



Notes

Common value experimentation [☆]

Jan Eeckhout ^{a,b,*}, Xi Weng ^c

^a University College London, United Kingdom

^b Barcelona GSE-UPF-ICREA, Spain

^c Department of Applied Economics, Guanghua School of Management, Peking University, China

Received 30 May 2014; final version received 25 September 2015; accepted 5 October 2015

Available online 23 October 2015

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 or when unemployed is informative about her productivity in another job. We analyze a general setup of experimentation with common values, and show that 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 whenever the agent switches action. This condition holds generally whenever the stochastic process has continuous increments. The main appeal of our approach is its applicability, which is demonstrated with two applications featuring common value experimentation: strategic pricing, and job search with switching costs.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

JEL classification: D83; C02; C61

Keywords: Common values; Experimentation; Jump-diffusion process

[☆] 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 benefited 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 339186, and MINECO Grant ECO2012-36200. Weng acknowledges support from the National Natural Science Foundation of China (Grant No. 71303014) and Guanghua Leadership Institute (Grant No. 12-02).

* Corresponding author.

E-mail addresses: j.eeckhout@ucl.ac.uk (J. Eeckhout), wengxi125@gsm.pku.edu.cn (X. Weng).

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. Contrast 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. 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 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 common value experimentation, 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, i.e., equilibrium actions must be robust to deviations. 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 switching point, this gives rise to an inequality condition on the first order derivative of the value function. Since the inequality holds on both

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 et al. (2008). There, given a stream of signals, two agents want to learn the value of an unknown parameter, so 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.

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. 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, the differential equation for the value function depends on the second derivative, sometimes referred to as the marginal value of information. 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. 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. Equating the second derivative ensures that no such gains from deviation exist.

There is an alternative 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). That method long precedes ours and 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 more involved than ours. In addition, it is not always applicable (see the example in Section 4.2). 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.

The main appeal of our approach is its applicability, which is also the main contribution of this paper. Not only are such environments with common value experimentation prevalent in many economic contexts, the implementation of the one shot deviation principle is straightforward. We consider two applications. First, we consider a setting of strategic pricing with continued learning, and show how the second derivative condition affects the equilibrium allocation and its efficiency. Second, we apply common value experimentation to a setting with unemployment and show the generality of the method even in the presence of search frictions (or switching costs).

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 $r > 0$.

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:

$$dx_j(t) = \mu_j(x(t-))dT_j(t) + \sigma_j(x(t-))d\mathbb{Z}_j(T_j(t)) + \int_{\mathbb{R} \setminus \{0\}} G_j(x(t-), y)\mathbb{N}_j(dT_j(t), dy). \quad (1)$$

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 Wiener process and \mathbb{N}_j is a Poisson random measure that is independent of \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 $i = 1, \dots, m_j$. In this case,

$$\nu_j = \sum_{i=1}^{m_j} \lambda_i \delta_{h_i}, \quad (2)$$

where δ_h is a Dirac mass concentrated at h .² $G_j(x, y)$ 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 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)³ 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 $\mathcal{F}_t = \mathcal{F}_{T_1(t)}^1 \vee \mathcal{F}_{T_2(t)}^2$. 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:

$$\begin{aligned} v(x) = \sup_{a_t} \left\{ \mathbb{E} \int_{t=0}^{\infty} e^{-rt} f_{a_t}(x_t) dt \right\} \\ \text{s.t. } dx_t = \mu_{a_t}(x_t)dt + \sigma_{a_t}(x_t)d\mathbb{Z}_{a_t}(t) + \int_{\mathbb{R}-\{0\}} G_{a_t}(x(t-), y) \mathbb{N}_{a_t}(dt, dy). \end{aligned} \quad (3)$$

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 .

² 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).

³ 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 et al., 2005).

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](#))

$$dx_j(t) = \mu_j(x(t-))dT_j(t) + \sigma_j(x(t-))d\mathbb{Z}_j(T_j(t)) + \int_{\mathbb{R}-\{0\}} G_j(x(t-), y)\mathbb{N}_j(dT_j(t), dy). \quad (4)$$

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.

3. The main result

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 $\bar{x} > x > x^*$ chooses arm 1 and an agent with $\underline{x} < x < x^*$ 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 as⁴:

$$rv(x) = f_1(x) + \mu_1(x)v'(x) + \frac{1}{2}\sigma_1^2(x)v''(x) + \int_{\mathbb{R}-\{0\}} [v(x + G_1(x, y)) - v(x)]v_1(dy), \quad (5)$$

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

$$rv(x) = f_2(x) + \mu_2(x)v'(x) + \frac{1}{2}\sigma_2^2(x)v''(x) + \int_{\mathbb{R}-\{0\}} [v(x + G_2(x, y)) - v(x)]v_2(dy), \quad (6)$$

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

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^* 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^* . Moreover, we derive a new second derivative condition based on the Bellman's principle of optimality.

⁴ For the derivation of the value function, see [Applebaum \(2004\)](#) and [Cohen and Solan \(2013\)](#).

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

$$v(x^*+) \triangleq \lim_{x \searrow x^*} v(x) = v(x^*-) \triangleq \lim_{x \nearrow x^*} v(x). \quad (7)$$

Suppose on the contrary this condition is violated and $v(x^*+) > v(x^*-)$, then switching arms at x^* cannot be optimal. In particular, there exists sufficiently small ϵ such that choosing arm 1 at the interval $(x^* - \epsilon, x^*)$ 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^* : $v'(x^*+) = v'(x^*-)$. 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^* . 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, this implies a restriction on the first derivative.

In particular, instead of switching at x^* , the agent switches arms only when the state reaches $x^* + \epsilon$. Such a deviation may take arbitrarily long time or even be permanent in the sense that switching never happens. At x^* , 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 $x^* - \epsilon$ instead of at x^* . 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^* .

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^* is that $v''(x)$ is continuous at x^* (i.e., $v''(x^*+) = v''(x^*-)$) for any possible cutoff x^* .*

Proof. Without loss of generality, we assume that an agent with $\bar{x} > x > x^*$ chooses arm 1 and an agent with $\underline{x} < x < x^*$ chooses arm 2. Then the optimality of choosing arm 1 for $x \in (x^*, \bar{x})$ immediately implies that:

$$f_2(x) + \mu_2(x)v'(x) + \frac{1}{2}\sigma_2^2(x)v''(x) + \int_{\mathbb{R}-\{0\}} [v(x + G_2(x, y)) - v(x)]v_2(dy) \leq rv(x). \quad (8)$$

For $x \rightarrow x^*$, the above inequality implies:

$$f_2(x^*) + \mu_2(x^*)v'(x^*+) + \frac{1}{2}\sigma_2^2(x^*)v''(x^*+) + \int_{\mathbb{R}-\{0\}} [v(x^* + G_2(x^*, y)) - v(x^*+)]v_2(dy) \leq rv(x^*+) \quad (9)$$

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

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

$$\begin{aligned} f_2(x^*) + \mu_2(x^*)v'(x^*+) + \frac{1}{2}\sigma_2^2(x^*)v''(x^*+) \\ + \int_{\mathbb{R}-\{0\}} [v(x^* + G_2(x^*, y)) - v(x^*+)]v_2(dy) \\ \leq f_2(x^*) + \mu_2(x^*)v'(x^*-) + \frac{1}{2}\sigma_2^2(x^*)v''(x^*-) \\ + \int_{\mathbb{R}-\{0\}} [v(x^* + G_2(x^*, y)) - v(x^*-)]v_2(dy). \end{aligned} \quad (10)$$

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

The second derivative measures the change in value of having the option to switch arms and can be interpreted as a measure of the marginal value of information (as $\frac{1}{2}\sigma_j^2(x)v''$ is often called as “value of information”). A way to read the second derivative condition then is that it requires the marginal value of information 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.⁵

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

⁵ 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.

condition on the second derivative does not hold any longer. In particular, if $\sigma_1 = \sigma_2 = 0$, then Bellman's principle of optimality leads to the same condition as the smooth pasting condition: $v'(x^*+) = v'(x^*-)$.⁶

There is a close relation between our result in Theorem 1 and Wirl (2008). He derives the second derivative condition in the context of a common-value two-armed bandit problem with *only* diffusion processes, i.e. a pure Brownian motion in the absence of discrete increments. But his Brownian motion setting is very special as it requires that the variance term on the noise is *identical* on both arms: $\sigma_1(x) = \sigma_2(x)$ for all x . As shown by the Online Appendix, this assumption enables him to cancel the second derivative term of the value function, because by construction the value of information is the same on both arms, 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 $v''(x^*+) = v''(x^*-)$. This logic clearly does not hold under a minor difference in the Brownian noise terms. 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 show when the result does and when it does not hold (in the absence of a continuous increment component). Second, 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 that is generally applicable. This enables us to investigate problems with more interesting stochastic processes⁷ as well as environments with strategic and market interaction.

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.⁸

4. Applications

In this section, we apply our framework to two different settings, one with strategic interaction, and one with unemployment from search frictions, and hence switching costs.

4.1. Strategic pricing

Consider two firms that sell to $N \geq 1$ consumers whose preferences are identical but unknown. A representative consumer's valuation is determined by a common factor across the different sellers' products. 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. This setup is similar to that in Bergemann and Välimäki (1996) and Bergemann and Välimäki (2000) with

⁶ 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. The logic of proof implies that V is locally more convex than U : $V''(x^*) \geq U''(x^*)$, which is satisfied at x^* .

⁷ For example, in Eeckhout and Weng (2010), the signal-to-noisy ratios generically differ across firm types.

⁸ 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.

independent arms, except that here learning is about common values and therefore the arms are perfectly correlated.

Model set up. The market consists of $N \geq 1$ buyer indexed by $i = 1, \dots, N$ 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 values of both product are initially unknown to all players in the market, and depend on the type of the buyers. We assume that all buyers have the same type, which is either high or low, with a common prior that the type is high x_0 . If the type is high, a representative buyer receives expected value ξ_{1H} from consuming good 1 and ξ_{2H} from consuming good 2. If the type is low, a representative buyer receives expected value ξ_{1L} from consuming good 1 and ξ_{2L} from consuming good 2. We assume that a low type buyer prefers good 1, while a high type buyer prefers good 2: $\xi_{1L} > \xi_{2L}$ and $\xi_{2H} > \xi_{1H}$.

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:

$$du_{ji}(t) = \xi_{ji}dt + \sigma_j d\mathbb{Z}_j(t), \quad (11)$$

where \mathbb{Z}_1 and \mathbb{Z}_2 are independent standard Wiener processes.

Denote by $x \in [0, 1]$ the common belief that the type is high. Then each buyer's expected flow 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 $b_1 > b_2$. If n_1 buyers purchase product 1 and $n_2 = N - n_1$ buyers purchase product 2, previous results (see, e.g., Bolton and Harris, 1999, Bergemann and Välimäki, 2000) show that x is updated following a diffusion process with zero drift and instantaneous variance $n_1 \Sigma_1(x) + n_2 \Sigma_2(x)$, where $\Sigma_j(x) = x^2(1-x)^2(\frac{a_j}{\sigma_j})^2$. Afterwards, we will let $s_j \triangleq \frac{a_j}{\sigma_j}$.

Socially efficient allocation. The planner is facing a common value two-armed bandit problem. Denote the total social surplus function to be $v(x)$:

$$v(x) = \sup_{n_{1t}} \left\{ \mathbb{E} \int_{t=0}^{\infty} e^{-rt} [n_{1t} f_1(x_t) + (N - n_{1t}) f_2(x_t)] dt \right\} \quad (12)$$

$$\text{s.t. } dx_t = \sqrt{n_{1t}x_t(1-x_t)}s_1 d\bar{\mathbb{Z}}_1(t) + \sqrt{N - n_{1t}x_t(1-x_t)}s_2 d\bar{\mathbb{Z}}_2(t),$$

where $\bar{\mathbb{Z}}_1(t)$ and $\bar{\mathbb{Z}}_2(t)$ are independent innovation processes, and follow the standard Wiener process. As shown by Bolton and Harris (1999) and Bergemann and Välimäki (2000), the planner's optimal solution is a corner solution: n_{1t}^e is either N or 0 . $W_1(x)$ ($W_2(x)$) denotes the total social surplus if in a neighborhood of x , the planner optimally chooses product 1 (2). The value functions $W_1(x)$ and $W_2(x)$ are:

$$rW_1(x) = N(f_1(x) + \frac{1}{2}\Sigma_1(x)W_1''(x)), \quad rW_2(x) = N(f_2(x) + \frac{1}{2}\Sigma_2(x)W_2''(x)). \quad (13)$$

Since there is no drift term in the updating of x , Theorem 5 in the Online Appendix 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 $x < x^e$.

Therefore, for $x < x^e$,⁹

$$W_1(x) = \frac{Nf_1(x)}{r} + C_1 x^{\frac{\gamma_1+1}{2}} (1-x)^{-\frac{\gamma_1-1}{2}} \quad (14)$$

and for $x > x^e$,

$$W_2(x) = \frac{Nf_2(x)}{r} + C_2 x^{-\frac{\gamma_2-1}{2}} (1-x)^{\frac{\gamma_2+1}{2}}, \quad (15)$$

where $\gamma_j = \sqrt{1 + \frac{8r}{Ns_j^2}}$.

At the optimal cutoff x^e , the value functions W_1 and W_2 satisfy value matching, smooth pasting and second derivative conditions. Thus, we can explicitly derive an equation about $x = x^e$:

$$\begin{aligned} \Psi(x) \triangleq & \left[s_1^2 \left(\frac{\gamma_1+1}{2} - x \right) + s_2^2 \left(\frac{\gamma_2-1}{2} + x \right) \right] \frac{b_1 - b_2}{r} \\ & + \left[s_1^2 \frac{\gamma_1-1}{2} + s_2^2 \frac{\gamma_2+1}{2} \right] \frac{(a_1 - a_2)x}{r} = 0, \end{aligned} \quad (16)$$

which implies:

$$x^e = \frac{(b_1 - b_2)(s_1^2 \frac{\gamma_1+1}{2} + s_2^2 \frac{\gamma_2-1}{2})}{(a_2 - a_1) \left[s_1^2 \frac{\gamma_1-1}{2} + s_2^2 \frac{\gamma_2+1}{2} \right] + (b_1 - b_2)(s_1^2 - s_2^2)}. \quad (17)$$

Markov perfect equilibrium. We consider Markov perfect equilibria. At each instant of time t , the sellers simultaneously set prices. After observing the price vector, the buyers choose which product to buy. The natural state variable is state x for both sellers. The pricing strategy for seller $j = 1, 2$ is $p_j : [0, 1] \rightarrow \mathbb{R}$. 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 buyer i is $\alpha_i : [0, 1] \times \mathbb{R} \times \mathbb{R} \rightarrow \{1, 2\}$. A symmetric Markov perfect equilibrium is a triple (p_1, p_2, α) such that *i*) given p_1, p_2 and all other buyers' strategies α , α maximizes buyer i 's expected intertemporal value; *ii*) given α and p_{-j}, p_j maximizes seller j 's expected intertemporal profit.

Suppose n buyers purchase product 1, then the value of each seller satisfies the following equations:

$$rV_1(x) = np_1(x) + \frac{1}{2} (n\Sigma_1(x) + (N-n)\Sigma_2(x)) V_1''(x) \quad (18)$$

$$rV_2(x) = (N-n)p_2(x) + \frac{1}{2} (n\Sigma_1(x) + (N-n)\Sigma_2(x)) V_2''(x). \quad (19)$$

Each V_j can be decomposed into the flow revenue resulting from sales, $n_j p_j$, and the expected change in the competitive position generated by the sales of the new product, captured by $\frac{1}{2} (n\Sigma_1(x) + (N-n)\Sigma_2(x)) V_j''(x)$.

Since all buyers have identical preferences and they have access to the same information, their equilibrium value is the same in a symmetric equilibrium. Denote $U(x)$ to be the equilibrium value of each buyer, and $U(x)$ satisfies equation

⁹ For $x < x^e$, the general solution is $\frac{Nf_1(x)}{r} + C_1 x^{\frac{\gamma_1+1}{2}} (1-x)^{-\frac{\gamma_1-1}{2}} + C_1' x^{-\frac{\gamma_1-1}{2}} (1-x)^{\frac{\gamma_1+1}{2}}$, but since $x = 0$ is included in the domain, C_1' has to be zero to ensure the boundedness of the value function.

$$rU(x) = \max \left\{ \begin{aligned} & f_1(x) - p_1(x) + \frac{1}{2} [(n+1)\Sigma_1(x) + (N-n-1)\Sigma_2(x)] U''(x) \\ & f_2(x) - p_2(x) + \frac{1}{2} [n\Sigma_1(x) + (N-n)\Sigma_2(x)] U''(x) \end{aligned} \right\}, \quad (20)$$

when n other buyers purchase product 1. Due to price competition, each buyer has to be indifferent between the alternatives, which implies that:

$$f_1(x) - p_1(x) + \frac{1}{2} \Sigma_1(x) U''(x) = f_2(x) - p_2(x) + \frac{1}{2} \Sigma_2(x) U''(x). \quad (21)$$

Equation (21) is not enough to pin down p_1 and p_2 at the same time. Hence, following Bergemann and Välimäki (1996) and Bergemann and Välimäki (2000), we investigate a Markov perfect equilibrium with cautious strategies.¹⁰ In particular, in a symmetric equilibrium, all buyers will buy from the same seller: either $n_j = 0$ or $n_j = N$. A cautious strategy means that if $n_j = 0$, seller j will charge a price $p_j(x)$ such that she is just indifferent between selling to one buyer at price $p_j(x)$ and not selling:

$$p_j(x) + \frac{1}{2} \Sigma_j(x) V_j''(x) = \frac{1}{2} \Sigma_{-j}(x) V_j''(x) \quad \text{or} \quad p_j(x) = \frac{1}{2} [\Sigma_{-j}(x) - \Sigma_j(x)] V_j''(x). \quad (22)$$

Once p_j is determined, the other firm's price can be easily derived from equation (21). We focus on cutoff strategies. Suppose each buyer purchases product 2 if $x > \underline{x}$ and purchases product 1 if $x < \underline{x}$. Then the sellers' optimality conditions imply that at cutoff \underline{x} , we have the equilibrium conditions:

$$V_j(\underline{x}-) = V_j(\underline{x}+), \quad V_j'(\underline{x}-) = V_j'(\underline{x}+), \quad V_j''(\underline{x}-) = V_j''(\underline{x}+) \quad \text{for seller } j = 1, 2. \quad (23)$$

Moreover, for each buyer, the value matching condition has to hold at \underline{x} :

$$U(\underline{x}-) = U(\underline{x}+). \quad (24)$$

We can characterize the equilibrium cutoff from the above boundary conditions and obtain the following theorem. The proofs of the following and all subsequent results can be found in the Appendix A.

Theorem 2. *The Markov perfect equilibrium with cautious strategies is socially efficient if and only if $N = 1$ or $s_1 = s_2$. Moreover, $\underline{x} > x^e$ ($\underline{x} < x^e$) when $s_1 > s_2$ ($s_1 < s_2$).*

There are two sources of inefficiency in our model. First, when one seller sells to the buyer, this generates information about the product of the other seller. The other seller can thus utilize this externality, and free ride on information acquisition. Second, similar to Bergemann and Välimäki (2000), there is another externality due to the fact that each individual buyer fails to take into account the effect of his own purchases on all other buyers. This makes experimentation relatively inexpensive for the seller.

The Markov perfect equilibrium with cautious strategies is socially efficient in two special cases. When $N = 1$, the second source of inefficiency does not exist, and cautious strategies are crucial to ensure that the equilibrium is efficient. Under cautious strategies, the free-riding seller

¹⁰ This requirement captures the logic behind trembling hand perfection in this infinite time horizon framework (see Bergemann and Välimäki, 1996).

is indifferent, and hence the first source of externality is fully internalized.¹¹ When $N \geq 2$, we generalize the results in Bergemann and Välimäki (2000) by showing that the Markov perfect equilibrium with cautious strategies is efficient if and only if the signal-to-noise ratios of the two products are the same ($s_1 = s_2$), and if the ratios are different, there will be excessive experimentation towards the product with a higher signal-to-noise ratio. These results are closely related to the second derivative condition. In particular, when a product has a higher signal-to-noise ratio, then the seller enjoys a higher value of information at the equilibrium cutoff. Meanwhile, the second externality implies that experimentation is relatively inexpensive. As a result, the seller will take advantage of this inexpensive experimentation by conducting excessive experimentation. On the other hand, if the signal-to-noise ratios are the same, the value of information is the same at the equilibrium cutoff due to the second derivative condition. And hence no seller has an incentive to conduct excessive experimentation. In Bergemann and Välimäki (2000), the signal-to-noise ratios satisfy $s_1 = 0 < s_2$. Thus, their inefficiency result can be viewed as a special case of Theorem 2.¹²

4.2. Unemployment, learning, and efficient directed search

In this section, we consider another application in which, similar to Papageorgiou (2014), a worker learns his fixed ability in the labor market. Our model is different in the following ways. First, in Papageorgiou (2014), a worker can choose to work in one of two occupations or become unemployed. In our model, there is only one occupation, and a worker simply chooses whether to work or not. Second, to capture common value experimentation, we allow for learning when unemployed. This assumption is rationalized by the practice of unemployment training programs. Finally, search is directed, whereas Papageorgiou (2014) studies random search model. This allows us to study the planner's optimal allocation problem, which in directed search coincides with the equilibrium allocation.

In our context, the planner is essentially solving a common value two-arm bandit problem.¹³ The problem, however, is different from the previous problems due to the existence of search frictions: an employed worker can become unemployed without any cost, but an unemployed worker has to search for a vacant firm. We nonetheless show that the second derivative condition still holds in our context, and we use this condition to obtain a characterization of the planner's optimal solution.

Population of firms and workers. Time is continuous: $t \in [0, +\infty)$. The economy is populated with a unit measure of workers and a sufficiently large measure of long-lived firms to ensure free

¹¹ 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 model, the notion of cautious equilibrium is important to guarantee efficiency. In other words, non-cautious Markov perfect equilibria might be inefficient.

¹² In related work, Kehoe and Pastorino (2011) and Kehoe and Pastorino (2013) propose a different, discrete time model to study experimentation with common value arms and apply it to strategic pricing, where they investigate a duopoly model with one consumer. In this framework, they apply constrained efficiency results similar to those in Bergemann and Välimäki (1996) in order to fully characterize the equilibrium allocations. In our model, in contrast, we investigate a duopoly with many consumers. Now the Markov perfect equilibrium can be inefficient due to the externality caused by the fact that each individual buyer fails to take into account the effect of his own purchases on all other buyers. Our continuous-time technique enables us to explicitly characterize the equilibrium cutoffs, and discuss when the Markov perfect equilibrium is efficient and when it is not.

¹³ In Papageorgiou (2014), a worker's optimal decision is characterized by solving two one-armed bandit problems.

entry. Both firms and workers are *ex ante* homogeneous. The worker ability $\xi \in \{H, L\}$ is not observable, both to firms and workers, i.e., the information is symmetric. Nonetheless, both hold a common belief about the worker type, denoted by $x \in [0, 1]$. Upon entry, a newly born worker is of type H with probability $x_0 \in (0, 1)$ and of type L with probability $1 - x_0$. Workers die with exogenous probability ρ . New workers are born at the same rate.

Workers and firms are risk-neutral and discount future payoffs at a rate $r > 0$. Utility is perfectly transferable. Output is produced in pairs of one worker and one firm. The quality of each match is purely determined by the ability of the worker. We assume that more able workers are more productive, $\mu_H \geq \mu_L$. At any moment, the planner can separate a match immediately without paying any cost.

Information. Because worker ability is not observable to both the worker and the firm, parties face an information extraction problem. They observe a noisy measure of productivity, denoted by X_t . Cumulative output is assumed to be a Brownian motion with drift μ_ξ and variance σ^2

$$X_t = \mu_\xi t + \sigma \mathbb{Z}_t \quad (25)$$

where \mathbb{Z}_t is a standard Wiener process. The realized performance is public information, so both parties commonly update their beliefs about the quality of the match according to Bayes' rule. Specifically, denote $x_t \equiv \Pr(\xi = H | X^t)$ as the posterior belief about the worker's quality with duration t where $X^t = \{X_\tau\}_{\tau=0}^t$ is the realized path. The standard result by [Liptser and Shiryaev \(2001\)](#) implies that

$$dx_t = x_t(1 - x_t)s_1 \frac{[dX_t - x_t\mu_H dt - (1 - x_t)\mu_L dt]}{\sigma}, \quad (26)$$

where $[dX_t - x_t\mu_H dt - (1 - x_t)\mu_L dt]/\sigma$ is the innovation process and follows the standard Wiener process, and the signal-to-noise ratio $s_1 = (\mu_H - \mu_L)/\sigma$ measures the informativeness of the learning.

An unemployed worker enjoys a flow payoff b from leisure. Assume $b \in (\mu_L, \mu_H)$, so that a high-quality match is always desirable, but a low-quality match is not. Moreover, an unemployed worker can also learn from some other source about his ability (i.e., unemployment training programs). For simplicity, we assume that when a worker is unemployed, there is another public signal, denoted by Y_t , revealing the ability of this worker. The signal is assumed to follow a Brownian motion with drift v_ξ and variance ζ^2

$$Y_t = v_\xi t + \zeta \tilde{\mathbb{Z}}_t \quad (27)$$

where $\tilde{\mathbb{Z}}_t$ is a standard Wiener process independent of \mathbb{Z}_t . Under signal Y_t , the posterior belief is updated according to equation

$$dx_t = x_t(1 - x_t)s_0 \frac{[dY_t - x_t v_H dt - (1 - x_t)v_L dt]}{\zeta}, \quad (28)$$

where $s_0 = (v_H - v_L)/\zeta$.

Search and matching technology. Since the expected output is determined by the worker's exogenously given ability, there is no social value from searching on the job. However, when the worker receives favorable information about his ability, it may be optimal to let an unemployed worker reenter into the market. At any instant, unemployed workers and firms search for new matches at different locations.

At each location, the workers and the vacancies meet (and match) according to a constant-returns-to-scale matching technology that can be described in terms of the tightness of the location θ (i.e., the vacancy-to-worker ratio at the location). Specifically, at any instant, a worker meets a vacancy at a rate $p(\theta)$ where $p: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a twice continuously differentiable, strictly increasing, and concave function such that $p(0) = 0$, and $\lim_{\theta \rightarrow 0} p'(\theta) = \infty$. The flow cost of maintaining a vacancy is $k > 0$, which implies that the social cost of maintaining a vacancy also constantly returns to scale and can be described in terms of the tightness θ . The per-worker social cost of maintaining a tightness θ is given by $k\theta$.

Formulation of the Planner's problem. As shown in the directed search models of [Menzio and Shi \(2011\)](#) and [Li and Weng \(2015\)](#), it is sufficient to consider the planner's problem in the current context. Due to directed search, the planner's efficient plan can be implemented by a decentralized Block recursive market equilibrium if firms and workers can sign bilaterally efficient contracts. The posterior belief about the worker's ability, $x \in [0, 1]$, is a sufficient statistic for past history and thus is also a natural state variable of any worker. As a result, the space of individual state can be defined as a set $\Xi = [0, 1] \times \{0, 1\}$, where $\eta \in \{0, 1\}$ denotes the employment status of a worker, i.e., $\eta = 0$ when a worker is unemployed and $\eta = 1$ otherwise. For each worker, denote his expected "output" $y(x, \eta)$ as follows,

$$y(x, \eta) = \eta(x\mu_H + (1-x)\mu_L) + (1-\eta)b. \quad (29)$$

When a worker is unemployed, we interpret his unemployment benefit b as his "output".

Define $\Delta(\Xi)$ as the set of all feasible cross-sectional probability measures of the individual worker states. At any time t , denote $G_t^\eta: [0, 1] \rightarrow [0, 1]$ as the cumulative distribution function of employed/unemployed workers' posterior beliefs at time t . Feasibility requires that $G_t^0(1) + G_t^1(1) = 1$. At time t , the planner observes the aggregate state of the economy G_t . The planner then decides whether to separate each match. Denote $\delta(x) \in \{0, 1\}$ to be the planner's separation decision for a match whose quality is believed to be high with probability $x \in [0, 1]$, i.e., $\delta(x) = 1$ represents separation while $\delta(x) = 0$ represents no separation. In addition, the planner also chooses $\theta(x) \in \mathbb{R}_+$, the tightness at the location where the x -worker looks for new matches.¹⁴

For each individual, we allow the planner's allocation to potentially depend not only on the individual's state but also the calendar time and the aggregate state G_t . Formally, an admissible plan for the planner is a measurable function $(\theta, \delta): \mathbb{R}_+ \times [0, 1] \times \Delta(\Xi) \rightarrow \mathbb{R}_+ \times [0, 1]$ which is right-continuous in time. Namely, given time t and aggregate state G_t , $\theta(t, x, G_t)$ denotes the tightness of the location where the x -worker is sent to search, and $\delta(t, x, G_t)$ is the separation decision for a match with posterior x .

Fix an admissible plan and the initial distribution of the states is given by G_0 , the process of $\{G_t\}_{t \geq 0}$ is deterministic. The planner's payoff is given by

$$\int_0^\infty e^{-rt} \left[\sum_\eta \int_{x \in [0, 1]} y(x, \eta) dG_t^\eta(x) \right] dt - \int_0^\infty e^{-rt} \left[\int_{x \in [0, 1]} k\theta(t, x, G_t) dG_t^0(x) \right] dt, \quad (30)$$

¹⁴ We omit η in the future expressions, because it is well established that θ is the choice for an unemployed worker and δ is the choice of an employed worker.

where $\sum_{\eta} \int_{x \in [0,1]} y(x, \eta) dG_t^{\eta}(x)$ is the total social “output” at time t and $k \int_{x \in [0,1]} \theta(t, x) dG_t^0(x)$ is the total social cost of vacancy creation at time t .

Let $g^{\eta}(t, x)$ denote the probability density function corresponding with $G_t(\cdot)$ whenever it is well-defined. We can use Kolmogorov forward equation to describe the law of motion of $g^{\eta}(t, x)$, $\forall \eta \in [0, 1]$. For any $x \in [0, 1]$ such that both G_t^0 and G_t^1 do not have a mass point at x , we have

$$\begin{aligned} \frac{\partial g^0(t, x)}{\partial t} &= \frac{\partial^2}{\partial x^2} [\Sigma_0(x) g^0(t, x)] - [\rho + p(\theta(t, x, G_t))] g^0(t, x) \quad \text{where} \\ \Sigma_0(x) &= \frac{1}{2} s_0^2 x^2 (1-x)^2 \end{aligned} \quad (31)$$

and

$$\begin{aligned} \frac{\partial g^1(t, x)}{\partial t} &= \frac{\partial^2}{\partial x^2} [\Sigma_1(x) g^1(t, x)] - \rho g^1(t, x) + p(\theta(t, x, G_t)) g^0(t, x) \quad \text{where} \\ \Sigma_1(x) &= \frac{1}{2} s_1^2 x^2 (1-x)^2. \end{aligned} \quad (32)$$

The first term of the right-hand-side of equation (31) captures the density change due to arriving information and the second term reflects the change due to death and directed search. Equation (32) can be interpreted similarly.

If G_t^0 has a mass point at x but G_t^1 does not, then $g^1(t+, x)$ may have a kink corresponding to the inflow of workers. Denote $h(\tilde{t}, \tilde{x}) = \frac{\partial g^1(t, x)}{\partial x} \big|_{(t, x) = (\tilde{t}, \tilde{x})}$, then

$$\Sigma_1(x) \lim_{\tilde{t} \searrow t} \left(\lim_{\tilde{x} \nearrow x} h(\tilde{t}, \tilde{x}) - \lim_{\tilde{x} \searrow x} h(\tilde{t}, \tilde{x}) \right) = p(\theta(t, x, G_t)) \left(G_t^0(x) - G_t^0(x-) \right). \quad (33)$$

If G_t^1 has a mass point at x but G_t^0 does not and $\delta(t, x, G_t) = 1$, then

$$G_{t+}^0(x) - G_t^0(x) = G_t^1(x) - G_t^1(x-), \quad (34)$$

which implies that the measure of unemployed workers with posterior belief x jumps at time t .

The planner’s problem is to choose an admissible plan to maximize (30) subject to (31), (32), (33), and (34). Denote the planner’s value function as $S(t, G_t)$. In principle, the planner’s problem depends on the aggregate state of the economy G_t ; thus the planner’s problem has infinitely many dimensions of “state variables.” Borrowing the technique from Li and Weng (2015), Theorem 3 below shows that the planner’s problem can be broken down into a set of individual problems, and the optimal plan is distribution-free.

Theorem 3 (Separability of the Planner’s problem). *The planner’s value function is linear in $G_t(\cdot)$ and it depends on the calendar time only through G_t . That is,*

$$S(t, G_t) = \int_{x \in [0,1]} U(x) dG_t^0(x) + \int_{x \in [0,1]} V(x) dG_t^1(x), \quad (35)$$

for any $t \geq 0$ where $V(x)$ and $U(x)$ are the component value functions such that

$$rV(x) = \max_{\delta \in [0,1]} \left\{ (1-\delta)rU(x) + \delta \left[x\mu_H + (1-x)\mu_L + \rho(0 - V(x)) + \Sigma_1(x)V''(x) \right] \right\}, \quad (36)$$

and

$$(r + \rho)U(x) = \max_{\theta \geq 0} \{b + \Sigma_0(x)U''(x) + p(\theta)[V(x) - U(x)] - k\theta\}. \quad (37)$$

The planner's problem in equation (36) is associated with the value of an employed worker or match with a different belief; while the value function $U(x)$ in equation (37) is associated with the unemployed worker's value.

Planner's solution. For an employed worker, the decision of separation is essentially an optimal stopping problem. Suppose the planner separates the match when the posterior belief reaches x^* . For $x < x^*$, it is optimal for a worker to stay unemployed, and hence the value function $U(x)$ satisfies:

$$(r + \rho)U(x) = b + \Sigma_0(x)U''(x). \quad (38)$$

For $x > x^*$, it is optimal for a worker to stay employed, and hence the value function $V(x)$ satisfies:

$$(r + \rho)V(x) = x\mu_H + (1 - x)\mu_L + \Sigma_1(x)V''(x). \quad (39)$$

Since $V(x) > U(x)$ for $x > x^*$, the assumption $\lim_{\theta \rightarrow 0} p'(\theta) = \infty$ implies that $\theta^*(x) > 0$ for all $x > x^*$, and hence the value function $U(x)$ satisfies:

$$(r + \rho)U(x) = b + \Sigma_0(x)U''(x) + \max_{\theta \geq 0} \{p(\theta)[V(x) - U(x)] - k\theta\}. \quad (40)$$

At x^* , the following boundary conditions should be satisfied:

$$U(x^*-) = U(x^*+) = V(x^*+) \quad (\text{Value Matching}), \quad (41)$$

and

$$U'(x^*-) = V'(x^*+) \quad (\text{Smooth Pasting}). \quad (42)$$

Notice that equation (41) implies that $U''(x^*-) = U''(x^*+) \triangleq U''(x^*)$. But the above boundary conditions are not enough to pin down all of the value functions and the cutoff x^* . Furthermore, the argument in Theorem 1 (and the method of sub- and super-solutions as well) does not apply in the current context because of the search friction. Inspired by the proof of Theorem 1, we find a different way to prove the second derivative condition.

Corollary 1. A necessary condition for the optimal solution x^* is that

$$V''(x^*) = U''(x^*). \quad (43)$$

The second derivative condition (43) together with the other boundary conditions (41) and (42) enables us to fully characterize the planner's solution. We summarize the results into the following theorem.

Theorem 4. There is a unique triple $(V(x), U(x), x^*)$ such that both V and U are continuous and twice differentiable, $x^* \in (0, 1)$, and they solve (36), (37), (41), (42) and (43) and characterize the planner's component value function in Theorem 3. Moreover, x^* is independent of the job creation cost k and the matching function $p(\theta)$.

This application is novel not only for the economics but also because it combines common value experimentation with switching costs due to (directed) search. To our knowledge, nothing similar exists in the literature and due to the switching costs, the method of sub- and super-solutions cannot be applied in this context. It is precisely because of our method to derive the second derivative condition based on one shot deviations that we can completely characterize the planner's solution.

Another striking feature of this application is that an employed worker can become unemployed without any cost, but an unemployed worker has to search for a vacant firm. One may thus expect that as the search cost goes up, it becomes more attractive for the planner to let the worker stay employed. This logic then implies that x^* should decrease as k increases. This argument turns out to be incorrect, and [Theorem 4](#) states that the planner's optimal solution is obtained as if there were no search frictions. By doing this, the planner maximizes an employed worker's value $V(x)$, and from equation (40), an unemployed worker's value is also maximized no matter what the job creation cost k or the matching function $p(\theta)$ is. Furthermore, as we have explicit solutions to value functions (38) and (39), x^* can be explicitly characterized. Finally, although the cutoff x^* is independent of k and $p(\theta)$, the laws of motion (31) and (32) will indeed depend on k and $p(\theta)$. Therefore, a change in k and $p(\theta)$ will affect the unemployment rate without changing the cutoff, simply because the stationary distribution changes.

5. 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 or whether she is unemployed; 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. It can also account for switching cost. The value of the paper is to derive the second derivative condition from the one-shot-deviation principle (Bellman's principle of optimality) and to show that this method can easily be used to prove results in economically relevant applications.

Appendix A. Omitted proofs

Proof of Theorem 2. If $N = 1$, the single buyer is also facing a common value two-armed bandit problem. As a result, at the cutoff \underline{x} , we should have:

$$U(\underline{x}-) = U(\underline{x}+), \quad U'(\underline{x}-) = U'(\underline{x}+), \quad U''(\underline{x}-) = U''(\underline{x}+). \quad (44)$$

Define total social surplus function $\tilde{W}(x) = V_1(x) + V_2(x) + U(x)$, and it is immediate to see that $\tilde{W}_j(x)$ satisfies the following differential equation:

$$r\tilde{W}_j(x) = f_j(x) + \frac{1}{2}\Sigma_j(x)\tilde{W}_j''(x). \quad (45)$$

Since \tilde{W} shares the same differential equation as equation (13) and the boundary conditions are the same, we conclude that the equilibrium cutoff \underline{x} must be the same as x^e .

We will focus on the $N > 1$ afterwards. For $x > \underline{x}$, the cautious strategies imply

$$p_1(x) = \frac{1}{2}[\Sigma_2(x) - \Sigma_1(x)]V_1''(x), \quad (46)$$

and hence

$$p_2(x) = f_2(x) - f_1(x) + \frac{1}{2}[\Sigma_2(x) - \Sigma_1(x)](V_1''(x) + U''(x)). \quad (47)$$

This enables us to rewrite the value functions as:

$$\begin{aligned} rV_1(x) &= \frac{N}{2}\Sigma_2(x)V_1''(x) \\ rV_2(x) &= N\left[f_2(x) - f_1(x) + \frac{1}{2}[\Sigma_2(x) - \Sigma_1(x)](V_1''(x) + U''(x)) + \frac{1}{2}\Sigma_2(x)V_2''(x)\right] \\ rU(x) &= f_1(x) + \frac{N}{2}\Sigma_2(x)U''(x) - \frac{1}{2}[\Sigma_2(x) - \Sigma_1(x)](V_1''(x) + U''(x)). \end{aligned}$$

Following Lemma 1 in [Bergemann and Välimäki \(2000\)](#), we derive the general solutions to the above differential equations:

$$\begin{aligned} V_1(x) &= A_1x^{-\frac{\gamma_2-1}{2}}(1-x)^{\frac{\gamma_2+1}{2}} \\ V_2(x) &= N\frac{f_2(x) - f_1(x)}{r} + A_3x^{-\frac{\gamma_2-1}{2}}(1-x)^{\frac{\gamma_2+1}{2}} - NA_2x^{-\frac{\lambda_2-1}{2}}(1-x)^{\frac{\lambda_2+1}{2}} \\ U(x) &= \frac{f_1(x)}{r} + A_2x^{-\frac{\lambda_2-1}{2}}(1-x)^{\frac{\lambda_2+1}{2}} - A_1x^{-\frac{\gamma_2-1}{2}}(1-x)^{\frac{\gamma_2+1}{2}}. \end{aligned}$$

Similarly, for $x < \underline{x}$, we derive the general solutions as

$$\begin{aligned} V_1(x) &= N\frac{f_2(x) - f_1(x)}{r} + B_3x^{\frac{\gamma_1+1}{2}}(1-x)^{-\frac{\gamma_1-1}{2}} - NB_2x^{\frac{\lambda_1+1}{2}}(1-x)^{-\frac{\lambda_1-1}{2}} \\ V_2(x) &= B_1x^{\frac{\gamma_1+1}{2}}(1-x)^{-\frac{\gamma_1-1}{2}} \\ U(x) &= \frac{f_2(x)}{r} + B_2x^{\frac{\lambda_1+1}{2}}(1-x)^{-\frac{\lambda_1-1}{2}} - B_1x^{\frac{\gamma_1+1}{2}}(1-x)^{-\frac{\gamma_1-1}{2}}, \end{aligned}$$

where $\gamma_j = \sqrt{1 + \frac{8r}{Ns_j^2}}$ and $\lambda_j = \sqrt{1 + \frac{8r}{(N-1)s_j^2 + s_{-j}^2}}$.

Then boundary conditions (23) and (24) imply a system of 7 equations with 7 unknowns. At the cutoff $x = \underline{x}$, simplification yields

$$\frac{f_1(x) - f_2(x)}{r} = \frac{\Psi(x)}{G_1} + \frac{\Psi(x)}{G_2}, \quad (48)$$

where

$$\begin{aligned} \Psi(x) &= \left[s_1^2\left(\frac{\gamma_1+1}{2} - x\right) + s_2^2\left(\frac{\gamma_2-1}{2} + x\right)\right]\frac{b_1 - b_2}{r} \\ &\quad + \left[s_1^2\frac{\gamma_1-1}{2} + s_2^2\frac{\gamma_2+1}{2}\right]\frac{(a_1 - a_2)x}{r}, \end{aligned} \quad (49)$$

$$G_1 = \frac{1}{2}(N-1)s_1^2(\lambda_1 - \gamma_1) + \frac{1}{2}s_2^2(\lambda_1 + \gamma_2), \quad (50)$$

and

$$G_2 = \frac{1}{2}(N-1)s_2^2(\lambda_2 - \gamma_2) + \frac{1}{2}s_1^2(\lambda_2 + \gamma_1). \quad (51)$$

At the socially efficient cutoff x^e , $\Psi(x^e) = 0$ from equation (16), and hence the Markov perfect equilibrium is socially efficient if and only if $f_1(x^e) - f_2(x^e) = 0$, which is true only when $s_1 = s_2$. If $s_1 \neq s_2$, the equilibrium cutoff \underline{x} is different from x^e since $f_1(x^e) - f_2(x^e) \neq 0$. Moreover, when $s_2 > s_1$,

$$\Psi(x) = \left[s_1^2 \left(\frac{\gamma_1 + 1}{2} - x \right) + s_2^2 \left(\frac{\gamma_2 - 1}{2} + x \right) \right] \frac{f_1(x) - f_2(x)}{r} + (1-x)(s_2^2 - s_1^2)(a_1 - a_2) \frac{x}{r} \quad (52)$$

is decreasing in x , and

$$\frac{\Psi(0)}{G_2} > \frac{b_1 - b_2}{r} = \frac{f_1(0) - f_2(0)}{r}. \quad (53)$$

We should have $\underline{x} < x^e$ when $s_2 > s_1$. Similarly, we get $\underline{x} > x^e$ when $s_2 < s_1$. \square

Proof of Theorem 3. First, borrowing from Li and Weng (2015), we are able to show that the planner's value function at time 0 can be rewritten as:

$$\int_{x \in [0,1]} \hat{U}_0(x) dG_0^0(x) + \int_{x \in [0,1]} \hat{V}_0(x) dG_0^1(x),$$

where $\hat{U}_0(x)$ denotes the expected payoff of an unemployed worker with prior x at time 0:

$$\hat{U}_0(x) = \sup_{(\theta, \delta) \in \mathcal{A}} \mathbb{E} \left[\int_0^\tau e^{-rt} [b - k\theta_t(x_t)] dt + e^{-r\tau} \hat{V}_\tau(x_\tau) | x_0 = x \right], \quad (54)$$

and $\hat{V}_0(x)$ denotes the expected payoff of an employed worker with prior x at time 0:

$$\hat{V}_0(x) = \sup_{(\theta, \delta) \in \mathcal{A}} \mathbb{E} \left[\int_0^\tau e^{-rt} y(x_t) dt + e^{-r\tau} \hat{U}_\tau(x_\tau) | x_0 = x \right], \quad (55)$$

and $\hat{U}_\tau(x)$ and $\hat{V}_\tau(x)$ are defined in the same fashion at time $\tau \geq 0$. The proof is similar to the proof of Theorem 1 in Li and Weng (2015), and hence is omitted.

Second, we want to show that there admits an optimal solution with $\hat{V}_0(x) = \hat{V}_t(x)$ and $\hat{U}_0(x) = \hat{U}_t(x)$. Suppose on the contrary that in all of the optimal solutions, there exists some t such that $\hat{V}_t(x) \neq \hat{V}_0(x)$ for a positive measure of x . This allows for two possibilities. When $\hat{V}_t(x) > \hat{V}_0(x)$, since the planner's objective is to maximize

$$\int_{x \in [0,1]} \hat{U}_0(x) dG_0^0(x) + \int_{x \in [0,1]} \hat{V}_0(x) dG_0^1(x),$$

then at time 0, the planner can simply change the employed workers' strategies to be the same as the time t employed workers' strategies. By doing so, the planner obtains a higher value, which leads to a contradiction. Similarly, when $\hat{V}_t(x) < \hat{V}_0(x)$, the planner can increase her value by letting the time t employed workers mimic the time 0 employed workers. By doing so, the unemployed worker's expected payoff $\hat{U}_0(x)$ increases from equation (54). Hence, it follows

that there admits an optimal solution with $\hat{V}_0(x) = \hat{V}_t(x)$ for all $t \geq 0$. Using the same logic, we can also prove that there admits an optimal solution with $\hat{U}_0(x) = \hat{U}_t(x)$ for all $t \geq 0$.

Since this stochastic optimal control problem has a (stationary) Markovian solution, we can further express the value functions $V(x)$ and $U(x)$ as:

$$rV(x) = \max \{ rU(x), x\mu_H + (1-x)\mu_L - \rho V(x) + \Sigma_1(x)V''(x) \},$$

and

$$(r + \rho)U(x) = \max_{\theta \geq 0} \{ b + \Sigma_0(x)U''(x) + p(\theta)[V(x) - U(x)] - k\theta \}. \quad \square$$

Proof of Corollary 1. It is straightforward to show that $V''(x^*) \geq U''(x^*)$ using the fact that it is suboptimal to stay employed for $x < x^*$. Assume by contradiction that at the optimal solution x^* , $V''(x^*) > U''(x^*)$. The general solutions to value functions (38) and (39) are given by:

$$U(x) = \frac{b}{r + \rho} + k_0 x^{\beta_0} (1-x)^{1-\beta_0}$$

$$V(x) = \frac{x\mu_H + (1-x)\mu_L}{r + \rho} + k_1 x^{1-\beta_1} (1-x)^{\beta_1}$$

where $\beta_0 = \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2(r+\rho)}{s_0^2}} \geq 1$ and $\beta_1 = \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2(r+\rho)}{s_1^2}} \geq 1$. For a fixed x^* , we can solve k_0 and k_1 using the value matching condition (41) and the smooth pasting condition (42):

$$\frac{b}{r + \rho} + k_0 (x^*)^{\beta_0} (1-x^*)^{1-\beta_0} = \frac{x^* \mu_H + (1-x^*) \mu_L}{r + \rho} + k_1 (x^*)^{1-\beta_1} (1-x^*)^{\beta_1} \quad (56)$$

$$k_0 (x^*)^{\beta_0} (1-x^*)^{1-\beta_0} \frac{\beta_0 - x^*}{x^*(1-x^*)} = \frac{\mu_H - \mu_L}{r + \rho} - k_1 (x^*)^{1-\beta_1} (1-x^*)^{\beta_1} \frac{\beta_1 + x^* - 1}{x^*(1-x^*)}. \quad (57)$$

Total differentiation with respect to (56) and (57) implies:

$$U'(x^*)dx^* + (x^*)^{\beta_0} (1-x^*)^{1-\beta_0} dk_0 = V'(x^*)dx^* + (x^*)^{1-\beta_1} (1-x^*)^{\beta_1} dk_1 \quad (58)$$

and

$$\begin{aligned} U''(x^*)dx^* + (x^*)^{\beta_0} (1-x^*)^{1-\beta_0} \frac{\beta_0 - x^*}{x^*(1-x^*)} dk_0 \\ = V''(x^*)dx^* - (x^*)^{1-\beta_1} (1-x^*)^{\beta_1} \frac{\beta_1 + x^* - 1}{x^*(1-x^*)} dk_1. \end{aligned} \quad (59)$$

Since $U'(x^*) = V'(x^*)$ and $U''(x^*) < V''(x^*)$, we obtain $\frac{dk_0}{dx^*} > 0$ and $\frac{dk_1}{dx^*} > 0$. Suppose the planner chooses another cutoff $\tilde{x}^* = x^* + \epsilon$. Value matching and smooth pasting at this new cutoff \tilde{x}^* imply a new pair of parameters $(\tilde{k}_0, \tilde{k}_1)$. Then, $\frac{dk_0}{dx^*} > 0$ and $\frac{dk_1}{dx^*} > 0$ imply that $\tilde{k}_0 > k_0$ and $\tilde{k}_1 > k_1$ if $\epsilon > 0$ is sufficiently small as illustrated by Fig. 1. To show that deviating to \tilde{x}^* yields a higher value for the planner, it remains to prove that $\tilde{U}(x) \geq U(x)$ for all $x \geq \tilde{x}^*$.

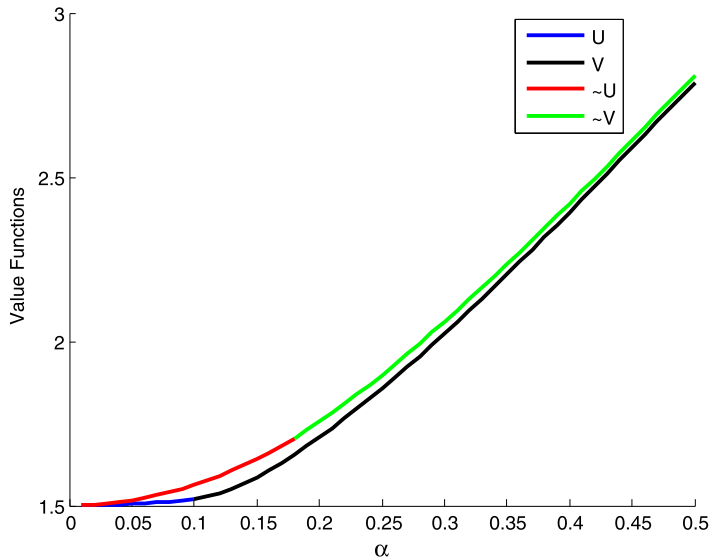


Fig. 1. Solutions for different cutoffs.

Notice that for $x \geq \tilde{x}^*$, equation (40) implies that:

$$(r + \rho)U(x) = b + \Sigma_0(x)U''(x) + \max_{\theta \geq 0} \{p(\theta)[V(x) - U(x)] - k\theta\} \quad (60)$$

and

$$(r + \rho)\tilde{U}(x) = b + \Sigma_0(x)\tilde{U}''(x) + \max_{\theta \geq 0} \{p(\theta)[\tilde{V}(x) - \tilde{U}(x)] - k\theta\}. \quad (61)$$

Obviously, when $x = \tilde{x}^*$, $\tilde{U}(x) > U(x)$; and when $x = 1$ (the worker's ability is believed to be H for sure), $\tilde{U}(x) = U(x)$ since $\tilde{V}(x) = V(x)$. Now assume on the contrary that $\tilde{U}(x) < U(x)$ for some $x > \tilde{x}^*$. Then there must exist $x' > \tilde{x}^*$ such that $\tilde{U}(x') = U(x')$ and $\tilde{U}(x) > U(x)$ for $x < x'$. This implies that $\tilde{U}'(x') \leq U'(x')$. From equations (60) and (61), $U''(x) > \tilde{U}''(x)$ whenever $\tilde{U}(x) \leq U(x)$ since $\max_{\theta \geq 0} \{p(\theta)\Delta - k\theta\}$ is increasing in Δ . Therefore, for $x > x'$, we have $U''(x) > \tilde{U}''(x)$ and hence $U'(x) > \tilde{U}'(x)$. But this then implies $\tilde{U}(1) < U(1)$, which leads to a contradiction.

Finally, from equation (35), deviating to \tilde{x}^* yields a higher value $S(t, G_t)$ to the planner. But this contradicts with the definition of x^* . Therefore, we must have $V''(x^*) = U''(x^*)$. \square

Proof of Theorem 4. The general solutions to the value functions are given by:

$$U(x) = \frac{b}{r + \rho} + k_0 x^{\beta_0} (1 - x)^{1 - \beta_0}$$

$$V(x) = \frac{x\mu_H + (1 - x)\mu_L}{r + \rho} + k_1 x^{1 - \beta_1} (1 - x)^{\beta_1}$$

where $\beta_0 = \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2(r+\rho)}{s_0^2}} \geq 1$ and $\beta_1 = \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2(r+\rho)}{s_1^2}} \geq 1$. At the cutoff x^* , value matching, smooth pasting and the second derivative conditions are satisfied. Therefore, we can explicitly derive x^* :

$$x^* = \frac{(b - \mu_L)(s_0^2 \beta_0 + s_1^2 (\beta_1 - 1))}{(a_2 - a_1) [s_0^2 (\beta_0 - 1) + s_1^2 \beta_1] + (b - \mu_L)(s_0^2 - s_1^2)},$$

and

$$k_0 = \frac{s_0^2 (b - x^* \mu_H - (1 - x^*) \mu_L)}{(s_1^2 - s_0^2)(r + \rho)},$$

$$k_1 = \frac{s_1^2 (b - x^* \mu_H - (1 - x^*) \mu_L)}{(s_1^2 - s_0^2)(r + \rho)}.$$

Finally, we want to show that for $x \geq x^*$, there exists a unique solution to differential equation:

$$(r + \rho)U(x) = b + \Sigma_0(x)U''(x) + \max_{\theta \geq 0} \{p(\theta)[V(x) - U(x)] - k\theta\} \quad (62)$$

with boundary conditions

$$U(x^*) = \frac{b}{r + \rho} + k_0(x^*)^{\beta_0}(1 - x^*)^{1-\beta_0},$$

and

$$U(1) = \frac{b}{r + \rho} + \frac{\max_{\theta \geq 0} \{p(\theta)[\frac{\mu_H}{r + \rho} - U(1)] - k\theta\}}{r + \rho}.$$

The existence and uniqueness proof is similar to the proof of Theorem 2 in Bonatti (2011). First, equation (62) can be rewritten as:

$$U''(x) = \min_{\theta \geq 0} \frac{(r + \rho)U(x) - b - \{p(\theta)[V(x) - U(x)] - k\theta\}}{\Sigma_0(x)}. \quad (63)$$

It is straightforward to verify that $U_1(x) = \frac{b}{r + \rho}$ is a strict subsolution of (63), and $U_2(x) = V(1) = \frac{\mu_H}{r + \rho}$ is a strict supersolution of (63). Moreover, for a given interval $J_\epsilon = [x^*, 1 - \epsilon]$, the boundary value problem (62) is regular with respect of $U_1(x)$ and $U_2(x)$ on J_ϵ since

$$\min_{\theta \geq 0} \frac{(r + \rho)U(x) - b - \{p(\theta)[V(x) - U(x)] - k\theta\}}{\Sigma_0(x)} < \max_{x \in J_\epsilon} \frac{(r + \rho)V(x) - b}{\Sigma_0(x)}.$$

Therefore, from Lemma 2 in Bonatti (2011), for all $\epsilon > 0$, the boundary value problem (62) admits a C^2 solution on J_ϵ . Let $\epsilon \rightarrow 0$, and the limit function solves (62) as shown by Bonatti (2011).

The solution to the boundary value problem (62) is unique. Suppose instead there were two solutions U_1 and U_2 with $U_1(x) > U_2(x)$ for all $x \in \mathcal{A}$. Then from equation (63), we must have $U_1''(x) > U_2''(x)$ for $x \in \mathcal{A}$. On the other hand, since these two solutions share the same boundary conditions, $U_1(x) - U_2(x)$ must attain its local maximum on \mathcal{A} . But this implies that $U_1''(x) < U_2''(x)$ for some $x \in \mathcal{A}$. This contradicts the assumption $U_1(x) > U_2(x)$. \square

Appendix B. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.jet.2015.10.002>.

References

- Applebaum, D., 2004. *Lévy Processes and Stochastic Calculus*. Cambridge University Press.
- Bergemann, D., Välimäki, J., 1996. Learning and strategic pricing. *Econometrica* 64 (5), 1125–1149.
- Bergemann, D., Välimäki, J., 2000. Experimentation in markets. *Rev. Econ. Stud.* 67 (2), 213–234.
- Bergemann, D., Välimäki, J., 2008. Bandit problems. In: *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke.
- Bolton, P., Harris, C., 1999. Strategic experimentation. *Econometrica* 67 (2), 349–374.
- Bonatti, A., 2011. Menu pricing and learning. *Am. Econ. J. Microecon.* 3 (3), 124–163.
- Cohen, A., Solan, E., 2013. Bandit problems with Lévy payoff. *Math. Oper. Res.* 38 (1), 92–107.
- Cripps, M., Ely, J., Mailath, G., Samuelson, L., 2008. Common learning. *Econometrica* 76 (4), 909–933.
- Dumas, B., 1991. Super contact and related optimality conditions. *J. Econ. Dyn. Control* 15, 675–685.
- Eeckhout, J., Weng, X., 2010. *Assortative Learning*. Mimeo.
- Gittins, J., Jones, D., 1974. A dynamic allocation index in the sequential design of experiments. In: *Progress in Statistics*. North Holland, Amsterdam, pp. 241–266.
- Karatzas, I., 1984. Gittins indices in the dynamic allocation problem for diffusion processes. *Ann. Probab.* 12 (1), 173–192.
- Kaspi, H., Mandelbaum, A., 1995. Lévy bandits: multi-armed bandits driven by Lévy processes. *Ann. Appl. Probab.* 5, 541–565.
- Kehoe, P., Pastorino, E., 2011. Individual and market experimentation with multi-armed dependent bandits. Mimeo, University of Minnesota.
- Kehoe, P., Pastorino, E., 2013. Pricing and efficiency in markets with differentiated experience goods. Mimeo, University of Minnesota.
- Keller, G., Rady, S., 1999. Optimal experimentation in a changing environment. *Rev. Econ. Stud.* 66 (3), 475–507.
- Keller, G., Rady, S., Cripps, M., 2005. Strategic experimentation with exponential bandits. *Econometrica* 73 (1), 39–68.
- Li, F., Weng, X., 2015. Efficient learning and job turnover in the labor market. Mimeo.
- Liptser, R.S., Shiryaev, A.N., 2001. *Statistics of Random Processes II: Applications*. Springer.
- Menzio, G., Shi, S., 2011. Efficient search on the job and the business cycle. *J. Polit. Econ.* 119 (3), 468–510.
- Papageorgiou, T., 2014. Learning your comparative advantages. *Rev. Econ. Stud.* 81, 1263–1295.
- Peskir, G., Shiryaev, A., 2006. *Optimal Stopping and Free-Boundary Problems*. Birkhäuser.
- Samuelson, P.A., 1965. Rational theory of warrant pricing. *Ind. Manage. Rev.* 6, 13–31.
- Whittle, P., 1980. Multi-armed bandits and gittins index. *J. R. Stat. Soc., Ser. B* 42, 143–149.
- Wirl, F., 2008. Reversible stopping (“switching”) implies super contact. *Comput. Manag. Sci.* 5, 393–401. Published online in 2007.