Common Value Experimentation*

Jan Eeckhout† and Xi Weng‡

May 2014

Abstract

In many economic environments, agents often continue to learn about 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 analyze a general setup of experimentation with common values, 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 that the second derivatives of the value function must be equal at the switching point. This condition holds generally whenever the stochastic process has continuous increments.

Keywords. Experimentation. Jump-diffusion Process. Common Values. Correlated Values.

JEL. D83. C02. C61.

*We thank Martin Cripps, Pauli Murto, Yuliy Sannikov, Robert Shimer, Lones Smith, Nancy Stokey, Marko Terviö and Juuso Välimäki for insightful discussions and comments. We also benefitted from the feedback of seminar audiences. Most of the results in this paper first appeared under the title "Assortative Learning" (2008). Eeckhout acknowledges support from the ERC, Grant 208068. Weng acknowledges support from the National Natural Science Foundation of China (Grant No. 71303014) and Guanghua Leadership Institute (Grant No. 12-02).

†University College London and Barcelona GSE-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. Common value experimentation is particularly important for promotion decisions. Likewise, common value experimentation is prevalent in consumer choice, for example, patients who learn about the effectiveness of different drugs.

In this paper we analyze optimal experimentation problems in environments of common values. 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 common value 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 generalized jump-diffusion process, that incorporates continuous increments like a Brownian motion, as well as Poisson jumps. This setup includes among many others pure Bayesian learning where the belief is

[1] See Bergemann and Välimäki (2008) for a survey on bandit problems. At this point, we must clarify the terminology we use. Common value experimentation is not common learning as in Cripps, Ely, Mailath, and Samuelson (2008). There, given a stream of signals, two agents want to learn the value of an unknown parameter. When each agent's signal space is finite, the true value of the parameter eventually becomes common knowledge. There is no experimentation, just belief updating. The term common learning refers to the fact that the value eventually becomes common knowledge to both agents. In our setting, one agent experiments (this includes non-martingale stochastic processes) over two arms, with an underlying type that is common value.

updated as a martingale, non-Bayesian learning where the state variable follows a Brownian motion with a drift, as well as experimentation in strategic and market settings.

Our main result is to establish a simple equilibrium condition on the value function that must be satisfied whenever the common value experimentation problem has a continuous increment component (Brownian motion). This condition imposes equalization of the second derivative of the value functions at each arm in the neighborhood of the cutoff. Our condition adds to the well-known conditions at the cutoff of value matching (the value function is continuous) and smooth pasting (the first derivative is continuous). In the presence of continued learning, equilibrium must now also satisfy that the second derivative is continuous at the cutoff.

In experimentation problems we typically cannot explicitly solve for the value function, yet its properties can be studied using Bellman's principle of optimality. An optimal policy is such that, whatever the initial state, the remaining decisions must constitute an optimal policy. At any point, equilibrium actions must therefore be robust to deviations. There are different aspects of the principle of optimality. In particular, the celebrated smooth pasting condition considers a *deviation in state space*: given a candidate equilibrium cutoff, the agent postpones switching arms at this cutoff, possibly deviating forever. Equilibrium requires this is suboptimal, and when the value of such a deviation is derived in the neighborhood of the equilibrium allocation, this gives rise to an inequality condition on the first order derivative of the value function. Since the inequality holds on both sides of the candidate equilibrium cutoff, this implies equalization of the first derivative of the value function.

The condition we derive here instead ensures in addition there is no *one-shot deviation*, i.e., no deviation in time space. The deviating agent switches arms for a short instance, and then reverts to the candidate equilibrium arm. When there are incremental changes in the stochastic process due to the Brownian motion component, the differential equation for the value function depends on the second derivative, sometimes referred to as the value of experimentation. The suboptimality of the one-shot deviation therefore implies an inequality on the second derivative. Near the cutoff, this has to be satisfied on both sides and with opposite inequality, which implies equalization of the second derivative of the value function.

If this condition is not satisfied, an experimenting agent close to 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 learning 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 learning ensures that no such gains from deviation exist.

When there is no continuous increment component (Brownian motion) and there are only discrete increments in the process, then the one-shot deviation principle follows directly from value matching and smooth pasting and imposes no additional restriction on the value function.

With reference to the benchmark of the one-armed bandit problem, it is important to note that the second derivative of the value function is not equal to zero (the value of the outside option is a constant). Yet, a one-shot deviation is trivially deterred when value matching and smooth pasting is satisfied. Because there is no experimentation value to taking the safe option, the one-shot deviation condition is one-sided and is given by the inequality that the second derivative is positive, which is trivially satisfied from the convexity of the value function.

Superficially, the second derivative condition appears similar to the well-known super-contact condition due to 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 the result of deterring a one-shot deviation, but rather a version of the smooth pasting condition in a setting with frictions.[2]

There is however a close relation between our result and the one by Wirl (2008). He derives the second derivative condition in the context of a pure Brownian motion in the absence of discrete increments. His Brownian motion setting however is very special as it requires that the variance term on the noise is *identical* on both arms. As a result, the common variance term cancels when applying the principle of optimality. This implies the first derivative at the cutoff can be calculated explicitly, and simply differentiating gives the second derivative condition. This logic clearly does not hold under a minor difference in the Brownian noise terms. This is clearly important for example when the rate of learning at two jobs differs. Our result adds to Wirl's in two ways. First we derive the condition in a very general context that includes general jump-diffusion processes, and then we show when the result does and when it does not hold (in the absence of a continuous increment component). Second, we prove the result by means of the one-shot deviation principle that

[2]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 also derived from considering a deviation in the state space, not in time space.

is generally applicable.

The main appeal of this simple condition is its applicability. Not only are such environments with continued learning prevalent in many economic contexts, the implementation of the one shot deviation principle 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 learning and show how the second derivative condition affects the equilibrium allocation and its efficiency.

We end the paper by deriving 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 learning 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. It can be checked that the Karatzas solution to the value function satisfies the second derivative condition.

Finally, it is important to point out that in the literature, there is another way of deriving the second derivative condition by applying the method of sub- and super-solutions (see, e.g., Keller and Rady (1999) and Bonatti (2011)). The method is used to prove the existence of solutions for many classes of boundary value problems involving ordinary and partial differential equations by showing the existence of continuously differentiable sub- and super-solutions and verifying a regularity condition. While the method is well known, it is much more involved than ours. In addition, it is not always applicable. In the one-dimensional state space case, the method of sub- and super-solutions implies that the value function is continuously twice differentiable, which coincides with our second derivative condition. However, this may not be true in the multi-dimensional state space case as shown in Section 5. Related to our general applicable method is the independent work by Strulovici and

Szydlowski (2014). Yet the models are different. Their setup has a large number of arms and Brownian uncertainty. Our setup is for two arms with Levy processes as well as Brownian uncertainty, and we also have results for correlated arms.

2 The Basic Model

Consider one agent and a bandit with two arms $j = 1, 2$. Time is continuous and denoted by $t \geq 0$. In the basic model, we will consider the case where there is only one real-valued state $x(t) \in \mathcal{X}$ and $\mathcal{X} \subset \mathbb{R}$ is a connected set. The state x determines the instantaneous flow payoffs of each arm $f_j(x)$. Future payoffs are discounted at rate $r > 0$. Extension of this basic model can be found in Section 5.

For each arm j, there is a probability space $\{\Omega^j, \mathcal{F}^j, P^j\}$ endowed with filtration $\{\mathcal{F}_t^j, t \geq 0\}$. It is assumed that the state x follows a jump-diffusion process under P^j. In other words, denote $T_j(t)$ to be the total measure of time to date t that arm j has been chosen. Then, the updating of x in arm j satisfies:

$$dx_j(t) = \mu_j(x(t-))dT_j(t) + \sigma_j(x(t-))d\mathbb{Z}_j(T_j(t)) + \int_{\mathbb{R}-\{0\}} G_j(x(t-), y)\mathbb{N}_j(dT_j(t), dy).$$

The state x is updated either in arm 1 or in arm 2 and hence $dx(t) = dx_1(t) + dx_2(t)$. In the updating formula, $\mathbb{Z}_j(t)$ is a standard Brownian motion process and $\mathbb{N}_j$ is a Poisson random measure that is independent of the Brownian motion $\mathbb{Z}_j$. For simplicity, we assume that $\mathbb{N}_j$ has finite intensity measure ν_j, i.e., $\mathbb{N}_j$ can be the sum of m_j independent Poisson processes, which are also independent of $\mathbb{Z}_j$. 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.[3] $G_j(x, y)$ denotes the change of the state when there is a Poisson jump y at state x. Furthermore, we assume that $\mathbb{N}_1$, $\mathbb{N}_2$, $\mathbb{Z}_1$ and $\mathbb{Z}_2$ are mutually independent of each other. The process considered by us covers a lot

[3]This jump-diffusion process is an important class of general Lévy process, which can be analyzed similarly (see, e.g., Cohen and Solan (2013) and Kaspi and Mandelbaum (1995)).

of interesting applications in the literature. For example, the standard diffusion process $dx = \mu_j(x)dt + \sigma_j(x)d\mathbb{Z}_j(t)$ is a special case without any jump in the process. And 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. Depending on the applications, there are many different interpretations of the state x. Observe that we allow for a general process for x and that a priori the martingale assumption is not made. This includes belief updating as a special case (the agent sees output and updates beliefs x using Bayes rule)[4] but also human capital accumulation where output changes over time (human capital x is accumulated stochastically and determines the realized output).

In this two-armed bandit problem, the stochastic process $\{x_t\}$ can be constructed on the product space $\{\Omega, \mathcal{F}\} = \{\Omega^1, \mathcal{F}^1\} \times \{\Omega^2, \mathcal{F}^2\}$ with filtration $\mathcal{F}_t = \mathcal{F}^1_{T_1(t)} \vee \mathcal{F}^2_{T_2(t)}$. Given $x_0 = x$, the agent is choosing an allocation rule $a_t \in \{1, 2\}$ adapted to filtration $\{\mathcal{F}_t\}_{t\geq 0}$ to solve the following optimal control problem:

$$v(x) = \sup_{a_t} \left\{ \mathbb{E} \int_{t=0}^{\infty} e^{-rt} f_{a_t}(x_t) dt \right\}$$
$$\text{s.t. } dx_t = \mu_{a_t}(x_t)dt + \sigma_{a_t}(x_t)d\mathbb{Z}_{a_t}(t) + \int_{\mathbb{R}-\{0\}} G_{a_t}(x(t-), y)\mathbb{N}_{a_t}(dt, dy).$$

To solve the problem, two technical assumptions are required on the functions of $f_j(x), \mu_j(x), \sigma_j(x)$ and $G_j(x, \cdot)$.

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

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 $|\frac{\partial G_j(x,y)}{\partial x}|$ for each y are all less than K.*

The above assumptions are standard in the literature. In particular, Assumption 2 guarantees Lipschitz continuity, which is crucial to guarantee that there exists a unique solution to stochastic differential equation (see, e.g., Applebaum (2004))

$$dx_j(t) = \mu_j(x(t-))dT_j(t) + \sigma_j(x(t-))d\mathbb{Z}_j(T_j(t)) + \int_{\mathbb{R}-\{0\}} G_j(x(t-), y)\mathbb{N}_j(dT_j(t), dy).$$

[4]It is well known that in the Bayesian learning case, the posterior belief follows a martingale stochastic process (see, e.g., Bolton and Harris (1999) and Keller, Rady, and Cripps (2005)).

Moreover, these conditions are usually satisfied in the applied literature.

Obviously, this common value experimentation problem is different from the standard optimal stopping problem. The reward function at stopping is not pre-specified, but endogenously determined. Moreover, we can no longer apply Gittins' logic in this situation because the state is perfectly correlated, and hence experimentation in one arm also changes the value of pulling the other arm.

A Motivating Example

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 follow 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 Bayes' rule:

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

where $\mathbb{Z}_{y,t}$ is a standard Brownian motion process. If we relabel the state x to be posterior belief p and arm 1 (2) to be firm H (L), the worker's dynamic job decision can be transformed into a common-value experimentation problem.

In such a market, each worker faces a common-value experimentation problem where the payoffs are determined in a competitive market. 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 (2010) investigate stationary competitive equilibria where each worker is making an optimal job-switching decision and where the market clears. The basic model discussed above hence is a building block to analyze more complicated equilibria.

3 Results

Following the literature on optimal stopping problem, we assume that the optimal strategy is an interval strategy. Potentially, we can partition the space $\mathcal{X}$ into (possibly) infinitely

many intervals such that arm j is chosen on disjoint intervals. This interval strategy allows us to derive the value function on each interval. Without loss of generality, 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. Then from Ito's lemma and Assumptions 1, 2, the value function $v(x)$ is at least $\mathcal{C}^2$ on each interval, which can be written as:[5]

$$rv(x) = f_1(x) + \mu_1(x)v'(x) + \frac{1}{2}\sigma_1^2(x)v''(x) + \int_{\mathbb{R}-\{0\}} [v(x + G_1(x, y)) - v(x)]\nu_1(dy).$$

for $x \in (x^\star, \bar{x})$; and

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

for $x \in (\underline{x}, x^\star)$.

Although we can explicitly derive the value function on each interval, we still need boundary conditions to pin down the cutoffs determining the intervals. Consider $x^\star$ to be any cutoff in the interior of $\mathcal{X}$. The traditional way to solve this continuous time bandit problem is to apply the "Whittle reduction" technique (Whittle (1980)) and transform the problem to a standard optimal stopping problem (see, i.e., Karatzas (1984) and Kaspi and Mandelbaum (1995)). Unfortunately, the "Whittle reduction" technique is valid only when the arms are independent and hence cannot be applied to our common value experimentation problem. However, we are able to show that the well-known properties of value matching and smooth pasting still hold at the cutoff $x^\star$. Moreover, we derive a new second derivative condition based on the Bellman's principle of optimality.

3.1 Value Matching and Smooth Pasting

Value matching and smooth pasting conditions are standard boundary conditions in the optimal stopping literature, where the reward function is exogenously given. It is natural to conjecture that these two conditions still hold even when the reward function is endogenously determined by the equilibrium value function.

Value matching condition implies that the value function $v(x)$ is continuous at $x^\star$. Since

[5]For the derivation of the value function, see Applebaum (2004) and Cohen and Solan (2013).

$v(\cdot)$ is not well defined at $x^\star$, this condition actually means

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

Suppose on the contrary this condition is violated and $v(x^\star+) > v(x^\star-)$, then switching arms at $x^\star$ cannot be optimal. In particular, there exists sufficiently small ϵ such that choosing arm 1 at the interval $(x^\star - \epsilon, x^\star)$ can lead to higher payoffs.

Smooth pasting condition is another standard condition in the optimal stopping models. In the context of our model, the condition implies that the first derivative of the value function is continuous at $x^\star$: $v'(x^\star+) = v'(x^\star-)$. 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 switching is postponed at the 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, instead of switching at $x^\star$, the agent switches arms only when the state reaches $x^\star + \epsilon$. Such a deviation may take arbitrarily long time or even be permanent in the sense that switching never happens. At $x^\star$, the induced payoff from this deviation is lower than the equilibrium payoff. As ϵ becomes small, this inequality transforms into an inequality of the first derivatives. Likewise, we consider another deviation from the candidate equilibrium where the agent switches at $x^\star - \epsilon$ instead of at $x^\star$. This again induces another inequality, which in the limit is the opposite of the first inequality, therefore implying equality of the first derivatives at the optimal cutoff $x^\star$.

3.2 Bellman's Principle of Optimality

We now establish that Bellman's principle of optimality imposes the second derivative condition, an additional constraint on the equilibrium allocation. 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 $\mathcal{X}$. On the contrary, 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 arbitrarily small. The deviation in time space is similar to the one-shot deviation in discrete time. Instead, the deviation in state space could be a permanent deviation.

Theorem 1 *(Second Derivative Condition) 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 that $v''(x)$ is continuous at $x^\star$ (i.e., $v''(x^\star+) = v''(x^\star-)$) 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. Then the optimality of choosing arm 1 for $x \in (x^\star, \bar{x})$ immediately 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 $v''(x^\star+) \triangleq \lim_{x \searrow x^\star} v''(x)$.

At $x^\star$, we have $v(x^\star+) = v(x^\star-)$, from the value matching condition. This implies that:

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

From the smooth pasting condition, $v'(x^\star+) = v'(x^\star-)$ and hence we should have: $v''(x^\star+) \leq v''(x^\star-)$. 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: $v''(x^\star+) \geq v''(x^\star-)$. Therefore, it must be the case that $v''(x^\star+) = v''(x^\star-)$. ■

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 values of experimentation ensures that at the optimal cutoff, no gains from switching exist.[6]

The key assumption of Theorem 1 is that both arms contain a non-trivial diffusion process. If this condition is violated and there are only discrete jump changes, it is quite easy to see that this condition on the second derivative does not hold any longer. In particular, if $\sigma_1 = \sigma_2 = 0$, then Belman's principle of optimality leads to the same condition as the smooth pasting condition: $v'(x^\star+) = v'(x^\star-)$.[7]

As we mentioned in the introduction, Wirl (2008) derived the same result as Theorem 1 in a common-value two-armed bandit problem with *only* diffusion processes. 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 value function and to derive an explicit formula of the first derivative of the value function at the cutoff. Using this first derivative formula, Wirl (2008) writes down expressions of the second derivative of the value function and shows algebraically that $v''(x^\star+) = v''(x^\star-)$. Our result establishes that the second derivative condition holds for the more general jump-diffusion process as long as the diffusion components are non-trivial. Furthermore, our proof is not constructive and does not hinge on the restrictive assumption $\sigma_1(x) = \sigma_2(x)$. Instead we use the one-shot deviation principle. This enables us to investigate problems with more interesting stochastic processes[8] as well as environments with strategic and market interaction.

4 Applications

In this section we illustrate how to use the three boundary conditions to solve equilibrium 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 extend the model to a two-player strategic interaction problem.[9] Finally, we analyze a strategic

[6] We can more generally consider extensions with endogenous choices by the agents, for example, where the agent chooses effort e to change f_j, μ_j and σ_j at cost $c(e)$. It is straightforward to show that the second derivative condition still holds in that situation.

[7] 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 $x^\star$.

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

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

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.

Notice that the value matching, smooth pasting and second derivative conditions are all necessary conditions satisfied at the cutoff. The general procedure is to first use these conditions to derive a candidate solution to the optimal control problem, and then verify the candidate solution indeed is optimal. In each of these settings, the second derivative condition plays a crucial role to pin down the equilibrium.

4.1 Linear Payoffs

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 $dx = \mu_y dt + \sigma_y d\mathbb{Z}_y$ 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.

To simplify the notations, we will 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. Moreover, the "+" and "–" signs will be omitted if no confusion results.

If it is optimal to pull arm 1, the value function satisfies the following 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, $k_{12} = 0$ 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 $k_{22} \geq 0$. With the help of the value matching, smooth pasting and second derivative conditions, we are able to solve the three unknowns k_{12}, k_{21} and $x^\star$ simultaneously and establish the following 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. ■

The common value experimentation problem can be potentially very complicated because there may exist multiple cutoffs. The above theorem shows that this cannot be the case if $(a_i - a_j)(\mu_i\sigma_j^2 - \mu_j\sigma_i^2) \geq 0$. An immediate implication is that for pure Bayesian learning illustrated in the motivating example, i.e., in the absence of a drift term in the Brownian motion ($\mu_i = \mu_j = 0$), 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 learning 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.[10]

Finally, $x^\star$ solved by the three boundary conditions is just a candidate cutoff. The following theorem explicitly verifies that switching at $x^\star$ indeed is the optimal solution to the two-armed bandit problem. The theorem focuses on the case that $a_1 > a_2$. The opposite case that $a_1 < a_2$ can be proved similarly.

Theorem 3 *Suppose the parameter values satisfy: $a_1 > a_2$ and $\mu_1\sigma_2^2 \geq \mu_2\sigma_1^2$. 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. ■

Figure 1 illustrates the equilibrium value function when $a_1 = 1$, $a_2 = b_1 = b_2 = \mu_1 = \mu_2 = 0$, $\sigma_1 = 2$, $\sigma_2 = 1$. In this case, the unique cutoff is negative: it is optimal to pull arm 1 even when the instantaneous payoff is smaller. This is because the learning rate is higher in arm 1 ($\sigma_1 > \sigma_2$) and hence, the agent is facing the tradeoff between experimentation and exploitation. When x is not very negative, it is optimal to sacrifice instantaneous payoff for future payoff.[11]

4.2 Strategic Interaction

Now we consider strategic interaction in the linear payoff model. There are symmetric two

[10] 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.

[11] It is straightforward to verify that the opposite is true when $\sigma_1 > \sigma_2$.

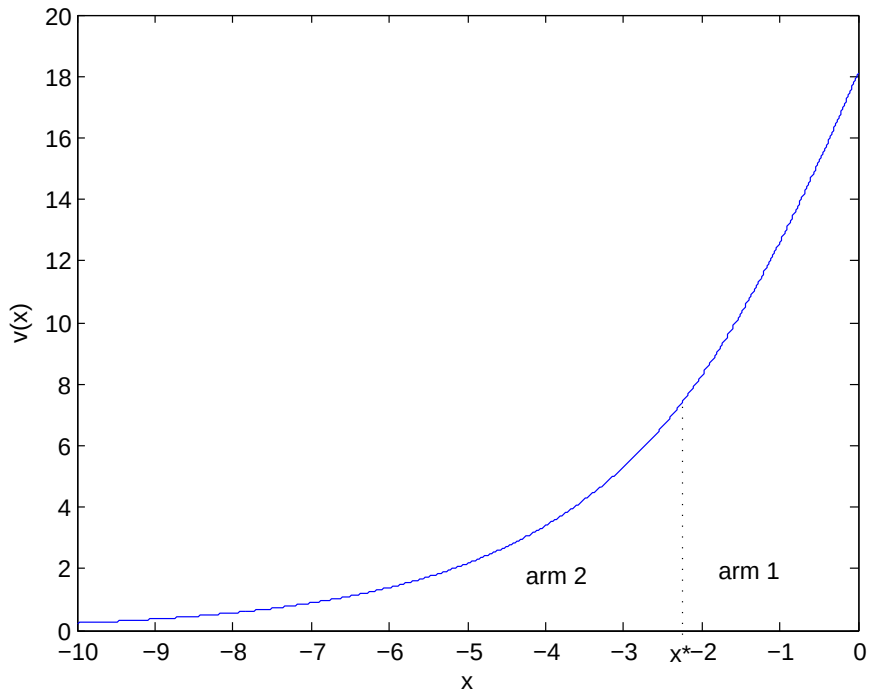

Figure 1: $v(x)$ in the single agent's decision problem.

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:

$$dx = (\theta_1 + \theta_2)\mu_1 dt + (2 - \theta_1 - \theta_2)\mu_2 dt + \sqrt{(\theta_1 + \theta_2)}\sigma_1 d\mathbb{Z}_1(t) + \sqrt{(2 - \theta_1 - \theta_2)}\sigma_2 d\mathbb{Z}_2(t),$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ 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:

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

We say $\{\theta_1(x), \theta_2(x)\}$ constitute a (Markov) equilibrium if $\theta_i(x)$ maximizes $V_i(x)$ given $\theta_j(x)$.

As shown by Bolton and Harris (1999), three possibilities can happen in a symmetric equilibrium: both players pull arm 1 only (value is denoted by V), both players pull arm 2 only (value is denoted by U) and each player randomizes between arm 1 and arm 2 (value is denoted by W). 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:

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

To fully characterize the symmetric equilibrium, we normalize the instantaneous payoff of arm 2 to be zero and $f_1(x) = ax + b$ with $a > 0$. 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. By straightforward calculations, the equilibrium cutoffs can be characterized by the following theorem:

Theorem 4 *Assume $f_1(x) = ax + b$ (with $a > 0$), $f_2(x) = 0$, $\mu_1 = \mu_2 = 0$ 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 $x \in (x_2^\star, x_1^\star)$, both players randomize between arm 1 and arm 2. $x_1^\star > x_2^\star$ and $x_2^\star$ solve the following system of equations:*

$$\frac{a}{r} + \frac{\sigma_1}{\sqrt{r}}\frac{2(ax_1+b)}{\sigma_1^2-\sigma_2^2} + \frac{ax_1^2}{\sigma_1^2-\sigma_2^2} + \frac{2bx_1}{\sigma_1^2-\sigma_2^2} = -\frac{\sigma_2}{\sqrt{r}}\frac{2(ax_2+b)}{\sigma_1^2-\sigma_2^2} + \frac{ax_2^2}{\sigma_1^2-\sigma_2^2} + \frac{2bx_2}{\sigma_1^2-\sigma_2^2},$$

and

$$\frac{ax_1+b}{r} - \frac{\sigma_1^2}{r}\frac{2(ax_1+b)}{\sigma_1^2-\sigma_2^2} + \frac{ax_1^3}{3(\sigma_1^2-\sigma_2^2)} + \frac{bx_1^2}{\sigma_1^2-\sigma_2^2}$$
$$= C(x_1 - x_2) - \frac{\sigma_2^2}{r}\frac{2(ax_2+b)}{\sigma_1^2-\sigma_2^2} + \frac{ax_2^3}{3(\sigma_1^2-\sigma_2^2)} + \frac{bx_2^2}{\sigma_1^2-\sigma_2^2},$$

where

$$C = -\frac{\sigma_2}{\sqrt{r}} \frac{2(ax_2 + b)}{\sigma_1^2 - \sigma_2^2} + \frac{ax_2^2}{\sigma_1^2 - \sigma_2^2} + \frac{2bx_2}{\sigma_1^2 - \sigma_2^2}.$$

Theorem 4 implies that with the help of the second derivative condition, we can extend the analysis of experimentation 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 in the previous section, if the extension is based on an exponential bandit or a Poisson bandit, then the second derivative condition is not needed and typically does not hold at the optimal cutoff.

Figure 2 illustrates the equilibrium value function when $a_1 = 1$, $a_2 = b_1 = b_2 = \mu_1 = \mu_2 = 0$, $\sigma_1 = 2$, $\sigma_2 = 1$. As in Figure 1, the cutoffs are negative since $\sigma_1 > \sigma_2$. However, compared with the cutoff in the single-agent decision problem, the equilibrium cutoffs in Figure 2 become larger. When facing the tradeoff between experimentation and exploitation, the players sacrifice the instantaneous payoff for future payoff. In the strategic interaction environment, each player can free ride and update x from the other player's actions. As in Bolton and Harris (1999), this free riding opportunity undermines each player's incentive to conduct experimentation. Hence, in equilibrium the players switch to arm 2 earlier.

4.3 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 sequential optimality condition that equates the value of learning 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 ibuprofen 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

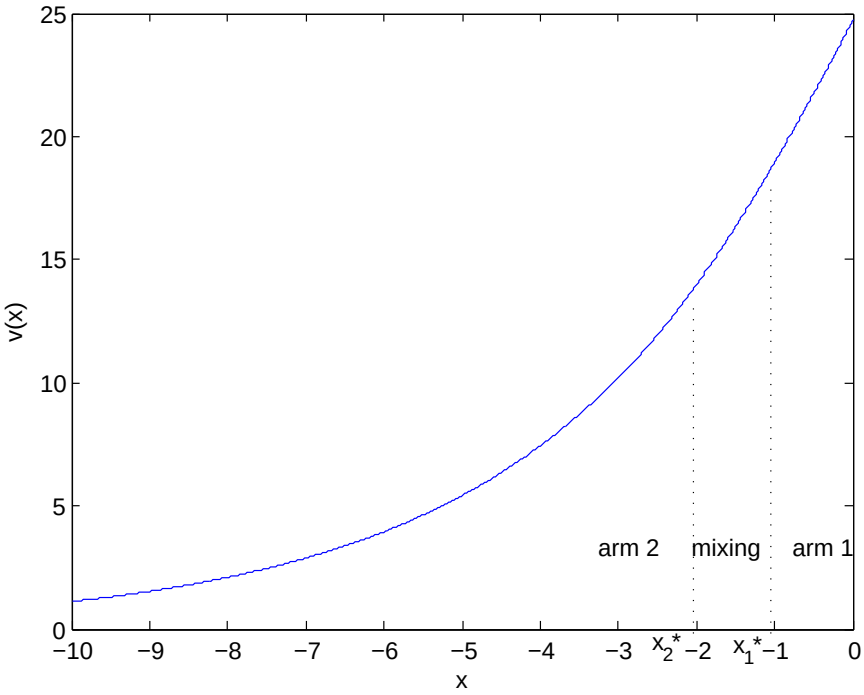

Figure 2: $v(x)$ in the strategic interaction environment.

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:

$$du_{ji}(t) = \xi_{ji}dt + \sigma_j d\mathbb{Z}_j(t),$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ 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 $dx = x(1-x)s_i d\bar{\mathbb{Z}}_i$ where $s_i = \frac{a_i}{\sigma_i}$ and $\bar{\mathbb{Z}}_i$ 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 $v(x)$:

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

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

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 deriva-

tive 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, seller i's value is $K_i(x)$. From Ito's Lemma, it is immediately seen that:

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

and

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

where

$$\Sigma_i(x) = s_i^2 x^2(1 - x)^2, \quad \text{and} \quad \Sigma_j(x) = s_j^2 x^2(1 - x)^2.$$

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

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

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

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

[13]In Bergemann and Välimäki (1996), all Markov perfect equilibria are efficient. The notion of cautious

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.

5 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

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

where both $\mathbb{Z}_{jx}(t)$ and $\mathbb{Z}_{jy}(t)$ 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.*

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 $\mathbb{Z}_{jx}, \mathbb{Z}_{jy}$ are independent. If they were not, there is a covariance term $\text{cov}(d\mathbb{Z}_{jx}, d\mathbb{Z}_{jy}) = g_j(x, y)dt$ 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 independent Brownian motion processes.

As before, the value function can be written as:

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

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:

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

[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.

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

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

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

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

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 learning is the weighted sum of the learning 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]

$$\sigma_1^2(x)V_{11}(x,y) + \omega_1^2(y)V_{22}(x,y) = \sigma_1^2(x)U_{11}(x,y) + \omega_1^2(y)U_{22}(x,y) \tag{1}$$

and

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

Proof. In Appendix. ■

6 Concluding Remarks

Learning about the common value by means of different choices is common in many economic environments: information about a worker's ability is revealed whether she takes one job or another; a patient and his doctor's belief about his illness evolves whether he takes one drug or another; firms learn about consumers' preferences whether they buy one good or another. In this paper, we have proposed a general setup that allows us to analyze common value experimentation for a general stochastic framework. We have shown how this setup can be applied as a decision problem, in a strategic setting or in a market economy.

Experimentation problems similar to ours are used as a building block to investigate many important economic issues. A non-exhaustive list of related papers includes Bolton and Harris (1999), Bonatti (2011), Daley and Green (2012), Faingold and Sannikov (2011), Hörner and Samuelson (2013), and Strulovici (2010). Almost all of the existing papers use a one-armed bandit framework: agents stop learning after switching to the safe arm. Our setup and the second derivative condition enables us to investigate these same environments, but where agents continue to learn about the common underlying state variable, even if they switch action.

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

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

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

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

and

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

The subtraction of the above two equations leads to

$$\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} r(V(x_2) - V(x_1)) + \frac{\sigma_2^2}{\sigma_1^2}(f_1(x_2) - f_1(x_1)) + (\mu_1 - \mu_2 - \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2}\mu_1)(V'(x_2) - V'(x_1)) = f_2(x_2) - f_2(x_1). \quad (3)$$

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 3 implies

$$\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} r(V(x_2) - V(x_1)) + \frac{\sigma_2^2}{\sigma_1^2}(f_1(x_2) - f_1(x_1)) \leq f_2(x_2) - f_2(x_1). \tag{4}$$

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 4 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 4 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:

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

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 is indeed 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

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

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, $v(x_t, y_t) = V(x_t, y_t)$ with probability $1 - o(t)$. This implies that at $(x - \epsilon/2, y)$:

$$f_2 + \mu_2 V_2 + \nu_2 V_2 + \frac{1}{2}\sigma_2^2 V_{11} + \frac{1}{2}\omega_2^2 V_{22} - \left[f_1 + \mu_1 V_1 + \nu_1 V_2 + \frac{1}{2}\sigma_1^2 V_{11} + \frac{1}{2}\omega_1^2 V_{22}\right] \leq 0. \quad (5)$$

Take ϵ goes to zero and inequality (5) 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:

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

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

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

Sequential 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 $(x - \epsilon/2, y)$:

$$f_2 + \mu_2 V_1 + \nu_2 V_2 + \frac{1}{2}\sigma_2^2 V_{11} + \frac{1}{2}\omega_2^2 V_{22} - \left[f_2 + \mu_2 \hat{U}_1 + \nu_2 \hat{U}_2 + \frac{1}{2}\sigma_2^2 \hat{U}_{11} + \frac{1}{2}\omega_2^2 \hat{U}_{22}\right] \geq 0. \quad (6)$$

Meanwhile, as ϵ goes to zero, $\hat{U}$ converges to U. Therefore, inequality (6) 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:

$$\sigma_2^2(x) V_{11}(x, y) + \omega_2^2(y) V_{22}(x, y) = \sigma_2^2(x) U_{11}(x, y) + \omega_2^2(y) U_{22}(x, y). \quad (7)$$

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. (2011): "Menu Pricing and Learning," *American Economic Journal: Microeconomics*, 3(3), 124–163.

Cohen, A., and E. Solan (2013): "Bandit Problems with Lévy Payoff," *Mathematics of Operations Research*, 38(1), 92–107.

Cripps, M., J. Ely, G. Mailath, and L. Samuelson (2008): "Common Learning," *Econometrica*, 76(4), 909–933.

Daley, B., and B. Green (2012): "Waiting for News in the Dynamic Market for Lemons," *Econometrica*, 80(4), 1433–1504.

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 (2011): "Reputation in Continuous-Time Games," *Econometrica*, 79(3), 773–876.

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 (2013): "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 (1999): "Optimal Experimentation in a Changing Environment," *Review of Economic Studies*, 66(3), 475–507.

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

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.

Strulovici, B., and M. Szydlowski (2014): "On the Smoothness of Value Functions and the Existence of Optimal Strategies," mimeo.

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.