$ satisfies Condition (5.5) we check separately the cases $p \leq p^*$ and $p > p^*$. Fix a Markovian control strategy $k \in \Upsilon_M$ and $p \leq p^*$. Then,

$$\mathbb{L}^k F(p) + rW(p, k(p)) - rF(p)$$
$$= -(\bar{\nu}_1 - \bar{\nu}_0)p(1-p)F'(p)k(p) + \frac{1}{2}\beta^2\sigma^2 F''(p)p^2(1-p)^2 k(p) \tag{6.6}$$
$$+ \int_{\mathbb{R}\setminus\{0\}} \left(F\left(\frac{p}{p + (1-p)\frac{\nu_0}{\nu_1}(h)}\right) - F(p)\right)(p\nu_1(dh) + (1-p)\nu_0(dh))k(p)$$
$$+ r[(\mu_1 p + \mu_0(1-p))k(p) + \varrho(1-k(p))] - rF(p)$$
$$= \int_G \left(F_1\left(\frac{p}{p + (1-p)\frac{\nu_0}{\nu_1}(h)}\right) - \varrho\right)(p\nu_1(dh) + (1-p)\nu_0(dh))k(p) \tag{6.7}$$
$$+ r[(\mu_1 p + \mu_0(1-p))k(p) + \varrho(1-k(p))] - r\varrho$$
$$\leq \int_G \left(F_1\left(\frac{p^*}{p^* + (1-p^*)\frac{\nu_0}{\nu_1}(h)}\right) - \varrho\right)(p^*\nu_1(dh) + (1-p^*)\nu_0(dh))k(p) \tag{6.8}$$
$$+ r[(\mu_1 p^* + \mu_0(1-p^*))k(p) + \varrho(1-k(p))] - r\varrho$$

$$\leq \int\limits_{\mathbb{R}\setminus\{0\}} \left(F_1\left(\frac{p^*}{p^* + (1-p^*)\frac{\nu_0}{\nu_1}(h)} \right) - \varrho \right) (p^*\nu_1(dh) + (1-p^*)\nu_0(dh))k(p) \tag{6.9}$$

$$+ r[(\mu_1 p^* + \mu_0(1-p^*))k(p) + \varrho(1-k(p))] - r\varrho$$

$$= 0, \tag{6.10}$$

where $G = \left\{ h \mid \frac{p}{p+(1-p)\frac{\nu_0}{\nu_1}(h)} > p^* \right\}$. Eq. (6.6) follows from Eq. (3.12). Eq. (6.7) follows since for $p < p^*$ we have $F(p) = \varrho$, $F'(p) = F''(p) = 0$, and for $p > p^*$ we have $F(p) = F_1(p)$. Inequality (6.8) follows from the monotonicity of F_1 and Assumptions A5 and A6. Inequality (6.9) follows since $F_1(q) > \varrho$ for every $q > p^*$. Eq. (6.10) is satisfied for every $k(p)$ by the definition of p^* and F_1.

Fix a Markovian control strategy $k \in \Upsilon_M$ and $p > p^*$. Then $F(p) = F_1(p)$ and $F\left(\frac{p}{p+(1-p)\frac{\nu_0}{\nu_1}(h)} \right) = F_1\left(\frac{p}{p+(1-p)\frac{\nu_0}{\nu_1}(h)} \right)$, since by Assumption A6, for every $0 \leq p < 1$ we have $P_p\left(\frac{p}{p+(1-p)\frac{\nu_0}{\nu_1}(h)} > p \right) = 1$. Therefore,

$$\mathbb{L}^k F(p) + rW(p, k(p)) - rF(p)$$

$$= -(\bar{\nu}_1 - \bar{\nu}_0)p(1-p)F'(p)k(p) + \frac{1}{2}\beta^2\sigma^2 F''(p)p^2(1-p)^2 k(p) \tag{6.11}$$

$$+ \int\limits_{\mathbb{R}\setminus\{0\}} \left(F\left(\frac{p}{p+(1-p)\frac{\nu_0}{\nu_1}(h)} \right) - F(p) \right) (p\nu_1(dh) + (1-p)\nu_0(dh))k(p)$$

$$+ r[(\mu_1 p + \mu_0(1-p))k(p) + \varrho(1-k(p))] - rF(p)$$

$$= -(\bar{\nu}_1 - \bar{\nu}_0)p(1-p)F_1'(p)k(p) + \frac{1}{2}\beta^2\sigma^2 F_1''(p)p^2(1-p)^2 k(p) \tag{6.12}$$

$$+ \int\limits_{\mathbb{R}\setminus\{0\}} \left(F_1\left(\frac{p}{p+(1-p)\frac{\nu_0}{\nu_1}(h)} \right) - F_1(p) \right) (p\nu_1(dh) + (1-p)\nu_0(dh))k(p)$$

$$+ r[(\mu_1 p + \mu_0(1-p))k(p) + \varrho(1-k(p))] - rF_1(p)$$

$$\leq 0, \tag{6.13}$$

where the last inequality is satisfied for every $k(p)$ by the definition of F_1.

By the definition of the Markovian control strategy k^*, for every $p \leq p^*$ we have $k^*(p) = 0$ and therefore Eqs. (6.6)–(6.10) hold with equality, and for every $p > p^*$ we have $k^*(p) = 1$ and therefore Eqs. (6.11)–(6.13) hold with equality by the definition of $F(p)$. This proves condition (5.6).

Without Assumption A6 there may be a set B that satisfies $\nu_0(B) > 0$, such that for every $h \in B$ and every $0 \leq p < 1$ one has $P_p\left(\frac{p}{p+(1-p)\frac{\nu_0}{\nu_1}(h)} < p \right) = 1$. Thus, for every $p \in \left[p^*, \frac{p^*}{p^*+(1-p^*)\frac{\nu_0}{\nu_1}(h)} \right)$ we need to substitute $F_1(p)$ for $F(p)$, and ϱ for $F\left(\frac{p}{p+(1-p)\frac{\nu_0}{\nu_1}(h)} \right)$. This problem has a higher level of complexity and it is not clear how to approach it using the tools introduced here.

REMARK 7. Since the process (p_t) has no negative jumps, it enters the interval $[0, p^*]$ continuously. Therefore, we expect the value function to be C^1 at the cut-off point p^*. In a model where the process (p_t) has negative jumps, it can enter the interval $[0, p^*]$ with a jump. In this case we expect that the value function will not be C^1 at the cut-off point p^*. For simple cases of Lévy processes (such as when the High (resp. Low) type is a jump process with height h_1 and rate λ_1 (resp. height h_0 and rate λ_0), where $h_1\lambda_1 > \varrho > h_0\lambda_0$ and $\lambda_1 < \lambda_0$, so, in particular, Assumption A3 fails) the method introduced in Peskir and Shiryaev (2000) [20] may be useful to characterize the optimal strategy and the value function. In the general setup, a sample path method may be helpful to approximate the value function via iterations (see Dayanik and Sezer (2006) [8]). This investigation is left for future research.

∎

6.1. Comparative statics The explicit forms of the cut-off point p^* and the value function U allow us to derive simple comparative statics of these quantities. As is well known, a DM who plays optimally switches to the safe arm later than a myopic DM, and indeed p^* is smaller than the myopic cut-off point $p^m := \frac{\varrho - \mu_0}{\mu_1 - \mu_0}$.

Note that the cut-off point p^* is an increasing function of α^*. As can be expected, α^* (and therefore also p^*) increases with the discount rate r and with $\nu_0(dh)$, and decreases with $\nu_1(dh)$ and with $\mu_1 - \mu_0$. That is, the DM switches to the safe arm earlier in the game as the discount rate increases, as jumps provide less information, or as the difference between the drifts of the two types increases.[13] Furthermore, as long as $p > p^*$ the value function $p \mapsto U(0, p)$ decreases in α^*. Thus, decreasing the discount rate, increasing the informativeness of the jumps, or increasing the difference between the drifts is beneficial to the DM.

6.2. Generalization In our model the parameters μ_0 and μ_1 have two roles. By the definition of the Lévy process $(X(t))$ they play the role of the unknown drift. In Eq. (4.1) they determine the instantaneous expected payoff. Here we separate these two roles; that is, we assume that the parameters that determine the instantaneous expected payoff are not μ_0 and μ_1, but rather two other parameters, g_0 and g_1 respectively. Formally, in the definition of $W(p, l)$ in Eq. (4.1) we substitute μ_0 and μ_1 with other parameters, g_0 and g_1, and observe the change in the optimal strategy and the optimal payoff. This formulation allows us to separate the information-relevant parameters from the payoff-relevant parameters. It also supplies an optimal strategy and an optimal payoff in a model where the DM receives a general linear function of the process $(X(t))$.

If we replace $W(p, l) = (\mu_1 p + \mu_0(1 - p))l + \varrho(1 - l)$ with $\widehat{W}(p, l) = (g_1 p + g_0(1 - p))l + \varrho(1 - l)$, where g_0 and g_1 are constants that satisfy $g_1 > \varrho > g_0$, then the solution of the optimization problem

$$\widehat{U}(t, p) = \sup_{t \le \tau, \, k \in \Upsilon_M} E^{t,p} \left[\int_t^\tau r e^{-rs} \widehat{W}(p_{s-}, k(s, p_{s-})) ds + \varrho e^{-r\tau} \right]$$

has a similar form to the one given in Theorem 1. Denote $\widehat{p}^* := \frac{\widehat{\alpha}^*(\varrho - g_0)}{(\widehat{\alpha}^* + 1)(g_1 - \varrho) + \widehat{\alpha}^*(\varrho - g_0)}$, where $\widehat{\alpha}^* = \alpha^*$. Under Assumptions A1–A6, there is a unique optimal strategy that is time-homogeneous and is given by

$$\widehat{k}^* = \begin{cases} 0 & \text{if } p \le \widehat{p}^*, \\ 1 & \text{if } p > \widehat{p}^*. \end{cases}$$

The expected payoff under $\widehat{k}^*$ is

$$\widehat{U}(0, p) = \widehat{V}_{\widehat{k}^*}(0, p) = \begin{cases} \varrho & \text{if } p \le \widehat{p}^*, \\ p g_1 + (1 - p) g_0 + \widehat{C}_{\alpha^*}(1 - p)(\frac{1-p}{p})^{\alpha^*} & \text{if } p > \widehat{p}^*, \end{cases}$$

where $\widehat{C}_{\alpha^*} = \frac{\varrho - g_0 - \widehat{p}^*(g_1 - g_0)}{(1 - \widehat{p}^*)\left(\frac{1 - \widehat{p}^*}{\widehat{p}^*}\right)^{\alpha^*}}$.

The significance of this result is that it separates the information-relevant parameters of the model from the payoff-relevant parameters. The quantity α^* summarizes all the information-relevant parameters, whereas g_1 and g_0 are the only payoff-relevant parameters. For beliefs above the cut-off, the optimal payoff is the sum of the expected payoff if the DM always continues, $p g_1 + (1 - p) g_0$, and the option value of experimentation, which is given by $\widehat{C}_{\alpha^*}(1 - p)(\frac{1-p}{p})^{\alpha^*}$.

---

[13] Moreover, $\alpha^*(r = 0) = 0$ and $\alpha^*(r = \infty) = \infty$.

Appendix

The following lemma states that a time-changed martingale under an allocation strategy remains a martingale.

LEMMA 3. *Let $(M(t))$ be a martingale with respect to $\mathcal{F}_t$, and let $(\boldsymbol{K}(t))$ be an allocation strategy that satisfies (K1), (K2), and (K3). Then $(M(K(t)))$ is a martingale with respect to $\mathcal{F}_{K(t)}$.*

Proof. Fix $0 \le s \le t$. Then $K(s)$ and $K(t)$ are bounded stopping times with $K(s) \le K(t)$. The *optional stopping theorem* implies that $M(K(s)) = E[M(K(t)) \,|\, \mathcal{F}_{K(s)}]$, and therefore, $(M(K(t)))$ is indeed an $(\mathcal{F}_{K(t)}, P)$-martingale. ∎

The following lemma presents the (predictable) compensator of a process under a change of measure; see Jacod and Shiryaev (1987, Ch. I, Theorem 3.18) [11].

LEMMA 4. *Under the notations of Section 3.2, let $(H(t))$ be a stochastic process, and let θ be an independent Bernoulli random variable with parameter p, such that given θ, the process $(H(t) - a_\theta t)$ is a martingale with respect to $\mathcal{F}_t^H$ under P_θ. Let $\tilde{p}_t := P_p\{\theta = 1 | \mathcal{F}_t^H\}$ be the posterior belief that $\theta = 1$ given $(H(s))_{s \le t}$ under the probability measure P_p. Then the process $\left(\int_0^t (\tilde{p}_{s-} a_1 + (1 - \tilde{p}_{s-}) a_0) ds \right)$ is the (predictable) compensator of the process $(H(t))$ with respect to $\mathcal{F}_t^H$ under the probability measure P_p.*

Proof. Plainly we have

$$
H(t) - \int_0^t (\tilde{p}_{s-} a_1 + (1 - \tilde{p}_{s-}) a_0) ds
$$
$$
= \theta(H(t) - a_1 t) + (1 - \theta)(H(t) - a_0 t) + \int_0^t (\theta - \tilde{p}_{s-})(a_1 - a_0) ds.
$$

Fix $0 \le u \le t$. The expectation with respect to P_p is

$$
E\left[H(t) - \int_0^t (\tilde{p}_{s-} a_1 + (1 - \tilde{p}_{s-}) a_0) ds - H(u) + \int_0^u (\tilde{p}_{s-} a_1 + (1 - \tilde{p}_{s-}) a_0) ds \,\Big|\, \mathcal{F}_u^H \right]
$$
$$
= E\left[H(t) - H(u) - \int_u^t (\tilde{p}_{s-} a_1 + (1 - \tilde{p}_{s-}) a_0) ds \,\Big|\, \mathcal{F}_u^H \right]
$$
$$
= E\left[\theta(H(t) - a_1 t - H(u) + a_1 u)) + (1 - \theta)(H(t) - a_0 t - H(u) + a_0 u) | \mathcal{F}_u^H \right]
$$
$$
+ E\left[\int_u^t (\theta - \tilde{p}_{s-})(a_1 - a_0) ds \,\Big|\, \mathcal{F}_u^H \right]
$$
$$
= E\left[E\left[\theta(H(t) - a_1 t - H(u) + a_1 u) + (1 - \theta)(H(t) - a_0 t - H(u) + a_0 u) | \theta, \mathcal{F}_u^H \right] \big| \mathcal{F}_u^H \right]
$$
$$
+ E\left[\int_u^t E\left[(\theta - \tilde{p}_{s-}) | \mathcal{F}_{s-}^H \right] (a_1 - a_0) ds \,\Big|\, \mathcal{F}_u^H \right] = 0.
$$

It follows that the process $\left(H(t) - \int_0^t (\tilde{p}_{s-} a_1 + (1 - \tilde{p}_{s-}) a_0) ds \right)$ is a martingale with respect to $\mathcal{F}_t^H$ under the probability measure P_p. Therefore, the predictable process $\left(\int_0^t (\tilde{p}_{s-} a_1 + (1 - \tilde{p}_{s-}) a_0) ds \right)$ is the (predictable) compensator of the process $(H(t))$ with respect to $\mathcal{F}_t^H$ under the probability measure P_p. ∎

Lemmas 3 and 4 yield the following corollary:

COROLLARY 1.
C1. The (predictable) compensator of the process $(X(K(t)))$ is $\left(\int_0^t (p_{s-} \mu_1 + (1 - p_{s-}) \mu_0) dK(s) \right)$.
C2. The (predictable) compensator of the process is

$\left(X(K(t)) - \int_{\mathbb{R}\setminus\{0\}} h\widetilde{N}_{\nu_0}(K(t), dh) - \mu_0 K(t) \right)$ is $\left(-\beta\sigma^2 \int_0^t p_{s-} dK(s) \right)$.

C3. The (predictable) compensator of the random measure $N(dK(t), dh)$ is $(p_{s-}\nu_1(dh) + (1 - p_{s-})\nu_0(dh))dK(s)$; see Jacod and Shiryaev (1987, Ch. II, Theorem 1.8) [11].

The following lemma states the the posterior process (p_t) spends zero time at any given positive contour-line lower than 1.

LEMMA 5. *For every $t \geq 0$, $p \in [0, 1]$, and every $0 < \delta < 1$,*

$$E^{0,p}\left[\int_0^t \chi_{\{p_{s-}=\delta\}} ds\right] = 0.$$

Proof.

$$
\begin{aligned}
E^{0,p}\left[\int_0^t \chi_{\{p_{s-}=\delta\}} ds\right] &= E^{0,p}\left[\int_0^t \chi_{\{p_s=\delta\}} ds\right] \\
&= E\left[\int_0^t \chi_{\{p_s=\delta\}} ds \,\Big|\, \theta = 1\right] p + E\left[\int_0^t \chi_{\{p_s=\delta\}} ds \,\Big|\, \theta = 0\right](1-p) \\
&= E\left[\int_0^t \chi_{\{\ln\left(\frac{1-p_s}{p_s}\right)=\ln\left(\frac{1-\delta}{\delta}\right)\}} ds \,\Big|\, \theta = 1\right] p + E\left[\int_0^t \chi_{\{\ln\left(\frac{1-p_s}{p_s}\right)=\ln\left(\frac{1-\delta}{\delta}\right)\}} ds \,\Big|\, \theta = 0\right](1-p) = 0.
\end{aligned}
$$

The last equation follows since, as long as jumps from B_∞ do not appear (see Remark 2), the process $\left(\ln\left(\frac{1-p_s}{p_s} \right) \right)$ is a time change of a Lévy process whose jump process part has finite variation and has no positive jumps, given the type θ (See Bertoin (1996, Ch. V, Theorem 1) [4]). In case a jump from B_∞ appears, from that time onwards the posterior process (p_t) remains at level 1. ∎

The following lemma assures that α^* is well defined.

LEMMA 6. *Eq. (6.1) admits a unique solution in the interval $(0, \infty)$.*

Proof.
The function f is a continuous function that satisfies $f(0) < 0$ and $f(\infty) = \infty$. To show that $f(\alpha) = 0$ has a unique solution, it is therefore sufficient to prove that f is increasing in α. Note that (if $\sigma \neq 0$ then) $\frac{1}{2}(\alpha+1)\alpha \left(\frac{\mu_1 - \mu_0}{\sigma} \right)^2 - r$ is increasing in α. It remains to prove that if $\bar{\nu}_1 - \bar{\nu}_0 > 0$, i.e., $\nu_1(\mathbb{R}\setminus\{0\}) - \nu_0(\mathbb{R}\setminus\{0\}) > 0$, then $\int_{\mathbb{R}\setminus\{0\}} \left(\left(\frac{\nu_0}{\nu_1}(h) \right)^\alpha - 1 \right) \nu_0(dh) + \alpha(\bar{\nu}_1 - \bar{\nu}_0)$ is increasing in α. Since

$$
\begin{aligned}
&\int_{\mathbb{R}\setminus\{0\}} \left[\left(\left(\frac{\nu_0}{\nu_1}(h) \right)^\alpha - 1 \right) \nu_0(dh) + \alpha(\nu_1(dh) - \nu_0(dh)) \right] \\
&= \int_{\mathbb{R}\setminus\{0\}} \left[\left(\frac{\nu_0}{\nu_1}(h) \right) \left(\left(\frac{\nu_0}{\nu_1}(h) \right)^\alpha - 1 \right) + \alpha \left(1 - \frac{\nu_0}{\nu_1}(h) \right) \right] \nu_1(dh)
\end{aligned}
$$

and

$$
\int_{\{h \mid \frac{\nu_0}{\nu_1}(h)=1\}} \left[\left(\frac{\nu_0}{\nu_1}(h) \right) \left(\left(\frac{\nu_0}{\nu_1}(h) \right)^\alpha - 1 \right) + \alpha \left(1 - \frac{\nu_0}{\nu_1}(h) \right) \right] \nu_1(dh) = 0
$$

it is sufficient to prove that for ν_1-a.e. $h \in \{h \mid \frac{\nu_0}{\nu_1}(h) \neq 1\}$,

$$
g_h(\alpha) = \left(\frac{\nu_0}{\nu_1}(h) \right) \left(\left(\frac{\nu_0}{\nu_1}(h) \right)^\alpha - 1 \right) + \alpha \left(1 - \frac{\nu_0}{\nu_1}(h) \right)
$$

is increasing[14] in α. Now,

$$g_h'(\alpha) = \left(\frac{\nu_0}{\nu_1}(h)\right)^{\alpha+1} \ln\left(\frac{\nu_0}{\nu_1}(h)\right) + \left(1 - \frac{\nu_0}{\nu_1}(h)\right)$$
$$> \left(\frac{\nu_0}{\nu_1}(h)\right) \ln\left(\frac{\nu_0}{\nu_1}(h)\right) + \left(1 - \frac{\nu_0}{\nu_1}(h)\right) > 0,$$

where the first inequality holds since $\alpha > 0$ and the second inequality holds since $x\ln(x) + 1 - x > 0$ for every $x \neq 1$. Therefore, $g_h(\alpha)$ is increasing, as desired. ∎

References

[1] Applebaum, D. 2004. *Lévy Processes and Stochastic Calculus*. Cambridge University Press: Cambridge.

[2] Bergemann, D., J. Välimäki. 2006. Bandit problems. HECER Discussion Paper 93.

[3] Berry, D. A., B. Fristedt. 1985. *Bandit Problems: Sequential Allocation of Experiments*. Chapman and Hall: New York.

[4] Bertoin, J. 1996. *Lévy Processes*. Cambridge University Press: Cambridge.

[5] Besanko, D., J. Wu. 2008. The impact of market structure on R&D competition with learning. Preprint.

[6] Bolton, P., C. Harris. 1999. Strategic experimentation. *Econometrica* **67** 349–374.

[7] Chernoff, H. 1972. *Sequential Analysis and Optimal Design*. Society for Industrial and Applied Mathematics: Philadelphia.

[8] Dayanik, S., O. S. Sezer. 2006. Compound Poisson disorder problem. *Math. Oper. Res.* **31** 649–672.

[9] Gapeev, P. V. 2005. The disorder problem for compound Poisson processes with exponential jumps. *Annals of Applied Probability* **15** 487–499.

[10] Gittins, J. C., D. M. Jones. 1979. A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika* **66** 561–565.

[11] Jacod, J., A. S. Shiryaev. 1987. *Limit Theorems for Stochastic Processes*. Springer: Berlin.

[12] Keller, G., S. Rady. 2010. Strategic experimentation with Poisson bandits. *Theoretical Economics* **5** 275–311.

[13] Keller, G., S. Rady, M. Cripps. 2005. Strategic experimentation with exponential bandits. *Econometrica* **73** 39–68.

[14] Klein, N., S. Rady. 2011. Negatively correlated bandits. *Review of Economic Studies* **78** 693–732.

[15] Kobayashi, K. 2011. Stochastic calculus for a time-changed semimartingale and the associated stochastic differential equations. *Journal of Theoretical Probability* **24** 789–820.

[16] Lai, T. L., H. Robbins. 1984. Asymptotically optimal allocation of treatments in sequential experiments. T. J. Santner, A. C. Tamhane, eds., *Design of Experiments: Ranking and Selection*. Society for Industrial and Applied Mathematics: Philadelphia.

[17] Mandelbaum, A., L. A. Shepp, R. J. Vanderbei. 1990. Optimal switching between a pair of Brownian motions. *Annals of Probability* **18** 1010–1033.

[18] Moscarini, G., F. Squintani. 2010. Competitive experimentation with private information. *Journal of Economic Theory* **145** 639–660.

---

[14] In our model $\frac{\nu_0}{\nu_1}(h) \neq 1$ is actually $\frac{\nu_0}{\nu_1}(h) < 1$ ν_1-a.s. Yet, the proof works in the more general case of inequality.

[19] Øksendal, B. 2000. *Stochastic Differential Equations: An Introduction with Applications*. 5th ed. Springer: Berlin.

[20] Peskir, G., A. N. Shiryaev. 2000. Sequential testing problems for Poisson processes. *Annals of Statistics* **28** 837–859.

[21] Peskir, G., A. N. Shiryaev. 2002. Solving the Poisson disorder problem. K. Sandmann, P. Schönbucher, eds., *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann*. Springer: Berlin, 295–312.

[22] Revuz, D., M. Yor. 1999. *Continuous Martingales and Brownian Motion*. 3rd ed. Springer: Berlin.

[23] Robbins, H. 1952. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society* **58** 527–535.

[24] Roberts, K., M. L. Weitzman. 1981. Funding criteria for research, development, and exploration projects. *Econometrica* **49** 1261–1288.

[25] Rothschild, M. 1974. A two-armed bandit theory of market pricing. *Journal of Economic Theory* **9** 185–202.

[26] Sato, K. 1999. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press: Cambridge.

[27] Shiryaev, A. N. 1978. *Optimal Stopping Rules*. Springer: Berlin.

[28] Weitzman, M. L. 1979. Optimal search for the best alternative. *Econometrica* **47** 641–654.