Research Article
A Common Value Experimentation with Multiarmed Bandits

Xiujuan Gao ⬤, Hao Liang, and Tong Wang ⬤

The School of Economic Mathematics, Southwestern University of Finance and Economics, 611130 Wenjiang, Chengdu, China

Correspondence should be addressed to Tong Wang; 2160202z1008@2016.swufe.edu.cn

We study a value common experimentation with multiarmed bandits and give an application about the experimentation. The second derivative of value functions at cutoffs is investigated when an agent switches action with multiarmed bandits. If consumers have identical preference which is unknown and purchase products from only two sellers among multiple sellers, we obtain the necessary and sufficient conditions about the common experimentation. The Markov perfect equilibrium and the socially effective allocation in K-armed markets are discussed.

1. Introduction

A financial debate has arisen when we need to choose the best goods among multiple products. In Robbins [1], this problem is described as a decision maker facing N slot machines (called arms) and the maker has to choose one of them at each instantaneous time. Gittins and Jones [2] and Michael et al. [3] calculate the value of pulling an arm, i.e., Gittins index, in discrete time. Comparing the value to the Gittins index of all other arms, Michael et al. [3] present that the optimal strategy for an N-armed problem is an N-dimensional discounted Markov decision chain and the value pulling each arm itself is independent of the cutoff. Karatzas [4], Kaspi, and Mandelbaum [5] transform the problem into a standard optimal stopping problem. Bolton and Harris [6] and Bergemann and Välimäki [7] show that when there are $K \geq 2$ sellers who sell different products and N consumers whose preferences are identical (but unknown) in the market, the optimal strategies of consumers are to buy the products from the same seller; i.e., it is symmetric equilibria. Cohen and Solan [8] study two-armed bandit problems in the continuous time with the property of *Lévy* processes and obtain the Hamilton-Jacobi-Bellman (HJB) equation for the problem. They conclude that the optimal strategy is a cutoff strategy when the arms have two types. For other optimal strategies and control approaches, the reader is referred to [9–11] and the references therein.

Jan and Xi [12] investigate that second derivatives of value functions are equal at the cutoff with value matching and smooth pasting (assume that the value function is $V(x) \in C^2(R^+)$ and the cutoff is $x^* \in R^+$. Value matching: $V(x^*-) = V(x^*+)$ and smooth pasting: $V'(x^*-) = V'(x^*+)$ which is discussed in [13, 14], where $V(x-) := \lim_{x \to x^-} V(x)$ and $V(x+) := \lim_{x \to x^+} V(x)$), when there are two arms with different types, and conclude that the optimal strategy is an interval strategy in the market. In [12], an application is given about strategic pricing of two vendors in a competitive market. There are $N \geq 2$ consumers who have the same type either H or L. Two vendors produce two different kinds of goods for the two types, respectively. Jan and Xi [12] describe the socially efficient allocation and pricing strategies of two vendors in the market. Moreover, they use value matching, smooth pasting, and second derivatives of value functions to discuss the Markov perfect equilibrium and the socially efficient allocation.

In this article, we investigate multiarmed bandits, while two-armed bandits are studied in [12]. We consider multiple sellers instead of two sellers. In general, there are multiple sellers to sell the same type of goods in the market, but the quality, utilities, and the prices of goods sold by each seller are different. Therefore, changing the number of arms from two to $K \geq 2$ is reasonable in the market. People believe that purchasing two products has the best effects to type H or L in the market. For example, the first seller is the highest utility to type H but is the lowest utility to type L. On the contrary, the second seller is the highest utility to type L but is the lowest utility to type H and other utilities are between about the two

sellers to types H and L. Instinctively, we think that a rational consumer chooses goods only from the first two sellers but this is not perfect or this strategy needs some conditions. We discuss the necessary and sufficient conditions for this strategy.

In order to obtain our conclusions under certain assumptions of the model for multiarmed bandits, using the methods used in [12], we calculate the optimal cutoff points by solving the corresponding ordinary differential equations. Then, we obtain the necessary and sufficient conditions for the strategy of which consumers purchase products from only two sellers among multiple sellers.

The rest of the paper is as follows. In Section 2, we introduce a multiarmed model under certain assumptions and show a conclusion about the model. In Section 3, we give an application for the model. In Section 3.1, we discuss optimal choice of consumers in the case of market equilibrium. Based on the optimal choice of consumers, we derive HJB equations for their utilities functions. Using solutions of the HJB equations, value matching, and smooth pasting, we get the cutoff point at Markov perfect equilibrium and the necessary and sufficient conditions about the common experimentation. In Section 3.2, we get the cutoff point at socially efficient allocation with the same way in Section 3.1. The relationship between the Markov perfect equilibrium and the socially efficient allocation in K-armed markets is discussed.

2. The Model

Jin and Xi [12] consider one agent and a bandit with 2 arms. We study multiarmed bandits and consider the case where there is only one real-valued state $x(t) \in \Upsilon$ and $\Upsilon \in R$ is a connect set. The instantaneous flow payoff of each arm is $f_j(x)$ at state x, $j = 1, 2, \ldots, K$. Let $r > 0$ be the discount rate. For each arm j, there is a probability space $\{\Omega^j, \mathscr{F}^j, P^j\}$ endowed with filtration $\{\mathscr{F}_t^j, t \geq 0\}$ and $T_j(t)$ is the total measure of time to time t when arm j has been chosen. From [12], we know that the updating of x in arm j, when there are K arms in the market, satisfies

$$dx_j(t) = \mu_j(x(t-)) dT_j(t) + \sigma_j(x(t-)) dW_j(T_j(t))$$
$$+ \sum_{i \neq j} \int_{\mathbb{R}-\{0\}} J_j(x(t-), x_i) \mathbb{N}_j(dT_j(t), dx_i), \quad (1)$$

where $W_j(t)$ is a standard Brownian motion, $W_j(t)$ is independent of $W_i(t)$ ($i \neq j$), and $\mathbb{N}_j$ is a Poisson random measure that is independent of W_j. The state x is updated from arm i, $i = 1, 2, \ldots, K$ and $dx(t) = \sum_{i=1}^{K} dx_i(t)$. Let $J_j(x(t-), x_i)$ be the change of the state when there is a Poisson jump. Equation (1) shows that the state changes from x_j to x_i ($i \neq j$). Thus, (1) does not contain the case that state x_j jump from other states. Because state x_j can jump to any other states and any other states can jump back to x_j, (1) contains all jump processes that describe the changes of state.

In multiarmed bandit problem, the stochastic process $\{x_t\}$ is constructed on the product space $\{\Omega, \mathscr{F}\} := \{\Omega^1, \mathscr{F}^1\} \times$

$\{\Omega^2, \mathscr{F}^2\} \times \cdots \times \{\Omega^K, \mathscr{F}^K\}$ with filtration $\mathscr{F}_t = \bigvee_{i=1}^{K} \mathscr{F}_{T_i(t)}^i$. If $x = x_0$, the agent chooses an allocation rule $m_t = \{1, 2, \ldots, K\}$ adapted to filtration $\{\mathscr{F}_t\}_{t \geq 0}$ to solve the following optimal control problem:

$$v(x) = \sup_{m_t} \left\{ \mathbb{E} \int_{t=0}^{\infty} e^{-rt} f_{m_t}(x_t) dt \right\}, \quad (2)$$

$$\text{s.t} \quad dx_t$$
$$= \mu_{m_t}(x_t) dt + \sigma_{m_t}(x_t) dW_t$$
$$+ \sum_{i \neq m_t} \int_{\mathbb{R}-\{0\}} J_{m_t}(x(t-), x_i) \mathbb{N}_{m_t}(dt, dx_i). \quad (3)$$

Assumption 1 (see [12]). Let Υ be a connected set. Assume that $f_j(x), \mu_j(x), \sigma_j(x)$, and $J_j(x, x_i)$ for $i \neq j$ are C^2 functions of $x \in \Upsilon$.

Assumption 2 (see [12]). Assume that the first derivatives of $f_j(x), \mu_j(x), \sigma_j(x)$, and $J_j(x, \cdot)$ with respect to x are bounded. Namely, there exists $M > 0$ such that for any $x \in \Upsilon$, $|f_j'(x)| \leq M$, $|\mu_j'(x)| \leq M$, $|\sigma_j'(x)| \leq M$, and $|\partial J_j(x, x_i)/\partial x| \leq M$ for each $i \neq j$. Using (2) and the dynamic programming principle (the detailed process is in [15]), for any $h > 0$, we obtain

$$v(x) = \sup_{m_t} \left\{ \mathbb{E} \int_{t=0}^{h} e^{-rt} f_{m_t}(x_t) dt + e^{-rh} v(x_h) \right\}. \quad (4)$$

Using Ito's lemma for $e^{-rt} v(x_t)$ and property of Poisson random measure (the property of Poisson random measure is in [13]), we get

$$de^{-rt} v(x_t) = -r e^{-rt} v(x_t) dt + e^{-rt} v'(x_t) dx_t^c + \frac{1}{2}$$
$$\cdot e^{-rt} v''(x_t) d[x, x]_t = -r e^{-rt} v(x_t) dt$$
$$+ e^{-rt} v'(x_t) dx_t^c + \frac{1}{2} e^{-rt} \sigma_{m_t}^2(x_t) v''(x_t) dt$$
$$+ \sum_{i=1, i \neq m_t}^{K-1} \left[\int_{\mathbb{R}-\{0\}} J_{m_t}(x(t-), x_i) \mathbb{N}_{m_t}(dt, dx_i) \right]^2$$
$$+ 2 \sum_{1 \leq p < q \leq K-1} \int_{\mathbb{R}-\{0\}} J_{m_t}(x(t-), x_p) \mathbb{N}_{m_t}(dt, dx_p)$$
$$\times \int_{\mathbb{R}-\{0\}} J_{m_t}(x(t-), x_q) \mathbb{N}_{m_t}(dt, dx_q)$$
$$= -r e^{-rt} v(x_t) dt + e^{-rt} v'(x_t) dx_t^c + \frac{1}{2}$$
$$\cdot e^{-rt} \sigma_{m_t}^2(x_t) v''(x_t) dt$$
$$+ \sum_{i=1, i \neq m_t}^{K-1} \int_{\mathbb{R}-\{0\}} \left[v\left(x + J_{m_t}(x(t-), x_{m_t})\right) - v(x) \right]$$
$$\cdot \mathbb{N}_{m_t}(dt, dx_i), \quad (5)$$

where $dx_t^c = \mu_{m_t}(x_t) dt + \sigma_{m_t}(x_t) dW_t$.

Then, we have

$$e^{-rh} v(x_h) = v(x) + \int_{t=0}^{h} -re^{-rt} v(x_t)\, dt$$

$$+ e^{-rt} v'(x_t)\, dx_t^c + \int_{t=0}^{h} \frac{1}{2} e^{-rt} \sigma_{m_t}^2 (x_t) v''(x_t)\, dt$$

$$+ \sum_{i=1, i \neq m_t}^{K-1} \int_{t=0}^{h} \left[v\left(x + J_{m_t} \left(x(t-), x_{m_t} \right) \right) - v(x) \right]$$

$$\cdot \mathbb{N}_{m_t} (dt, dx_i).$$

(6)

Substituting (6) into (4), using the mean value theorem of integrals and sending $h \longrightarrow 0$, we get the HJB equation

$$rv(x) = \sup_{m_t} \left\{ f_{m_t}(x) + \mu_{m_t}(x) v'(x) + \frac{1}{2} \sigma_{m_t}^2 (x) v''(x) \right.$$

$$\left. + \sum_{i=1, i \neq m_t}^{K-1} \int_{\mathbb{R}-\{0\}} \left[v\left(x + J_{m_t}\left(x(t-), x_i \right) \right) - v(x) \right] \nu_{m_t}(dx_i) \right\},$$

(7)

where ν_{m_t} is the finite intensity measure of $\mathbb{N}_{m_t}$. From [13], we know that there exists a unique solution to (1).

We assume that optimal strategy of consumers is an interval strategy (see [14]). Namely, there exists $K - 1$ points (cutoffs) $x^1, x^2, \ldots, x^{k-1}$ to partition the space Υ into $K - 1$ intervals, where $x^1 \leq x^2 \leq \cdots \leq x^{k-1}$. When $i, j = 1, 2, \ldots, K - 2$ and $i \neq j$, $(x^i, x^{i+1}) \cap (x^j, x^{j+1}) = \phi$, we assume that if $x \in (x^{j-1}, x^j)$, the agent chooses arm j, where $j = 1, 2, \ldots, K$ and $x^0 := \inf\{x\}$, $x^K := \sup\{x\}$. Thus, we have

$$rv(x) = f_j(x) + \mu_j(x) v'(x) + \frac{1}{2} \sigma_j^2 (x) v''(x)$$

$$+ \sum_{i=1, i \neq j}^{K-1} \int_{\mathbb{R}-\{0\}} \left[v\left(x + J_j\left(x(t-), x_i \right) \right) - v(x) \right]$$

$$\cdot \nu_j(dx_i).$$

(8)

Using the conclusions in [12, 14], we have the value matching

$$v\left(x^j - \right) = v\left(x^j + \right)$$

(9)

and the smooth pasting

$$v'\left(x^j - \right) = v'\left(x^j + \right).$$

(10)

Now, in the light of the value matching and smooth pasting, we have the following conclusion.

Theorem 3. *If $\sigma_j(x) > 0$ for all $x \in \Upsilon$, $j = 1, 2, \ldots, K$, a necessary condition for optimal solution x^j is that $v''(x)$ satisfies $v''(x^j-) = v''(x^j+)$ for any possible cutoffs x^j.*

Proof. On the basis of assumption $x \in (x^{j-1}, x^j)$, an agent chooses arm j, $j = 1, 2, \ldots, K$. It derives that

$$f_j(x) + \mu_j(x) v'(x) + \frac{1}{2} \sigma_j^2 (x) v''(x)$$

$$+ \sum_{i=1, i \neq j}^{K-1} \int_{\mathbb{R}-\{0\}} \left[v\left(x + J_j\left(x(t-), x_i \right) \right) - v(x) \right] \nu_j(dx_i)$$

(11)

$$\leq rv(x).$$

For $x \longrightarrow x_j$ and $x > x_j$, inequality (11) becomes

$$f_j\left(x_j \right) + \mu_j\left(x_j \right) v'\left(x_j + \right) + \frac{1}{2} \sigma_j^2 \left(x_j \right) v''\left(x_j + \right)$$

$$+ \sum_{i=1, i \neq j}^{K-1} \int_{\mathbb{R}-\{0\}} \left[v\left(x_j + J_j\left(x_j, x_i \right) \right) - v\left(x_j + \right) \right]$$

(12)

$$\cdot \nu_j(dx_i) \leq rv\left(x_j + \right).$$

Due to the value matching, we have

$$f_j\left(x_j \right) + \mu_j\left(x_j \right) v'\left(x_j + \right) + \frac{1}{2} \sigma_j^2 \left(x_j \right) v''\left(x_j + \right)$$

$$+ \sum_{i=1, i \neq j}^{K-1} \int_{\mathbb{R}-\{0\}} \left[v\left(x_j + J_j\left(x_j, x_i \right) \right) - v\left(x_j + \right) \right]$$

(13)

$$\cdot \nu_j(dx_i) \leq rv\left(x_j - \right)$$

and

$$rv\left(x_j - \right) = f_j\left(x_j \right) + \mu_j\left(x_j \right) v'\left(x_j - \right) + \frac{1}{2} \sigma_j^2 \left(x_j \right)$$

$$\cdot v''\left(x_j - \right)$$

$$+ \sum_{i=1, i \neq j}^{K-1} \int_{\mathbb{R}-\{0\}} \left[v\left(x_j + J_j\left(x_j, x_i \right) \right) - v\left(x_j - \right) \right]$$

(14)

$$\cdot \nu_j(dx_i).$$

From the smooth pasting, we obtain $v''(x_j+) \leq v''(x_j-)$. In the same way, we get $v''(x_j-) \leq v''(x_j+)$. Thus, we have $v''(x_j-) = v''(x_j+)$ for any $j = 1, 2, \ldots, K$. $\square$

Theorem 3 gives a necessary condition under which second derivatives of value functions are equal at every cutoff when there are K arms in the market.

When $K = 2$, i.e., there are two arms in the market; the agent has two states x_1 and x_2 in the model. In this case, we only need to consider that an agent jumps between the two states, i.e., $m_t = \{1, 2\}$. Thus, there is one cutoff, which is discussed in [12].

3. Application

The application of Theorem 3 is similar to that in [12]. The difference is that there are $K \geq 2$ sellers offering different

products in the market. We index these sellers with $j = 1, 2, \ldots, K$. We assume that all consumers have the same type i, either H or L. Let β_{jH} and β_{jL} be the utilities of consumers who buy good j with type H or L, respectively. We assume $\beta_{1H} > \beta_{2H} > \cdots > \beta_{KH}$ and $\beta_{1L} < \beta_{2L} < \cdots < \beta_{KL}$, i.e., the more likely the consumers are to buy type H, the more tendency they choose the previous sellers. Therefore, we denote that $x \in [0, 1]$ is the common belief that the type is high and then the expected utility of controlling the seller j is $f_j(x) := x\beta_{jH} + (1 - x)\beta_{jL}$. If we denote $a_j := \beta_{jH} - \beta_{jL}$ and $b_j := \beta_{jL}$, the utility is represented by

$$f_j(x) = a_j x + b_j, \tag{15}$$

where $a_j + b_j < a_{j+1} + b_{j+1}$ and $b_j > b_{j+1}$.

At any time, all market participants observe all previous outcomes. Because of the influence caused by uncertain external factors, the flow utility $u_{ji}(t)$ has a noisy signal of the true value (the detailed introduction can be found in [12]).

$$du_{ji}(t) = \xi_{ji}dt + \sigma_j d\widetilde{W}_j(t), \tag{16}$$

where $\widetilde{W}_j(t)$ is independent of $\widetilde{W}_i(t)$ $(i \neq j)$.

x is related to time and it is described as a learning process in [16], denoted by x_t. Without loss of generality, we assume

that there are n_j consumers choosing seller j and $\sum_{j=1}^{K} n_j = N \geq 2$. From the statements in [6, 7], we have that x_t satisfies

$$dx_t = \sum_{j=1}^{K} n_j \Xi_j(x)\, dW_j(t), \tag{17}$$

where $W_j(t)$ is independent of $W_i(t)$ $(i \neq j)$ and $\Xi_j(x) := x^2(1 - x)^2(a_j/\sigma_j)$. We denote $s_j := a_j/\sigma_j$.

In the next subsection, the Markov perfect equilibrium and the socially efficient allocation in K-armed market are discussed.

3.1. Markov Perfect Equilibrium.
Let p_j denote the price of goods of seller j. The price is related to x at instantaneous time t. So p_j is a mapping $[0, 1] \longrightarrow \mathbb{R}$, $j = 1, 2, \ldots, K$. We denote α_i as the choice of ith consumer which is related not only to his common belief x, but also to the prices of the sellers' goods, i.e., $\alpha_i : [0, 1] \times \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{\text{The number of } \mathbb{R} \text{ is } K} \longrightarrow \{1, 2, \ldots, K\}$. A symmetric Markov perfect equilibrium is $\{\alpha, p_1, p_2, \ldots, p_K\}$.

When the choice of the previous $N - 1$ consumers is observed, the utility of the Nth consumer is maximized. Let $U(x)$ denote the maximum utility of this consumer and n'_j denote the number of choosing the jth seller. We have $\sum_{j=1}^{K} n'_j = N - 1$. There exists $d \in \{1, 2, \ldots, K\}$ subject to $n'_d = n_d - 1$ and $n'_i = n_i, i \neq d$. Then, we have

$$rU(x) = \max \left\{ \begin{array}{c} f_1(x) - p_1(x) + \dfrac{1}{2}\left[\displaystyle\sum_{i=1}^{K} n_i\Xi_i(x) + \Xi_1(x) - \Xi_d(x)\right]U''(x) \\[2ex] \vdots \\[2ex] f_{d-1}(x) - p_{d-1}(x) + \dfrac{1}{2}\left[\displaystyle\sum_{i=1}^{K} n_i\Xi_i(x) + \Xi_{d-1}(x) - \Xi_d(x)\right]U''(x) \\[2ex] f_d(x) - p_d(x) + \dfrac{1}{2}\left[\displaystyle\sum_{i=1}^{K} n_i\Xi_i(x)\right]U''(x) \\[2ex] f_{d+1}(x) - p_{d+1}(x) + \dfrac{1}{2}\left[\displaystyle\sum_{i=1}^{K} n_i\Xi_i(x) + \Xi_{d+1}(x) - \Xi_d(x)\right]U''(x) \\[2ex] \vdots \\[2ex] f_K(x) - p_K(x) + \dfrac{1}{2}\left[\displaystyle\sum_{i=1}^{K} n_i\Xi_i(x) + \Xi_K(x) - \Xi_d(x)\right]U''(x) \end{array} \right\}. \tag{18}$$

Due to the price competition, the consumer chooses equal utility for other sellers. Thus, we get

$$f_1(x) - p_1(x)$$

$$+ \frac{1}{2}\left[\sum_{i=1}^{K} n_i\Xi_i(x) + \Xi_1(x) - \Xi_d(x)\right]U''(x) = \cdots$$

$$= f_{d-1}(x) - p_{d-1}(x)$$

$$+ \frac{1}{2}\left[\sum_{i=1}^{K} n_i\Xi_i(x) + \Xi_{d-1}(x) - \Xi_d(x)\right]U''(x)$$

$$= f_d(x) - p_d(x) + \frac{1}{2}\left[\sum_{i=1}^{K} n_i\Xi_i(x)\right]U''(x)$$

$$= f_{d+1}(x) - p_{d+1}(x)$$

$$+ \frac{1}{2}\left[\sum_{i=1}^{K} n_i\Xi_i(x) + \Xi_{d+1}(x) - \Xi_d(x)\right]U''(x) = \cdots$$

$$= f_K(x) - p_K(x)$$
$$+ \frac{1}{2}\left[\sum_{i=1}^{K} n_i \Xi_i(x) + \Xi_K(x) - \Xi_d(x)\right] U''(x). \tag{19}$$

Therefore, we obtain

$$f_i(x) - p_i(x) + \frac{1}{2}\Xi_i(x) U''(x)$$
$$= f_j(x) - p_j(x) + \frac{1}{2}\Xi_j(x) U''(x), \tag{20}$$

where $i, j = 1, 2, \ldots, K$ and $i \neq j$.

Now, we discuss the pricing of goods of sellers. Denote $V_j(x)$ as the jth seller's utility. If there are n_j consumers buying goods j when the price is p_j, we have

$$V_j(x) = \mathbb{E}\left[\int_0^\infty e^{-rt} n_j p_j(x_t)\, dt\right], \tag{21}$$

where $dx_t = \sum_{j=1}^{K} \sqrt{n_j} x_t (1 - x_t) s_j dW_j(t)$.

From (21), we derive that

$$rV_j(x) = n_j p_j(x) + \frac{1}{2}\left(\sum_{i=1}^{K} n_i \Xi_i(x)\right) V_j''(x). \tag{22}$$

We get $n_j = 0$ or N because all consumers choose only one seller (see [6, 7]). When $n_j = 0$, the utility of seller j is presented in the form

$$rV_j(x) = \frac{N}{2}\Xi_d(x) V_j''(x), \tag{23}$$

where we assume that all consumers choose seller d, i.e., $j \in \{1, 2, \ldots, K\} - \{d\}$ and $n_d = N$.

When only one consumer chooses seller j, i.e., $n_j = 1$, the utility of seller j is

$$rV_j(x) = p_j(x)$$
$$+ \frac{1}{2}\left[(N-1)\Xi_d(x) + \Xi_j(x)\right] V_j''(x). \tag{24}$$

As a rational market participant, when no consumer buys goods, the seller adjusts the price so that the payoff in this case is equal to the payoff when only one consumer chooses this seller. We obtain the price of goods of seller j in the form

$$p_j(x) = \frac{1}{2}\left(\Xi_d(x) - \Xi_j(x)\right) V_j''(x). \tag{25}$$

From (20), we get the price of seller d in the form

$$p_d(x) = f_d(x) - f_j(x)$$
$$+ \frac{1}{2}\left(\Xi_d(x) - \Xi_j(x)\right)\left(U''(x) + V_j''(x)\right). \tag{26}$$

We have $N - 1$ cutoffs x^j, $j = 1, 2, \ldots, N - 1$, where $0 < x^1 \le x^2 \le \cdots \le x^{j-1} < 1$. When common belief $x \in (x^{i-1}, x^i)$, consumer chooses seller i, $i = 1, 2, \ldots, K$ ($x^0 := 0$, $x^K := 1$). If $x = x^i$, the utilities of consumers are indifferent when they choose seller i or $i + 1$ due to value matching. When $x \neq x^i$, we have the conclusion

Theorem 4. *All consumers only choose the first seller or the last seller in the market if and only if cutoffs x^j, $j = 1, 2, \ldots, N - 1$, are the same and equal to $(rD_0 - (b_1 - b_k))/((a_1 - a_K) - rD_1)$, where*

$$D_0 := \left(\frac{\gamma_1 + 1}{2} s_1^2 + \frac{\gamma_K - 1}{2} s_K^2\right)\frac{b_1 - b_K}{r}\frac{C_1 + C_K}{C_1 \cdot C_K},$$

$$D_1$$
$$:= \left(s_K^2 - s_1^2\right)\frac{b_1 - b_K}{r}$$
$$+ \left(s_1^2 \frac{\gamma_1 - 1}{2} + s_K^2 \frac{\gamma_K + 1}{2}\right)\frac{a_1 - a_K}{r}\frac{C_1 + C_K}{C_1 \cdot C_K},$$

$$C_1 := \frac{1}{2}(N-1) s_1^2 (\lambda_1 - \gamma_1) + \frac{1}{2} s_K^2 (\lambda_1 + \gamma_K),$$

$$C_K := \frac{1}{2}(N-1) s_K^2 (\lambda_K - \gamma_K) + \frac{1}{2} s_1^2 (\lambda_K + \gamma_1), \tag{27}$$

$$\gamma_j := \sqrt{1 + \frac{8r}{N s_j^2}},$$

$$\lambda_1 := \sqrt{1 + \frac{8r}{(N-1) s_1^2 + s_K^2}},$$

$$\text{and } \lambda_K := \sqrt{1 + \frac{8r}{(N-1) s_K^2 + s_1^2}}.$$

Proof. Firstly, we prove the sufficiency. When all cutoffs are equivalent and equal to $(rD_0 - (b_1 - b_k))/((a_1 - a_K) - rD_0)$, all consumers only choose the first seller or the last seller in the market.

Letting $x^* := x^1 = \cdots = x^{K-1}$, we have $x^* = x^1 = x^{K-1}$. When $x \in (0, x^1)$, all consumers choose the first seller. From (25) and (26), we have

$$p_K(x) = \frac{1}{2}\left(\Xi_1(x) - \Xi_K(x)\right) V_K''(x) \tag{28}$$

and

$$p_1(x) = f_1(x) - f_K(x)$$
$$+ \frac{1}{2}\left(\Xi_1(x) - \Xi_K(x)\right)\left(U''(x) + V_K''(x)\right). \tag{29}$$

Substituting (29) into (22) yields

$$rV_1(x) = N\left[f_1(x) - f_K(x)\right.$$
$$\left. + \frac{1}{2}\left(\Xi_1(x) - \Xi_K(x)\right)\left(V_K''(x) - U''(x)\right)\right] + \frac{N}{2} \tag{30}$$
$$\cdot \Xi_1 V_1''(x)$$

Substituting (28) and (29) into (22) gives rise to

$$rU(x) = f_K(x)$$
$$- \frac{1}{2}\left(\Xi_1(x) - \Xi_K(x)\right)\left(V_K''(x) + U''(x)\right) \quad (31)$$
$$+ \frac{N}{2}\Xi_1 U''(x).$$

From (23), (30), and (31), for $x \in (0, x^1)$, we have

$$V_1(x) = N\frac{f_1(x) - f_K(x)}{r}$$
$$+ P_3 x^{(\gamma_1+1)/2}(1-x)^{-(\gamma_1-1)/2}$$
$$- NP_2 x^{(\lambda_1+1)/2}(1-x)^{-(\lambda_1-1)/2}$$
$$V_K(x) = P_1 x^{(\gamma_1+1)/2}(1-x)^{-(\gamma_1-1)/2} \quad (32)$$
$$U(x) = \frac{f_K(x)}{r} + P_2 x^{((\lambda_1+1)/2)(1-x)^{-(\lambda_1-1)/2}}$$
$$- P_1 x^{(\gamma_1+1)/2}(1-x)^{-(\gamma_1-1)/2}.$$

Similarly, for $x \in (x^{K-1}, 1)$, it has

$$V_1(x) = Q_1 x^{-(\gamma_K-1)/2}(1-x)^{(\gamma_K+1)/2}$$
$$V_K(x) = N\frac{f_K(x) - f_1(x)}{r}$$
$$+ Q_3 x^{-(\gamma_K-1)/2}(1-x)^{(\gamma_K+1)/2} \quad (33)$$
$$- NQ_2 x^{-(\lambda_K-1)/2}(1-x)^{(\lambda_K+1)/2}$$
$$U(x) = \frac{f_1(x)}{r} + Q_2 x^{((\lambda+1)/2)(1-x)^{-(\lambda-1)/2}}$$
$$- Q_1 x^{(\gamma_K+1)/2}(1-x)^{-(\gamma_K-1)/2},$$

where $\gamma_j := \sqrt{1 + 8r/Ns_j^2}$, $\lambda_1 := \sqrt{1 + 8r/((N-1)s_1^2 + s_K^2)}$, $\lambda_K := \sqrt{1 + 8r/((N-1)s_K^2 + s_1^2)}$, and P_ξ and Q_ξ are constants, $\xi = 1, 2, 3$.

Using value matching, smooth pasting, the second derivative condition, and $x^* = x^1 = x^{K-1}$, we have $V_1(x^*-) = V_K(x^*+)$, $V_1'(x^*-) = V_K'(x^*+)$, $V_1''(x^*-) = V_K''(x^*+)$, and $U(x^*-) = U(x^*+)$. Therefore, we obtain

$$\frac{f_1(x^*) - f_K(x^*)}{r} = \frac{\varphi(x^*)}{C_1} + \frac{\varphi(x^*)}{C_K}, \quad (34)$$

where

$$\varphi(x) := \left(\frac{\gamma_1+1}{2}s_1^2 + \frac{\gamma_K-1}{2}s_K^2\right)\frac{b_1 - b_K}{r}$$
$$+ \left[\left(s_K^2 - s_1^2\right)\frac{b_1 - b_K}{r}\right.$$
$$+ \left.\left(s_1^2\frac{\gamma_1-1}{2} + s_K^2\frac{\gamma_K+1}{2}\right)\frac{a_1 - a_K}{r}\right]x, \quad (35)$$
$$C_1 := \frac{1}{2}(N-1)s_1^2(\lambda_1 - \gamma_1) + \frac{1}{2}s_K^2(\lambda_1 + \gamma_K)$$

and

$$C_K := \frac{1}{2}(N-1)s_K^2(\lambda_K - \gamma_K) + \frac{1}{2}s_1^2(\lambda_K + \gamma_1). \quad (36)$$

Define

$$D_0 := \left(\frac{\gamma_1+1}{2}s_1^2 + \frac{\gamma_K-1}{2}s_K^2\right)\frac{b_1 - b_K}{r}\frac{C_1 + C_K}{C_1 \cdot C_K} \quad (37)$$

and

$$D_1 := \left(s_K^2 - s_1^2\right)\frac{b_1 - b_K}{r}$$
$$+ \left(s_1^2\frac{\gamma_1-1}{2} + s_K^2\frac{\gamma_K+1}{2}\right)\frac{a_1 - a_K}{r}\frac{C_1 + C_K}{C_1 \cdot C_K}. \quad (38)$$

Thus, we get

$$\frac{f_1(x^*) - f_K(x^*)}{r} = D_o + D_1 x^* \quad (39)$$

and

$$x^* = \frac{rD_0 - (b_1 - b_k)}{(a_1 - a_K) - rD_0}. \quad (40)$$

When $x^* := x^1 = \cdots = x^{K-1}$, $(x^i, x^{i+1}) = \phi$, $i = 1, 2, \ldots, K-2$, all consumers only choose the first seller or the last seller in the market.

The sufficiency of Theorem 4 is proved.

Now, we prove the necessity. That all consumers choose only the first seller or the last seller in the market means that these cutoffs are the same and equal to $(rD_0 - (b_1 - b_k))/((a_1 - a_K) - rD_0)$, $x \neq x^*$. We prove it by contradiction.

Assuming $x^i \neq x^{i+1}$, we get $x^i < x^{i+1}$, $i = 1, 2, \ldots, K-2$. So $(x^i, x^{i+1}) \cap \{[0, x^*) \cup (x^*, 1]\} \neq \phi$, i.e., $\exists x_0 \neq x^*$, $x_0 \in (x^i, x^{i+1})$. When the common belief is x_0, the optimal choice is the $(i+1)$th and $i+1 \neq 1$ and $i+1 \neq K$. This is contradiction to the proposition of which all consumers choose only the first seller or the last seller. So $x^i = x^{i+1}$, $i = 1, 2, \ldots, K-2$. According to the proof of the sufficiency, we have $x^* = x^1 = x^{K-1} = (rD_0 - (b_1 - b_k))/((a_1 - a_K) - rD_0)$. The necessity of Theorem 4 is proved. □

Theorem 4 shows the necessary and sufficient conditions for the consumers' choice. We find that when the consumers only choose the first seller and the last seller, the multiarmed

bandits problem would be attributed to the two-armed bandits. In other words, other sellers gradually disappear in the market because they have no sales. The multiarmed bandit problem is transformed into the two-armed bandits in the situation discussed in [12].

3.2. Socially Efficient Allocation. We consider the optimal choice of planners when they face multiarmed bandit problem. Let the total social surplus function be $w(x)$. We have

$$w(x) = \sup_{n_1, n_2, \ldots, n_K} \left\{ \mathbb{E} \int_0^\infty e^{-rt} \sum_{j=1}^K n_{jt} f_j(x_t) dt \right\} \quad (41)$$

$$\text{s.t.} \quad dx_t = \sum_{j=1}^K \sqrt{n_{jt}} x_t (1 - x_t) s_j dW_j(t). \quad (42)$$

Assume that $\Pi_j(x)$ is the total social surplus in a neighborhood of x if a planner choose seller j. From (41), we have

$$r\Pi_j(x) = N \left[f_j(x) + \frac{1}{2} \Xi_j(x) \Pi_j''(x) \right], \quad (43)$$

where $j = 1, 2, \ldots, K$.

Solving the ordinary differential equation (43), we get

$$\Pi_1(x) = N \frac{f_1(x)}{r} + C_1^1 x^{(\gamma_1 + 1)/2} (1 - x)^{-(\gamma_1 - 1)/2},$$

$$\Pi_j(x) = N \frac{f_j(x)}{r} + C_j^1 x^{(\gamma_j + 1)/2} (1 - x)^{-(\gamma_j - 1)/2}$$

$$+ C_j^2 x^{-(\gamma_j - 1)/2} (1 - x)^{(\gamma_j + 1)/2}$$

$$j = 2, \ldots, K - 1,$$

$$\Pi_K(x) = N \frac{f_K(x)}{r} + C_K^2 x^{-(\gamma_K - 1)/2} (1 - x)^{(\gamma_K + 1)/2}, \quad (44)$$

where C_i^1 and C_i^2, $i = 1, 2, \ldots, K$, are constants.

We denote $\hat{x}^1, \hat{x}^2, \ldots, \hat{x}^{K-1}$ subjecting to $0 < \hat{x}^1 \le \hat{x}^2 \le \cdots \le \hat{x}^{K-1} < 1$ as the cutoffs of the choice for the planner. When $x \in (\hat{x}^{i-1}, \hat{x}^i)$, the optimal choice is the ith seller, $i = 1, 2, \ldots, K$ and $\hat{x}^0 = 0$, $\hat{x}^K = 1$. Due to value matching, smooth pasting, and the second derivative conditions, one has

$$\Pi_j(\hat{x}_j-) = \Pi_{j+1}(\hat{x}_j+),$$

$$\Pi_j'(\hat{x}_j-) = \Pi_{j+1}'(\hat{x}_j+), \quad (45)$$

$$\Pi_j''(\hat{x}_j-) = \Pi_{j+1}''(\hat{x}_j+),$$

where $j = 1, 2, \ldots, K - 1$.

There are $(3K - 3)$ unknown parameters and $(3K - 3)$ equations in system (45) with $[x^\beta (1-x)^{1-\beta}]' \ne x^\beta (1-x)^{1-\beta}$, for all $\beta > 1$, i.e., the coefficient matrix of system (45) is nonsingular. Thus, system (45) has a unique solution.

The planner chooses from the first seller or the $K - th$ seller when $\hat{x}^1 = \hat{x}^2 = \cdots = \hat{x}^{K-1}$. Let $x^\dagger := \hat{x}^1 = \hat{x}^2 = \cdots = \hat{x}^{K-1}$. System (45) becomes

$$\Pi_1(x^\dagger) = \Pi_K(x^\dagger),$$

$$\Pi_1'(x^\dagger) = \Pi_K'(x^\dagger), \quad (46)$$

$$\Pi_1''(x^\dagger) = \Pi_K''(x^\dagger).$$

We obtain

$$x^\dagger = \frac{(b_1 - b_k)\left(((\gamma_1 + 1)/2) s_1^2 + ((\gamma_K - 1)/2) s_K^2\right)}{(a_K - a_1)\left(s_1^2 ((\gamma_1 - 1)/2) + s_K^2 ((\gamma_K + 1)/2)\right) + (b_1 - b_K)(s_1^2 - s_K^2)}. \quad (47)$$

The results in [12] introduce the necessary and sufficient condition under which the Markov perfect equilibrium with cautious strategies is socially efficient with two-armed bandits.

Corollary 5. *When $K \ge 2$, $N \ge 2$, the consumers' cutoffs are x^*, and the planner's cutoffs are $x^\dagger$, the Markov perfect equilibrium with cautious strategies is socially efficient if and only if $s_1 = s_K$. Moreover, $x^* > x^\dagger$ ($x^* < x^\dagger$) when $s_1 > s_K$ ($s_1 < s_K$).*

The proof of Corollary 5 is similar to that of Theorem 2 in [12]. We omit its proof.

Corollary 5 shows that the necessary and sufficient conditions under which the Markov perfect equilibrium with cautious strategies is socially efficient when the cutoffs are multiarmed bandits. Jin and Xi [12] present the conditions in

the case of two-armed market. Thus, Theorem 4 extends parts of results in [12].

According to the condition in Corollary 5, when $(\beta_{1H} - \beta_{1L})/\sigma_1 = (\beta_{KH} - \beta_{KL})/\sigma_K$, we obtain that the Markov perfect equilibrium with cautious strategies is socially efficient. If $\sigma_1 = \sigma_k$, we obtain $\beta_{1H} - \beta_{KH} = \beta_{1L} - \beta_{KL}$. In the light of our initial hypothesis, we have $\beta_{1H} = \beta_{KH}$ and $\beta_{1L} = \beta_{KL}$; i.e., all sellers are identical in the market.

4. Conclusion

We study a common value experimentation with multiarmed bandits and present its application. This extends two-armed bandits in [12] to multiarmed bandits. We derive the HJB equation with multiarmed bandits. In the application, we get the necessary and sufficient conditions for the choices of consumers from two sellers. The necessary and sufficient

conditions guarantee that the Markov perfect equilibrium with cautious strategies is socially efficient. In future, we need to solve all the cutoffs in system (45) when these cutoffs are different and give general solutions about these cutoffs.

Data Availability

No data are used to support the study. Using theoretical derivation and proof, the authors get their conclusions.

Conflicts of Interest
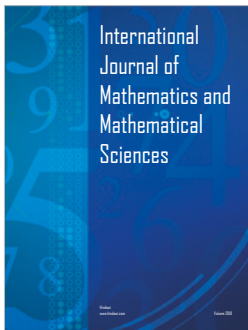
The authors declare that they have no conflicts of interest.
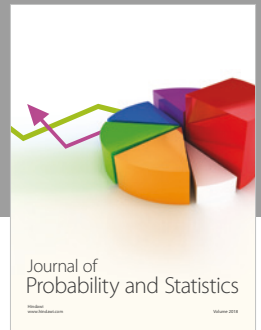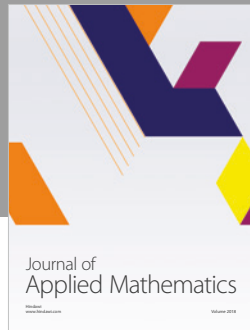
Authors' Contributions

The article is a joint work of three authors who contributed equally to the final version of the paper. The authors read and approved the final manuscript.

Acknowledgments

References

[1] H. Robbins, "Some aspects of the sequential design of experiments," *Bulletin (New Series) of the American Mathematical Society*, vol. 58, pp. 527–535, 1952.

[2] J. C. Gittins and D. M. Jones, "A dynamic allocation index in the sequential design of experiments," in *Progress in Statistics*, pp. 241–266, North-Holland, Amsterdam, 1974.

[3] M. N. Katehakis and J. Veinott, "The multi-armed bandit problem: decomposition and computation," *Mathematics of Operations Research*, vol. 12, no. 2, pp. 262–268, 1987.

[4] I. Karatzas, "Gittins indices in the dynamic allocation problem for diffusion processes," *Annals of Probability*, vol. 12, no. 1, pp. 173–192, 1984.

[5] H. Kaspi and A. Mandelbaum, "Lévy bandits: multi-armed bandits driven by Lévy processes," *The Annals of Applied Probability*, vol. 5, no. 2, pp. 541–565, 1995.

[6] P. Bolton and C. Harris, "Strategic experimentation," *Econometrica*, vol. 67, no. 2, pp. 349–374, 1999.

[7] D. Bergemann and J. Valimaki, "Experimentation in Markets," *Review of Economic Studies*, vol. 67, no. 2, pp. 213–234, 2000.

[8] A. Cohen and E. Solan, "Bandit problems with Lévy payoff," *Mathematics of Operations Research*, vol. 38, no. 1, pp. 92–107, 2013.

[9] X. Yu, X. J. Xie, and N. Duan, "Small-gain control method for stochastic nonlinear systems with stochastic iISS inverse dynamics," *Automatica*, vol. 46, no. 11, pp. 1790–1798, 2010.

[10] Z.-J. Wu, X.-J. Xie, P. Shi, and Y.-Q. Xia, "Backstepping controller design for a class of stochastic nonlinear systems with Markovian switching," *Automatica*, vol. 45, no. 4, pp. 997–1004, 2009.

[11] X. Xie and M. Jiang, "Output feedback stabilization of stochastic feedforward nonlinear time-delay systems with unknown output function," *International Journal of Robust and Nonlinear Control*, vol. 28, no. 1, pp. 266–280, 2018.

[12] J. Eeckhout and X. Weng, "Common value experimentation," *Journal of Economic Theory*, vol. 160, pp. 317–339, 2015.

[13] D. Applebaum, *Lévy Processes and Stochastic Calculus*, vol. 93 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge, UK, 2004.

[14] G. Peskir and A. Shiryaev, *Optimal stopping and free-boundary problems*, Birkhauser Verlag, Basel, Switzerland, 2006.

[15] J. Yong and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, vol. 43 of *Applications of Mathematics*, Springer, New York, NY, USA, 1999.

[16] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes II: Applications*, vol. 6, Springer-Verlag, Berlin, 2001.