The Value of Continued Experimentation*

Jan Eeckhout† and Xi Weng‡

April 2012

Abstract

In many economic environments, agents often continue to face the same underlying state variable, even if they switch action. For example, a worker's ability revealed in one job is informative about her productivity in another job. We model these environments by a general class of *common value* two-armed bandit problems with *Lévy* process and show that the value of experimentation must be equal whenever the agent switches action. In addition to the well-known conditions of value matching (level) and smooth pasting (first derivative), this implies a new equilibrium condition. It captures the Bellman's Principle of Optimality and restricts the second derivatives of the value function.

Keywords. Bandit Problem. Bellman's Principle of Optimality. Common and Correlated Values. Lévy process.

JEL. D83. C02. C61.

*We thank numerous colleagues for insightful discussions and comments. We also benefitted from the feedback of several seminar audiences. Eeckhout acknowledges support from the ERC, Grant 208068.

†University College London and UPF, j.eeckhout@ucl.ac.edu.

‡Department of Applied Economics, Guanghua School of Management, Peking University, wengxi125@gsm.pku.edu.cn.

1 Introduction

Consider a firm that hires a young, promising recruit and wishes to find out about her productive ability. Even if she is assigned to a junior position, the firm will nonetheless also learn a great deal about that worker's ability to perform in an executive position. Most likely the productivity and the learning rates will be different in both jobs, yet whatever the firm learns when the worker is in one position affects the value and beliefs about her productivity in other positions. Continued experimentation is particularly important for promotion decisions. Likewise, continued experimentation is prevalent in consumer choice, for example, patients who learn about the effectiveness of different drugs.

In this paper we analyze optimal allocation problems in environments of continued experimentation. Compare our setting to the canonical experimentation problem (Gittins and Jones (1974)) in discrete time and with independent arms.[1] Gittins' seminal insight is to calculate the value of pulling an arm (denoted by the so-called Gittins index) and compare the value to the Gittins index of all other arms. We can proceed in this manner because each of the value functions is independent of the stopping rule. In other words, the value of pulling each arm itself is not a function of the cutoff.

Instead, when there is continued experimentation, the underlying states are no longer independent and pulling any given arm affects the value of the other arms. As we learn about the ability of a given worker in one job, we also update our information about her ability in all other jobs. The immediate implication is that the decision to pull any given arm affects the value of pulling all other arms. As a result, we have to solve for the value of each of the arms *and* the cutoffs *simultaneously* and we can no longer apply Gittins' logic. To our knowledge, there is no known solution to deal with this problem in discrete time. We can, however, analyze this problem in continuous time. We think here of a general experimentation setup where the stochastic component follows a Lévy process. This includes among many others pure Bayesian learning, learning where the state variable has a drift, and general experimentation.

Our main result is to establish a simple equilibrium condition that we call *second order condition* and that must be satisfied for any problem with continued experimentation. This condition imposes equalization of the second derivative of each of the value functions at the cutoff. It is a condition of sequential rationality because it stems from the requirement that an agent has no incentive to

[1] See Bergemann and Välimäki (2008) for a survey on bandit problems.

deviate for a short period of time by choosing the other arm. Denote the value of arm one at the cutoff $x^\star$ by $V(x^\star)$ and at arm two by $U(x^\star)$. Our condition adds to the well-known conditions at the cutoff of value matching ($V(x^\star) = U(x^\star)$) and smooth pasting ($V'(x^\star) = U'(x^\star)$). In the presence of continued experimentation, equilibrium must now also satisfy $V''(x^\star) = U''(x^\star)$.

The second derivative can be interpreted as a measure of the value of experimentation. It measures the change in value of having the option to switch arms. Clearly, the payoff from the option of switching arms depends also on the noise term of the Brownian motion – in particular, how likely it is that the switch will persist – but conditional on switching the second derivative is a measure of how much better off the agent is from taking the better of the arms. The convexity of the value function indicates the change in the marginal return to a better posterior.

The equality of the value of experimentation on both arms at the cutoff ensures that there is no incentive to deviate from the equilibrium path. We derive the condition by considering a deviating agent who chooses the alternative arm for a short period of time and who then switches back to the equilibrium strategy, akin to the one-shot deviation in discrete time. This requirement of sequential optimality imposes a simple restriction on the value of experimentation across different options, which must be equal. Intuitively, an experimenting agent at the cutoff would be better off briefly deviating to capture the gains from increased experimentation value. The condition is so stark because at the cutoff there is no gain from switching permanently, from value matching and smooth pasting, but the experimentation trajectories will differ with a periodical deviation, and in the limit, the only difference in payoff is due to the experimentation value. Equating the value of experimentation ensures that no such gains from deviation exist.

Our result is derived using a standard principle of optimization. Our approach is related to the Bellman's Principle of Optimality, which is derived in a discrete time setting. In the continuous time setting here, we employ the dynamic programming approach to handle the optimal control problem and to derive the Hamilton-Jacobi-Bellman equation. What we show is that Bellman's Principle of Optimality in our setting implies more than the HJB equation. It also has implications for the boundary conditions!

With reference to the benchmark of the one-armed bandit problem, it is important to note that the second derivative condition does not hold in that case. Because there is no experimentation value to taking the safe option, the condition is one-sided and is given by the inequality $V''(x^\star) \geq 0$, which is trivially satisfied from the convexity of the value function.

The main appeal of this simple condition is its applicability. Not only are such environments

with continued experimentation prevalent in many economic contexts, the implementation of the second derivative condition is straightforward. Moreover, it is not only applicable to decision problems where payoffs are exogenous, it can easily be embedded in a strategic setting or a market setting where payoffs are determined endogenously. We consider three applications. First, we fully characterize a decision problem with linear payoffs. Second, we illustrate the applicability of continued experimentation in a situation with strategic interactions. Third, we analyze a model of strategic pricing in the vein of Bergemann and Välimäki (1996) and Bergemann and Välimäki (2000) with continued experimentation and show how the second derivative condition affects the equilibrium allocation and its efficiency.

Finally, we derive the condition in full generality for imperfectly correlated arms, which we model by means of multi-dimensional signals. The generalized second derivative condition now involves equating the weighted sum of all partial derivatives in each dimension between different arms. The interpretation is as before in that summing over all dimensions, we equate the total value of experimentation between different arms. It is well established that solving for the value functions of multi-dimensional problems involves systems of partial differential equations for which no general solution exists. Nonetheless, we can analyze special cases. In particular, we use the solution provided by Karatzas (1984) (and further extended in Felli and Harris (1996)) for the case of independent arms. The Karatzas solution is the continuous time version of the standard Gittins index. We establish that the Karatzas solution to the value function satisfies the second derivative condition.

2 The Basic Model

Setup. Consider one agent and a bandit with two arms $j = 1, 2$. Time is continuous and is denoted by t. There is one state variable $x \in \mathcal{X}$, with $\mathcal{X} \subset \mathbb{R}$ is a connected set and two choices $j \in \{1, 2\}$. The state variable x determines the instantaneous flow payoffs of each arm $f_i(x)$. Future payoffs are discounted at rate $r > 0$.

State Updating. The state variable x evolves according to a *Lévy* process in arm j (see Applebaum (2004)):[2]

[2]The Lévy bandit has already analyzed by Kaspi and Mandelbaum (1995). Different from our paper, the authors only consider independent arms.

$$dx(t) = \mu_j(x(t-))dt + \sigma_j(x(t-))dZ_j(t) + \int_{\mathbb{R}-\{0\}} G_j(x(t-), y)N_j(dt, dy),$$

where $Z_j(t)$ is a standard Brownian motion process and N_j is a Poisson random measure that is independent of the Brownian motion Z_j. For simplicity, we follow Cohen and Solan (2009) and assume that N_j has finite intensity measure ν_j.[3] Furthermore, we assume that N_1, N_2, Z_1 and Z_2 are mutually independent of each other. This jump-diffusion process is a special case of the *Lévy* process but it indeed covers a lot of interesting applications in the literature. For example, the standard diffusion process $dx = \mu_j(x)dt + \sigma_j(x)dZ_j(t)$ is a special case without any jump in the process. On the other hand, if we assume that $\sigma_j(x) = 0$, then the path of x is determined by the drift $\mu(x)$, interspersed with jumps taking place at random times. We allow for a general process for x and in general the martingale assumption is not made. This includes belief updating as a special case (the agent sees output and updates beliefs)[4] but also human capital accumulation where output changes over time.

Two technical assumptions are required on the functions of $f_j(x), \mu_j(x), \sigma_j(x)$ and $G_j(x, \cdot)$. In particular, the assumption 2 guarantees Lipschitz continuity, which is common in the literature. In general, we don't need to impose such strong restrictions. But these conditions are usually satisfied in the applied literature.

Assumption 1 *$f_j(x), \mu_j(x), \sigma_j(x)$ and $G_j(x, \cdot)$ are $\mathcal{C}^2$ of x, for any $x \in \mathcal{X}$ where $\mathcal{X}$ is a connected set.*

Assumption 2 *The first derivatives of $f_j(x), \mu_j(x), \sigma_j(x)$ and $G_j(x, \cdot)$ with respect to x are bounded: there exists $K > 0$ such that for any $x \in \mathcal{X}$, $|f_j'(x)|, |\mu_j'(x)|, |\sigma_j'(x)|$ and $|\frac{\partial G_j(x,\cdot)}{\partial x}|$ are all less than K.*

We consider the Markovian strategy where the agent's decision depends only on the state x. The value function can be written as:

$$v(x) = \sup_{a:\mathcal{X}\to\{1,2\}} \left\{ \mathbb{E} \int_{t=0}^{\infty} e^{-rt} f_{a_t}(x_t)dt \right\}$$

[3]For example, N_j can be the sum of m_j independent Poisson processes, which are also independent of Z_j. Each Poisson process has intensity λ_j and takes value in h_i for $i = 1, \cdots, m_j$. In this case,

$$\nu_j = \sum_{i=1}^{m_j} \lambda_i \delta_{h_i},$$

where δ_h is a Dirac mass concentrated at h.

[4]See, for example, Bolton and Harris (1999), Keller, Rady, and Cripps (2005), Keller and Rady (2010), etc.

$$\text{s.t. } dx_t = \mu_{a_t}(x_t)dt + \sigma_{a_t}(x_t)dZ_{a_t}(t) + \int_{\mathbb{R}-\{0\}} G_{a_t}(x(t-), y)N_{a_t}(dt, dy) \quad \text{and} \quad a_t \triangleq a(x_t).$$

If the value function is sufficiently smooth, (i.e., $v(x)$ is at least $\mathcal{C}^2$), then by Bellman's Principle and Optimality and Ito's lemma, the value function can be further written as:[5]

$$v(x) = \max_{j\in\{1,2\}} \left[f_j(x) + \mu_j(x)v'(x) + \frac{1}{2}\sigma_j^2(x)v''(x) + \int_{\mathbb{R}-\{0\}} [v(x + G_j(x, y)) - v(x)]\nu_j(dy) \right].$$

Denote $V(x)$ $(U(x))$ to be the value function of an agent with state in a neighborhood of x optimally choosing arm 1 (2). Obviously, $v(x) = V(x)$ if arm 1 is optimally chosen and $v(x) = U(x)$ otherwise.

3 Motivating Examples

Consider, for example, an assignment problem in the presence of learning (see Eeckhout and Weng (2010)). There are two types of workers $x \in \{H, L\}$ and two types of firms $y \in \{H, L\}$. The type y is observable to all agents in the economy but the worker ability x is not observable, both to firms and workers. Cumulative output of a worker-firm pair is assumed to be a Brownian motion with drift μ_{xy} and variance σ_y^2. The worker and the firm y face the same information extraction problem based on the noisy information of cumulative output. The common belief p about the worker's type being H is updated according to Bayesian rule:

$$dp_t = p_t(1 - p_t)\frac{\mu_{Hy} - \mu_{Ly}}{\sigma_y} d\bar{Z}_{y,t},$$

where $Z_{y,t}$ is a standard Brownian motion process.[6] Therefore, at each type of firm, a worker learns at a different rate about his common ability measured by the signal-noise ratio $(\mu_{Hy} - \mu_{Ly})/\sigma_y$. As a result of different learning experiences, the worker expects different wage paths across firms.

In such a market, each worker faces a common-value bandit problem given the market-determined competitive payoffs. Moreover, there is a continuum of agents who experiment simultaneously. Workers are able to switch jobs costlessly as beliefs about his ability change. Eeckhout and Weng

[5] The derivation of the value function can be seen in Cohen and Solan (2009) and Kaspi and Mandelbaum (1995).

[6] Instead of assuming that the cumulative output follows the standard diffusion process, we can allow jumps in the cumulative output. Then the belief updating follows a Lévy process as shown by Cohen and Solan (2009).

(2010) investigate stationary competitive equilibria where each worker is making an optimal job-switching decision and where the market clears.

In the Bayesian learning envirenment, belief as the state variable evolves according to a martingale process. However, there are other applications where the state variable does not follow a martingale process. Examples include the CO_2 emission problem discussed in Wirl (2008) and the investment problems without switching costs (see Dixit and Pindyck (1994)).

4 Results

Throughout this paper, we will consider cutoff strategies: there is a set of (countably many) cutoffs such that the agent switches action once state x reaches a cutoff. Consider $x^\star$ to be any cutoff in the interior of $\mathcal{X}$. Other than the well-known properties of value matching and smooth pasting, we derive a new second derivative condition based on the Bellman's Principle of Optimality.

4.1 Value Matching

Value matching is a standard condition in the optimal stopping literature: $V(x^\star) = U(x^\star)$. It implies that the value function $v(x)$ is continuous at $x^\star$. Notice the expression uses a short-hand notation since $V(\cdot)$ and $U(\cdot)$ are not well defined at $x^\star$. Suppose it is optimal to choose arm 1 (2) in a right (left) neighborhood of $x^\star$. Then the above condition means

$$V(x^\star+) \triangleq \lim_{x \searrow x^\star} V(x) = U(x^\star-) \triangleq \lim_{x \nearrow x^\star} U(x).$$

In the remainder of the paper, the "+" and "–" signs will be omitted if no confusion results.

The value matching condition is driven by the continuity of value functions. If this condition is violated and $V(x^\star) > U(x^\star)$, then the decision by the agent cannot be optimal. The agent can choose a new cutoff $x^\star - \epsilon$. Under this new decision rule, the value at state $x^\star - \epsilon/2$ is $\tilde{V}(x^\star - \epsilon/2)$ instead of $U(x^\star - \epsilon/2)$. By continuity,

$$\lim_{\epsilon \to 0} \tilde{V}(x^\star - \epsilon/2) = V(x^\star) \quad \text{and} \quad \lim_{\epsilon \to 0} U(x^\star - \epsilon/2) = U(x^\star).$$

By choosing ϵ sufficiently small, we can make $\tilde{V}(x^\star - \epsilon/2)$ arbitrarily close to $V(x^\star)$ and $U(x^\star - \epsilon/2)$ arbitrarily close to $U(x^\star)$ such that $\tilde{V}(x^\star - \epsilon/2) > U(x^\star - \epsilon/2)$. This implies that the original policy is not optimal at state $x = x^\star - \epsilon/2$.

4.2 Smooth Pasting

The smooth pasting condition is another standard condition in models of continuous time optimal stopping problems. In the context of our model, the condition is: $V'(x^\star) = U'(x^\star)$.

Smooth pasting was first proposed by Samuelson (1965) as a first-order condition for optimal solution. The proof of smooth pasting condition can be found in Peskir and Shiryaev (2006) and is omitted here. The logic of the proof builds on the notion of a deviation in the state space $\mathcal{X}$. A candidate equilibrium prescribes the optimal switching of action at the cutoff $x^\star$. Optimality of a cutoff implies that the value is lower if the cutoff is slightly moved away in the neighborhood of the optimal cutoff. This implies we can rank the value functions V and U. In the limit as we get arbitrarily close to the cutoff, this implies a restriction on the first derivative.

In particular, suppose instead of at the optimal cutoff $x^\star$, the agent switches action in the neighborhood, say, at $x^\star + \epsilon$. The induced payoff from switching at $x^\star + \epsilon$ is lower than the equilibrium payoff. As ϵ becomes small, this inequality transforms into an inequality of the first derivatives. Likewise, we consider a deviation from the candidate equilibrium where the agent switches at $x^\star - \epsilon$ instead of at $x^\star$. This again induces an inequality, which in the limit is the opposite of the first inequality, therefore implying equality of the first derivatives at the optimal cutoff $x^\star$.

4.3 Bellman's Principle of Optimality

We now establish that Bellman's Principle of Optimality imposes an additional constraint on the equilibrium allocation. We derive the second derivative condition by considering a deviating agent who chooses the alternative arm for a short period of time and then switches back to the equilibrium strategy. Conceptually, the key difference between the smooth pasting and second derivative conditions is that we consider different kinds of deviations. For the smooth pasting condition, the deviation is in the state space $\mathcal{X}$. For the second derivative condition, the deviation is in the time space. The deviating agent chooses the other arm for a duration dt, and then switches back. We consider the value of such deviations as dt becomes small. The deviation in time space is similar to the one-shot deviation in discrete time. Instead, the deviation in state space is similar to a permanent deviation, since it changes decision rules permanently.

Theorem 1 *If $\sigma_1(x) > 0$ and $\sigma_2(x) > 0$ for all $x \in \mathcal{X}$, a necessary condition for the optimal*

solution $x^\star$ is:

$$V''(x^\star) = U''(x^\star) \quad \textit{(Second Derivative Condition)} \tag{1}$$

for any possible cutoff $x^\star$.

Proof. Without loss of generality, we assume that an agent with $\bar{x} > x > x^\star$ chooses arm 1 and an agent with $\underline{x} < x < x^\star$ chooses arm 2. Consider one possible one-shot deviation: at some $\bar{x} > x > x^\star$, the agent chooses arm 2 for t length of time and then switches back. The value associated with this deviation can be written as:

$$\tilde{U}(x;t) = \mathbb{E}\left\{\int_0^t e^{-rs} f_2(x_s)ds + \int_0^t \int_{\mathbb{R}-\{0\}} e^{-rs} v(x_s + G_2(x_s, y))\nu_2(dy)ds + e^{-rt}(1-\bar{\nu}_2 t)v(x_t) + o(t)\right\}. \tag{2}$$

In the above expression, $\bar{\nu}_i = \nu_i(\mathbb{R} - \{0\})$ and $(1 - \bar{\nu}_2 t)$ is an approximation of the probability that no jumps happen during the deviation period since strictly more than one jumps happen with probability $o(t)$. $v(x_t)$ is the value function at t with state x_t when there is no jump during the deviation period. Then we know that x_t is most likely to be between $x^\star$ and $\bar{x}$ and $v(x_t) = V(x_t)$.[7]

Since it is optimal to choose arm 1 in a neighborhood of x, there exists $\bar{t}$ such that for all $t \leq \bar{t}$, $\tilde{U}(x;t) \leq V(x)$ and hence

$$\lim_{t\to 0} \frac{\tilde{U}(x;t) - V(x)}{t} \leq 0.$$

This implies that:

$$f_2(x) + \mu_2(x)V'(x) + \frac{1}{2}\sigma_2^2(x)V''(x) + \int_{\mathbb{R}-\{0\}} [v(x + G_2(x,y)) - V(x)]\nu_2(dy) \leq rV(x).$$

For $x \to x^\star$, the above inequality implies:

$$f_2(x^\star) + \mu_2(x^\star)V'(x^\star+) + \frac{1}{2}\sigma_2^2(x^\star)V''(x^\star+) + \int_{\mathbb{R}-\{0\}} [v(x^\star + G_2(x^\star, y)) - V(x^\star+)]\nu_2(dy) \leq rV(x^\star+)$$

where $V'(x^\star+) \triangleq \lim_{x \searrow x^\star} V'(x)$ and $V''(x^\star+) \triangleq \lim_{x \searrow x^\star} V''(x)$.

[7] It is possible that x_t goes below $x^\star$ or above $\bar{x}$. But the probability of this event is $o(t)$ for the Brownian motion.

At $x^\star$, we have $V(x^\star+) = U(x^\star-)$, from the value matching condition where $U(x^\star-) \triangleq \lim_{x \nearrow x^\star} U(x)$. This implies:

$$f_2(x^\star) + \mu_2(x^\star)V'(x^\star+) + \frac{1}{2}\sigma_2^2(x^\star)V''(x^\star+) + \int_{\mathbb{R}-\{0\}} [v(x^\star + G_2(x^\star, y)) - V(x^\star+)]\nu_2(dy)$$
$$\leq f_2(x^\star) + \mu_2(x^\star)U'(x^\star-) + \frac{1}{2}\sigma_2^2(x^\star)U''(x^\star-) + \int_{\mathbb{R}-\{0\}} [v(x^\star + G_2(x^\star, y)) - U(x^\star-)]\nu_2(dy).$$

From the smooth pasting condition, $V'(x^\star+) = U'(x^\star-)$ and hence we should have: $V''(x^\star+) \leq U''(x^\star-)$. Similarly, we can consider a one-shot deviation on the other side of $x^\star$ (i.e., at $\underline{x} < x < x^\star$). By the same logic we get: $V''(x^\star+) \geq U''(x^\star-)$. Therefore, it must be the case that $V''(x^\star) = U''(x^\star)$. ■

The second derivative measures the change in value of having the option to switch arms and can be interpreted as a measure of the value of experimentation. An interpretation of the second derivative condition is that it requires the value of experimentation to be the same at the optimal cutoff. At the optimal cutoff there is no gain from switching permanently, from the value matching condition. But the experimentation trajectories will differ with a periodical deviation, and in the limit the only difference in payoff is due to the experimentation value. Equating the value of experimentation ensures that at the optimal cutoff, no gains from switching exist.

The key assumption of theorem 1 is that both arms contain a non-trivial diffusion process. If this condition is violated, it is quite easy to see that this nice condition on the second derivative does not hold any longer. In particulare, if $\sigma_1 = \sigma_2 = 0$, then the Belman's Principle of Optimality leads to the same condition as the smooth pasting condition: $V'(x^\star) = U'(x^\star)$.[8]

Wirl (2008) derived the same result as Theorem 1 in a common-value two-armed bandit problem with *only* diffusion processes. The proof is also based on Bellman's Principle of Optimality. However, Wirl (2008) makes the key assumption that $\sigma_1(x) = \sigma_2(x)$ for all x. This assumption enables him to cancel the second derivative term of the vale function and to derive an explicit formula of the first derivative of the value function at the cutoff. Using this the first derivative formula, Wirl (2008) writes down expressions of the second derivative of the value function and shows algebraically that $V''(x^\star) = U''(x^\star)$. Our result indicates that the second derivative condition holds for the more general Lévy process as long as the diffusion components are non-trivial.

[8] In the context of one-armed bandit problems, let V be the value of pulling the risky arm and U be the value of pulling the safe arm. Then the logic of proof implies that V is locally more convex than U: $V''(x^\star) \geq U''(x^\star)$, which is satisfied at $x^\star$.

Furthermore, our proof does not hinge on the restrictive assumption $\sigma_1(x) = \sigma_2(x)$, which enables us to investigate more interesting problems.[9]

5 Applications

In this section we illustrate the second derivative condition in three applications. We consider a decision problem with linear payoffs for which we can explicitly derive the value functions and calculate the equilibrium cutoff. We then extends the model to a two-player strategic interaction problem.[10] Finally, we analyze a strategic game in which two firms set prices in order to induce a buyer to experiment by buying from either of the firms. Now the buyer's payoffs are no longer exogenously given but determined in equilibrium. In each of these settings, the second derivative condition plays a crucial role.

5.1 Linear Payoffs

In this section, we consider a standard bandit problem with linear payoffs. The common state is $x \in (-\infty, \infty)$. The payoffs are linear: $f_1(x) = a_1 x + b_1$ and $f_2(x) = a_2 x + b_2$. x is updated by $dx = \mu_y dt + \sigma_y dZ_y$ in arm y. We assume $a_1 \neq a_2$ to avoid the trivial situation that one arm is always better than the other one.

The differential equation

$$rV(x) = f_1(x) + \mu_1(x)V'(x) + \frac{1}{2}\sigma_1^2(x)V''(x)$$

can be solved explicitly:

$$V(x) = \frac{r(a_1 x + b_1) + a_1\mu_1}{r^2} + k_{11}e^{\beta_1 x} + k_{12}e^{-\gamma_1 x}$$

where $\beta_1 = \frac{\sqrt{\mu_1^2 + 2r\sigma_1^2} - \mu_1}{\sigma_1^2}$ and $\gamma_1 = \frac{\sqrt{\mu_1^2 + 2r\sigma_1^2} + \mu_1}{\sigma_1^2}$. In the above expression, $k_{11}e^{\beta_1 x} \geq 0$ measures the option value that the agent switches arms as x goes up; and $k_{12}e^{-\gamma_1 x} \geq 0$ measures the option value that the agent switches arms as x goes down. If $+\infty$ is included in the domain of V, $k_{11} = 0$ since then the agent would never switch as x goes up; and if $-\infty$ is included in the domain, $k_{12} = 0$

[9] For example, in the basic model of Eeckhout and Weng (2010), it is generic that the signal-to-noisy ratios are different in different types of firms.

[10] Wirl (2008) claims that the second derivative condition is not applicable in this situation without using mixing strategies.

since then the agent would never switch as x goes down.

Similarly, we get

$$U(x) = \frac{r(a_2 x + b_2) + a_2\mu_2}{r^2} + k_{21}e^{\beta_2 x} + k_{22}e^{-\gamma_2 x}$$

with $\beta_2 = \frac{\sqrt{\mu_2^2+2r\sigma_2^2}-\mu_2}{\sigma_2^2}$ and $\gamma_2 = \frac{\sqrt{\mu_2^2+2r\sigma_2^2}+\mu_2}{\sigma_2^2}$. Also, we require that $k_{21} \geq 0$ and $k_{22} \geq 0$. This allows us to establish the uniqueness result:

Theorem 2 *Suppose the parameter values satisfy:* $(a_i - a_j)(\mu_i\sigma_j^2 - \mu_j\sigma_i^2) \geq 0$. *Then there must be a unique* $x^\star$ *satisfying the three equilibrium conditions:* $V(x^\star) = U(x^\star)$ *(value matching),* $V'(x^\star) = U'(x^\star)$ *(smooth pasting), and* $V''(x^\star) = U''(x^\star)$ *(second derivative).*

Proof. In Appendix. ■

An immediate implication is that for pure Bayesian learning, i.e., in the absence of a drift term in the Brownian motion ($a_i = a_j = 0$), there is a unique cutoff. The possible source of multiplicity stems from the role of the drift terms of the Brownian motion. Consider an extreme case where $a_2 < a_1$ is very close to a_1 but μ_2 is sufficiently larger than μ_1. Then in some intermediate region of the state x, the agent may want to choose arm 2 to accelerate the change of x. To guarantee uniqueness, we impose a condition such that if arm i has a higher slope a_i, then the drift μ_i in arm i is also sufficiently higher than the drift μ_j in arm $j \neq i$.

Under the assumption that $a_1 > a_2$ and $(a_i - a_j)(\mu_i\sigma_j^2 - \mu_j\sigma_i^2) \geq 0$, we are able to completely characterize the optimal cutoff $x^\star$ and value functions as:

$$\begin{aligned}
x^\star &= \tfrac{r(b_2-b_1)+a_2\mu_2-a_1\mu_1}{r(a_1-a_2)} + \tfrac{\gamma_1-\beta_2}{\gamma_1\beta_2} \\
V(x) &= \tfrac{r(a_1x+b_1)+a_1\mu_1}{r^2} + k_1e^{-\gamma_1 x}, \quad k_1 = e^{\gamma_1 x^\star}\tfrac{\beta_2(a_1-a_2)}{r\gamma_1(\gamma_1+\beta_2)} \\
U(x) &= \tfrac{r(a_2x+b_2)+a_2\mu_2}{r^2} + k_2e^{\beta_2 x}, \quad k_2 = e^{-\beta_2 x^\star}\tfrac{\gamma_1(a_1-a_2)}{r\beta_2(\gamma_1+\beta_2)}.
\end{aligned}$$

The one-armed bandit problem can be viewed as a special case where $a_2 = \mu_2 = \sigma_2 = 0$. The optimal cutoff in that problem can be written as:

$$x^{o\star} = \frac{r(b_2 - b_1) - a_1\mu_1}{ra_1} - \frac{1}{\gamma_1}.$$

However, at the optimal cutoff $x^{o\star}$, $U = b_2$ is a constant but V is a strictly convex function. Therefore, $V''(x^{o\star}) > U''(x^{o\star}) = 0$! This is because in the one-armed bandit problem, there is no experimentation value to taking the safe option. As a result, the second derivative is one-sided and

is given by the above inequality. Although the second derivative condition does not hold at $x^{o\star}$, the optimal cutoff exhibits an interesting continuity property: as a_2, μ_2, σ_2 all go to 0, the limit of the optimal cutoff in the two-armed bandit problem, $x^\star(a_2, \mu_2, \sigma_2)$ indeed converges to $x^{o\star}$, the optimal cutoff in the one-armed bandit problem[11].

Finally, we can show explicitly that switching at $x^\star$ indeed is the optimal solution to the two-armed bandit problem. The next result focuses on the case that $a_1 > a_2$. The case that $a_1 < a_2$ can be proved similarly.

Theorem 3 *Suppose the parameter values satisfy: $a_1 > a_2$ and $(a_i - a_j)(\mu_i\sigma_j^2 - \mu_j\sigma_i^2) \geq 0$. Then the optimal solution to the two-armed bandit problem is to choose arm 1 for $x > x^\star$ and arm 2 for $x < x^\star$.*

Proof. In Appendix. ■

5.2 Strategic Interaction

In this section, we consider an interesting extension of the linear payoff model. There are symmetric two players 1 and 2. At each instant of time, each player i decides to split one unit of resource between arm 1 and arm 2. Let θ_i denote the fraction of the resource put in arm 1 by player i. The updating rule of the state x is written as:

$$dx = (\theta_1 + \theta_2)\mu_1 dt + (2 - \theta_1 - \theta_2)\mu_2 dt + \sqrt{(\theta_1 + \theta_2)}\sigma_1 dZ_1(t) + \sqrt{(2 - \theta_1 - \theta_2)}\sigma_2 dZ_2(t),$$

where Z_1 and Z_2 are independent standard Brownian motion processes. This formula generalizes the updating rule of our baseline model. The divisible choice setting is a common way to introduce mixed strategy in such environment. Each player i has instantaneous payoff $\theta_i f_1(x) + (1 - \theta_i) f_2(x)$ and maximizes the discounted future payoff given the strategy of her opponent:

[11] As a_2, μ_2, σ_2 all go to 0, it is easy to check that $\frac{r(b_2-b_1)+a_2\mu_2-a_1\mu_1}{r(a_1-a_2)}$ converges to $\frac{r(b_2-b_1)-a_1\mu_1}{ra_1}$. Also, the limit of $\frac{1}{\beta_2}$ is

$$\lim_{(\mu_2,\sigma_2^2)\to 0} \frac{\sqrt{\mu_2^2 + 2r\sigma_2^2} + \mu_2}{2r},$$

which converges to 0.

$$rV_i(x) = \max_{\theta_i \in [0,1]} \{\theta_i f_1(x) + (1-\theta_i) f_2(x) + \left[(\theta_i + \theta_{-i})\mu_1 + (2 - \theta_i - \theta_{-i})\mu_2\right] V_i'(x) + \frac{1}{2}\left[(\theta_i + \theta_{-i})\sigma_1^2 + (2 - \theta_i - \theta_{-i})\sigma_2^2\right] V_i''(x)\}.$$

We say $\{\theta_1(x), \theta_2(x)\}$ constitute a (Markov) equilibrium if $\theta_i(x)$ maximizes $V_i(x)$ given $\theta_j(x)$. Three possibilities can happen in a symmetric equilibrium: both players pull arm 1 only, both players pull arm 2 only and each player randomizes between arm 1 and arm 2. When both players pull arm 1, the value function of player 1 satisfies:

$$rV(x) = f_1(x) + 2\mu_1 V'(x) + \sigma_1^2 V''(x);$$

when both players pull arm 2, the value function of player 1 satisfies:

$$rU(x) = f_2(x) + 2\mu_2 U'(x) + \sigma_2^2 U''(x);$$

when both players randomize, the value function of player 1 satisfies:

$$f_1(x) - f_2(x) + (\mu_1 - \mu_2) W'(x) + (\sigma_1^2 - \sigma_2^2) W''(x) = 0.$$

To fully characterize the symmetric equilibrium, we normalize the instantaneous payoff of arm 2 to be zero and $f_1(x) = ax + b$ with $a > 0$ and $b > 0$. Moreover, we assume that $\mu_1 = \mu_2 = 0$ and $\sigma_1 \neq \sigma_2$. All of the three equations can be solved explicitly and the value matching, smooth pasting, second derivative conditions apply at the cutoffs. The equilibrium cutoffs are characterized by the following theorem:

Theorem 4 *Assume $f_1(x) = ax + b$ (with $a > 0$ and $b > 0$), $f_2(x) = 0$, $\mu_1 = \mu_2 = 0$ and $\sigma_1 \neq \sigma_2$. Then there exists a symmetric equilibrium such that both players pull arm 1 for $x \geq x_1^\star$ and arm 2 for $x \leq x_2^\star$. For $x \in (x_1^\star, x_2^\star)$, both players randomize between arm 1 and arm 2. $x_1^\star = x_2^\star + \Delta$, where $\Delta > 0$ is the unique strictly positive root to equation:*

$$\frac{ay^4}{3} + \frac{4a(\sigma_1 + \sigma_2)y^3}{3\sqrt{r}} + \frac{[2a(\sigma_1^2 + \sigma_2^2) + 2b\sqrt{r}\sigma_1]y^2}{r} + \frac{2\sigma_1(\sigma_1 + \sigma_2)(ay^2 + by)}{r} + \frac{2a\sigma_1(\sigma_1 + \sigma_2)^2 y}{r\sqrt{r}} = \frac{a(\sigma_1^2 - \sigma_2^2)^2}{r^2}$$

and

$$x_2^\star = -\frac{a\Delta^2 + 2b\Delta + \frac{2a\sigma_2\Delta}{\sqrt{r}} + \frac{2b(\sigma_1+\sigma_2)}{\sqrt{r}} + \frac{a(\sigma_1^2-\sigma_2^2)}{r}}{2a(\Delta + \frac{\sigma_2+\sigma_2}{\sqrt{r}})}.$$

Theorem 4 implies that with the help of the second derivative condition, we can extend the analysis of experimentaion problem in Bolton and Harris (1999) to situation where learning occurs in both arms. However, unlike the smooth pasting condition, the application of the second derivative condition must be consistent with the stochastic processes used by the model. As shown by the previous section, if the extension is based on exponential bandit (Keller, Rady, and Cripps (2005)) or Poisson bandit (Keller and Rady (2010)), then the second derivative condition is not needed and usually does not hold at the optimal cutoff.

5.3 Strategic Pricing

In many economic situations, the payoffs associated with each arm are not exogenously given. Instead, they are determined by strategic interactions among players. We can apply the second derivative condition that equates the value of experimentation across different options to a setting with strategic interaction. Consider two firms that sell to one consumer whose preferences are unknown. The consumer's valuation is common across the different sellers' products instead of independent. A real life example is that of a patient who does not know whether the consumption of a painkiller is effective. Buying iboprufen products from different sellers obviously generates information about a common underlying state, the effectiveness of the painkiller.

Bergemann and Välimäki (1996) consider a similar setup with independent arms. In their model, the Gittins index can be used to represent the value of each firm. From Bertrand competition, the firm with a higher Gittins index will always undercut the firm with a lower Gittins index. Therefore, the equilibrium is always efficient in the sense that the firm with a higher Gittins index will always be chosen by the buyer.

When the consumer's valuation is common, it is impossible to use the Gittins index to represent the value of each firm. Clearly, there is an externality in our framework: when one firm sells to the buyer, this also generates information about the product of the other seller. At first, it appears that the equilibrium cannot be efficient. Surprisingly, this intuition turns out to be incorrect as we show below.

Model Setting. The market consists of one buyer and two sellers indexed by $j = 1, 2$. The two sellers offer differentiated products and compete in prices in a continuous time model with an

infinite horizon. The production costs of these two products are both zero. The type of the buyer is either high or low. If the buyer is a high type, the expected value is ξ_{1H} from consuming good 1 and ξ_{2H} from consuming good 2. If the buyer is a low type, the expected value is ξ_{1L} from consuming good 1 and ξ_{2L} from consuming good 2. We assume that the low type buyer prefers good 1, while the high type buyer prefers good 2: $\xi_{1L} > \xi_{2L}$ and $\xi_{2H} > \xi_{1H}$.

Initially, all market participants hold the same prior belief about the buyer's type. At each instant of time, all market participants are also informed of all the previous outcomes. The performance of the products is, however, subject to random disturbances. If a type $i = H, L$ buyer purchases from a type j seller, the flow utility resulting from this purchase provides a noisy signal of the true underlying valuation:

$$du_{ji}(t) = \xi_{ji}dt + \sigma_j dZ_j(t),$$

where Z_1 and Z_2 are independent standard Brownian motion processes.

Denote by x the common belief that the buyer is a high type. Then the payoffs are linear in x: $f_1(x) = a_1x + b_1$ and $f_2(x) = a_2x + b_2$ where $a_j = \xi_{jH} - \xi_{jL}$ and $b_j = \xi_{jL}$ satisfying $a_1 + b_1 < a_2 + b_2$ and $b_1 > b_2$. Previous results (see, e.g., Bergemann and Välimäki (2000), Eeckhout and Weng (2010), Felli and Harris (1996)) show that x is updated by $dx = x(1-x)s_i d\bar{Z}_i$ where $s_i = \frac{a_i}{\sigma_i}$ and $\bar{Z}_i$ is a standard Brownian motion.

Socially Efficient Allocation. The social planner is facing a two-armed bandit problem with dependent arms. Denote the total social surplus function to be $v(x)$:

$$v(x) = \sup_{a:\mathcal{X}\to\{1,2\}} \left\{ \mathbb{E} \int_{t=0}^{\infty} e^{-rt} f_{a_t}(x_t)dt \right\}$$

$$\text{s.t. } dx_t = x_t(1-x_t)s_{a_t}d\bar{Z}_{a_t}(t) \quad \text{and} \quad a_t \triangleq a(x_t).$$

Denote $V_P(x)$ ($U_P(x)$) to be the optimal value if in a neighborhood of x, the social planner optimally chooses arm 1 (2). The general solutions to value functions are given by:

$$V_P(x) = \frac{f_1(x)}{r} + k_{11}x^{\alpha_1}(1-x)^{1-\alpha_1} + k_{12}x^{1-\alpha_1}(1-x)^{\alpha_1}$$
$$U_P(x) = \frac{f_2(x)}{r} + k_{21}x^{\alpha_2}(1-x)^{1-\alpha_2} + k_{22}x^{1-\alpha_2}(1-x)^{\alpha_2}$$

where $\alpha_1 = \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2r}{s_1^2}} \geq 1$ and $\alpha_2 = \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2r}{s_2^2}} \geq 1$. Since there is no drift term in the updating of x, Theorem 2 immediately implies that there is one unique socially optimal cutoff, denoted by x^e. $a_1 + b_1 < a_2 + b_2$ and $b_1 > b_2$ imply that arm 2 is chosen if $x > x^e$ and arm 1 is chosen if $x < x^e$. As a result, we must have $k_{12} = k_{21} = 0$ to guarantee that the value functions are bounded away from infinity.

The planner's optimal cutoff satisfies value matching, smooth pasting and second derivative and given linear payoffs, we can explicitly calculate x^e:

$$x^e = \frac{(b_1 - b_2)(\frac{s_1^2}{s_2^2}\alpha_1 + \alpha_2 - 1)}{(a_2 - a_1)[\frac{s_1^2}{s_2^2}(\alpha_1 - 1) + \alpha_2] + (b_1 - b_2)(\frac{s_1^2}{s_2^2} - 1)}.$$

Markov Perfect Equilibrium. We consider Markov perfect equilibria. At each instant of time t, the sellers set prices. After observing the price vector, the buyer chooses which product to buy. The natural state variable is state x for both sellers. The pricing strategy for seller $i = 1, 2$ is $p_i : [0, 1] \to \mathbb{R}$. Prices can be negative to allow for the possibility that the seller subsidizes the buyer to induce her to purchase the product. The acceptance strategy for the buyer is $\alpha : [0, 1] \times \mathbb{R} \times \mathbb{R} \to \{1, 2\}$. A Markov perfect equilibrium is a triple (p_1, p_2, α) such that *i)* given p_1 and p_2, α maximizes the buyer's expected intertemporal value; *ii)* given α and p_i, p_j maximizes seller $j \neq i$'s expected intertemporal profit.

As before, we denote $V(x)$ $(U(x))$ to be the equilibrium value of the buyer if the buyer purchases good 1 (2) in a neighborhood of x. Denote the value of the sellers by J_i, K_i: if the buyer purchases from seller i in a neighborhood of x, seller i's value is $J_i(x)$; otherwise, the value is $K_i(x)$. From the Hamilton-Jacobi-Bellman equation and Ito's Lemma, it is immediately seen that:

$$rJ_i(x) = p_i(x) + \frac{1}{2}\Sigma_i(x)J_i''(x),$$

and

$$rK_i(x) = \frac{1}{2}\Sigma_j(x)K_i''(x).$$

As in Bergemann and Välimäki (1996), Felli and Harris (1996), and Bergemann and Välimäki (2000), we investigate a Markov perfect equilibrium with cautious strategies.[12] A cautious strategy means that if the buyer purchases from seller i on the equilibrium path, seller $j \neq i$ will charge a

[12]This requirement captures the logic behind trembling hand perfection in this infinite time horizon framework (see Bergemann and Välimäki (1996)).

price $p_j(x)$ such that she is just indifferent between selling at price $p_j(x)$ and not selling:

$$p_j(x) + \frac{1}{2}\Sigma_j(x)K_j''(x) = \frac{1}{2}\Sigma_i(x)K_j''(x)$$

and hence

$$p_j(x) = \frac{1}{2}[\Sigma_i(x) - \Sigma_j(x)]K_j''(x).$$

On the other hand, the dominant seller i's price $p_i(x)$ is set such that the buyer is just indifferent between the two products (suppose $i = 2$):

$$f_2(x) - p_2(x) + \frac{1}{2}\Sigma_2(x)U''(x) = f_1(x) - p_1(x) + \frac{1}{2}\Sigma_1(x)U''(x)$$

and hence

$$p_2(x) = f_2(x) - f_1(x) + \frac{1}{2}[\Sigma_2(x) - \Sigma_1(x)](K_1''(x) + U''(x)).$$

We focus on cutoff strategies. Suppose $\underline{x} \in (0, 1)$ is an arbitrary equilibrium cutoff. Then at this cutoff, we have the equilibrium conditions:

$$J_i(\underline{x}) = K_i(\underline{x}), \quad J_i'(\underline{x}) = K_i'(\underline{x}), \quad \text{for seller } i = 1, 2.$$

For the buyer, the equilibrium conditions are:

$$U(\underline{x}) = V(\underline{x}), \quad U'(\underline{x}) = V'(\underline{x}), \quad U''(\underline{x}) = V''(\underline{x}).$$

When writing down the equilibrium conditions, we require only that the second derivative condition holds for the buyer, $U''(\underline{x}) = V''(\underline{x})$. This is because given the pricing strategies, the buyer is basically facing a two-armed bandit problem. Since the sellers are not facing a bandit problem, it may not be the case that $J_1''(\underline{x}) = K_1''(\underline{x})$ or $J_2''(\underline{x}) = K_2''(\underline{x})$. We can characterize the equilibrium cutoff from the above boundary conditions:

Theorem 5 *The equilibrium cutoff $x^\star$ is unique and is the same as the socially optimal cutoff x^e.*

Proof. In Appendix. ■

It turns out that at the unique equilibrium cutoff $x^\star$, the second derivative is equalized also for the sellers: $J_1''(x^\star) = K_1''(x^\star)$ and $J_2''(x^\star) = K_2''(x^\star)$. While second derivative is not imposed for

the sellers, a deviation by the seller would induce a different response from the buyer for whom the second derivative condition is indeed imposed.

Both the second derivative condition and cautious price are crucial to ensure that the equilibrium is efficient.[13] The externality is the possible source of inefficiency: a seller who does not sell nonetheless benefits from the experimentation of the other seller. The seller uses the cautious price to internalize this externality. However, the cautious price is not sufficient for the externality to be fully internalized. In particular, if the price path is discontinuous at the equilibrium cutoff, the price will not fully internalize the externality and there is inefficiency. The second derivative condition guarantees the price path is continuous at the equilibrium cutoff and that the externality is fully internalized.

6 Imperfectly Correlated Arms

Now we establish the second derivative condition to a context in which there is general but not perfect correlation between the arms. Consider one agent and a bandit with two arms $j = 1, 2$. Time is continuous and is denoted by t. Now there are two state variables $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, with $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$ and two choices $j \in \{1, 2\}$. The state variables (x, y) determine the instantaneous flow payoffs of each arm $f_i(x, y)$. Future payoffs are discounted at rate $r > 0$.

The state variables are independent and evolve according to the following process in arm j

$$dx = \mu_j(x, y)dt + \sigma_j(x, y)dZ_{jx}(t)$$
$$dy = \nu_j(x, y)dt + \omega_j(x, y)dZ_{jy}(t),$$

where both $Z_{jx}(t)$ and $Z_{jy}(t)$ are independent standard Brownian motion processes.[14] For simplicity, we don't include jumps in the stochastic processes.

Assumption 3 *$f_j(x, y), \mu_j(x, y), \sigma_j(x, y), \omega_j(x, y)$ are $\mathcal{C}^2$ with bounded first order derivatives for $x \in \mathcal{X}, y \in \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ are connected sets.*

[13] In Bergemann and Välimäki (1996), all Markov perfect equilibria are efficient. The notion of cautious equilibrium is introduced to guarantee that the equilibrium is unique. However, in our Theorem 5, the notion of cautious equilibrium is important to guarantee efficiency. In other words, non-cautious Markov perfect equilibria might be inefficient.

[14] It is without loss of generality to assume that Z_{jx}, Z_{jy} are independent. If they were not, there is a covariance term $\text{cov}(dZ_{jx}, dZ_{jy}) = g_j(x, y)dt$ that enters the value function. We can always redefine two new state variables $\tilde{x}, \tilde{y}$ adding the covariance term to the drift and substituting for an independent Wiener process.

As before, the value function can be written as:

$$v(x,y) = \sup_{a:\mathcal{X}\times\mathcal{Y}\to\{1,2\}} \left\{\mathbb{E}\int_{t=0}^{\infty} e^{-rt} f_{a_t}(x_t,y_t)dt\right\}$$
$$\text{s.t. } dx_t = \mu_{a_t}(x_t)dt + \sigma_{a_t}(x_t)dZ_{1a_t}(t), \quad dy_t = \nu_{a_t}(y_t)dt + \omega_{a_t}(y_t)dZ_{2a_t}(t), \quad \text{and } a_t \triangleq a(x_t,y_t).$$

Let $V(x,y)$ ($U(x,y)$) be the value function of an agent with state in a neighborhood of (x,y) optimally choosing arm 1 (2). The optimal stopping decision is characterized by set

$$\mathcal{S} = \{(x^\star, y^\star) : \text{agent switches arms at } (x^\star, y^\star)\}.$$

Some conditions have to be imposed on $\mathcal{S}$:

Assumption 4 *$\mathcal{S}$ satisfies the following conditions:*

1. *for any $(x,y) \in \mathcal{S}$ and any $\epsilon > 0$, there exists $(x',y') \in B_\epsilon(x,y)$ and $(x',y') \neq (x,y)$ such that $(x',y') \in \mathcal{S}$;*[15]

2. *there exists $\eta > 0$ such that for any $(x,y) \in \mathcal{S}$, $(x, y+\epsilon) \notin S$ and $(x+\epsilon, y) \notin S$ for all $\epsilon \in (-\eta, \eta)$;*

3. *for any $(x,y) \in \mathcal{S}$, there exists $\bar{\epsilon}$, such that if $(x',y') \in B_{\bar{\epsilon}}(x,y)$ and $(x',y') \notin \mathcal{S}$, then $v(\cdot)$ is at $\mathcal{C}^2$ at (x',y').*

The first two conditions guarantee that the optimal stopping decisions can be characterized by isolated stopping curves and the last condition is made to guarantee that the first and second derivatives of v exist in a neighborhood of $\mathcal{S}$ (a similar condition is also imposed in Shiryaev (1978)).

For any (x,y), denote $\tau(x,y)$ to be the first time that (x_t,y_t) is in $\mathcal{S}$ beginning from (x,y). Then we can rewrite the value function as:

$$v(x,y) = \mathbb{E}\int_{t=0}^{\tau(x,y)} e^{-rt} f_{a_t}(x_t,y_t)dt + e^{-r\tau(x,y)} v(x_{\tau(x,y)}, y_{\tau(x,y)}).$$

In particular, if it is optimal to choose arm 1 in a neighborhood of (x,y), then it must be the case

[15] $B_\epsilon(x,y)$ is defined as:

$$B_\epsilon(x,y) = \{(x',y') : ||(x',y') - (x,y)|| \leq \epsilon\},$$

where $||\cdot||$ is the Euclidean norm.

that:

$$V(x,y) = \mathbb{E}\int_{t=0}^{\tau(x,y)} e^{-rt} f_1(x_t, y_t)dt + e^{-r\tau(x,y)} U(x_{\tau(x,y)}, y_{\tau(x,y)});$$

and if it is optimal to choose arm 2 in a neighborhood of (x,y), then it must be the case that:

$$U(x,y) = \mathbb{E}\int_{t=0}^{\tau(x,y)} e^{-rt} f_2(x_t, y_t)dt + e^{-r\tau(x,y)} V(x_{\tau(x,y)}, y_{\tau(x,y)}).$$

For the remainder, we will represent partial derivatives on the value functions by subscripts, for example, $V_1(x,y) = \frac{\partial}{\partial x}V(x,y)$ and $U_{22}(x,y) = \frac{\partial^2}{\partial y^2}U(x,y)$.

Consider any $(x^\star, y^\star)$ in the interior of $\mathcal{S}$. At $(x^\star, y^\star)$, Peskir and Shiryaev (2006) and Shiryaev (1978) show that the logic of proof of value matching and smooth pasting in the one-dimensional case can be extended to the multi-dimensional setting. In particular, the value matching condition also comes from the continuity of value functions. The smooth pasting condition can also be proved by considering deviations in the state space.Therefore, we can get the value matching and smooth pasting conditions are the same as in the one-dimensional case along each dimension.

Value Matching.

$$V(x^\star, y^\star) = U(x^\star, y^\star)$$

Smooth Pasting.

$$V_x(x^\star, y^\star) = U_x(x^\star, y^\star)$$
$$V_y(x^\star, y^\star) = U_y(x^\star, y^\star)$$

Furthermore, we show that the generalized second derivative condition involves equating the weighted sum of all partial derivatives in each dimension between different arms. The total value of experimentation is the weighted sum of the experimentation value along each of the dimensions x and y. This must be equated at both arms with values V and U.

Theorem 6 *Suppose $(x^\star, y^\star)$ is in the interior of $\mathcal{S}$. Then, a necessary condition is at $(x,y) =$*

$(x^\star, y^\star)$:[16]

$$\sigma_1^2(x)V_{11}(x,y) + \omega_1^2(y)V_{22}(x,y) = \sigma_1^2(x)U_{11}(x,y) + \omega_1^2(y)U_{22}(x,y) \tag{3}$$

and

$$\sigma_2^2(x)V_{11}(x,y) + \omega_2^2(y)V_{22}(x,y) = \sigma_2^2(x)U_{11}(x,y) + \omega_2^2(y)U_{22}(x,y). \tag{4}$$

Proof. In Appendix. ■

7 Concluding Remarks

The bandit problem is used as a building block to investigate many broader issues. A nonexhaustive list of related papers includes Bolton and Harris (1999), Bonatti (2009), Daley and Green (2008), Faingold and Sannikov (2007), Hörner and Samuelson (2009), Moscarini (2005), and Strulovici (2010). Almost all of the existing papers use a one-armed bandit framework: agents stop experimentation after switching to the safe arm. The second derivative condition enables us to investigate the situation in which agents continue to learn about the same underlying state variable, even if they switch action.

If there are costs from switching, for example, due to search frictions, the same logic of second derivative applies. Now, there will be bounds on the value of experimentation because the cost of switching will drive a wedge. The implication is that the value of experimentation on one arm can differ from that on the other arm but the difference between the two cannot be too big.

Superficially, our condition appears similar to the super contact condition (Dumas (1991)), yet there is no relation. The setting is fundamentally different because in Dumas there is only one arm and a cost is paid to stay on that arm. More important, the super contact condition is not a condition of second derivative, but rather a version of the smooth pasting condition in a setting with frictions. The experimenter has to pay a flow cost to stop the control from moving beyond the cutoff. When recalculating the smooth pasting condition in the presence of a proportional cost, this condition implies a restriction on the second derivative. This restriction is derived from considering a deviation in the state space. In the case of continued experimentation with multiple dependent

[16] In the special independent arm case ($\omega_1 = \sigma_2 = 0$), Karatzas (1984) applies the Whittle reduction technique developed in Whittle (1980) and chracterizes the optimal stopping rule. It is straightforward to show that at the optimal stopping boundary, the second derivative conditions are also satisfied:

$$V_{11}(x,y) = U_{11}(x,y), \quad V_{22}(x,y) = U_{22}(x,y).$$

arms, the second-order condition is derived from imposing no deviation for a short duration, i.e., in time space. Moreover, as we have shown for the case of multiple dimensions, such second derivative implies a condition that is generally very different from the super contact condition. This illustrates that the similarity is a coincidence.

Appendix

Proof of Theorem 2

Proof. Without loss of generality, assume $a_1 > a_2$ and $(\mu_1 - \frac{\sigma_1^2}{\sigma_2^2}\mu_2) \geq 0$. The case for $a_1 < a_2$ can be proved similarly. $a_1 > a_2$ implies that as x goes to $+\infty$, arm 1 must be optimally chosen and as x goes to $-\infty$, arm 2 must be optimally chosen. Consider a cutoff strategy where the agent chooses the same arm y on an interval between the cutoffs. Since $k_{y1} \geq 0$ and $k_{y2} \geq 0$, v_y is a convex function. Therefore, the smooth pasting condition implies that $v'(x)$ is a continuously increasing function on $(-\infty, +\infty)$ satisfying $\lim_{x\to+\infty} v'(x) = \frac{a_1}{r}$, $\lim_{x\to-\infty} v'(x) = \frac{a_2}{r}$ and $v'(x) \in (\frac{a_2}{r}, \frac{a_1}{r})$.

Suppose by contradiction there are two cutoffs $x_1 < x_2$ such that arm 1 is chosen for $x \in (x_1, x_2)$ and arm 2 is chosen in a neighborhood of $x < x_1$ and $x > x_2$. At cutoffs x_1 and x_2, the value matching condition implies:

$$f_1(x_1) + \mu_1 V'(x_1) + \frac{1}{2}\sigma_1^2 V''(x_1) = f_2(x_1) + \mu_2 U'(x_1) + \frac{1}{2}\sigma_2^2 U''(x_1)$$

and

$$f_1(x_2) + \mu_1 V'(x_2) + \frac{1}{2}\sigma_1^2 V''(x_2) = f_2(x_2) + \mu_2 U'(x_2) + \frac{1}{2}\sigma_2^2 U''(x_2).$$

Smooth pasting and second derivative conditions imply that

$$V'(x_1) = U'(x_1), \quad V''(x_1) = U''(x_1), \quad V'(x_2) = U'(x_2), \quad V''(x_2) = U''(x_2).$$

Rearranging terms yields

$$\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} rV(x_1) + \frac{\sigma_2^2}{\sigma_1^2} f_1(x_1) + (\mu_1 - \mu_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2}\mu_1)V'(x_1) = f_2(x_1)$$

and

$$\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} rV(x_2) + \frac{\sigma_2^2}{\sigma_1^2} f_1(x_2) + (\mu_1 - \mu_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2}\mu_1)V'(x_2) = f_2(x_2).$$

The subtraction of the above two equations leads to

$$\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} r(V(x_2) - V(x_1)) + \frac{\sigma_2^2}{\sigma_1^2}(f_1(x_2) - f_1(x_1)) + (\mu_1 - \mu_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2}\mu_1)(V'(x_2) - V'(x_1)) = f_2(x_2) - f_2(x_1). \tag{5}$$

Notice $\mu_1 - \mu_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2}\mu_1 \geq 0$ and $V'(x_2) - V'(x_1) \geq 0$ since $V(\cdot)$ is a convex function. Equation

5 implies

$$\frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2}r(V(x_2)-V(x_1))+\frac{\sigma_2^2}{\sigma_1^2}(f_1(x_2)-f_1(x_1))\leq f_2(x_2)-f_2(x_1). \tag{6}$$

There are three cases in total.

Case 1, $\sigma_1=\sigma_2$ and then we have

$$(a_1-a_2)(x_2-x_1)\leq 0.$$

This is impossible given $a_1>a_2$ and $x_2>x_1$.

Case 2, $\sigma_1>\sigma_2$. Since $V(\cdot)$ is a convex function with $V'\in(\frac{a_2}{r},\frac{a_1}{r})$. The left-hand side of inequality 6 is strictly larger than

$$\frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2}a_2(x_2-x_1)+\frac{\sigma_2^2}{\sigma_1^2}a_1(x_2-x_1),$$

and hence $\frac{\sigma_2^2}{\sigma_1^2}(a_1-a_2)(x_2-x_1)<0$, which leads to a contradiction.

Case 3, $\sigma_1<\sigma_2$. Then the left-hand side of inequality 6 is strictly larger than

$$\frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2}a_1(x_2-x_1)+\frac{\sigma_2^2}{\sigma_1^2}a_1(x_2-x_1)$$

and hence $(a_1-a_2)(x_2-x_1)<0$, which also leads to a contradiction.

Therefore, there must be a unique optimal cutoff. ■

Proof of Theorem 3

Proof. The policy implies a well-defined law of motion for x. The value functions and $x^\star$ can be written as:

$$\begin{aligned}
x^\star &= \tfrac{r(b_2-b_1)+a_2\mu_2-a_1\mu_1}{r(a_1-a_2)}+\tfrac{\gamma_1-\beta_2}{\gamma_1\beta_2}\\
V(x) &= \tfrac{r(a_1x+b_1)+a_1\mu_1}{r^2}+k_1e^{-\gamma_1 x},\quad k_1=e^{\gamma_1x^\star}\tfrac{\beta_2(a_1-a_2)}{r\gamma_1(\gamma_1+\beta_2)}\\
U(x) &= \tfrac{r(a_2x+b_2)+a_2\mu_2}{r^2}+k_2e^{\beta_2 x},\quad k_2=e^{-\beta_2x^\star}\tfrac{\gamma_1(a_1-a_2)}{r\beta_2(\gamma_1+\beta_2)}.
\end{aligned}$$

To prove the policy is optimal, it suffices to prove that

$$1=\text{argmax}_{a\in\{1,2\}}\left\{f_a(x)+\mu_a(x)V'(x)+\frac{1}{2}\sigma_a^2(x)V''(x)\right\}$$

for $x > x^\star$, and

$$2 = \text{argmax}_{a\in\{1,2\}} \left\{ f_a(x) + \mu_a(x)U'(x) + \frac{1}{2}\sigma_a^2(x)U''(x) \right\}$$

for $x < x^\star$. Then it is equivalent to show that

$$\Phi_1(x) = f_2(x) - f_1(x) + \mu_2(x)V'(x) - \mu_1(x)V'(x) + \frac{1}{2}\sigma_2^2(x)V''(x) - \frac{1}{2}\sigma_1^2(x)V''(x) < 0$$

for $x > x^\star$ and

$$\Phi_2(x) = f_2(x) - f_1(x) + \mu_2(x)U'(x) - \mu_1(x)U'(x) + \frac{1}{2}\sigma_2^2(x)U''(x) - \frac{1}{2}\sigma_1^2(x)U''(x) > 0$$

for $x < x^\star$. We only need to show that the first inequality and the proof of the second inequality is similar. Since $\Phi_1(x^\star) = 0$, showing $\Phi_1(x) < 0$ is equivalent to proving $\Phi_1'(x) < 0$. Rewrite Φ_1 as:

$$\Phi_1(x) = f_2(x) - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} rV(x) - \frac{\sigma_2^2}{\sigma_1^2} f_1(x) - (\mu_1 - \mu_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2}\mu_1)V'(x)$$

and

$$\Phi_1'(x) = a_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} rV'(x) - \frac{\sigma_2^2}{\sigma_1^2} a_1 - (\mu_1 - \mu_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2}\mu_1)V''(x).$$

$(\mu_1 - \mu_2 - \frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2}\mu_1)V''(x) \geq 0$ since $\mu_1 - \mu_2 - \frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2}\mu_1 \geq 0$ and V is convex. Therefore, we only need to show $\phi_1(x) = a_2 - \frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2} rV'(x) - \frac{\sigma_2^2}{\sigma_1^2} a_1 < 0$. There are three cases in total: if $\sigma_1 = \sigma_2$ and then $\phi_1(x) = a_2 - a_1 < 0$; if $\sigma_1 > \sigma_2$ and then

$$\phi_1(x) < a_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} a_2 - \frac{\sigma_2^2}{\sigma_1^2} a_1 = \frac{\sigma_2^2}{\sigma_1^2}(a_2 - a_1) < 0;$$

if $\sigma_1 < \sigma_2$ and then

$$\phi_1(x) < a_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} a_1 - \frac{\sigma_2^2}{\sigma_1^2} a_1 = a_2 - a_1 < 0.$$

$\Phi_1(x) < 0$ and $\Phi_2(x) > 0$ imply that the cutoff policy and the value functions V, U solve the HJB equation. Therefore, the cutoff policy is optimal. ■

Proof of Theorem 5

Proof. Consider an arbitrary equilibrium cutoff $\underline{x}$. Without loss of generality, suppose it is the case that the buyer chooses good 2 in a neighborhood of x such that $x > \underline{x}$ and chooses good 1 in

a neighborhood of x such that $x < \underline{x}$. For $x > \underline{x}$, the buyer's value satisfies:

$$rU(x) = f_2(x) - p_2(x) + \frac{1}{2}\Sigma_2(x)U''(x) = f_1(x) - p_1(x) + \frac{1}{2}\Sigma_1(x)U''(x),$$

while for $x < \underline{x}$,

$$rV(x) = f_1(x) - p_1(x) + \frac{1}{2}\Sigma_1(x)V''(x) = f_2(x) - p_2(x) + \frac{1}{2}\Sigma_2(x)V''(x).$$

At $x = \underline{x}$, since $V(x) = U(x)$ and $V''(x) = U''(x)$, it must be the case that both $p_1(\cdot)$ and $p_2(\cdot)$ are continuous at $\underline{x}$. Consider the value functions for the sellers. For $x > \underline{x}$,

$$rJ_2(x) = p_2(x) + \frac{1}{2}\Sigma_2(x)J_2''(x);$$

and for $x < \underline{x}$,

$$rK_2(x) = p_2(x) + \frac{1}{2}\Sigma_2(x)K_2''(x),$$

from the assumption of cautious equilibrium. The fact that $J_2(\underline{x}) = K_2(\underline{x})$ and $p_2(x)$ is continuous at $\underline{x}$ immediately implies that $J_2''(x) = K_2''(x)$. Similarly, we have: $J_1''(x) = K_1''(x)$. Define

$$W_1(x) = V(x) + J_1(x) + K_2(x) \quad \text{and} \quad W_2(x) = U(x) + J_2(x) + K_1(x).$$

Obviously, W_i denotes the social surplus from purchasing good i in equilibrium. Moreover, the social surplus function $W_i(x)$ satisfies the following differential equation:

$$rW_i(x) = f_i(x) + \frac{1}{2}\Sigma_i(x)W_i''(x).$$

The value matching, smooth pasting and second derivative conditions imply that at $x = \underline{x}$, the following boundary conditions are satisfied:

$$W_1(x) = W_2(x), \quad W_1'(x) = W_2'(x), \quad W_1''(x) = W_2''(x).$$

Notice that at the socially efficient cutoff x^e, it must be the case that

$$V_P(x^e) = U_P(x^e), \quad V_P'(x^e) = U_P'(x^e), \quad V_P''(x^e) = U_P''(x^e).$$

Meanwhile, V_P and W_1 (U_P and W_2) satisfy the same differential equation. Since the socially efficient cutoff x^e is unique, there exists a unique equilibrium cutoff $x^\star$. Furthermore, since $x^\star$ and x^e share the same boundary conditions, it must be the case that $x^\star = x^e$. ■

Proof of Theorem 6

Proof. Pick an arbitrary point $(x, y) \in \mathcal{S}$. Assuption 3 implies that $\mathcal{S}$ is isolated. As a result, there exists $\bar{\epsilon}$ such that for all $\epsilon < \bar{\epsilon}$, it is optimal to choose arm 1 at $(x - \epsilon/2, y)$ and arm 2 at $(x + \epsilon/2, y)$. Without loss of generality, suppose that $\bar{\epsilon} > 0$. This implies that there exists $\bar{t}$ such that for all $t \leq \bar{t}$, a one-shot deviation with length t is not optimal. Denote

$$\tilde{V}(x - \epsilon/2, y; t) = \mathbb{E}\left\{\int_0^t e^{-r\tau} f_2(x_\tau, y_\tau) d\tau + e^{-rt} V(x_t, y_t)\right\}$$

and

$$dx_t = \mu_2(x_t)dt + \sigma_2(x_t)dZ_{12}(t) \quad dy_t = \nu_2(y_t)dt + \omega_2(y_t)dZ_{22}(t).$$

It must be the case: $\frac{\tilde{V}(x-\epsilon/2,y;t)-V(x-\epsilon/2,y)}{t} \leq 0$. As t goes to zero, this implies that at $(x - \epsilon/2, y)$:

$$f_2 + \mu_2 V_2 + \nu_2 V_2 + \frac{1}{2}\sigma_2^2 V_{11} + \frac{1}{2}\omega_2^2 V_{22} - \left[f_1 + \mu_1 V_1 + \nu_1 V_2 + \frac{1}{2}\sigma_1^2 V_{11} + \frac{1}{2}\omega_1^2 V_{22}\right] \leq 0. \quad (7)$$

Take ϵ goes to zero and inequality (7) implies that at (x, y),

$$f_2 + \mu_2 V_1 + \nu_2 V_2 + \frac{1}{2}\sigma_2^2 V_{11} + \frac{1}{2}\omega_2^2 V_{22} - \left[f_2 + \mu_2 U_1 + \nu_2 U_2 + \frac{1}{2}\sigma_2^2 U_{11} + \frac{1}{2}\omega_2^2 U_{22}\right] \leq 0.$$

Since $V_1 = U_1$ and $V_2 = U_2$ at $(x^\star, y^\star)$ from the smooth pasting condition, it must be the case that at (x, y),

$$\sigma_2^2 V_{11} + \omega_2^2 V_{22} - \sigma_2^2 U_{11} - \omega_2^2 U_{22} \leq 0.$$

This establishes one side of the equality.

Next, pick any $\epsilon \in (0, \bar{\epsilon})$. Consider another set of stopping cutoffs:

$$\hat{\mathcal{S}} = \{(x^\star - \epsilon, y^\star) \in \mathcal{X} \times \mathcal{Y} : (x^\star, y^\star) \in \mathcal{S}\}.$$

Under this new stopping rule, the agents take arm 2 in a neighborhood of $(x - \epsilon/2, y)$. For any (x, y), denote $\hat{\tau}(x, y)$ to the first time that (x_t, y_t) is in $\hat{\mathcal{S}}$ beginning from (x, y). Then we define

the value function associated with the new stopping rule as:

$$\hat{v}(x,y) = \mathbb{E}\int_{t=0}^{\hat{\tau}(x,y)} e^{-rt} f_{a_t}(x_t, y_t)dt + e^{-r\hat{\tau}(x,y)}\hat{v}(x_{\hat{\tau}(x,y)}, y_{\hat{\tau}(x,y)}).$$

In particular, at $(x-\epsilon/2, y)$,

$$\hat{U}(x-\epsilon/2,y) = \mathbb{E}\int_{t=0}^{\hat{\tau}(x-\epsilon/2,y)} e^{-rt} f_{a_t}(x_t, y_t)dt + e^{-r\hat{\tau}(x-\epsilon/2,y)}\hat{v}(x_{\hat{\tau}(x-\epsilon/2,y)}, y_{\hat{\tau}(x-\epsilon/2,y)}).$$

Bellman's Principle of Optimality implies that if it is optimal to choose arm 1 at $(x-\epsilon/2, y)$, then a one-shot deviation is better than deviating forever. In other words, there exists $\bar{t}'$ such that for all $t \leq \bar{t}'$,

$$\tilde{V}(x-\epsilon/2, y; t) \geq \hat{U}(x-\epsilon/2, y) \Longrightarrow \frac{\tilde{V}(x-\epsilon/2, y; t) - \hat{U}(x-\epsilon/2, y)}{t} \geq 0.$$

As t goes to zero, this implies at $(x-\epsilon/2, y)$:

$$f_2 + \mu_2 V_1 + \nu_2 V_2 + \frac{1}{2}\sigma_2^2 V_{11} + \frac{1}{2}\omega_2^2 V_{22} - \left[f_2 + \mu_2 \hat{U}_1 + \nu_2 \hat{U}_2 + \frac{1}{2}\sigma_2^2 \hat{U}_{11} + \frac{1}{2}\omega_2^2 \hat{U}_{22}\right] \quad \geq \quad 0. \quad (8)$$

Meanwhile, as ϵ goes to zero, $\hat{U}$ converges to U. Therefore, inequality (8) implies that at (x, y),

$$\sigma_2^2 V_{11} + \omega_2^2 V_{22} - \sigma_2^2 U_{11} - \omega_2^2 U_{22} \geq 0.$$

Then we should have an equality:

$$\sigma_2^2(x) V_{11}(x,y) + \omega_2^2(y) V_{22}(x,y) = \sigma_2^2(x) U_{11}(x,y) + \omega_2^2(y) U_{22}(x,y). \quad (9)$$

Apply the same procedure for $(x+\epsilon/2, y)$ and it is similar to get the other equation stated in the theorem. ■

References

Applebaum, D. (2004): *Lévy Processes and Stochastic Calculus*. Cambridge University Press.

Bergemann, D., and J. Välimäki (1996): "Learning and Strategic Pricing," *Econometrica*, 64(5), 1125–1149.

——— (2000): "Experimentation in Markets," *Review of Economic Studies*, 67(2), 213–234.

——— (2008): "Bandit Problems," in *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke.

Bolton, P., and C. Harris (1999): "Strategic Experimentation," *Econometrica*, 67(2), 349–374.

Bonatti, A. (2009): "Menu Pricing and Learning," mimeo.

Cohen, A., and E. Solan (2009): "Bandit problems with Lévy payoff," mimeo.

Daley, B., and B. Green (2008): "Waiting for News in the Dynamic Market for Lemons," mimeo.

Dixit, A., and R. Pindyck (1994): *Investment under Uncertainty*. Princeton University Press.

Dumas, B. (1991): "Super contact and related optimality conditions," *Journal of Economic Dynamics and Control*, 15, 675–685.

Eeckhout, J., and X. Weng (2010): "Assortative Learning," mimeo.

Faingold, E., and Y. Sannikov (2007): "Reputation Effects and Equilibrium Degeneracy in Continuous-Time Games," Yale mimeo.

Felli, L., and C. Harris (1996): "Learning, Wage Dynamics and Firm-Specific Human Capital," *Journal of Political Economy*, 104(4), 838–868.

Gittins, J., and D. Jones (1974): "A dynamic allocation index in the sequential design of experiments," in *Progress in Statistics*, pp. 241–266. North Holland, Amsterdam.

Hörner, J., and L. Samuelson (2009): "Incentives for Experimenting Agents," Yale mimeo.

Karatzas, I. (1984): "Gittins Indices in the Dynamic Allocation Problem for Diffusion Processes," *Annals of Probability*, 12(1), 173–92.

Kaspi, H., and A. Mandelbaum (1995): "Lévy Bandits: Multi-Armed Bandits Driven by Lévy Processes," *Annals of Applied Probability*, 5, 541–565.

Keller, G., and S. Rady (2010): "Strategic Experimentation with Poisson Bandits," *Theoretical Economics*, 5(2), 275–311.

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

Moscarini, G. (2005): "Job Matching and the Wage Distribution," *Econometrica*, 73(2), 481–516.

Peskir, G., and A. Shiryaev (2006): *Optimal stopping and free-boundary problems*. Birkhauser.

Samuelson, P. A. (1965): "Rational Theory of Warrant Pricing," *Industrial Management Review*, 6, 13–31.

Shiryaev, A. (1978): *Optimal Stopping Rules*. Springer-Verlag.

Strulovici, B. (2010): "Learning While Voting: Determinants of Collective Experimentation," *Econometrica*, 78(3), 933–971.

Whittle, P. (1980): "Multi-armed bandits and Gittins index," *J. Roy. Statist. Soc., Ser. B.*, 42, 143–149.

Wirl, F. (2008): "Reversible stopping ("switching") implies super contact," *Computational Management Science*, 5, 393–401, Published online in 2007.